OUTLINES
OF
BIOCHEMISTRY

OUTLINES
OF
BIOCHEMISTRY

By

Ross Aiken Gortner

Late Chief of the Division of Agricultural Biochemistry
University of Minnesota

Third Edition

edited by

Ross Aiken Gortner, Jr.

Professor of Biochemistry, Wesleyan University
Middletown, Connecticut

and

Willis Alway Gortner

Head, Department of Chemistry
Pineapple Research Institute, Honolulu, T. H.
Formerly Associate Professor of Biochemistry
School of Nutrition, Cornell University

John Wiley & Sons, Inc., New York
Chapman & Hall, Limited, London

THIRD EDITION
Second Printing, February, 1950

PRINTED IN THE UNITED STATES OF AMERICA

THIS VOLUME IS DEDICATED TO THOSE PERSONS WHO WERE ASSOCIATED WITH THE LATE ROSS AIKEN GORTNER ON THE STAFF OF THE DIVISION OF AGRICULTURAL BIOCHEMISTRY IN THE UNIVERSITY OF MINNESOTA, FOR IT WAS ONLY THE SPLENDID COOPERATIVE SPIRIT OF THIS GROUP THAT MADE THIS BOOK POSSIBLE.

CONTRIBUTORS

The cooperative efforts of the following persons, nearly all former students or colleagues of the late Ross Aiken Gortner, made possible the present edition of this book.

SAMUEL I. ARONOVSKY, *Head of Pulp and Paper Section, Agricultural Residues Division, Northern Regional Research Laboratories, Peoria, Ill.* ~ *revision of Chapter 29 on lignin and the tannins.*

PAUL D. BOYER, *Associate Professor of Agricultural Biochemistry, University of Minnesota* ~ *contribution of Chapter 26 on carbohydrate metabolism.*

DAVID R. BRIGGS, *Professor of Agricultural Biochemistry, University of Minnesota* ~ *revision of Chapters 6 through 10 on the colloids.*

HENRY B. BULL, *Professor of Biochemistry, Northwestern University Medical School* ~ *revision of Chapters 1 through 4 on the colloids, and contribution of Chapters 5 (Oxidation-Reduction) and 18 (Protein Denaturation).*

GEORGE O. BURR, *Head of Department of Physiology and Biochemistry, Experiment Station, Hawaii Sugar Planters' Association, Honolulu; formerly Professor of Physiological Chemistry, University of Minnesota* ~ *revision of Chapters 34 and 35 on plant pigments.*

WILLIAM F. GEDDES, *Chief of the Division of Agricultural Biochemistry, University of Minnesota* ~ *revision of Chapters 22 through 25 and 27 on carbohydrates.*

ROSS A. GORTNER, JR., *Professor of Biochemistry, Wesleyan University, Middletown, Conn.* ~ *revision of Chapters 19 through 21 on proteins and Chapter 37 on hormones.*

WILLIS A. GORTNER, *Head of Department of Chemistry, Pineapple Research Institute, Honolulu; formerly Associate Professor of Biochemistry, School of Nutrition, Cornell University* ~ *revision of Chapters 30, 31, and 33 on the lipids and Chapter 36 on vitamins; contribution of Chapter 32 on metabolism of the lipids.*

W. M. SANDSTROM, *Professor of Agricultural Biochemistry, University of Minnesota* ~ *revision of Chapters 11 through 17 on proteins and Chapter 38 on enzymes.*

J. J. WILLAMAN, *Chief of Biochemistry Division, Eastern Regional Research Laboratories, Philadelphia, Pa.* ~ *contribution of Chapter 28 on pectic substances.*

PREFACE TO THE THIRD EDITION

During the eleven years that have elapsed since this book was last revised, great strides have been made in nearly all phases of biochemistry. Of necessity, some of the chapters have had to be largely rewritten and many others have been extensively changed. New chapters dealing with protein denaturation, carbohydrate metabolism, and the metabolism of lipids have been added. Certain other chapters which appeared in the previous edition have been omitted from the present volume, but their subject matter has been partially reincorporated in logical fashion into existent chapters. Throughout this extensive revision, however, we have tried to keep the aims of the book and its approach to the subject matter as much as possible like those of the original author.

With so much new information at hand it was indeed difficult to keep the size of this new edition within reasonable limits. It was, of course, necessary to treat some sections more fully than in the past, but it was possible to make a roughly comparable reduction in certain other sections. In chapters dealing with subjects that are rather thoroughly dealt with in medical biochemistry textbooks, no attempt was made to cover the field extensively.

There have been rather prolonged, unavoidable delays in the preparation of this Third Edition of *Outlines*. World War II, coming at about the time that a new edition was being planned, held up preparation for several years. The post-war years placed unusually heavy teaching, research, and administrative responsibilities on the shoulders of those contributing to the present volume, with the result that several "deadlines" for receipt of manuscript had to be set. Thus, some parts of the manuscript were received well over a year before the last instalments were in. Despite many difficulties, we found the task inspiring and trust that the result is one of which the original author would be justly proud.

We are especially indebted to the following persons, each of whom handled the revision of one or more chapters: Drs. S. I. Aronovsky, P. D. Boyer, D. R. Briggs, H. B. Bull, G. O. Burr, W. F. Geddes, W. M. Sandstrom, and J. J. Willaman. We wish to thank Drs. Thomas Schoch and H. O. Halvorson for contributing material on starch and the bacterial metabolism of carbohydrates, respectively; Drs. M. G.

Burford, G. A. Hill, J. W. Sease, and Fred Smith for their helpful criticisms of certain parts of the manuscript; Rachel Rude Gortner for invaluable assistance in the early stages of the revision; Priscilla Cahill Gortner, Florence Pickett Gortner, and Dr. Robert W. Leeper for their help in connection with proofreading; and all those persons and publishers who so willingly permitted us to reproduce material from other sources.

Ross A. Gortner, Jr.
Willis A. Gortner

May, 1949

PREFACE TO THE FIRST EDITION

"When I have the honor of being consulted by a young man who has not yet found himself intellectually but who is filled with the desire to devote his life to some branch of medicine, be it clinical medicine, pathology, hygiene, bacteriology, physiology or pharmacology, my advice always is, 'Study chemistry at least three years. Try with all your power to master enough of this great science to start you on your career.'"

JOHN J. ABEL

It is rather generally agreed among the scientists that the actions and reactions of a biological organism are expressions of the energy relationships due to chemical and physicochemical processes taking place within the cells and tissues which comprise the organism.

The biological organism can be looked upon as a complex system of chemical constituents, composed mainly of proteins, carbohydrates, fats and lipids, mineral elements, and water, which are organized by the mysterious forces which we call "life," and the actions and reactions of this protoplasmic mass are in turn determined by the energy interchanges of molecular transformations and surface and interfacial forces.

In most of the universities of America the development of the field of biochemistry has been left very largely to the group interested in the medical aspects. Accordingly, in a very large measure the biochemistry of the American universities is not biochemistry in its strictest sense, but rather leans more and more toward the field of human pathology. Most medical school biochemistry could be more correctly designated as human pathological chemistry.

It is obvious that there should be strong departments of physiological chemistry associated with the medical schools. However, it should likewise be recognized that there is a necessity for a study of the fundamental reactions underlying the broader field of biology, the primary object of which is to study and investigate the chemical and physicochemical reactions which take place in the normal biological organism, whether that organism be animal or plant.

The greatest advance in the biological sciences can take place only when the chemists are fully aware of certain of the biological problems and the biological point of view, and only when the biologists appreciate the assistance which chemical knowledge and chemical technic can offer to the solution of the major problems.

When, in 1913, Professor R. W. Thatcher was called to the University

of Minnesota, he inaugurated a course of lectures on "phytochemistry" and taught the course until he became Dean of the Department of Agriculture in 1917, at which time Dr. C. A. Morrow assumed charge of the lecture work in "Phytochemistry" and supervision of the parallel laboratory course in "Biochemical Laboratory Methods."

Dr. Morrow remained in charge of both courses until his unfortunate illness in the spring of 1922, following which illness the responsibility for the lecture work was assumed by the present author, and upon resumption of active duties Dr. Morrow was left free to devote his entire energies to the development of the associated laboratory phases of the work. His success in this task is attested by the practical completion of the manuscript of the laboratory manual, *Biochemical Laboratory Methods*, which appeared posthumously from the press of John Wiley & Sons, Inc., in 1927.

The course of lectures upon which the present volume is based must not therefore be regarded as the product of a single individual, for it embodies primarily the efforts of three men who have actually taught the course, and in addition the advice, cooperation, and criticism of all other persons who from time to time have been members of the Staff of the Division of Agricultural Biochemistry.

During the year 1925–1926 the present author prepared a series of mimeographed lecture outlines as an aid to the students in the classroom. These outlines were somewhat revised in the summer of 1927 and issued to the students in bound mimeographed form. The present volume follows, in general, the topics as presented in these outlines, although the scope of the field has been greatly expanded.

All of the reactions and interactions which we call life take place in a colloid system, and the author believes that much of the "vital energy" can in the last analysis be traced back to energies characteristic of surface films and interfaces. This belief is the justification for the detailed consideration of colloid systems which forms the first section of the volume.

In the subsequent sections dealing with proteins, carbohydrates, fats, etc., those organic compounds characteristic of living tissues, particular attention has been paid to structural organic chemistry and organic and physicochemical reactions. No attempt has been made to produce a "handbook" of biochemical compounds or a "descriptive biochemistry" detailing the chemical properties, crystal structure, etc., of the various compounds. There are already many volumes covering these fields, and no good purpose would be served by adding to the list.

The purpose of the present volume is that those students who are interested in biological phenomena may have an insight into the roles which organic chemistry and physical chemistry play in living processes.

It may be regarded as an attempt to interpret some of the reactions characteristic of the normal cell. Although a great many of the illustrations have been drawn from plant material, it must be remembered that in the last analysis the chemistry of the cell is essentially the same both for plants and for animals. There is no sharp distinction between "phytochemistry" and "zoochemistry." The same general reactions of protoplasm apply to both. If the student interested in some pure or applied field of biology or of chemistry is assisted, ever so little, in the clarification of the problems in his special field, then the object of the author will have been attained.

Toward this end the exact titles have been included in the literature citations. With but few exceptions (and those are noted in the text) the citations have been checked back against the original publication so as to preclude bibliographic errors. A sincere attempt has been made to give to other investigators the proper credit for data or for ideas which have been utilized. On the other hand, the text does not pretend to be an exhaustive treatment of the subjects, nor is the literature cited necessarily in an historical sequence. It is recognized that in many instances the references cited are not the first publications on the subject, but it is felt that the references which are cited contain something which should be called to the attention of the reader. The author may perhaps be pardoned for the numerous references to work done in his own laboratories, for obviously he is most familiar with the details of such work and the conditions under which it was carried out.

In conclusion, the author wishes to thank his colleagues, Dr. J. J. Willaman and Dr. Leroy S. Palmer, who have contributed Chapters XXVII and XXXIV, respectively; The Williams & Wilkins Company, and Prof. W. Mansfield Clark; the editors of the *Journal of Biological Chemistry*; The Chemical Catalog Company, Prof. Harry N. Holmes, Prof. O. L. Sponsler, and Prof. W. H. Dore; The Carnegie Institution of Washington, and Dr. D. T. MacDougal; Carl Zeiss, Inc.; Dr. Karl Mez; Prof. John J. Abel; Prof. E. F. Burton; Jerome Alexander; Prof. Francis Lloyd; Dr. Robert Newton; Dr. G. E. Holm; Dr. P. V. Wells; Dr. William Robinson, and others for permission to use copyrighted material or for photographs or data.

Parts of the manuscript have been carefully read and criticized by my colleagues, Dr. W. M. Sandstrom, Dr. J. J. Willaman, Dr. L. S. Palmer, Dr. C. H. Bailey, Dr. W. B. Sinclair, Dr. David R. Briggs, and Mr. Charles F. Rogers. To them I express my thanks.

And lastly I wish to express my deepest appreciation of Miss Rachel Rude for invaluable assistance in the task of preparing the final manuscript and in proofreading.

Ross Aiken Gortner

CONTENTS

I. COLLOIDS

II. PROTEINS

III. CARBOHYDRATES AND RELATED SUBSTANCES

IV. THE LIPIDS AND ESSENTIAL OILS

V. PLANT PIGMENTS

VI. BIOCHEMICAL REGULATORS

I

COLLOIDS

The colloidal is, in fact, the dynamical state of matter; crystalloidal being the statical condition. The colloid possesses ENERGIA. *It may be looked upon as the probable primary source of the force appearing in the phenomena of vitality.*

THOMAS GRAHAM (1861)

The colloid field today presents the most promising realm in medical research.

W. G. MAYO (1928)

If the colloid structures did not display highly specialized molecular structures at their surface, no reactions would occur; for here catalysis occurs. Were it not equipped with catalysts, every living unit would be a static system.

SIR FREDERICK GOWLAND HOPKINS (1933)

Progress in our attack against disease, in our better understanding of the human body, of bacteria, viruses, and other vectors of disease, and of their interaction with each other depends on a better understanding of intermolecular forces and interactions.

Usually specific physiological properties are determined not by strong intramolecular forces but by the weak forces—van der Waals forces, hydrogen bonds, and so on—which operate between molecules. Physiological activity seems related to size, shape and structure of the interacting molecules.

LINUS PAULING (1946)

CHAPTER 1

The Colloid State of Matter

It seems appropriate to begin the study of biochemistry with a consideration of the colloid state of matter, because in the last analysis many of the reactions of biological systems are dependent on the colloidal phenomena operating in such systems. There are a number of useful text and reference books on colloid chemistry. Among these may be mentioned those of Hauser,[1] Weiser,[2] Hartman,[3] Alexander,[4] and Freundlich.[5]

We often hear the statement that such and such a material is a colloid. A statement of this kind is not strictly correct. Colloid phenomena deal with a *state* of matter, not a kind of matter. Early workers in this field published many papers indicating that certain materials could under certain specific conditions of manipulation be brought into the colloidal state. We now know that, if the proper technic is used, any material can be brought into the colloidal state. Consequently there is no sharp line of demarcation between substances which can be regarded as colloids and substances which are generally non-colloids.

The first real work in the field of colloid chemistry was due to the activities of Thomas Graham (1805–1869). All of Graham's forty-six research papers dealt with the single phenomenon of diffusion.[6] Graham[7] published important summaries in 1861 and 1864, and in these introduced many new terms. Thus, speaking of the properties of colloids, Graham states, "Their peculiar physical aggregation with the chemical indifference referred to appears to be required in substances that can intervene in the organic processes of life. The plastic elements of the animal body are found in this class. As gelatine appears to be

[1] E. A. Hauser, *Colloidal Phenomena*, McGraw-Hill Book Co., New York, 1939.

[2] H. B. Weiser, *Colloid Chemistry*, 2nd ed., John Wiley & Sons, New York, 1949.

[3] R. J. Hartman, *Colloid Chemistry*, Houghton Mifflin Co., Boston, 1939.

[4] J. Alexander, *Colloid Chemistry; Principles and Applications*, 4th ed., D. Van Nostrand Co., New York, 1937.

[5] H. Freundlich, *Kapillarchemie*, Bd. 1 and 2, Akademische Verlagsgellschaft M.B.H., Leipzig, 1930, 1932.

[6] Cf. R. A. Gortner, *J. Chem. Educ.*, **11**, 279 (1934).

[7] Thomas Graham, *Phil. Trans.*, 1861, p. 183; *J. Chem. Soc.*, **17**, 318 (1864).

its type, it is proposed to designate substances of the class as *colloidal*[8] and to speak of their peculiar form of aggregation as the *colloidal condition of matter.* Opposed to the colloidal is the crystalline condition. Substances affecting the latter form will be classed as *crystalloids.* The discussion is no doubt one of intimate molecular constitution."

Graham recognized that there was no sharp discontinuity between colloids and crystalloids and that one and the same substance may under different sets of conditions be either colloidal or crystalloidal. Thus, he states (1861), "A departure from its normal condition appears to be presented by a colloid holding so high a place in its class as albumen. In the so-called blood-crystals of Funke, a soft and gelatinous albuminoid body is seen to assume a crystalline contour. Can any facts more strikingly illustrate the maxim that in nature there are no abrupt transitions, and that distinctions of class are never absolute?"

We now know that any substance, even including such substances as sodium chloride, can by proper technic and by the proper choice of medium be brought into the colloidal state. Some substances may be colloidal in one liquid medium and truly crystalloidal in another. Tannin dispersed in water gives rise to a colloidal system, whereas it dissolves in acetic acid to form a crystalloidal solution. Many other substances show similar behavior. *Only the physical properties of the resulting system will show whether or not a given material is colloidal or crystalloidal.*

A substance cannot be strictly spoken of as a colloid, because colloid implies (1) a state of matter and (2) at least two components or phases. Thus, we have not a colloid material but rather colloidal systems.

Water has three states: vapor, liquid, and solid. Each of these may exist in colloidal systems. Barnes[9] has discussed certain natural phenomena produced by colloidal water and colloidal ice, and we shall have occasion later to consider instances in which water vapor is involved.

Inasmuch as a colloid system is a heterogeneous system, it is necessary to distinguish between the substance which is dispersed and the medium in which the material is dispersed. Various terms have been employed by different writers. Some authors speak of the *disperse phase* and *dispersions medium;* others of the *discontinuous phase* and the *continuous phase;* others of the *internal phase* and the *external phase;* and still others of the *micelles* and the *intermicellar liquid.* The last group of terms appears to be coming more and more into general use and in some respects appears preferable.

[8] From the Greek, κολλα, kolla, meaning glue or gelatin, and ειδος, eidos, meaning like.

[9] Howard T. Barnes, Colloid Symposium Monograph, Vol. III, p. 103, Chemical Catalog Co., New York, 1925.

In the older literature a distinction is drawn between colloids and crystalloids on the basis that colloids are amorphous. This distinction is to be discouraged as having little or no meaning; x-ray studies have shown that many colloidal particles have a crystalline structure. The only satisfactory way to define a colloid is on the basis of particle size. This has been done arbitrarily as shown in the following diagram.

Molecules and Ions	Colloids	Matter in Mass
Not Visible in Ultramicroscope	Visible in Ultramicroscope	Visible in Microscope
1 mμ	0.1 μ	

It will be noted that a boundary has been placed at 1 mμ (0.000001 mm.) diameter, defining the lower limit of the colloid realm, and another boundary at approximately 0.1 μ (0.0001 mm.), defining the upper limit of the colloid realm. All particles within this range of dimensions are known as colloidal whether they are molecularly dispersed or not. It cannot be too strongly emphasized that *these are arbitrary boundaries.* However, they seem to have been rather wisely chosen. Most compounds which are crystalline in the solid state and which yield monodisperse solutions of molecules or ions do not contain molecules which exceed 1 mμ in diameter. On the other hand, monomolecular solutions of proteins fall within this size range and exhibit properties of colloidal systems.

The upper limit of the colloid realm, 0.1 μ in diameter, was chosen because this represents the extreme range of the ordinary microscope, with an oil-immersion objective. The colloid realm is thus concerned with systems containing subdivisions of matter lying between the limits of visibility of the usual laboratory microscope and those solutions which are strictly crystalloidal. It must be understood at the outset that the properties of the colloidal systems do not strictly coincide with these arbitrary boundaries but that *there is a continual gradation in properties from truly crystalloidal systems to coarsely divided suspensions, and that the optimum zone of colloidality lies somewhere near the center of the colloid realm, as noted in the diagram above.* In other words a coarsely divided suspension will show to a minor degree certain of the phenomena characteristic of the colloidal state, and in the same way certain of the phenomena of the colloidal state reflect in a minor degree the characteristics of true solutions.

Siedentopf and Zsigmondy classify the three states of matter noted in the above diagram as *microns* visible in the microscope, *submicrons* or *ultramicrons* visible in the ultramicroscope, and *amicrons* not visible in the ultramicroscope. According to their terminology all systems

showing characteristic colloidal properties contain submicrons. We are probably nearer the truth when we place the lower limit of colloid systems at 1 mμ than when we place the upper limit at 0.1 μ, and there has been a general tendency in recent years to raise the upper limit to perhaps 0.5 μ, inasmuch as fine suspensions possess to a large degree certain characteristic properties of colloidal systems.

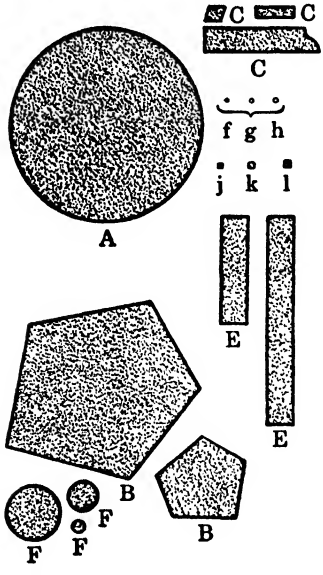


FIG. 1. *A* = red blood cells; *B* = starch grains; *C* = clay particles; *E* = *Anthrax bacilli*; *F* = *Cocci*; *fgh* = colloidal gold (*f* = 10 mμ); *jkl* = gold coagula. (After Zsigmondy. By permission from Alexander's *Colloid Chemistry*, 1924.)

Perhaps everyone has viewed a red blood corpuscle under the microscope and accordingly has some idea of the relative size of a red blood corpuscle as compared with other familiar objects. Figure 1 illustrates the size relationships of some rather common materials as compared with the size of colloidal particles. It should be noted, however, that the smallest colloidal particle noted at (*f*) is ten times the diameter of the particles which fix the lower limit of the colloidal realm. Hackh [10] has tabulated some very interesting data comparing size, time, mass, velocity, and temperature relationships for common objects and systems. A study of these tabulations will greatly assist the student in orienting himself in the field of natural phenomena.

According to Wo. Ostwald one can conveniently classify colloid systems into eight great groups:

Solid-in-Solid. Examples are the ruby glass of the cathedral windows which is a colloid system of metallic gold dispersed in glass; blue rock salt, a colloid system of sodium dispersed in sodium chloride; the black diamond, which is carbon dispersed in crystalline carbon. The colors of many precious stones are examples of solid-in-solid colloidal systems. This type of system is important in metallurgy, but from the biological standpoint it is relatively unimportant.

Solid-in-Liquid. Faraday, in 1857, exhibited before the Royal Society a brilliant red liquid which he had prepared by reducing an aqueous solution of auric chloride. Inasmuch as this liquid showed a beautiful red color by transmitted light and a golden sheen by reflected light, he suggested that the color might be due to particles of solid gold suspended

[10] I. W. D. Hackh, *Sci. Monthly*, **36**, 55 (1933).

in the liquid. This experiment was forgotten for many years, but we now know that Faraday had prepared one of the first examples of a solid-in-liquid colloidal system.

However, the definition of "a solid-in-liquid" does not at all times tell the whole story. In the gold-in-water systems we have a *lyophobic* (solvent-hating) system in which *the disperse phase is insoluble in the dispersions medium and the dispersions medium does not dissolve in the disperse phase.* Thus, we may have systems of gold and water, sulfur and water, rubber and water, sodium chloride and alcohol, etc. This type of system is of rather general importance in biochemical studies, and certain of its characteristics will be considered in greater detail in the following pages.

On the other hand, when gelatin is dispersed in water a *lyophilic* (solvent-loving) system results. This is defined by Martin Fischer as *"a system in which the disperse phase and the dispersions medium are mutually more or less soluble one in the other."* This indicates that *hydration* (when water is the dispersions medium) or *solvation* (a term which applies to any dispersions medium) takes place. Gelatin swells in water; some of the gelatin disperses in the water, and a considerable amount of the water "dissolves" in the gelatin. The gelatin is *hydrated.* Gelatin in alcohol or benzine is a lyophobic system. Rubber in alcohol or benzine is a lyophilic system, in water a lyophobic system. In biological problems the lyophilic colloids constitute by far the most important class.

Solid-in-Gas. In this system we are dealing with such phenomena as the smoke-vapors of ammonia and hydrochloric acid, the "blue haze" of the forest fire areas that hangs for weeks in the air without settling, and the "blue" smoke of the cigarette. The darker colored smokes usually contain the more coarsely divided suspensions of carbon. When these larger particles have settled out, the blue colloidal haze still persists and forms a very stable system. From the military standpoint such systems constitute a rather important field of study involving smoke screens, etc., for the reason that one desires the maximum density of smoke which can be formed from the minimum amount of original material, the degree of subdivision thus playing a very important role. Bacterial and fungus spores present in spray from coughing and sneezing may be considered examples of solid-in-gas systems which are of importance from the biological and medical standpoint.

Liquid-in-Solid. The principal examples of this system are found in minerals and gems. The opal is a system of silicon dioxide and water, the pearl a system of calcium carbonate and water. An opal rather readily loses its "fire" and the pearl its "life" or luster if kept for a con-

siderable period of time in a dry atmosphere. Instances have been known where valuable pearls placed for years in a safe-deposit box have been rendered practically valueless. Both the opal and the pearl are most beautiful when worn often near the skin, *i.e.*, in a region of fairly high humidity which tends to keep the equilibrium amount of water in the gem.

Liquid-in-Liquid. This class may be called the *emulsions.* An emulsion, in its simplest form, may be defined as a mixture of two mutually insoluble liquids. It is obvious from this definition that *two different types of emulsions are possible depending on which liquid forms the disperse phase and which liquid forms the dispersions medium.* When water is one of the components we distinguish these two forms of emulsions by oil-in-water or water-in-oil emulsions. Emulsions met with in biological phenomena almost invariably have water as one of the phases. In many instances the dispersions medium contains a lyophilic colloid to assist in stabilizing the emulsion.

Emulsions are not always truly colloidal systems, if we limit our definition to the size of the particle concerned. Very dilute emulsions can be classified as liquid-in-liquid systems, providing all the particles fall within the range of diameters characteristic of colloidal systems. Most emulsions, however, contain oil or water droplets very much larger than even the 0.5 μ which forms the upper limit of the truly colloidal field. Mayonnaise, milk, and egg yolk are typical examples of emulsions. In each of the examples we are dealing not only with oil droplets suspended in an aqueous medium, but we also have present at the same time lyophilic colloids in the form of proteins which stabilize the emulsion, so that a study of emulsions involves not only a study of the dispersions medium and the disperse phase, but likewise, and probably of greater importance, an investigation of the nature and properties of the lyophilic colloid which acts as the stabilizer.

Protoplasm has been spoken of as an emulsion. Certainly there are droplets of fats and oils in living protoplasm which are stabilized by the lyophilic colloids which are present. It is equally true that there are solid particles in living protoplasm, so that protoplasm can be looked upon as a complex colloid system, the dispersions medium being water and the disperse phases consisting of lyophilic colloids, lyophobic colloids, and microscopically visible fat droplets in the form of an emulsion, as well as other microscopically visible particles.

Liquid-in-Gas. *Fogs, Mists.* This class is important from the standpoint of meteorology. In fogs and mists we sometimes, but certainly not always, have to deal with a solid-in-gas, with water particles condensed on the solid surfaces, inasmuch as dust particles cause mist

formation at lower degrees of supersaturation than would otherwise be effective.

Owens [11] studied conditions which bring about fog in the London area. He finds that on a clear day there is approximately 1 mg. of solid material per cubic meter of air. In a dense fog the amount of solid material rises to the neighborhood of 5 mg. of solid material per cubic meter of air. This does not seem a large amount, but over the London area it amounts to 193 tons of solid material on the 120 square miles of area to a height of 122 meters. The size of the solid particles varies from 0.00013 to 0.00026 mm. in diameter. The water film condensed on the surface of these solid particles may be as great as 0.0014 mm. in thickness. Owens notes that at 6 A.M. the air above London may be perfectly clear and at 9 A.M. there may be a dense fog, and he ascribes the onset of the fog to the smoke rising from the fires of the homes and factories. In the *Ninth Report on Atmospheric Pollution* [12] it is pointed out that on a clear day there are approximately 100 particles of solid per cubic centimeter of air, whereas in a dense fog there may be as many as 80,000 particles. Incidentally it may be noted that 10,000 particles of solid per cubic centimeter are equivalent to a weight of 1 mg. per cubic meter. The diameters of the particles range from 1.5 μ down to ultramicroscope size.

Gas-in-Solids. From the biological standpoint this field is relatively unimportant. From the standpoint of mineralogy, metallurgy, and industrial chemistry it offers a fruitful field of research. The form in which gas is present in solids is still in many instances a debatable question. In some minerals gas probably occurs in the form of very finely divided bubbles, the bubbles perhaps being of colloidal dimensions. In other instances the gas is probably adsorbed or "fixed" on the surface of the solid material. Many solid materials possess the property of condensing or adsorbing gas upon their surfaces, and a great variety of industrial applications have been based upon this phenomenon. Possibly such systems could be logically spoken of as gas-in-solid systems. However, this is not the place for an extended discussion of such systems, which will be reserved for consideration in Chapter 7 under adsorption.

Gas-in-Liquid. Such systems are generally spoken of as foams or systems in which gas bubbles are finely divided and suspended in liquids. The liquid in which the gas bubbles are suspended is usually not a pure

[11] Owens, *Seventh Report of the Committee for the Investigation of Atmospheric Pollution*, No. 249, Meteorological Office, Air Ministry, London, 1922. (See *Science*, **55**, 596, 1922.)

[12] *Ninth Report of the Committee for the Investigation of Atmospheric Pollution* (Report on Observations in the year ended March 31, 1923), Meteorological Office, London, 1924. Abstracted in *Science*, **60**, 136 (1924).

material but is more often already a colloidal system containing lyophilic colloids. Typical foams are produced when egg white is beaten or when cream is whipped, the lyophilic colloids which are present serving to stabilize the resulting foam. In many instances foams which are extremely difficult to break develop in industrial processes. Accordingly a study of such systems is necessary in order to devise methods for the destruction of foams at points where they are undesirable.

Gas-in-Gas. All the foregoing eight classes of colloidal systems have been realized experimentally, and all occur in nature. The ninth hypothetical system, gas-in-gas, is theoretically impossible to attain, inasmuch as gas is always regarded as occurring in the molecular state, so that large aggregates of molecules are theoretically impossible, for no known gas molecule is of colloidal dimensions. Possibly if someone were to work with mixtures of gases at their critical state, this system might be experimentally realized. It is doubtful, however, whether it would be of any technical importance. Certainly it would have no biological significance.

Nomenclature. It is necessary at this point to consider a few general terms of special nomenclature.

Graham designated liquid colloidal systems as *sols*. A sol may be defined as a colloidal system which to the eye appears as a solution, *i.e.*, it is fluid and appears to be homogeneous. It can be more or less readily poured from one beaker to another, but it differs from a solution in that the sizes of particles suspended in the liquid are of colloidal dimensions. Occasionally in speaking of sols, it is desirable to define the dispersions medium. Thus, following Graham's terminology, we have hydrosols where water is the dispersions medium, alcosols, benzosols, etc., with organosols a special term limited to organic solvents as a dispersions medium.

Graham designated the more or less rigid colloidal systems as *gels*. A gel may be defined as a colloidal system possessing more or less the properties of a solid. Gels logically fall into two general subdivisions.

1. The true gels or "jellies," as some prefer to call them, are formed by the characteristically lyophilic colloids and may be represented by such familiar materials as gelatin, thick custards, the ordinary fruit jellies and jams of the household, muscle tissue, heat-coagulated egg white; from the biological standpoint they constitute the important group of gels.

2. Another subdivision, sometimes referred to as gels, is the precipitates which are formed by the coagulation of lyophobic colloid systems. These are preferably referred to as *coagula* rather than as gels, although

there is no general agreement in regard to this distinction in nomenclature. When a gold hydrosol is precipitated by the addition of an electrolyte, a purplish precipitate is thrown down. This precipitate is relatively bulky, and, when this precipitate is ignited, it is often found that an almost unweighable amount of gold is actually present. The precipitate contains the gold particles originally dispersed in the form of a sol, and adhering to these coagulated particles is a relatively large amount of the dispersions medium (water). Inasmuch as such a precipitate does not show, in general, properties similar to those exhibited by the "jellies," it seems preferable to differentiate between the two types and to refer to the precipitates formed in lyophobic systems as *coagula.*

In the gels as in the sols, we use such special terminology as hydrogels, alcogels, benzogels, organogels, etc.

The terms lyophilic and lyophobic have already been defined; they are general terms referring to the affinity or lack of affinity of the disperse phase for the dispersions medium. A lyophilic system in general produces a gel which contains a high percentage of the dispersions medium. A lyophobic system on the other hand yields a coagulum which contains a relatively small amount of the dispersions medium. A few exceptions to the above rules (for example, dibenzoyl-L-cystine gel) will be taken up in Chapter 9.

CHAPTER 2

Methods of Preparation

The methods of preparation of colloidal systems are referred to in the literature under two separate systems of nomenclature. We have the classification due to Svedberg, which is used by Wo. Ostwald.

	Crystals or Matter in Mass	Colloid Systems	Condensation Methods ← True Solution
Dispersion Methods →			

Or we have the classification due to Von Weimarn.

	Crystals or Matter in Mass	Colloid Systems	Crystallization Methods ← True Solution
Solution Methods →			

These classifications are essentially identical, the significant feature of both being to illustrate the fact that there is a continuous gradation from matter in mass to true solutions. Every crystal which grows in a true solution has at one time possessed dimensions which would bring it within the colloid realm. Likewise every crystal which dissolves to form a true solution must eventually yield a particle of colloidal size before it completely disappears into true solution. Either Svedberg's or Von Weimarn's classification is satisfactory.

To prepare a stable colloidal sol one must stop when the particles are within the colloidal realm. It is also reasonably desirable to have fairly high concentration of the disperse phase in the dispersions medium.

When condensation methods are employed, the stability and the concentration of the resulting system depend upon two factors, (*a*) the velocity of the formation of nuclei, and (*b*) the rate of crystal growth. Von Weimarn has expressed these two phenomena in some rather general equations.

$$W = K \frac{\text{Condensation pressure}}{\text{Condensation resistance}} \tag{1}$$

where W is the velocity of nuclei formation and K is a constant.

If C is the total amount in solution and L is the solubility, $C - L$ is the condensation pressure or the supersaturation, and L, the solubility, is the condensation resistance. The equation then becomes

$$W = K\frac{C - L}{L} \tag{2}$$

This equation means that there is a certain tendency for the material to remain in a supersaturated solution expressed by condensation resistance and that the velocity is dependent on the ratio of these two factors to each other and to a constant. It is obvious that, the more nuclei formed, the greater will be the surface area of the disperse phase, and likewise the smaller will be the diameter of the resulting particles per unit weight of material. To form a stable colloidal system it is, therefore, desirable to have as great a number of nuclei formed as is practicable.

Once the nuclei are formed, they grow according to the Noyes-Nernst equation for crystal growth:

$$V = \frac{\Delta}{l}S(C - L) \tag{3}$$

where V = velocity of growth
Δ = diffusion coefficient
l = length of diffusion path
S = surface area of disperse phase
C = concentration of solution
L = solubility of disperse phase of a given size
$C - L$ = absolute supersaturation.

As an illustration let us consider silver chloride and sodium chloride; for sodium chloride $C - L$, the supersaturation, may be relatively great, and $(C - L)/L$ or the excess of sodium chloride in a solution of the very soluble sodium chloride will make for a low velocity of nuclei formation and for the formation of relatively large crystals of sodium chloride. Accordingly a hydrosol of sodium chloride is not attainable. With the same values for $C - L$ for silver chloride we get at once a dense, curdy precipitate (a gel or coagulum) of silver chloride. For sodium chloride L is large, and slow crystallization results. For silver chloride L is extremely small, and instant precipitation occurs. If instead of using an aqueous medium we form the sodium chloride by the interaction of sodium ethylate dissolved in absolute alcohol and hydrogen chloride dissolved in absolute alcohol, we will form either a curdy precipitate, similar to the precipitate of silver chloride, or a stable colloidal system of a sodium chloride alcosol, depending on the concentra-

tions of the materials which were used. *The smaller the solubility of a substance in a solvent, the easier it is to prepare a colloidal system, and there is a maximum solubility above which a stable colloidal system, at least for lyophobic colloids, is impossible.*

Adding nuclei on which crystal growth will take place has been suggested.[1] In the instance cited a small drop of a sol of potassium-zirconium sulfate was added to the solution under test in order to induce crystallization and turbidity sufficient to measure. By this modification the sensitivity of the zirconium sulfate reagent for potassium was increased from 0.48 mg. potassium to 0.32 mg. potassium per 2 ml. of reaction mixture.

The concentrations of the solutions determine in a large measure the nature of the system which will result. If one deals with concentrated solutions of very soluble substances which form a very insoluble precipitate, a colloidal gel is likely to result. Intermediate concentrations of the same materials may yield granular or crystalline precipitates which can be easily filtered off and which show no typical colloidal properties. Very dilute solutions of the same materials when mixed may yield an excellent colloidal sol. Figure 2 illustrates these generalizations in a diagrammatic way.

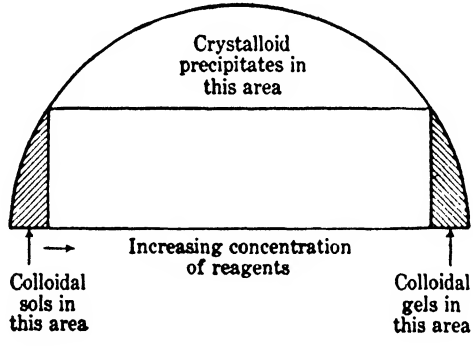


FIG. 2. A diagrammatic representation of the role of reagent concentration on the properties of the resulting systems.

Typical examples are laboratory experiments in which ferric chloride and potassium ferrocyanide are used to form Prussian blue. If one employs equimolecular concentrated solutions and mixes these solutions practically instantaneously, the resulting Prussian blue forms a dense gel, and if a small portion of this gel is stirred into a large volume of pure water, a stable colloidal sol of Prussian blue results. Such a sol will readily pass unchanged through a filter paper. If now one dilutes these

[1] R. D. Reed and J. R. Withrow, *J. Am. Chem. Soc.*, **52**, 2666 (1930).

concentrated solutions eight or ten times and again mixes them, a precipitate of Prussian blue is formed which can be readily filtered off. If one further dilutes the initial solutions to have only a few milligrams of Prussian blue formed per 100 ml. of the resulting mixture, one obtains a stable sol of Prussian blue which again will readily pass through a filter paper. Accordingly if one wishes to filter off a precipitate of Prussian blue, the initial concentration of the reagents which are employed may be a determining factor in the complete retention of the precipitate on the filter. In the colorimetric estimation of hydrocyanic acid, it is essential that all the Prussian blue remain in the form of a colloidal sol. In such cases the initial concentration of the hydrocyanic acid solution must not exceed a certain value; the limit is probably near 4 mg. of hydrocyanic acid per 100 ml. Under such conditions the Prussian blue sol will be stable, and its depth of color can be accurately measured in a colorimeter. If much greater amounts of hydrocyanic acid are present, a granular or crystalline precipitate is likely to form, and such a precipitate will settle rather readily and give low results when the color is measured.

Similar experiments can be conducted by using ferric chloride and sodium arsenate or ferric chloride and sodium phosphate as the reacting materials. A series of beakers arranged with progressive dilutions of the reagents will show at one end of the series clear and apparently homogeneous gels and at the other end transparent, apparently homogeneous sols, and the intervening dilutions will show more or less complete precipitation grading in properties from one end of the series to the other. It is obvious that such phenomena are of extreme importance to the analytical chemist, and many of the instances where the precipitate "goes through the filter paper" can be avoided by choosing the proper concentration prior to making the precipitation.

Before we consider the various methods by which colloidal sols may be formed, it is necessary to note one additional property characteristic of stable colloidal systems. *The micelles present in colloidal systems possess either a positive or a negative electric charge.* The electric charge may arise from several causes such as the *direct ionization* of the material comprising the micelle, *the capture of an ion by the micelle* (adsorption, see Chapter 7), in which case the micelle attracts to it an ion of an electrolyte and this, becoming fixed on the surface, gives rise to either a positive or a negative charge, or perhaps in rare instances *electrification by contact* with the dispersions medium, in the same way that a wax rod becomes charged when rubbed with a woolen cloth.

Most hydrosols contain negatively charged micelles, whereas sols in which turpentine is the dispersions medium ordinarily contain posi-

tively charged micelles. The reversal of charge when we pass from water to turpentine has been ascribed to a change in the dielectric constant of the dispersions medium and has given rise to the general rule that *micelles are usually negative when in contact with a liquid having a high dielectric constant, and positive when in contact with a liquid having a low dielectric constant.* The dielectric constant of water is 78.5, which is very high compared to that of most liquids. Hydrogen peroxide (46 per cent) is somewhat higher. Certain of the dielectric constants of common liquids are shown in Table 1.

TABLE 1. DIELECTRIC CONSTANTS OF SOME COMMON SUBSTANCES

Substance	*Dielectric Constant*	*Temperature, °C.*
Vacuum	Unity by definition	
Air	1.00	..
Water	78.54	25
Deuterium oxide	78.25	25
Nitrobenzene	35.8	20
Methanol	32.4	20
Ethanol	25.0	20
n-Propanol	20.81	20
Acetone	19.6	20
Isopropanol	18.62	20
n-Butanol	17.0	20
Turpentine	2.2	20
Petroleum oil	2.1	20
Benzene	2.28	20

As a general rule only the oxides and hydroxides of metals and basic organic compounds are positively charged in water.[2] In these instances we are probably dealing with either the phenomena of ionization or of the adsorption of an ion. Thus, silver hydroxide or ferric hydroxide might be regarded as ionizing according to the scheme

$$AgOH \rightarrow Ag^+ \text{ and } OH^-$$

$$[Fe(OH)_3]_x \rightarrow [Fe(OH)_2]_x{}^+ \text{ and } OH^-$$

Or we might postulate that the ferric hydroxide micelle still contained a small amount of ferric chloride, in which case we could have an ionization as follows:

$$[Fe(OH)_3]_x(FeCl_3)_y \rightarrow [Fe(OH)_3]_x(FeCl_2)_y{}^+ \text{ and } Cl^-$$

In each one of these instances we have a negative ion given off into the solution with a corresponding residual positive charge on the micelle.

[2] There are some exceptions to this rule, but as a generalization it is fairly satisfactory.

Other discussions of the electrical charge will be given in greater detail in the following pages. It may not be amiss, however, to discuss briefly the meaning of the term *dielectric constant.*

Hildebrand [3] states, "Suppose that we have two metal plates that can be charged electrically, one positive, the other negative. If this is done in a vacuum, there is a definite amount of electricity stored up on applying a given electric potential; however, if the space between the plates is filled with a given liquid, a larger amount of electricity is absorbed before the same potential is reached. *The ratio of the amount of electricity for this liquid to the amount in a vacuum is called its dielectric constant.*" Hildebrand then goes on to consider how this extra charge can be absorbed and indicates that there are three measurements which may be involved: (1) the electrons within the molecules may be displaced in their relative position to the positive atomic nuclei; (2) the molecules may be stretched, twisted, or bent by the displacement of the atoms within them; and (3) the molecules may become oriented if they already possess permanent electron displacements or dipoles, and in this orientation the positive portion of the molecule may be brought closer to the negative plate, and *vice versa.*

Table 2, from Hildebrand's paper, distributes the total polarization of four chemical compounds between these three measurements. Car-

TABLE 2. THE CONTRIBUTION OF ELECTRON DISPLACEMENT, ATOMIC DISPLACEMENT, AND MOLECULAR ORIENTATION TO THE MOLAL POLARIZATION OF CERTAIN CHEMICAL COMPOUNDS

Compound	*Electron*	*Atomic*	*Orientation*
HCN	6	7	140
HCl	8	1.2	22
HI	15	0.7	3
CCl_4	28	3	0

bon tetrachloride, being completely symmetrical (as are benzene, methane, etc.), has no dipole moment and shows no effect of orientation. The magnitude of the dipole moment may be taken as a measure of the molecular dissymmetry. Smyth [4] lists the dipole moments of many organic compounds. Since the energy of an electric field is directly proportional to charge and inversely proportional to distance separating charges, on the assumption that a single electron is involved in the dipole moment of an organic molecule, then the charge e would be 4.77×10^{-10} esu. If we take the diameter of a molecule to be approximately $1 \times$

[3] J. H. Hildebrand, *Science*, **80**, 125 (1934).

[4] C. P. Smyth, *Dielectric Constant and Molecular Structure*, American Chemical Society Monograph 55, Chemical Catalog Co., New York, 1931.

10^{-8} cm., the dipole moment of such a molecule will approach 4.77×10^{-18} as an upper limiting value. Most organic molecules have dipole moments somewhat less than this value. They may be thought of as bar magnets, a negative charge residing in one end of the molecule and a corresponding positive charge at distance δ toward the other end. When we later discuss molecular orientation we shall make extensive use of this electrical description of molecules.

PRACTICAL METHODS FOR THE PREPARATION OF COLLOIDAL SYSTEMS

As already noted at the beginning of this chapter, the methods of preparation may be divided into crystallization or condensation methods, where we pass from a true solution to colloidal systems; or solution or dispersion methods, where we pass from matter in mass to colloidal micelles.

Crystallization or Condensation Methods. These methods may be divided into six general groups.

The Cooling of a Liquid. Typical examples are water in ether, cooled by carbon dioxide snow, giving rise to a colloidal sol of ice-ether. Referring to equations (2) and (3), we can see that the lowering of the temperature has (*a*) reduced the solubility of the water in the ether, so that the supersaturation $C - L$ is much greater, and likewise (*b*) it has reduced the diffusion coefficient Δ of the Noyes-Nernst equation and increased the viscosity which will have the same effect as increasing the length of the diffusion path l. We have thus made for the formation of a greater number of nuclei and have a much slower rate of crystal growth. Accordingly an ice-ether sol is relatively stable at a very low temperature. Other examples can be cited with the same general argument.

Replacement of Solvent. This is a very common method for the formation of colloidal systems, and the success of the method depends again on factors noted in equations (2) and (3). If we have two solvents, one more or less soluble in the other, one liquid being an excellent solvent for some material which is at the same time practically insoluble in the other liquid, we have conditions where the replacement-of-solvent method can be employed. Thus, sulfur is relatively soluble in carbon disulfide but extremely insoluble in water. If a few drops of a solution of sulfur in carbon disulfide is rapidly stirred into a large volume of water, a beautiful hydrosol of sulfur results. We have here suddenly decreased L and increased the initial values of $C - L$, so that many nuclei are formed. The final value of L is so extremely low that almost

instantaneously the initial supersaturation $C - L$ falls to practically zero, owing to the formation of nuclei. Accordingly when crystal growth takes place $C - L$ is essentially zero, and therefore the sol is stable for a long period of time.

Sulfur hydrosols and colloidal sulfur are becoming of increasing importance as fungicides. In some instances the sulfur hydrosols are used as sprays; in others the colloidal sulfur has been reduced to dry form, either as the pure sulfur itself or as colloidal sulfur associated with carrier material, and is used as a dust.

A very common method for the purification of such materials as gums, proteins, and enzymes is to take advantage of their insolubility in alcohol in order to precipitate them in a form which can be filtered. If one is dealing with a water-soluble protein, it is possible to concentrate it by vacuum evaporation down to a rather viscous sol. If one pours this sol into absolute alcohol, the protein is, as a rule, precipitated in the form of flocs which can be readily gathered and in many instances even filtered off. As the purification progresses, however, and the above process of alcohol precipitation is repeated several times, it frequently happens that no precipitate is formed when the aqueous sol is poured into the alcohol, but instead we have only a limpid, opalescent sol resulting. Then arises the problem of "breaking" such a sol to cause the formation of the desired precipitate. This can usually be effected by the addition of a small amount of electrolyte, such as a drop of saturated sodium chloride or preferably lithium chloride. Some sols of this sort are very stable and require a considerable amount of experimentation before the dispersed phase is precipitated.

The beginner in biochemical preparations is probably less likely to lose valuable material when a complete sol results, in an attempted alcohol precipitation, than he is when only a partial sol is formed. It very often happens that part of the material precipitates in the expected manner, while a greater or a smaller portion remains in the form of a slightly opalescent alcosol. Under such conditions the worker is very likely to decant and discard the supernatant liquor, retaining only the precipitated portion. Many instances have been known in which 90 per cent of valuable material went down the drain as the result of the decantation of supernatant sols.

Reduction Methods. These methods depend on the reduction of a soluble metallic salt to the insoluble metal. In this group of methods in particular the purity of the water is all-important. Distilled water which has stood for several hours in an ordinary soft-glass bottle usually contains sufficient electrolytes to render the production of the desired sol uncertain. Likewise the purity of the reagents employed becomes a

very important factor. Typical examples of colloidal systems formed by the reduction methods are the reduction of auric chloride by formaldehyde, phosphorus, phenylhydrazine, or tannin, as noted by Morrow and Sandstrom [5] or by Holmes.[6] Similarly a beautiful silver sol can be prepared by reduction with hydrogen. If a stream of hydrogen is passed into a hot solution of freshly precipitated silver hydroxide, reduction takes place, and an intensely yellow, essentially electrolyte-free sol of silver results. The reaction which takes place can be expressed as

$$2AgOH + H_2 \rightarrow 2Ag + 2H_2O$$

The sols that are formed by reduction methods are, as a rule, very sensitive to the action of electrolytes and form, from the standpoint of laboratory experimentation, one of the most interesting groups of colloidal systems.

Oxidation Methods. Relatively few examples of colloidal systems formed by the use of oxidation processes are recorded. The outstanding example perhaps is the oxidation of hydrogen sulfide and the resulting formation of a sulfur hydrosol.

$$2H_2S + O_2 \rightarrow 2S + 2H_2O$$

This sol should be familiar to everyone who has had any laboratory chemistry. Those who are not interested in colloid phenomena are usually exasperated when they return to the laboratory and find that their solution of hydrogen sulfide has turned to a milky appearing liquid and the active hydrogen sulfide has disappeared from the solution. If the bottle in which the hydrogen sulfide solution was stored had been completely filled with the solution and tightly stoppered, this oxidation would not have occurred, but under ordinary laboratory conditions it is an inevitable occurrence. A later stage is the coagulation of the sulfur hydrosol and the precipitation of crystalline sulfur in the bottom of the bottle. Precipitation, however, is rarely complete, and the supernatant liquid over such a sulfur precipitate is usually a dilute sulfur hydrosol. Selenium and tellurium form sols similar to the sulfur hydrosols.

Hydrolysis Methods. Many materials which are readily soluble in water can be made to undergo hydrolysis with the production of an extremely insoluble residue. In such instances appropriate technic will

[5] C. A. Morrow and W. M. Sandstrom, *Biochemical Laboratory Methods for Students of the Biological Sciences*, 2nd ed., pp. 1–5, John Wiley & Sons, New York, 1935 (out of print).

[6] H. N. Holmes, *Laboratory Manual of Colloid Chemistry*, 3rd ed., p. 33, John Wiley & Sons, New York, 1934 (out of print).

result in the formation of stable sols. Ferric chloride may be chosen as possibly the best example of such a material. The reaction is usually written

$$FeCl_3 + 3H_2O \rightarrow Fe(OH)_3 + 3HCl$$

Experiments based on this reaction are given by Morrow and Sandstrom [7] and by Holmes.[8] The above reaction is somewhat misleading in that it indicates the formation of micelles composed entirely of ferric hydroxide. Weiser and Milligan [9] have shown that those sols which we have normally designated as ferric hydroxide hydrosol and aluminum hydroxide hydrosol are in reality hydrous oxide sols, since the solid phase shows the x-ray diffraction patterns of the oxides. Weiser and Milligan suggest that the composition of the solid phase can best be represented by $[xFe_2O_3 \cdot yHCl \cdot zH_2O]Fe_m^{+++}H_n^{+}$, and that these micelles are in ionic equilibrium with chloride ions in the surrounding liquid. We have known for a long time that the micelles in these sols are not pure ferric hydroxide, but it has been believed that they are a mixture of ferric hydroxide and ferric chloride, present in varying ratios. Taylor [10] discusses them under the title, "The 'Complex' Theory of Colloids," and points out that pure ferric "hydroxide" sols, or as a matter of fact most of the lyophobic sols, are very unstable when they are entirely free from admixture with stabilizing ions. Weiser and Milligan's formula indicates that the solid phase is composed of hydrous ferric oxide, ferric chloride, and hydrochloric acid, and that it possesses a positive charge due to the ionization of the residual ferric chloride and hydrochloric acid. Browne [11] has made studies of the hydrogen-ion concentration and the chloride-ion concentration in such sols and finds that the chloride-ion concentration is more or less proportional to the stability of the sol. Sorum [12] reports a ferric oxide hydrosol which is chloride-free. Weiser and Milligan suggest that this must contain some other electrolyte or some protective material in order to account for its stability.

Hydrolysis methods are applicable to the formation of many hydrosols of the metallic oxides and hydroxides.

[7] C. A. Morrow and W. M. Standstrom, *Biochemical Laboratory Methods for Students of the Biological Sciences*, 2nd ed., pp. 5–6, John Wiley & Sons, New York, 1935 (out of print).

[8] H. N. Holmes, *Laboratory Manual of Colloid Chemistry*, 3rd ed., p. 37, John Wiley & Sons, New York, 1934 (out of print).

[9] H. B. Weiser and W. O. Milligan, *J. Phys. Chem.*, **40**, 1 (1936); *Colloid Symposium Monograph* (1935), p. 1.

[10] W. W. Taylor, *The Chemistry of Colloids*, p. 108, Longmans, Green and Co., New York, 1915.

[11] F. L. Browne, *J. Am. Chem. Soc.*, **45**, 297 (1923).

[12] C. H. Sorum, *J. Am. Chem. Soc.*, **50**, 1263 (1928).

Double Decomposition or Precipitation Methods. In many instances the product of the interaction of two soluble materials is an extremely insoluble precipitate. By proper manipulation the precipitate can usually be obtained in the form of a sol. Typical examples are hydrosols of barium sulfate or arsenous sulfide.

$$BaCl_2 + H_2SO_4 \rightarrow BaSO_4 + 2HCl$$

$$As_2O_3 + 3H_2S \rightarrow As_2S_3 + 3H_2O$$

Probably every analytical chemist has had the undesirable experience of a barium sulfate precipitate going through the filter paper. The conditions of precipitation were such that a more or less stable hydrosol was formed. Such conditions are usually brought about when there is a nearly exact equivalence of barium chloride and sulfuric acid and when the resulting concentration of hydrochloric acid is very low. Both precipitation at the boiling temperature and the general practice of allowing a barium sulfate precipitate to stand for a number of hours before filtering favor the growth of crystals and the subsequent increase in particle size, so that the precipitate is retained on the filter paper. After formation of the nuclei, the elevated temperature increases L, the solubility of the disperse phase, and $C - L$, the absolute supersaturation, at the same time increasing Δ, the diffusion coefficient, and decreasing viscosity which will likewise increase the rate of diffusion. The length of the diffusion path l will be decreased, owing to the convection currents and stirring at the higher temperature, so that it may become essentially negligible.

Trimble [13] points out that crystal growth ceases to be a factor when the particles of barium sulfate are larger than 2 μ apparent diameter, but that aggregation and cementing together of the particles is an important factor in the retention of the precipitate on the filter paper.

Kolthoff [14] and his students have been very active in investigating the changes which take place after the initial formation of a precipitate. Space in the present volume is too limited to permit a discussion of the findings, but the student of analytical chemistry should consult these papers, inasmuch as they have a profound bearing on the purity of analytical precipitates.

In the arsenous sulfide sols, no precipitating ion is formed by the reaction, and extremely stable sols can accordingly be prepared. Here again we are probably not dealing with pure As_2S_3 micelles, but ap-

[13] H. M. Trimble, *J. Phys. Chem.*, **31**, 601 (1927).

[14] I. M. Kolthoff, *et al.*, *J. Am. Chem. Soc.*, **56**, 1264, 1658 (1934); **57**, 597, 607, 2573, 2577 (1935); **58**, 116, 121 (1936); *J. Phys. Chem.*, **36**, 549, 860 (1932); **37**, 153, 443, 459, 723 (1933).

parently have a complex ion of arsenous sulfide stabilized by the sulfhydryl (SH^-) ion, owing to the presence of a slight excess of hydrogen sulfide, so that the micelle actually possesses the formula

$$(As_2S_3)_x SH^-$$

It is a common experience for the analytical chemist, in precipitating the sulfides of arsenic, antimony, tin, silver, etc., to find that they occasionally are not retained on a filter paper. In such instances hydrogen sulfide has been passed into the solution for too long a period of time, and the initial precipitate of sulfide has become stabilized by the adsorption of some of the excess hydrogen sulfide with a resulting hydrosol formation.

Solution or Dispersion Methods. These methods may be divided into three general groups: electrical dispersion, ultrasonic dispersion, and peptization.

Electrical Dispersion. Bredig,[15] in 1898, announced a general method for the preparation of metallic sols by the process of electrical pulverization. When an arc is passed between two electrodes, one of the electrodes becomes corroded. Bredig made use of this phenomenon by striking an arc between metallic electrodes inserted under the surface of a liquid. Figure 3 is a diagrammatic representation of such an apparatus. Using thick wires of gold, platinum, silver, etc., and a direct current of 4 to 10 amperes at 35 to 50 volts, he was able to prepare metallic sols not only in water as a dispersions medium but also in certain of the organic solvents, although the latter offer difficulties, owing to the separation of colloidal carbon by the decomposition of the organic liquid in the heat of the electric arc.

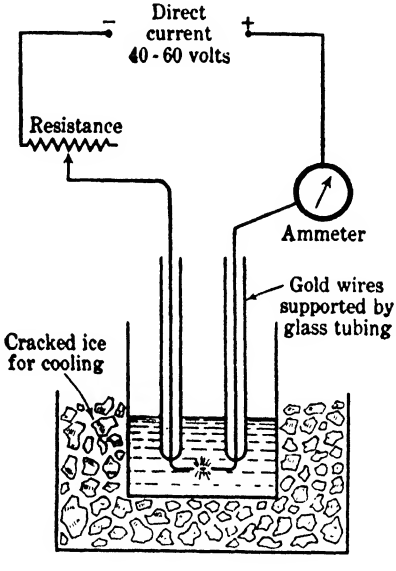


FIG. 3. A diagrammatic illustration of the Bredig method of preparing metallic sols by electrical pulverization.

When the points of two gold wires are momentarily brought together beneath the surface of the water and then separated, an arc is struck between the points. The tip of the cathode becomes molten, and a stream of fine gold particles is shot from the molten cathode toward the anode. In the heat of the arc and under the influence of the electric

[15] G. Bredig, *Z. angew. Chem.*, 951 (1898).

current some of the smaller of these particles may vaporize or become resubdivided; in either event part of the gold fails to reach the anode, instead being dispersed in a colloidal cloud which rises through the liquid from the vicinity of the arc with the appearance of a rising cloud of smoke or dust. The gold particles which do reach the anode fuse on its tip, so that the wire forming the anode lengthens as the wire forming the cathode shortens. A certain amount of coarse, granular gold dust, derived from larger particles of gold which failed to reach the anode, is deposited in the bottom of the beaker. It is preferable to bring the electrodes in contact and later to separate them to strike the arc, with a mechanical device of springs and set screws, rather than to attempt to operate the electrodes by hand. Kraemer and Svedberg [16] modified Bredig's original method, making use of a high-frequency alternating-current instead of a direct-current arc, and by use of this modified technic prepared metallic organosols containing minimal amounts of carbon. Svedberg [17,18] had earlier used an induction coil for the preparation of similar sols.

Both vaporized metal and finely divided molten metal contribute to the Bredig sols. Such sols contain a rather wide range of particle size, the sizes varying from particles which gradually settle out to particles near the lower limit of the colloid realm. The range in particle size in sols obtained by the Bredig method is much greater than it is in metallic sols obtained by the reduction methods, and as a general rule it is easier to obtain sols with small-sized particles (10–25 mμ) by reduction methods than it is to obtain similar sols by the Bredig method. The noble metals rather readily yield sols by the Bredig method. With other metals as electrodes the sols obtained are usually contaminated with larger or smaller amounts of oxides and hydroxides.

Bredig's method offers an extremely valuable technic for fundamental studies of the properties of colloid systems, since sols obtained from pure metallic electrodes and conductivity water contain only metallic particles suspended in a pure liquid and as a result the influence of minute traces of electrolytes on colloid stability can be investigated. Burton [19] has made extensive observations of the behavior of metallic sols, particularly silver sols prepared by Bredig's method.

Ultrasonic Dispersion. The use of ultrasonic waves is a more recent method for bringing about colloidal dispersion. Wood and Loomis [20]

[16] E. O. Kraemer and The Svedberg, *J. Am. Chem. Soc.*, **46**, 1980 (1924).
[17] The Svedberg, *Ber.*, **38**, 3616 (1905).
[18] The Svedberg, *Ber.*, **39**, 1705 (1906).
[19] E. F. Burton, *Phil. Mag.*, [6] **11**, 425 (1906).
[20] R. W. Wood and A. L. Loomis, *Phil. Mag.*, [7] **4**, 417 (1927).

appear to be the first workers who utilized this technic. The effect of ultrasonic waves in the preparation of colloidal systems and the influence of ultrasonic waves on colloidal systems have been extensively investigated by Freundlich, Söllner, and others.

A piezoelectric quartz plate is set into vibration by a high-frequency current. The waves so produced are in reality sound waves vibrating above the limit of audibility in the frequency range of approximately 200,000 cycles per second. The disturbances set up in the liquid cause fragmentation of gross particles suspended in the liquid with the production of particles of colloidal size. Here again it is possible by this technic to prepare sols in the essential absence of electrolytes. Ultrasonic waves fragment bacteria and destroy animal life in the liquid subjected to the vibration.

Peptization. Peptization is undoubtedly the most important of all dispersion methods. Graham,[21] in 1864, coined the term designating this phenomenon. Graham says, "The solution of these colloids, in such circumstances, may be looked upon as analogous to the solution of insoluble organic colloids witnessed in animal digestion. . . . Liquid silicic acid may be represented as the 'peptone' of gelatinous silicic acid; and the liquefaction of the latter by a trace of alkali, may be spoken of as the peptization of the jelly."

The older statements were to the effect that the peptizer must (*a*) have one ion in common with the material to be dispersed, (*b*) be capable of forming a soluble compound with the material to be dispersed, or (*c*) have one ion which is very strongly adsorbed by the material being dispersed. Kruyt [22] has clarified these statements by pointing out the actual conditions which must be met. His viewpoint is that lattice forces are responsible for the inner coat of the double layer, the silver halides being a model example. With AgI, the I ions from KI fit into the open spots of the lattice surface, or Ag^+, if $AgNO_3$ is used as a peptizer. "Although every ion with an isomorphic relation to those in the lattice is able to peptize, no other ion whatever can accomplish this. Other ions may increase the potential of the double layer, but they are not able to form the original double layer." An AgS sol formed by peptization with H_2S is enormously stabilized by $K_4Fe(CN)_6$ by reason of an increased zeta potential, but $K_4Fe(CN)_6$ is not able to peptize HgS; only H_2S or SH^- can do this, since these ions can distribute themselves in the lattice as well as in the surrounding liquid medium.

We have already noted the "complex theory of colloids," using the

[21] Thomas Graham, *J. Chem. Soc.*, **17**, 318 (1864).

[22] H. R. Kruyt, "The Modern Development of Colloid Chemistry," *Act. IX Congr. intern. quim. pura y aplicada*, Madrid, April, 1934 (separate 12 pp.).

example of a ferric oxide hydrosol where the micelle probably has the formula of $(Fe_2O_3 \cdot nH_2O)_x(FeCl_2)_y^+$. In this case the ferric chloride can be regarded as the peptizing agent of the hydrous ferric oxide. If precipitated hydrous ferric oxide is suspended in water and heated to near the boiling point (the heat is merely to hasten the reaction) and a small amount of hydrochloric acid added, it will be found that within a short time the precipitated flocculent hydrous ferric oxide has dispersed to a transparent yellow-brown sol which will readily pass through a filter paper. The amount of hydrochloric acid necessary to bring about such a peptization may be less than 1 per cent of the amount necessary to convert into ferric chloride the hydrous ferric oxide present. The small amount of ferric chloride formed by this reaction acts as a peptizing agent on the flocculent hydrous ferric oxide, dispersing the flocs into smaller aggregates of colloidal demensions and stabilizing these aggregates, probably by adsorption of a small amount of the ferric chloride.

A similar sol can be made by the peptization of hydrous aluminum oxide with hydrochloric acid. Peptization can likewise be induced by adding ferric chloride to hydrous ferric oxide or aluminum chloride to hydrous aluminum oxide. Even ammonium hydroxide may act as a peptizing agent, and it is for this reason that texts in analytical chemistry stress the fact that the ammonia used in precipitating the R_2O_3 group of metals must be boiled off before the precipitate is filtered. Otherwise a certain amount of hydrous aluminum oxide sol may pass through the filter paper, and a loss of aluminum will result.

Hydrogen sulfide is a very efficient peptizing agent for sulfides. In the arsenous sulfide sols, the reaction probably is

$$As_2S_3 + H_2S \rightarrow (As_2S_3)SH^-$$

It should be stressed that such micelles as

$$(As_2S_3)SH^-, \quad (Fe_2O_3 \cdot nH_2O)_xFe^{+++} \quad \text{or} \quad (Fe_2O_3 \cdot nH_2O)_x(FeCl_2)_y^+$$

are not chemical compounds in the sense that they represent a pure, chemical individual. Their composition is not constant but varies continuously through a wide range, depending on the amount of peptizing agent adsorbed.

The analytical chemist who ignores the phenomena of peptization is very likely to secure results which are seriously in error, for many of the precipitates of gravimetric analyses are readily peptized when in a moist condition. Figure 4 illustrates diagrammatically some general phenomena applicable to the gravimetric determination of silver as silver bromide and to the peptization of precipitated silver bromide to form stable silver bromide sols. If exactly equivalent solutions of silver nitrate and

potassium bromide are mixed, an isoelectric precipitate of silver bromide will be formed, as shown at point C on the diagram. *This is the only point on the diagram where strictly pure silver bromide is found and where the weight of the precipitate accurately represents the amount of silver present in the solution.* The precipitates obtained at points A and E will be heavier than the precipitate obtained at C, and accordingly will give high results for silver and bromine respectively, whereas precipitates obtained anywhere along the curve AC will be light in weight and give low results for bromine, or along the curve CE will give low results for silver.

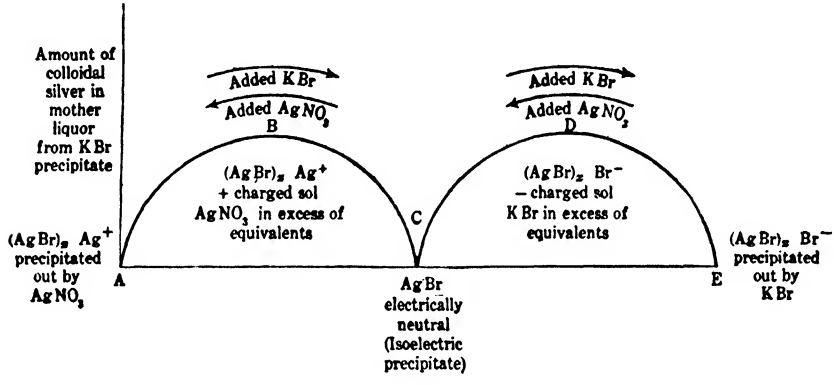


FIG. 4. A diagrammatic illustration of peptization, the influence of ions on the sign of the charge on the micelle, isoelectric precipitation, and "salting out" at high salt concentrations.

Anyone who is interested in a consideration of the nicety of technic necessary for the *accurate* determination of silver should read a paper dealing with the determination of atomic weights where silver is used as the reference standard. In such determinations extreme accuracy is necessary, and in such experiments the precipitation of isoelectric silver bromide or chloride is rigidly controlled.

If one adds an excess of potassium bromide to the precipitated silver bromide, the amount of silver in the mother liquor follows the curve CDE, the point reached depending on the amount of potassium bromide added in excess. Throughout the entire area under the curve CDE is formed a precipitate of silver bromide containing adsorbed bromide ions, and the supernatant liquid is a negatively charged silver bromide sol, probably of the composition $(AgBr)_xBr^-$. At point D on the curve we attain the greatest concentration of the negatively charged silver bromide micelles, a further addition of potassium bromide causing precipitation of the colloidal particles, so that finally at E there is no silver in the mother liquor, and a precipitate forms which appears to be silver

bromide but in reality is silver bromide containing a considerable excess of bromine, probably contaminated with potassium bromide.

A similar argument can be made for the region lying under curve *ABC* except that in this instance silver nitrate is added in excess. Throughout this area we have a positively charged sol in the mother liquor, the composition of the micelles being represented as $(AgBr)_xAg^+$. This sol reaches its maximum concentration at point *B* and is precipitated by the excess silver nitrate at point *A*, the precipitate having the composition noted on the diagram. The above example has been cited in detail because it is typical of the behavior of many precipitates encountered in the various fields of chemistry.

The change in the sign of the charge on the silver precipitate with excess halogen ions or excess silver ions has been made use of by Kolthoff, Lauer, and Sunde,[23] who used dichlorofluorescein as an indicator in the titration of chlorides with silver nitrate. As long as the precipitate is negatively charged (an excess of Cl^- is present), there is no adsorption of the dyestuff by the precipitate. The instant that there is an excess of Ag^+ in the solution, the colorless silver chloride becomes positively charged and adsorbs the anion of the dyestuff, the colorless precipitate instantly becoming an intense red. The advantage of this indicator over the usual chromate indicator is that it can be used even in solutions which are decidedly acidic.

The peptization of matter in mass may be brought about in many ways, chemically, mechanically, and electrically. In each instance, however, *energy must be added to the system.* As we shall see when we consider surface phenomena (Chapter 7), a given mass of material dispersed in small micelles has a much greater energy content than the same mass of material when present in large aggregates. This energy must be added to the system in some form if we are to peptize matter in mass.

Peptizing methods fall into four general classes.

THE WASHING OUT OF A PRECIPITATING AGENT. Everyone familiar with quantitative analysis recognizes the fact that certain precipitates cannot be washed with pure water without causing the precipitate to go through the filter paper. Thus, we wash the "yellow precipitate" of ammonium phosphomolybdate with a dilute solution of ammonium nitrate rather than with pure water. The ammonium nitrate prevents the precipitated ammonium phosphomolybdate from dispersing into a colloidal sol. Similarly, if we wash the precipitate of silver bromide obtained at points *A*, *C*, *E* on the diagram in Fig. 4, we find that the curves *ABC* and *CDE* are reversible and that, as the silver nitrate or potassium bromide is washed out of precipitates *A* and *E*, the liquid passing through

[23] I. M. Kolthoff, W. M. Lauer, and C. J. Sunde, *J. Am. Chem. Soc.*, **51**, 3273 (1929).

the filter contains silver bromide micelles positively charged in the case of precipitate *A* and negatively charged in the case of precipitate *E*. The isoelectric precipitate obtained at point *C* will not be peptized by the action of water, and *such isoelectric material is the only material which would give accurate information in regard to the solubility of silver bromide in water.* In the great majority of instances where a precipitate goes through the filter paper we are dealing with the peptization of a precipitated colloid system (a gel or a coagulum) by the washing out of the precipitating agent.

In the mechanical analysis of soils this phenomenon comes into play. If a soil is leached with hydrochloric acid and then washed free of chlorides with distilled water, it is the general experience that before the chlorides are completely removed clay begins to pass through the filter paper. The clay in the soil is usually present in a more or less precipitated condition, owing to the presence of inorganic salts. The removal of these salts permits the peptization of the clay aggregates to take place with the formation of a clay hydrosol. If the soil is a surface soil, the black humus likewise becomes peptized by washing out the calcium or other metallic salts which hold it in a precipitated condition, and it often happens that appreciable amounts of humus can be extracted from the soil in the form of a colloidal sol by first leaching the soil with hydrochloric acid to remove the calcium, and then washing with distilled water to the absence of chlorides.

Many organic substances are peptized by the direct addition of water. Familiar examples are gelatin, gum arabic, and dextrin. In certain instances it is necessary to add energy in the form of heat to cause complete dispersion. Thus, agar and gelatin form gels in cold water, but sols in hot water.

THE ADDITION OF A PEPTIZING AGENT. This has already been discussed fully in the sections above. Specific examples are: the addition of ammonium hydroxide to a soil in order to form a clay hydrosol; the peptization of china clays by alkalies or carbon dioxide; the removing of dirt from clothing by soap; stable sols of lampblack and graphite by the use of soaps, proteins, gums, and resins; emulsions, such as kerosene emulsion, so widely used in spraying, by the use of soaps. A great number of laboratory or technical applications have been made [24] involving peptization of one or another material by the addition of a peptizing agent.

[24] Compare, for example, W. D. Bancroft, *Report on Peptization and Precipitation*, Second Report on Colloid Chemistry, Brit. Assoc. for the Advancement of Science (1918); R. H. Bogue, *The Theory and Application of Colloidal Behavior*, Vol. II, McGraw-Hill Book Company, New York, 1924; and P. H. Fall, *J. Phys. Chem.*, **31**, 801 (1927).

MECHANICAL DISINTEGRATION. If one places a fragment of a microscope cover glass in a mortar and retains sufficient enthusiasm to grind this fragment vigorously for 20 to 30 minutes, it is found that a considerable part of the glass will remain as a stable hydrosol when water is added to the ground mass. Similarly, almost any practically insoluble material can be obtained in the form of a colloidal sol. Most insoluble materials form lyophobic sols. The formation of lyophobic sols requires a much greater expenditure of mechanical energy than does the formation of similar lyophilic sols. Most lyophilic sols form more or less spontaneously on the addition of the liquid.

Colloid mills have been devised to prepare finely divided materials for industrial uses. Most of these mills are misnamed "colloid" mills, inasmuch as the size of the suspended particle rarely falls within the limits fixed by the colloid realm. On the other hand, the disintegration produced by such mills is greater than that generally produced by other grinding methods, and such apparatus is coming more and more into use for the grinding of pigments, the making of emulsions, or finely dispersing any desired material in a liquid.

ELECTRICAL ENERGY. The formation of colloidal systems by the use of electrical energy may be regarded as a form of peptization. We have already discussed it under electrical dispersion, and it is referred to again, inasmuch as some authors make no distinction between electrical dispersion and peptization. Whether or not such a distinction is made will depend on how closely lines are drawn in a definition.

Emulsions. Possibly emulsions could be classified as a fifth group of colloidal systems formed by peptization. However, the range of particle size in emulsions generally exceeds the upper limit of the colloid realm, and many of the fat droplets in emulsions are visible in the ordinary microscope field.

It is extremely difficult to produce simple emulsions, *i.e.*, emulsions containing only one pure liquid dispersed in another pure liquid, which have any appreciable concentration and which are stable. In certain instances dilute emulsions (1:1,000) can be obtained where the above conditions are met. For example, in the preparation of aniline by steam distillation, the supernatant liquid in the receiving flask very often is a milky appearing liquid, a dilute emulsion of aniline in water. Such emulsions are often obtained in the distillation of volatile liquids which are relatively insoluble in water. If the bulk of such emulsions is small, this phenomenon does not constitute a serious loss in industrial processes unless the suspended material is extremely valuable. With certain essential oils, the material remaining in the emulsion must be recovered either by extracting with a solvent or by breaking the emulsion in some manner, causing the oil to separate.

Most emulsions used in the industries are stabilized by the presence of another disperse system, preferably a colloid which shows a marked affinity for either the oil or the water. Gums are the substances most generally used as stabilizers. Gum acacia (gum arabic) is the common stabilizer for oil-in-water emulsions, although proteins act similarly. Probably the reason that the emulsions used in the industries are stabilized with gum acacia instead of proteins is that proteins are readily attacked by putrefactive bacteria, whereas gum acacia is not. Gum acacia disperses in water to form a lyophilic hydrosol. It is a general rule that *those emulsion stabilizers which are soluble in water or which form lyophilic hydrosols produce emulsions of the oil-in-water type, whereas those stabilizers which are insoluble in water and soluble in oil produce emulsions of the water-in-oil type.* The sodium or potassium soaps, being water-soluble, act as stabilizers of oil-in-water emulsions, whereas the calcium soaps, being water-insoluble and oil-soluble, serve as efficient stabilizers for water-in-oil emulsions. Gum dammar, an oil-soluble gum, is an efficient stabilizer of water-in-oil emulsions.[25]

As a rule the natural emulsions, such as milk and the milky latex of plants, are stabilized by proteins. The latex of certain of the *Euphorbiaceae* is stabilized by sterols and in other instances by proteins.[26] The rubber hydrocarbon in *Hevea brasiliensis* is held in the latex by a protein which has more or less the characteristics of a globulin. A special protein [27] stabilizes the fat droplets in milk and cream. The egg-yolk proteins stabilize mayonnaise dressing, and a little gelatin, or better still gum dammar, will make a very stable French dressing emulsion from olive oil and vinegar.

There is apparently no limit to the amount of the disperse phase which may be present in an emulsion. *Ninety-nine per cent of oil can be dispersed in 1 per cent of water containing an emulsifying agent to form an oil-in-water emulsion.*[28] The liquid droplets apparently pack and deform into a more or less honeycomb structure with only a thin film of water between them. Obviously such emulsions possess a certain degree of rigidity or a rather high viscosity. It is doubtful that such concentrated emulsions can be formed without a distortion of the fat droplets, for a mixture of such concentration cannot be obtained in the case of a solid dispersed in water. For example, lead shot and water reach a limit of shot content depending on the size of the shot.

[25] H. N. Holmes and D. H. Cameron, *Science*, **56**, 724 (1922).

[26] L. S. Moyer, *Am. J. Bot.*, **22**, 609 (1935).

[27] L. S. Palmer and Hilda F. Wiese, *J. Dairy Sci.*, **16**, 41 (1933); Hilda F. Wiese and L. S. Palmer, *J. Dairy Sci.*, **17**, 29 (1934); C. E. Rimpila and L. S. Palmer, *J. Dairy Sci.*, **18**, 827 (1935).

[28] S. U. Pickering, *Kolloid-Z.*, **7**, 11 (1910); *J. Chem. Soc.*, **91**, 2001 (1907)

Pickering pointed out that various solid materials, such as barium sulfate, precipitated sulfur, Prussian blue, Trumbull's blue, and purple of Cassius, can be used to stabilize emulsions. In such instances the precipitate has an affinity for either the water or the oil and collects on the surface of the droplet, forming a more or less rigid membrane which prevents a later coalescence of the disperse phase. Schlaepfer [29] concluded that a finely divided solid which is more easily wetted by oil than by water should be capable of forming emulsions with water as a disperse phase. He used lampblack to test out his theory and formed stable water-in-oil emulsions. Talmud and Suchowolskaja [30] discuss three-phase emulsions of oil-water-insoluble powders with reference to emulsification, gelation, and syneresis. They do not believe that electrical charges play any appreciable role but that solvation is the important factor and that this in turn is due to the wettability of the solid phase, *i.e.*, the presence of polar or non-polar groups. They suggest that syneresis occurs when polar groups disappear through combination with some other surface and that consequently the ability to hold the dispersions medium is lost. Berl and Schmitt [31] have adopted this view, and they designate a powder hydrophilic, *i.e.*, preferentially wetted by water, if it stabilizes an oil-in-water emulsion, and hydrophobic if it stabilizes a water-in-oil emulsion, and apply the results of their investigation to the problems of ore flotation.

The stability of an existing emulsion depends principally on the state of the film formed at the oil-water interface by the emulsifier. Interfacial tension seems to have little influence on the stability. If the tension is low an emulsion is much more easily formed, but, once it is formed, a low interfacial tension does not increase the stability. The electrokinetic potential has little significance for the stability of such emulsions. Particle size does not seem to be of prime importance. As a matter of fact, it is almost impossible to form an emulsion whose particles are much below 0.5 μ in diameter, and small particles in an emulsion will usually rather quickly increase in size. The vigorous motion of the small-sized emulsion particles brings about coalescence.

Finkle, Draper, and Hildebrand [32] and Harkins [33] are undoubtedly correct in ascribing the action of efficient emulsifiers to the collecting of

[29] A. U. M. Schlaepfer, *J. Chem. Soc.*, **113**, 522 (1918).

[30] D. Talmud and S. Suchowolskaja, *Kolloid-Z.*, **55**, 48 (1931).

[31] E. Berl and B. Schmitt, *Kolloid-Z.*, **61**, 80 (1932).

[32] P. Finkle, H. D. Draper, and J. H. Hildebrand, Colloid Symposium Monograph, Vol. I, p. 196 (published by the Department of Chemistry, University of Wisconsin, Madison, 1923).

[33] W. D. Harkins, Colloid Symposium Monograph, Vol. II, p. 141, Chemical Catalog Co., Inc., New York, 1925.

the emulsifying agent in the interfacial film in such a manner that the oil droplets are surrounded by a "shell" of oriented molecules of the stabilizing agent. The theory which Finkle, Draper, and Hildebrand propose to explain the formation of oil-in-water emulsions when the alkali salts of the fatty acids are used, and the formation of water-in-oil emulsions when the fatty acid salts of the alkaline earths or of the heavy metals are employed, is incorrect, inasmuch as the size of a molecule of a stabilizing soap is not sufficiently great to exert the leverage action which is postulated; *cf.* Harkins. On the other hand, there is no doubt that, in emulsions stabilized by a soap, the hydrocarbon end of the molecule of the salt of a fatty acid is oriented toward the oil phase and the carbonyl group, with its metallic radical, is oriented toward the water phase. Figure 5 shows diagrammatically such orientation.

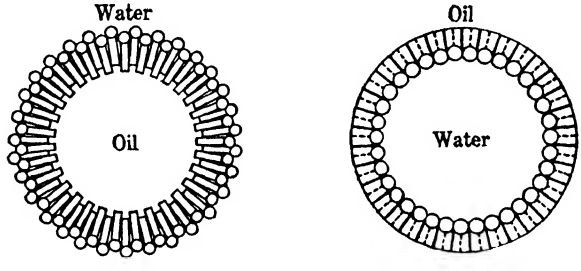


FIG. 5. A diagrammatic representation of molecular orientation of the stabilizing molecules in oil-in-water and water-in-oil emulsions. (After Harkins.)

This question of molecular orientation will be more fully considered later (p. 162). It is sufficient to note at this point that the suggested explanation for the progressive reduction of surface tension in aqueous solutions of the alkali salts of the fatty acids with progressive increase in the length of the carbon chain is accounted for by the eventual formation of a hydrocarbon interface between liquid and air. When such an interface is complete and we have a continuous, closely packed "skin" of oriented molecules on the surface of the water, we no longer have a water-air interface but rather a hydrocarbon-air interface which has a much lower surface tension than a water-air interface.

As we increase the length of the carbon chain past C_{14}, there is relatively little change in the effect on surface tensions or interfacial tension. This is to be expected, inasmuch as, if we are already dealing with a hydrocarbon surface, the addition of a $—CH_2—$ radical underneath that surface should have little effect on the properties of the surface.

Clowes [34] studied the effect of sodium chloride and calcium chloride

[34] G. H. A. Clowes, *J. Phys. Chem.*, **20**, 407 (1916).

on emulsification. Figure 6 and Table 3, taken from his data, represent the experimental results. The experimental technic was to allow slightly rancid olive oil to flow from the tip of a stalagmometer and to count the number of droplets which are formed by a given volume of the oil. It will be noted that sodium chloride solutions greatly decrease interfacial tension with the resulting formation of small droplets, whereas calcium chloride solutions increase interfacial tension with the resulting formation of large droplets. The action of the sodium chloride and calcium chloride in these experiments is explained by the formation of small amounts of sodium or calcium soaps, and, though no emulsion resulted in this case, it is obvious that the sodium salt increases the ease of dispersion of the olive oil in water, whereas the calcium salt would hinder dispersion.

It is of interest to note that sodium and calcium chlorides when mixed antagonize each other and nullify the effect of either, so that such solutions are without any great effect on interfacial tension. This phenomenon has become known in biological reactions as *antagonism of ions.* Harkins and Zollman [35] studied the systems investigated by Clowes in regard to interfacial tension changes, and the last two columns of Table 3 show their experimental data. They point out that, when the interfacial tension of a water-benzene system falls below 10 dynes per centimeter, the benzene can be easily emulsified in the aqueous phase, and that, when the interfacial tension falls below 1 dyne per centimeter, emulsions are formed spontaneously.

The antagonistic action of calcium and sodium is well illustrated by the work of Chambers and Reznikoff.[36] By means of Chambers' micromanipulation apparatus, amoebae were injected with dilute solutions of sodium chloride, potassium chloride, calcium chloride, and magnesium chloride. Injection of sodium chloride or potassium chloride caused a liquefaction of the protoplasm surrounding the injected area. In other words, the protoplasm shows a decreased viscosity, becoming more nearly a sol. On the other hand, when calcium chloride is injected, there is a local solidification of the injected protoplasmic region, which produces a more or less rigid mass of gel which the amoeba "pinches off" and rejects. The portion affected is in this way eliminated, leaving the remainder of the amoeba in an apparently normal condition. When magnesium chloride is injected, there is a local solidification similar to that occurring on the injection of calcium chloride. In this case, however, no "pinching off" reaction takes place, and the solidification proc-

[35] W. D. Harkins and Henrietta Zollman, *J. Am. Chem. Soc.*, **48**, 69 (1926).

[36] Robert Chambers and Paul Reznikoff, *Proc. Soc. Exptl. Biol. Med.*, **22**, 320 (1925).

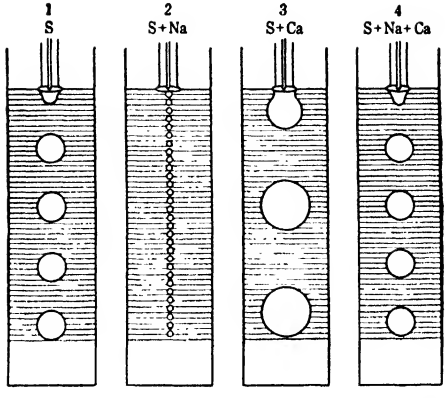


FIG. 6. Illustrating Expts. 1–4 in Table 3. In each case the number of drops in the diagram is one-tenth of the number actually obtained. (After Clowes.)

TABLE 3. NUMBER OF DROPS FORMED FROM A GIVEN VOLUME OF OLIVE OIL FLOWING FROM A STANDARD TIP INTO SOLUTIONS OF 0.001 *M* NaOH ALONE AND PLUS VARIOUS CONCENTRATIONS OF NaCl AND $CaCl_2$ SOLUTIONS *

Exp. No.	Concentration of			Number of Drops	Interfacial Tension	
	NaOH	NaCl	$CaCl_2$		Olive oil, ergs for each sq. cm.†	Paraffin oil, ergs for each sq. cm.†
1	0.001 *M*	.. ....	.....	44	7.3	7.2
2	0.001 *M*	0.15 *M*	.....	300	0.002	0.00
3	0.001 *M*	.....	0.0015 *M*	24	9.88	9.65
4	0.001 *M*	0.15 *M*	0.0015 *M*	44	6.88	7.48
5	0.001 *M*	0.30 *M*	0.003 *M*	43	6.36	7.12
6	0.001 *M*	0.45 *M*	0.005 *M* ‡	43	6.70	7.36
7	0.001 *M*	0.60 *M*	0.01 *M* §	43	7.31	8.20
8	.....	.... ..	.....	..	24.11	31.05

* Data of Clowes (1916) except for last two columns, which present data of Harkins and Zollman (1926) on the interfacial tensions of these solutions.

† With 0.001 *N* oleic acid in the oil.

‡ 0.0045 in Harkins' experiments.

§ 0.0060 in Harkins' experiments.

ess gradually spreads throughout the amoeba, causing complete gelation and death. When mixtures of sodium chloride and calcium chloride or mixtures of potassium chloride and calcium chloride, similar in concentration to the mixtures used by Clowes and by Harkins and Zollman in Table 3, are injected, no effect is observable, the sodium chloride neutralizing the solidifying effect of the divalent salt and the divalent salt neutralizing the liquefying action of the sodium chloride. Chambers and Reznikoff suggest that "at least one of the features of the antagonistic action of NaCl or KCl to $CaCl_2$ is the maintenance in protoplasm of a definite balance between its liquid and solid phases. This phenomenon possibly depends on the formation of a balanced proportion of Na and Ca or of K and Ca protein salts. It may also be due to the formation of Na or K and Ca soaps." In the light of the experiments of Harkins and Zollman we must conclude that interfacial tension phenomena and alteration in the type of protoplasmic emulsions are features in the experiments noted by Chambers and Reznikoff.

The antagonistic action of ions has been further studied by Chambers and co-workers [37] and by Heilbrunn and co-workers.[38] Heilbrunn studied changes in viscosity by use of the centrifuge. He found that with the protoplasm of *Amoeba dubia* NaCl alone increases the viscosity of the protoplasm by 41 per cent, KCl alone by 21 per cent, $MgCl_2$ decreases the viscosity by 32 per cent, and $CaCl_2$ decreases the viscosity by 35 per cent. He suggests that the protoplasm micelles are positively charged. With *A. proteus*, Ca^{++} increases viscosity of the plasma gel, whereas K^+, Mg^{++}, and Na^+ decrease viscosity in the order named. When solutions are mixed, antagonism is exhibited. Ca^{++} and K^+ especially neutralize the effects of each other, and he concludes that calcium favors gelation, whereas potassium and magnesium favor solation. He further concludes that magnesium anesthesia is brought about by a liquefaction of the plasma gel and that potassium likewise can induce anesthesia by a similar liquefaction. He finds that the potassium ion is antagonized by hydrogen ions so that at an increased acidity the anesthetic and liquefying action of the potassium ion is not observed.

Weiser [39] studied the antagonistic action of ions in the precipitation of colloid sols, with the interesting result that antagonistic actions similar to the observed biological phenomena were demonstrated. He sug-

[37] P. Reznikoff and R. Chambers, *J. Gen. Physiol.*, **10**, 731 (1927); R. Chambers and H. Pollack, *J. Gen. Physiol.*, **10**, 739 (1927); R. Chambers, M. J. Kopac, and C. G. Grand, *Ind. Eng. Chem., Anal. Ed.*, **9**, 143 (1937).

[38] L. V. Heilbrunn and K. Daugherty, *Physiol. Zool.*, **4**, 635 (1931); **5**, 254 (1932); L. V. Heilbrunn, *Proc. Soc. Exptl. Biol. Med.*, **29**, 467 (1932); L. V. Heilbrunn and K. Daugherty, *J. Cellular Comp. Physiol.*, **5**, 207 (1934).

[39] Harry B. Weiser, *J. Phys. Chem.*, **30**, 1527 (1926).

gested that possibly the antagonistic action of salt pairs (*e.g.*, Na *vs.* Ca) might be conditioned on the effect of the ions on the permeability of the colloidal film making up the cell membrane.

Bull and Gortner [40] investigated the zeta potential as a factor in ion antagonism in salt pairs and could not observe any antagonistic effects. However, when the data were recalculated [41] in terms of absolute charge on the surface, a marked ion antagonism was found between NaCl:KCl, KCl:$MgCl_2$, NaCl:$MgCl_2$, NaCl:$CaCl_2$, KCl:$CaCl_2$, and $CaCl_2$:$MgCl_2$, so that perhaps the ratio of the ions contributing to the absolute charge at the interface may be at least one of the determining factors in the ion antagonisms which have been observed in biological systems. The last word has not been written on the question of ion antagonism as exhibited in biological systems, but the solution of the phenomenon will undoubtedly be hastened when the methods of colloid chemistry are applied to such systems.

The breaking of emulsions and the inversion of emulsions involve changes in the interfacial tension between the two components and an alteration in the orientation of the surface film surrounding the suspended droplets. It sometimes happens that crude petroleum issues from the ground emulsified with the salt water associated with the petroleum fields. In such instances the emulsifying agent appears to be asphaltic residues. Obviously such emulsions are without commercial value, and the breaking of such emulsions becomes an important industrial problem. A knowledge of surface chemistry and of the role that interfacial tension plays in emulsion formation, as well as a knowledge of the antagonistic effects of ions and of polar and non-polar groups in organic compounds, has proved very helpful to the chemist who is faced with such industrial problems. Specific instances will be cited under molecular orientation (p. 162).

As noted earlier, only a dilute emulsion results when we have oil dispersed in pure water. This is due to the fact that we have present in such a system a relatively high interfacial tension. Such systems behave as true lyophobic sols. When the interfacial tension is lowered by the introduction of an emulsifying agent, an extremely high concentration of the disperse phase can be obtained. Such emulsions behave more or less as lyophilic systems and take on to a large degree the properties characteristic of the emulsifying agent. Such systems are usually called emulsions, and the fact that the stabilizing agent is generally a lyophilic colloid justifies the inclusion of emulsions in a consideration of colloid systems.

[40] H. B. Bull and R. A. Gortner, *J. Phys. Chem.*, **35**, 700 (1931).

[41] L. S. Moyer and H. B. Bull, *J. Gen. Physiol.*, **19**, 239 (1935).

CHAPTER 3

Certain Physical Properties Characteristic of Colloid Systems

Viscosity. Viscosity is the internal friction of a liquid, the resistance to shear or flow. The unit of viscosity is the *poise*, so named after the Frenchman, Poiseuille,[1] who first devised methods for the measurement of viscosity. The poise is defined as that viscosity where unit force (1 dyne) is required to cause two parallel liquid surfaces of unit area (1 sq. cm.) and unit distance apart (1 cm.) to slide past one another with unit velocity (1 cm. per sec.).

The fundamental law governing the flow of liquids through capillaries is Poiseuille's. In the integrated form it is expressed as

$$V = \frac{\pi r^4 P t}{8 \eta l} \tag{4}$$

where V = the volume in milliliters of liquid flowing through the capillary in t seconds
r = the radius of the capillary in centimeters
P = the difference in pressure between the two ends of the capillary expressed in dynes per square centimeter
η = the coefficient of viscosity in poises
l = the length of the capillary in centimeters.

Theory demands that the flow be streamline or laminar and that there be no turbulence or eddies in the liquid as it flows. There is a critical pressure in every case above which turbulence appears, and this critical pressure depends on the geometry of the system as well as on the density and viscosity of the liquid.

The determination of the rate of flow through a capillary is the basis for an important method for the measurement of viscosity. In this method the time of flow of a known volume of liquid through a capillary under the influence of gravity is determined. Such an apparatus is diagrammed in Fig. 7.

[1] J. L. M. Poiseuille, *Ann. chim. phys.*, 3 ser., **21**, 76 (1847).

Since the Ostwald viscometer is the most frequently used type of apparatus for the measurement of viscosity, it will be discussed in some detail. A definite volume of liquid is placed in the viscometer, and the level of the liquid is drawn above the top mark of the bulb by suction (see Fig. 7). The liquid is allowed to flow out freely, and the time required for the liquid level to drop from the upper mark to the lower mark is measured. The relation between viscosity and time of outflow is

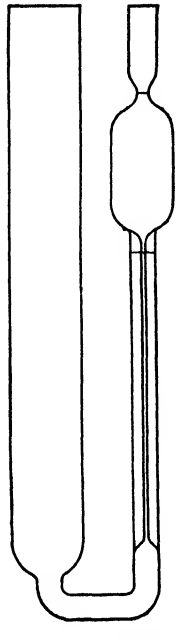


FIG. 7. Ostwald viscometer.

$$\eta = C\rho t - f\left(\frac{\rho v}{lt}\right) \tag{5}$$

where C = a constant which involves the length and radius of the capillary
ρ = the density of the liquid.

The second term on the right-hand side of the equation, $f(\rho v/lt)$, indicates that some function of $\rho v/lt$ must be subtracted from $C\rho t$ in order to obtain the viscosity. This term, the so-called kinetic correction, takes into account the motion of the liquid after it leaves the capillary. If t and l are sufficiently large, it can be neglected. The viscometer is calibrated with water or some other appropriate liquid whose viscosity is known exactly. If we neglect the kinetic correction, we have the *relative viscosity*

$$\eta_r = \frac{\eta}{\eta_0} = \frac{\rho t}{\rho_0 t_0} \tag{6}$$

where η = the coefficient of viscosity of the liquid whose viscosity we wish to measure
η_0 = the coefficient of viscosity of the standard liquid
ρ_0 = the density of the standard liquid
t_0 = the time of outflow of the standard liquid.

ρ and t are the corresponding quantities of the liquid whose viscosity is being measured.

The most reliable method to determine whether the kinetic correction $f(\rho v/lt)$ can be safely neglected is to measure the viscosity of two liquids whose viscosities are known. If the two measured viscosities bear the correct ratio to each other, and if the unknown viscosity has approxi-

mately the same value as that of the two known solutions, the kinetic correction is unnecessary. If not, it may be necessary to apply a kinetic correction although the error may be due to other causes. Indeed, if a kinetic correction is necessary, it is better to redesign the viscometer so that no such correction is required. It is well to emphasize that two absolute necessities for satisfactory viscosity measurements are cleanliness of the viscometer and an accurate temperature control.

The Couette viscometer employs a rotating cylinder containing the liquid. A smaller cylinder is suspended in the liquid by means of a fine wire. The torque produced on the suspended cylinder by the rotation of the outer cylinder is recorded by a beam of light thrown on a mirror attached to the suspension wire. The light is reflected from the mirror on a graduated scale. The type of flow obtained in a Couette viscometer is simpler than that in an Ostwald viscometer, and the theoretical treatment is less involved. Also, very much smaller flow gradients can be obtained conveniently. Unfortunately, the temperature control in this type of apparatus presents a real difficulty; also, the apparatus has to be very exactly made, the services of an expert machinist being required. Bjornstahl and Snellman[2] report the construction of a Couette viscometer. There are several commercial adaptations of the Couette principle. Great accuracy cannot, however, be expected from them.

The Hess viscometer[3] requires very small volumes of fluid and has found some favor in the measurement of the viscosity of blood. It consists of two horizontal capillaries. One of the capillaries is partly filled with water and the other partly filled with blood. The two capillaries are connected to a rubber bulb which can be squeezed. The distance traveled along the capillary by the water is compared with that traveled by the blood. The ratio of the two distances traveled is equal to the inverse ratio of the two viscosities. The Hess viscometer is commercially available but its accuracy is not of a high order.

Viscosity of Colloidal Solutions. It is evident that, if spherical colloidal particles are suspended in a liquid, the viscosity of the liquid will increase owing to the disruption of the uniform flow gradients of the liquid. On this basis Einstein[4] formulated his well-known equation for the viscosity of a suspension of spherical particles. The equation is

$$\eta = \eta_0 \left[\frac{1 + 0.5\phi}{(1 - \phi)^2} \right] \tag{7}$$

[2] Y. Bjornstahl and O. Snellman, *Kolloid-Z.*, **86**, 223 (1939).
[3] W. Hess, *Kolloid-Z.*, **27**, 154 (1920).
[4] A. Einstein, *Ann. Physik*, **19**, 289 (1906); **34**, 591 (1911).

where η = the coefficient of viscosity of the suspension
η_0 = that of the dispersion medium
ϕ = the relative volume concentration of the suspension; it is equal to the total volume of the suspended material divided by the total volume of the suspension.

Guth [5] has outlined the assumptions upon which the Einstein equation is based. They are: (1) the suspended spheres are large compared with the molecules of the dispersion medium but small compared with the dimensions of the viscometer; (2) the dispersion medium is incompressible; (3) the suspended particles are rigid and are completely wet by the dispersion medium; (4) gravitational forces acting on the particles can be neglected; (5) the velocity of flow of the medium is small; (6) the concentration of the suspension is so small that there is no interference of one particle with another; (7) the suspended particles are distributed at random throughout the medium; (8) it is assumed that the flow, with the exception of that in the immediate neighborhood of the particles, is static and does not change significantly except within a volume which is large compared with the radius of the particle.

The Einstein equation can be expanded into a power series which is

$$\frac{\eta}{\eta_0} = 1 + 2.5\phi + 4\phi^2 + 5.3\phi^3 \tag{8}$$

Owing to particle interaction, higher terms than the first are without significance, and the equation becomes

$$\frac{\eta}{\eta_0} = 1 + 2.5\phi \tag{9}$$

which is the form in which one frequently sees the equation written. Eirich, Bunzl, and Margaretha [6] have given adequate experimental demonstration that the Einstein equation is valid for very dilute suspensions of spherical particles, thus showing clearly that the first coefficient of the series is 2.5.

Guth and Gold [7] considered the interactions of the flow lines around neighboring particles. The attractive and repulsive forces between particles were considered insignificant. Thus, with the exception of

[5] E. Guth, *Kolloid-Z.*, **74**, 147 (1936).
[6] F. Eirich, M. Bunzl, and H. Margaretha, *Kolloid-Z.*, **74**, 276 (1936).
[7] E. Guth and O. Gold, *Phys. Rev.*, **53**, 322 (1938).

item 6 listed above, their assumptions were identical with those of Einstein. They obtained

$$\eta = \eta_0(1 + 2.5\phi + 14.1\phi^2 + \cdots) \tag{10}$$

This equation, tested [8] on emulsions of rubber latex, was found to be valid up to a concentration of 30 per cent. The second coefficient in the expansion series is thus 14.1 or close to it. We can write the equation

$$\frac{\eta}{\eta_0} - 1 = 2.5\phi + 14.1\phi^2 \tag{11}$$

The term $\eta/\eta_0 - 1$ or $\eta_r - 1$ is called the *specific viscosity* and is expressed by the symbol η_{sp}. Dividing equation (11) by ϕ, we have

$$\frac{\eta_{sp}}{\phi} = 2.5 + 14.1\phi \tag{12}$$

It is clear that, if the suspended particles are hydrated, Guth's equation offers a direct means for calculating the hydration. If we plot η_{sp}/ϕ against ϕ we obtain a straight line, at least for the lower concentrations. From the intercept on the η_{sp}/ϕ axis we obtain η_{sp}/ϕ at infinite dilution. If the particles are spherical, η_{sp}/ϕ will equal 2.5. We, therefore, calculate the additional volume, *i.e.*, hydration, which must be added to the volume of the suspended particles in their dry condition to yield 2.5 for the intercept. The difficulty is, of course, that the majority of suspensions do not have spherical particles, and the application of Guth's equation to non-spherical particles is meaningless.

Kraemer [9] calls the term η_{sp}/ϕ at infinite dilution the *volume intrinsic viscosity*, and he suggests that the readiest way of calculating it from experimental data is by the relation

$$\left(\frac{\eta_{sp}}{\phi}\right)_{\phi \to 0} = \frac{\log_e \eta_r}{\phi} \tag{13}$$

Correspondingly, the *weight intrinsic viscosity* is given by

$$\left(\frac{\eta_{sp}}{C}\right)_{C \to 0} = \frac{\log_e \eta_r}{C} \tag{14}$$

where C is the concentration in grams per 100 ml. of solution.

[8] H. F. Smith, *Rubber Chem. Technol.*, **15**, 301 (1942); see also E. Guth, *Proc. 5th Intern. Congr. Applied Mechanics* (1938).

[9] E. O. Kraemer, in The Svedberg and K. O. Pederson, *The Ultracentrifuge*, Oxford University Press, New York, 1940.

It was realized early that there must be a relation between the asymmetry of the suspended particles and the viscosity of a suspension. The particles are turning and twisting in Brownian motion, and they thus appear to occupy a larger volume than they actually do. The twisting of the particles interrupts the streamlines of flow and requires additional expenditure of work to maintain a given velocity of flow. This work appears as an increase in the viscosity.

The relation between particle asymmetry and viscosity is extraordinarily complicated; to date, no completely unambiguous relation between these two quantities has been formulated. The equation of Simha [10] appears, however, to be the most satisfactory.[11] Shown in Table 4 are the axial ratios for prolate (elongated ellipsoids of revolution) and for oblate (flattened ellipsoids of revolution) ellipsoids corresponding to

TABLE 4. AXIAL RATIOS OF SUSPENDED ELLIPSOIDS CORRESPONDING TO NUMERICAL VALUES FOR THE VOLUME INTRINSIC VISCOSITIES

(a/b is the axial ratio)

a/b	Intrinsic Viscosities		a/b	Intrinsic Viscosities	
	Prolate	Oblate		Prolate	Oblate
1.0	2.50	2.50	20.0	38.6	14.80
1.5	2.63	2.62	25.0	55.2	18.19
2.0	2.91	2.85	30.0	74.5	21.6
3.0	3.68	3.43	40.0	120.8	28.3
4.0	4.66	4.06	50.0	176.5	35.0
5.0	5.81	4.71	60.0	242.0	41.7
6.0	7.10	5.36	80.0	400.0	55.1
8.0	10.10	6.70	100.0	593.0	68.6
10.0	13.63	8.04	150.0	1222.0	102.3
12.0	17.76	9.39	200.0	2051.0	136.2
15.0	24.8	11.42	300.0	4278.0	204.1

the volume intrinsic viscosities of suspensions. These calculations are based on Simha's equation.

Diffusion in Colloid Systems. Graham's early distinction between colloidal and crystalloidal systems was largely based upon the pronounced difference in the rate of diffusion in the two instances. Graham

[10] R. Simha, *J. Phys. Chem.*, **44**, 25 (1940).

[11] M. A. Lauffer, *Chem. Revs.*, **31**, 561 (1942); *J. Am. Chem. Soc.*, **66**, 1188 (1944).

found that, if various materials were placed in bags of parchment paper or of animal membranes, such as dried bladder or goldbeater's skin, the various substances could be divided into two general classes: (1) those that readily passed through the parchment paper or animal membranes and (2) those that failed to pass through such membranes. The former he called crystalloids, the latter, colloids; and non-diffusion through a membrane was taken as the criterion of the colloid state.

It is obvious that if true solutions grade almost imperceptibly into colloidal systems and colloidal systems grade imperceptibly into coarse suspensions, there is no sharp line of demarcation, even with respect to rate of diffusion, between true solutions and colloid systems.

Diffusion of a substance is expressed in terms of the *diffusion coefficient*, which may be defined as the quantity of material which diffuses per second across a surface 1 sq. cm. in area under a unit concentration gradient. It is expressed in square centimeters per second.

There are two methods in common use for the measurement of the diffusion coefficient of a substance. One method is based upon the chemical analysis of two solutions of different concentrations of the solute separated by a porous disk.[12,13] The other method involves a measurement of variation of the index of refraction at the boundary between the solution and the solvent.[14] Table 5 lists the coefficient of diffusion of certain substances.

TABLE 5. DIFFUSION COEFFICIENTS IN WATER AT 25°C., EXPRESSED IN SQUARE CENTIMETERS PER SECOND

KCl	1.99×10^{-5}
Sucrose	5.35×10^{-6}
Egg albumin	8.8×10^{-7}
Serum albumin (horse)	7.0×10^{-7}
Serum globulin (horse)	4.4×10^{-7}

Sutherland [15] and Einstein [16] almost simultaneously published papers dealing with the diffusion of large molecules into a medium of small molecules, and they were able to derive an equation relating the diffusion coefficient of a spherical molecule to its size. The equation they give is

$$D = \frac{RT}{6\pi r\eta N} \tag{15}$$

[12] J. W. McBain and T. H. Liu, *J. Am. Chem. Soc.*, **53**, 59 (1931).
[13] J. Northrop and M. L. Anson, *J. Gen. Physiol.*, **12**, 543 (1928–29); **20**, 575 (1937).
[14] H. Neurath, *Chem. Revs.*, **30**, 357 (1942).
[15] W. Sutherland, *Phil. Mag.*, [6] **9**, 781 (1905).
[16] A. Einstein, *Ann. Physik*, [4] **17**, 549 (1905); *Z. Elektrochem.*, **14**, 337 (1908).

where D = the diffusion coefficient
R = the gas constant of 8.31×10^7 ergs per degree per mole
T = the absolute temperature
r = the radius of the spherical particle in centimeters
η = the coefficient of viscosity of the solvent
N = Avogadro's number.

Since we are dealing with spherical particles, r may be expressed in terms of the volume of the particle, and after substituting numerical values for the constants at 25°C. we have

$$D = \frac{33.06 \times 10^{-6}}{\sqrt[3]{\text{molecular volume}}} \tag{16}$$

The molecular volume is the molecular weight divided by the density.

The diffusion coefficient is thus inversely proportional to the cube root of the volume of the molecules. If the molecule is hydrated in solution, the rate of diffusion will be diminished and the degree of hydration of a spherical molecule can be calculated from its rate of diffusion, provided we know the volume of the unhydrated spherical molecule.

If the colloid particles are asymmetric, their diffusion coefficient will be less than that of a spherical molecule of the same volume. Qualitatively, it is easy to see why asymmetry should decrease the diffusion coefficient of a particle. The amount of surface exposed to the frictional forces of the medium by an asymmetric particle is greater than that exposed by a spherical particle of the same volume. Actually, nearly all particles show some degree of asymmetry as well as hydration, and accordingly diffusion problems can be very complicated.

Brownian Movement. In 1827, a botanist, Robert Brown, noted that pollen grains suspended in a liquid on the microscope stage were in continuous motion in the liquid. In order to ascertain whether or not such motion was characteristic of living pollen grains, he secured, from the herbarium, pollen and plant spores which were more than one hundred years old and found that the rate of motion was in no way diminished. It was later observed that any particle of small enough size to remain in more or less permanent suspension in a liquid would exhibit this characteristic motion, which has been called the Brownian movement. Particles up to 4 μ and accordingly easily visible in the microscope show the characteristic Brownian motion, but the motion is not so active or so rapid as in colloidal systems. Burton [17] has given an ex-

[17] E. F. Burton, *The Physical Properties of Colloidal Solutions*, 3rd ed., Longmans, Green and Co., London, 1938.

cellent presentation of the historical phases and theoretical significance of Brownian movement.

Because a particle in Brownian movement oscillates rapidly in a haphazard manner around a mean position, the velocity of movement must be calculated from observations occupying very short periods of time. This has been done by Svedberg, who employed a motion-picture camera to record changes in the position of a particle. Measurements of displacement on the film indicate that particles 40 to 50 mμ in diameter in a platinum hydrosol may reach a velocity as great as 320 μ per sec. This velocity in relation to the size of the platinum particle approximates the velocity of a low-speed rifle ball.

Various theories were devised to account for Brownian movement. These theories have been adequately discussed by Burton. It is now definitely known that Brownian movement is caused by collisions of the particle with molecules of the liquid in which the particle is suspended.

Many experiments have been performed to test the kinetic theory of Brownian motion. Perhaps the most striking experiments are those of Perrin,[18] who pointed out that particles of identical size and shape suspended in a liquid are distributed according to the same law that is observed by molecules of a gas acted upon by gravity; *i.e.*, the particles are less dense at the surface of a liquid and become more dense as the depth below the surface increases. Perrin counted in a single experiment as many as 13,000 particles at varying depths of 5 μ, 35 μ, 65 μ, and 95 μ below the surface and found that the number of particles increased approximately in geometrical progression. As Burton noted, the concentration dropped to about one-half value in a change of 0.03 mm. difference in level, whereas in the atmosphere a similar decrease would require a difference in level of 6,000 km.

Perrin, finding that the distribution ratios of the various particles at different depths in the surface layer obeyed the gas laws, determined Avogadro's constant for his colloidal sols, and from the equation

$$N = \frac{3}{2}\left(\frac{RT}{W}\right) \tag{17}$$

where N = the number of particles (molecules in the case of a gas) in 1 gram molecule

W = the mean kinetic energy of a particle (the mean kinetic energy of a molecule in a gas)

he found a value of $N = 70.5 \times 10^{22}$.

[18] J. Perrin, *Compt. rend.*, **147**, 530, 594 (1908); *Ann. chim. phys.*, [8] **18**, 5 (1909); *Z. Elektrochem.*, **15**, 269 (1909).

Smoluchowski devised an equation based on the displacement of a particle under Brownian movement, due to kinetic energy of the disperse phase, and Einstein independently developed a similar equation for the motion of small spheres suspended in a liquid medium. Their equations were based on the assumption that the individual particle behaved as a gas molecule and that the gas laws could accordingly be applied, and yielded values for N of 70.5×10^{22} and 65.0×10^{22}, respectively. These values are strikingly in agreement with the values of Millikan, calculated from the direct determination of the electrical charge on ions, *i.e.*, 60.61×10^{22}. We must accordingly note that *colloidal systems obey the gas laws of physical chemistry, provided that each particle of the disperse phase is regarded as behaving as a single molecule.*

The Ultracentrifuge. The rate of sedimentation of suspensions under the influence of gravity had been used by a number of early workers to estimate the particle size of the suspensions. Stokes' formula relating the rate of fall of a spherical particle to its radius was employed to give an approximation of particle size. In 1923 Svedberg and Nicols [19] employed for the first time a centrifuge to increase the gravity force and to speed up the rate of sedimentation for the purpose of measuring particle sizes. This pioneer work has been followed by a gradual development in the technique of ultracentrifugation until today this instrument is the single most powerful tool for physical research on proteins and on their colloidal molecules.

The ultracentrifuge rotates at very great speeds and necessitates very precise experimental control. The oil-driven turbine type employed by Svedberg as well as by a few other workers is very expensive—so expensive in fact that its cost is prohibitive for most laboratories. Beams and co-workers [20] have developed an air-driven ultracentrifuge which is not so elaborate as the Svedberg type, and its cost of construction is less. McBain [21] has described a very simple and relatively inexpensive type of ultracentrifuge which, however, cannot be expected to have the flexibility and accuracy of the more expensive instruments.

With the exception of the McBain opaque ultracentrifuge, which has a mechanical device for immobilizing the centrifuged material, some optical system is needed to visualize the boundary of the sedimentating material. The lens systems employed are modifications of those used in the study of electrophoresis and of diffusion.

With any type of ultracentrifuge or of optical system the determination of the molecular weight (or particle weight) can be carried out in

[19] The Svedberg and J. B. Nicols, *J. Am. Chem. Soc.*, **45**, 2910 (1923).
[20] J. W. Beams, F. W. Linke, and P. Sommer, *Rev. Sci. Instruments*, **9**, 248 (1938).
[21] J. W. McBain, *Chem. Revs.*, **24**, 289 (1939).

one of two ways. With one procedure the speed of the ultracentrifuge is made very large and the *rate of sedimentation* measured. Such a determination can usually be completed in a few hours. The other means of determination involves rotating the centrifuge at a more moderate speed and *allowing the solution to reach an equilibrium*, the solution being more concentrated toward the outer portion of the centrifuge tube. In rate measurements, forced diffusion occurs under the influence of the centrifugal gravity field, and the net force acting on a mole of particles at a distance X from the axis of rotation is simply the difference between the centrifugal weight and the buoyancy exerted by the displaced medium. The molecular weight M may be expressed as

$$M = \frac{RTS}{D(1 - \rho\overline{V}_2)} \tag{18}$$

where R = the gas constant
T = absolute temperature
S = the sedimentation rate per unit field of force
D = the diffusion coefficient
ρ = the density of the solution
$\overline{V}_2$ = the partial specific volume of the particles.

After the molecular weight has been determined, the frictional force for a spherical molecule of the same molecular volume can be calculated by means of the Sutherland-Einstein diffusion equation. This frictional force, known as the frictional coefficient, is denoted by f/f_0. The ratio f/f_0 was at one time known as the asymmetry ratio. It must be realized, however, that, although it is related, in a complicated way, to the molecular asymmetry, the relation is not a direct one.

The other way of proceding in a determination of the molecular weight, as was indicated above, is to allow an equilibrium condition to be established and to measure the concentration at distances X_1 and X_2 from the axis of rotation. The measurement of the concentration has to be made by optical means. At equilibrium we have

$$M = \frac{2RT \log_e \dfrac{C_2}{C_1}}{(1 - \rho\overline{V}_2)\omega^2(X_2^2 - X_1^2)} \tag{19}$$

where C_2 and C_1 are the concentrations at the distances X_2 and X_1, respectively, and ω is the angular velocity in radians per second.

The equilibrium-sedimentation method is in a sense equivalent to an osmotic pressure determination. Its theoretical background is some-

what clearer than that of the rate-sedimentation method. The time required for equilibrium to be reached, however, is rather extended, and from an experimental point of view the equilibrium-sedimentation method is not so satisfactory as the rate-sedimentation method.

Table 6 shows some of the molecular weights determined by Svedberg and co-workers.

TABLE 6. MOLECULAR WEIGHTS OF SOME PROTEINS AS DETERMINED BY THE ULTRACENTRIFUGE

(M_s is the molecular weight by rate sedimentation and M_e is by the equilibrium method)

Protein	M_s	M_e
Ribonuclease	13,000	13,000
Cytochrome *c*	15,600	.....
Myoglobin	16,900	17,500
Gliadin	27,500	27,000
Hordein	27,500	.....
Pepsin	35,500	39,000
Insulin	41,000	35,000
β-Lactoglobulin	41,500	38,000
Egg albumin (hen)	44,000	40,500
Hemoglobin (man)	63,000	.....
Hemoglobin (horse)	68,000	68,000
Serum albumin (horse)	70,000	68,000
Yellow enzyme	82,000	78,000
Serum globulin (horse)	167,000	150,000

Osmotic Pressure. If a colloidal solution is placed in a collodion sack or a sausage casing and the sack immersed in water, it will be found that water will flow into the sack and develop a hydrostatic pressure. The counter pressure which is just sufficient to prevent the flow of water into the sack is called the *osmotic pressure.*

It has been found that dilute solutions of proteins and of other colloids obey the ideal gas laws. That is,

$$PV = nRT \tag{20}$$

where P = the osmotic pressure
V = the volume of the solvent
n = the number of moles of solute
R = the gas constant
T = the absolute temperature.

V is evidently equal to the weight of the solvent (w) divided by its density ρ. n is equal to the number of grams of the solute (C) divided by

the molecular weight of the solute (M). Substituting this information in the above equation and rearranging, we have

$$P = \frac{\rho RTC}{wM} \tag{21}$$

Selecting T equal to 25°C. and w equal to 100 grams and realizing that R is equal to 8.48×10^4 cm. of water pressure, we have

$$P = \frac{2.528 \times 10^5 \rho C}{M} \tag{22}$$

which allows us to calculate the osmotic pressure. If 1 gram of a colloid with a molecular weight of 10,000 is dissolved in 100 grams of water at 25°C., the osmotic pressure developed will evidently be 25.2 cm. of water pressure; if the molecular weight is 20,000 or twice as great as that above, the osmotic pressure will be *half* of 25.2 cm. of water pressure.

Suitable methods for the measurement of the osmotic pressure of colloid systems have been described by a number of workers. That used by Bull and Currie [22] may be mentioned.

Perstillation and Pervaporation. Kober [23] described an important technic which appears to have been rather generally overlooked. He notes that, if one encloses liquid in a collodion (probably cellophane would be preferable) bag and suspends such a bag over a free flame or an electric heater, one has in reality a ball of water suspended in air with evaporation possible on all surfaces. Under these conditions it is practically impossible to raise the liquid inside of the bag to the boiling point, and evaporation is extremely rapid. This phenomenon he calls pervaporation. If the liquid contains both crystalloids and colloids, the crystalloids will diffuse through the membrane with the water and remain, after the evaporation of the water, on the outside of the membrane in crystal form completely free from colloidal contaminants which will remain inside of the membrane.

Filtration and Ultrafiltration of Colloid Systems. As a rule, colloid sols pass unchanged through filter paper or through the porcelain filters of the bacteriological laboratory. The pores in the usual filter paper are approximately 2 to 5 μ in diameter, and the pores of the finer Berkefeld and Chamberlain porcelain filters range in size from 0.2 to 0.6 μ. It is accordingly understandable why colloid sols should pass through such filters, inasmuch as the size of the colloidal particle is con-

[22] H. B. Bull and B. T. Currie, *J. Am. Chem. Soc.*, **68**, 742 (1946).
[23] P. A. Kober, *J. Am. Chem. Soc.*, **39**, 944 (1917).

siderably less than the size of the opening through which the particle must pass. Matter in mass can, therefore, be separated from colloidal sols by passage through a bacterial filter.

Ultrafilters, however, have been devised with pores sufficiently fine to retain the disperse phase and allow only the dispersions medium and its crystalloidal solutes to pass through. This process of separation is known as ultrafiltration. Since ultrafilters are invariably colloid gels, ultrafiltration may be defined as filtration of a sol through a gel.

Various types of ultrafilters have been devised. Those commonly used are membranes of collodion or gelatin, although rubber membranes have been used in some instances. The technic of ultrafiltration is largely due to the pioneer studies of Bechhold,[24] who used filter paper and cloth as the supporting membrane for collodion or gelatin films. The paper or cloth was coated with collodion or with gelatin, the gelatin subsequently being hardened by immersion in a dilute solution of formaldehyde.

The direct measurement of osmotic pressure of crystalloidal solutions is carried out by making use of a membrane of copper ferricyanide precipitated in the pores of a porous porcelain vessel. Such a membrane can be made which is truly semi-permeable and which will allow the passage of water molecules but retard the passage of most crystalloidal solutions, such as solutions of dextrose and sucrose. An ultrafilter differs only in degree from such a membrane, the pores of an ultrafilter being larger in diameter than the pores of a truly semi-permeable membrane. All gradations of pore size of ultrafilters can be obtained between the ranges of a copper ferricyanide membrane and ordinary filter paper, and numerous papers have been published which deal with the technic for preparing such graded ultrafilters. One of the earlier papers is that by Schoep,[25] who points out that almost any degree of permeability of collodion membranes can be obtained by mixing varying proportions of glycerol and castor oil with the collodion solutions. If it is desirable to conduct ultrafiltrations under high pressure, the membranes may be precipitated in a porous porcelain support. When so prepared, pressures up to 100 atmospheres may be applied to induce filtration.

Farmer [26] and Eggerth [27] both studied the procedure for producing ultrafilters of varying porosity. Eggerth controlled porosity by varying the ratio of alcohol to ether in the solvent which was used to dissolve the collodion. He designated the porosity of the collodion membrane

[24] H. Bechhold, *Z. physik. Chem.*, **60**, 257 (1907); **64**, 328 (1908).
[25] Alfred Schoep, *Kolloid-Z.*, **8**, 80 (1911).
[26] C. J. Farmer, *J. Biol. Chem.*, **32**, 447 (1917).
[27] A. H. Eggerth, *J. Biol. Chem.*, **48**, 203 (1921).

according to its "alcohol number," *i.e.*, the percentage of alcohol in the solvent used to dissolve the collodion. Figure 8 shows the relationship between alcohol number and the diameter of the pores. The diameter is proportional to $\sqrt[4]{QL}$, where Q is the volume of liquid passing through the filter in unit time, and L is the thickness of the membrane.

Duclaux and Errera [28] studied the mechanism of ultrafiltration and pointed out that the velocity of flow of liquid through the pores of the membrane is inversely proportional to the viscosity of the liquid and directly proportional to the pressure, the membrane behaving exactly

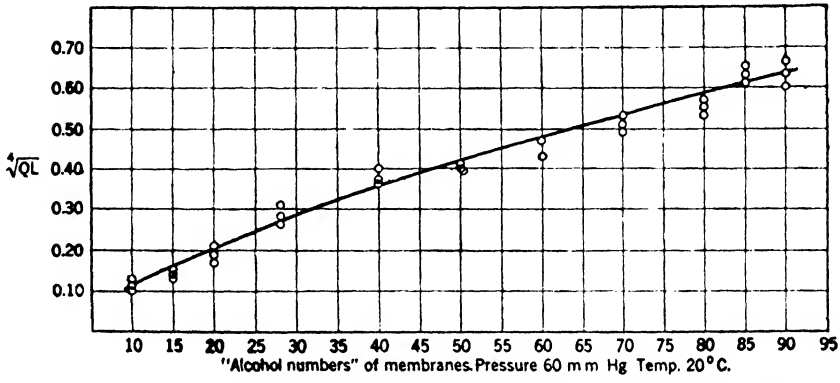


FIG. 8. Showing the relation between permeability of ultrafilters and Eggerth's alcohol numbers. (Data of Eggerth.)

like a bundle of fine capillary tubes. On the other hand, Brinkman and Szent-Györgyi [29] pointed out that surface tension or interfacial tension may profoundly alter the characteristics of a collodion ultrafilter. They found that, when a hemoglobin sol was placed in a collodion bag and subjected to a pressure of 3 atmospheres, the hemoglobin was completely retained and only water passed the membrane. If, however, a dilute solution of sodium oleate was first passed through the filter, the membrane became permeable to the hemoglobin. The hemoglobin which passed the sodium oleate-treated membrane would not subsequently pass through a second untreated membrane. Accordingly, the size of the micelles in the hemoglobin sol which had passed through the treated membrane had not been affected by the sodium oleate. The size of the pores in the treated membrane had not been altered, as was shown by rate of water filtration through the pores before and after the sodium oleate treatment. Sodium caproate, which has little or no effect on interfacial tension, did not alter the properties of the collodion mem-

[28] J. Duclaux and J. Errera, *Rev. gén. colloïdes*, **2**, 130 (1924).

[29] R. Brinkman and A. v. Szent-Györgyi, *Biochem. Z.*, **139**, 261 (1923).

brane. Sodium linoleate, sodium glycocholate, digitonin, α-monooleinglycerol ester, and Witte's peptone, which lower the interfacial tension in varying degrees, affect the permeability of the collodion membrane in the same ratio as they affect interfacial tension.

In a second paper [30] they studied the effects of alkaloids and purine bases on permeability of ultrafilters, using atropine, pilocarpine, caffeine, strychnine, quinine, and morphine. A change in the permeability of the collodion membrane toward hemoglobin was noted. Biologically inactive codeine did not alter the permeability; neither did cocaine nor novococaine. These results are very striking, and, as Brinkman and Szent-Györgyi suggest, may have a fundamental bearing on reactions which take place in living organisms.

Varney and Bronfenbrenner [31] studied Kendall's K medium in which Kendall claims to secure a "filterable form" of *Bacillus typhosus*. They grew the bacteria on agar slants, suspended them in sterile K medium, and filtered it at once. The bacteria in the K medium passed through the filter; those in broth did not. If sterile K medium was first passed through the filter and then the organisms in broth or saline filtered through the same ultrafilter, the organisms passed through. This is another illustration of the importance of surface relationships in ultrafiltration or in ordinary filtration.

Lundsgaard and Holbøll [32] investigated the preparation, standardization, and calibration of collodion membranes. They used a rather novel method of determining porosity. A glucose solution of known concentration was placed in a collodion bag or tube and allowed to dialyze against pure water for a definite interval of time. Analysis for glucose in the inner and outer liquids was made at intervals. When the volume of liquid inside the membrane was equal to the volume of liquid outside the membrane, the following equation could be applied:

$$\frac{dx}{dt} = \Delta\left(\frac{K}{2} - x\right) \tag{23}$$

where K = the original concentration of glucose in the inner liquid
x = the increase in glucose concentration in the outer liquid in time, t
Δ = diffusion coefficient of glucose for a given membrane.

Accordingly,

$$\Delta = \frac{1}{t} \log \frac{K}{K - 2x} \tag{24}$$

[30] R. Brinkman and A. v. Szent-Györgyi, *Biochem. Z.*, **139**, 270 (1923).
[31] P. L. Varney and J. Bronfenbrenner, *Proc. Soc. Exptl. Biol. Med.*, **29**, 804 (1932)
[32] C. Lundsgaard and S. A. Holbøll, *J. Biol. Chem.*, **68**, 439 (1926).

It was found from experiment that the original concentration of glucose (within limits ranging from 0.1 per cent to 0.3 per cent concentration) made no difference in the value obtained for the diffusion coefficient. Lundsgaard and Holbøll prepared membranes of definite porosity by coating glass tubes with collodion and immersing the collodion membrane, when partly dry, in alcohol of known concentration for a fixed period of time, and then immersing it in water. When alcohol of 70 per cent concentration was employed, the diffusion coefficient of glucose through the membrane ranged from $\Delta = 0.0133$ to $\Delta = 0.0122$ as maximum and minimum ranges (eight membranes tested). When 80 per cent alcohol was used, the maximum and minimum ranges were $\Delta = 0.0097$ to $\Delta = 0.0093$; and when 90 per cent alcohol was used, similar values were $\Delta = 0.0058$ to $\Delta = 0.0057$. They stated that the diffusion coefficient was not altered after the membranes had been aged for one week. Hitchcock [33] applied the viscosity equation to the measurement of the size of pores in collodion membranes. In the membranes with which he worked, the radius of the pores ranged from 20.8 mμ to 2.7 mμ. He pointed out that in the former case there were 70 billion capillary tubes per square centimeter and in the latter case 2700 billion capillary tubes per square centimeter.

Bechhold and Heymann [34] used ultrafiltration to concentrate gelatin sols, the dispersions medium passing through and the gelatin micelles being retained on the filter. In this way the ash content was greatly reduced. They stated that by using membranes of varying porosity they were able to separate gelatin into two fractions, one of which passed through the more porous membrane.

It is often desirable to test the porosity of a membrane by more or less qualitative methods. This can be done by making use of a series of colloidal sols ranging from a Prussian blue sol to a truly crystalloidal solution. The following list of materials has been suggested by various workers as offering a graded series of particle size:

Prussian blue > Bredig's platinum sol > casein in milk > As_2S_3 sol > Zsigmondy's red gold sol > 1 per cent gelatin sol > hemoglobin sol > litmus > nuclear gold sol > crystalloids.

Krueger and Ritter [35] and Bauer and Hughes [36] give detailed directions for preparing graded ultrafiltration membranes and standardizing them for the study of biological problems. Table 7, taken from the paper

[33] D. I. Hitchcock, *J. Gen. Physiol.*, **9**, 755 (1926).
[34] H. Bechhold and E. Heymann, *Biochem. Z.*, **171**, 33 (1926).
[35] A. P. Krueger and R. C. Ritter, *J. Gen. Physiol.*, **13**, 409 (1930).
[36] J. H. Bauer and T. P. Hughes, *J. Gen. Physiol.*, **18**, 143 (1934).

of Bauer and Pickels [37] lists the approximate particle sizes of various viruses as determined by ultrafiltration technic. The ultrafiltration apparatus described by Bauer and Hughes has been found to be very serviceable in the author's laboratories. Other papers that may be advantageously consulted are that by Ferry [38] which contains a detailed literature review of the various types of ultrafilters, their structure, and their application to specific problems (296 papers are cited), and that

TABLE 7. APPROXIMATE PARTICLE SIZE OF VIRUSES AS DETERMINED BY FILTRATION THROUGH GRADED COLLODION MEMBRANES

(By permission, *J. Exptl. Med.*)

Virus	Estimated Particle Size, millimicrons	Virus	Estimated Particle Size, millimicrons
Vaccinia	125–175	Vesicular stomatitis	70–100
Canary pox	125–175	Fowl plague	60– 90
Lymphogranuloma inguinale	125–175	Rift Valley fever	23– 35
Rous sarcoma 1	100–150	Equine encephalomyelitis	20– 30
Ectromelia	100–150	St. Louis encephalitis	20– 30
Pseudorabies	100–150	Yellow fever	17– 25
Herpes	100–150	Louping ill	15– 20
Borna disease	85–125	Poliomyelitis	10– 15
Influenza, swine and human	80–120	Foot-and-mouth disease	8– 12

by Höber,[39] which contains data on membranes of various pore sizes with particular reference to the rate at which ions may pass across membranes.

It is sometimes impossible to use ultrafiltration procedure because collodion or gelatin membranes are altered by the liquid which one desires to filter. Since such membranes are colloid gels, they would be affected by hot solutions and by solutions which are definitely acid or alkaline. Manning [40] prepared metallic ultrafilters by plating nickel on phosphorbronze or nickel gauze under such conditions that the plated nickel is deposited in a porous state. His paper gives directions for current density and concentration of solutions which should be employed.

[37] J. H. Bauer and E. G. Pickels, *J. Exptl. Med.*, **64**, 503 (1936).
[38] J. D. Ferry, *Chem. Revs.*, **18**, 373 (1936).
[39] R. Höber, *Physiol. Rev.*, **16**, 52 (1936).
[40] John Manning, *J. Chem. Soc.*, May, 1926, 1127.

Such ultrafilters can be used with organic solvents and are not affected by relatively high temperatures.

An important contribution to the literature of ultrafiltration is the observation by Kramer,[41] who emphasized the role of the charge on the colloid membrane in studies of filtration. Kramer was interested particularly in the nature of the so-called filterable microorganisms and viruses which pass unaltered through the porcelain filters of the bacteriological laboratory. He pointed out that all Berkefeld and Pasteur-Chamberlain filters were made from siliceous materials and consequently possessed a negative charge. He found that colloidal dyestuffs possessing a positive charge would not pass such filters, whereas those which were negatively charged passed readily. For example, Victoria blue (+) was retained in a Berkefeld filter, whereas Congo red (−) readily passed through. He, accordingly, prepared filters similar to the Berkefeld filter but possessing a positive charge. Using plaster of Paris ($CaSO_4$) filters, he found no differentiation between acid and basic dyestuffs. When, however, a small amount (± 5 per cent) of calcium carbonate was added to the plaster of Paris prior to forming the filter, he obtained positively charged filters which retained the Congo red (−) sols and allowed the Victoria blue (+) to pass through, thus reversing the retainability of the Berkefeld filter. A still more striking observation of Kramer is that the bacteriophage of *Staphylococcus aureus*, the *Vibrio percolans* of Mudd, vaccine virus, and rabies virus are retained by the positively charged plaster of Paris filters, whereas they readily pass through the ordinary bacteriological filters. As we shall see later in a discussion of the electrical properties of colloids, the retention is probably due to the mutual precipitation of oppositely charged micelles. A further illustration is found in the observations of Mulvania [42] that the virus of tobacco mosaic will pass through a collodion ultrafilter normally impermeable to the virus, provided that the *p*H of the virus solution is changed to the acid side of the virus isoelectric point.

Vividiffusion. Abel, Rowntree, and Turner [43] applied the methods of dialysis and ultrafiltration to a study of the crystalloids present in the blood stream. Figure 9 shows the apparatus which they devised. Using a glass cylinder open at both ends, they inserted within the cylinder a number of collodion tubes. These tubes were connected in series by glass U-tubes, together with in-flow and out-flow tubes which passed through rubber stoppers closing the ends of the glass cylinder. With the tubes in place and filled with physiological salt solution and surrounded

[41] S. P. Kramer, *J. Gen. Physiol.*, **9**, 811 (1926).

[42] M. Mulvania, *Phytopath.*, **16**, 853 (1926).

[43] J. J. Abel, L. G. Rowntree, and B. B. Turner, *J. Pharmacol.*, **5**, 275 (1914)

with physiological salt solution, the in-flow tube was connected to the carotid artery of the experimental animal, the out-flow tube being connected to the femoral vein. Hirudin (leach extract, an anti-blood-coagulant) is injected into the blood before it passes from the carotid artery into the system of collodion tubes, thus preventing the blood from coagulating when it comes in contact with the glass connecting tubes. The liquid surrounding the collodion tubes is kept at blood temperature by means of a thermostat. Thus, it is possible to pass blood from the

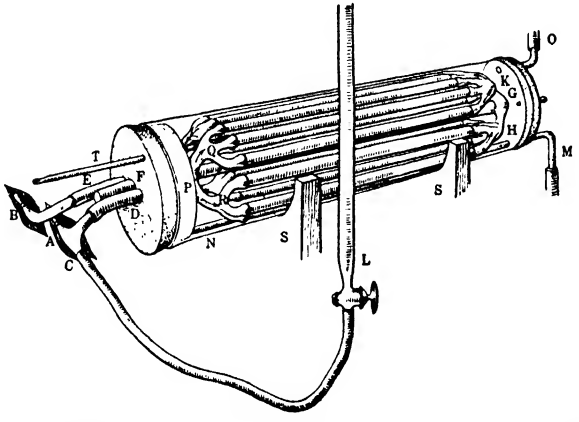


FIG. 9. Abel's vividiffusion apparatus. Vein at *B*, artery at *C*, anticoagulant added at *L*.

heart of an animal outside of the body through a dialyzing and ultrafiltration apparatus, back through the venous circulation into the animal body, and continue this cycle of blood circulation over relatively long periods of time.

The liquid on the outside of the collodion tubes may be a physiological salt solution or distilled water. If it is water, there is, of course, a rapid loss of the inorganic constituents of the blood. The student familiar with osmotic phenomena realizes that a crystalloid tends to collect in equal concentrations both within and without a membrane to which it is permeable. Accordingly the crystalloidal constituents of the blood pass through the collodion membrane into the outer liquid which can be drawn off at intervals and replaced with fresh liquid. In this way the crystalloidal constituents can be separated from the colloidal constituents of the blood.

Abel and his co-workers constructed various vividiffusion cells, the number of collodion tubes ranging from 2 to 32 or more. In some instances vividiffusion was kept up continuously for a period in excess of 10 hours. In preliminary experiments 1 gram of sodium salicylate was

slowly injected into the femoral vein of a dog weighing 7 kg. Of the total amount of salicylic acid injected, 19.1 per cent was recovered in the diffusion liquid of the vividiffusion apparatus, whereas only 17.5 per cent was eliminated in the urine during the same period of time, thus showing that the apparatus can compete with the kidneys on favorable terms, at least during a part of the dialysis period.

This apparatus gave definite proof that amino acids circulated in the blood stream in the free state. Abel demonstrated that there was a marked increase in amino acids in the diffusion liquid after the feeding of a meat diet to a dog which was being used for the vividiffusion experiment. This was apparently the first demonstration that proteins, on digestion, passed into the blood in the form of free amino acids and were carried in this form from the blood to the various cells and tissues of the body.

Electrodialysis. Dialysis which depends only upon diffusion is often an extremely slow process, and it is sometimes impossible to remove the last traces of adsorbed electrolytes. The removal of electrolytes from colloid sols and gels may be hastened and the colloidal material obtained in a much higher degree of purity by dialyzing with the aid of an electric current. Electrolytes in aqueous solutions are dissociated into ions, and these ions move under the influence of a direct current to the anode and the cathode. If the colloidal sol is restrained by suitable membranes, from moving to the anode and cathode compartments, electrolytes can be then almost completely removed. Sheppard, Sweet, and Benedict [44] used this method to prepare ash-free gelatin. Electrodialysis was carried out in a large wooden cell, in the middle of which was placed a porous inner compartment of silica which extended to and was imperviously cemented to both the bottom and the sides of the wooden cell. The silica container was filled with 5 per cent gelatin, the two end compartments of the wooden cell were filled with distilled water, and the whole electrodialyzed, using 110 volts, direct current, and a gold anode and a silver cathode. By this procedure gelatin containing from 0.02 to 0.03 per cent ash was obtained.

Knaggs, Manning, and Schryver [45] electrodialyzed gelatin by allowing a 10 to 20 per cent gelatin gel to set in the lower part of a bell jar, suspending this in water over a mercury cathode and inserting a platinum anode in water above the gelatin gel. The electrolysis was conducted with 100 to 220 volts, direct current, the water being changed periodically, and the process being continued until there was no further ap-

[44] S. E. Sheppard, S. S. Sweet, and A. J. Benedict, *J. Am. Chem. Soc.*, **44**, 1857 (1922).

[45] J. Knaggs, A. B. Manning, and S. B. Schryver, *Biochem. J.*, **17**, 473 (1923).

pearance of alkali at the cathode or acid at the anode. The ash content in this way was reduced to 0.02 per cent or less.

Hoffman and Gortner [46] used electrodialysis to study the composition of agar. They showed that the sulfur content of agar is organically bound in the molecule in the form of a substituted sulfuric acid, whereas the agar is freed from other inorganic elements. As usually obtained, agar is neutral in reaction. Electrodialyzed agar, on the other hand, is a strong acid, a 1 per cent solution having a *p*H of approximately 2.5. The silica which was present in the original agar did not pass through the membrane, indicating that it probably was in colloidal form. Thomas and Murray [47] electrodialyzed gum acacia and prepared an arabic acid, a 1 per cent solution of which had a *p*H value of 2.70. They found, however, that arabic acid is not a sulfuric acid ester, as is agar acid.

Since these studies, electrodialysis has come into almost universal use as a method through which colloidal systems can be readily and rapidly freed from electrolyte contamination. Pauli [48] was the first worker to use the technic extensively in the purification of proteins, in which field it is now a routine procedure. It has been applied extensively to the study of soil colloids and to the purification of cellulose, and in general it offers the best method for the purification of lyophilic colloids. Since electrolytes are frequently adsorbed on the surface of the colloid micelles and are held there by electrical forces, the electrolytes become to all intents and purposes a part of the colloidal micelle. It is only by means of an impressed electrical potential that the ions can be removed from the surfaces of the colloidal particles. Probably there is a characteristic "decomposition voltage" which must be impressed upon any particular system in order completely to remove the electrolytes from that system. Accordingly electrolytes which cannot be removed by ordinary dialysis or by ultrafiltration can be removed by electrodialysis.

Optical Properties of Colloid Systems. It is a common observation that a beam of light becomes visible when passed through a smoky atmosphere. The same is true when light is passed through a gold sol and through various other sols. The light is scattered from individual particles, and the cone of light which is visible is known as the *Tyndall cone.*

The scattering of light by high molecular weight substances has assumed considerable importance in relation to the determination of the molecular weight of such materials from the measurement of the amount

[46] W. F. Hoffman and R. A. Gortner, *J. Biol. Chem.*, **65**, 371 (1925).

[47] A. W. Thomas and H. A. Murray, *J. Phys. Chem.*, **32**, 676 (1928).

[48] Wo. Pauli and E. Valko, *Kolloidchemie der Eiweisskörper*, Theodor Steinkopff, Dresden and Leipzig, 1933.

of scattered light.[49-52] For colorless particles which are suspended in a colorless medium and which have a particle diameter $\frac{1}{10}$ or less that of the wave length of the incident light we have

$$\tau = \frac{32\pi^3}{3} \frac{\mu_0^2(\mu - \mu_0)^2}{\lambda^4} \frac{1}{n} \tag{25}$$

where τ = the turbidity and is defined as the fractional decrease of the intensity of the light passing through unit distance
μ_0 = the refractive index of the solvent
μ = the refractive index of the solution
λ = the wave length of light used
n = the number of particles per milliliter of solution.

Evidently, if we know the weight of the suspended matter per milliliter, we can calculate from n the molecular weight of the substance.

When the particle diameter is $\frac{1}{10}$ or less that of the wave length of the incident light, the intensity of the scattered light is the same in all directions. On the other hand, if the particle diameter is of the same order as or greater than the wave length of the light, the intensity of the scattered light is not uniform over all angles, but is greater in the forward direction than in the backward direction. This is a more complex situation, and equation (25) cannot be applied to this condition.

Equation (25) shows that, the shorter the wave length of the light, the greater is the amount of scattered light; it is a common experience that suspended particles frequently exhibit a blue tint.

The amount of scattered light also increases as the difference between the index of refraction of the solvent and that of the solution increases. If the two have the same index of refraction, no scattering of light occurs. Lyophobic colloids, as a rule, show intense Tyndall cones. On the other hand, lyophilic sols range from those which show a weak Tyndall cone to those which are practically optically clear. The reason for their transparency lies in the fact that they are highly solvated. A part of the dispersions medium is associated with the disperse phase; the disperse phase is greatly swollen by the dispersions medium which has "dissolved" in it, and accordingly its index of refraction has been brought very close to the index of refraction of the pure dispersions medium. There is therefore very little light refraction from the surface of the particles.

[49] M. von Smoluchowski, *Ann. Physik*, **25**, 205 (1908).
[50] A. Einstein, *Ann. Physik*, **33**, 1275 (1910).
[51] P. Debye. *J. Applied Phys.*, **15**, 338 (1944).
[52] P. M. Doty, B. H. Zimm, and H. Mark, *J. Chem. Phys.*, **13**, 159 (1945).

Live fresh-water medusae may contain as much as 98 per cent of water. When such an organism is floating in the water, it is in many instances practically invisible, because the rays of light pass nearly equally well through the water and through the organism. When it is removed from the water, it is, of course, plainly visible, owing to the difference in refractive index between the air and the organism.

The human eye is sensitive to only a narrow band of radiant energy. Accordingly, the Tyndall cones which we actually see are derived only from that portion of the light to which the retina is sensitive. When ultraviolet light is used as a source of illumination we are unable in many instances to see the Tyndall cone which results. Many systems which are optically empty to the eye may show a marked Tyndall cone when photographed under ultraviolet light. Certain lyophilic systems, which are optically empty with ordinary light, show markedly visible Tyndall cones when a beam of ultraviolet light is projected into them. This phenomenon is known as *fluorescence*, the ultraviolet light being transformed into visible wave lengths. Protein sols and gels in particular show such fluorescence. Svedberg and Tiselius [53] used the fluorescence of proteins to render visible the boundary between an egg-albumin sol and the surface of a solution which does not contain protein.

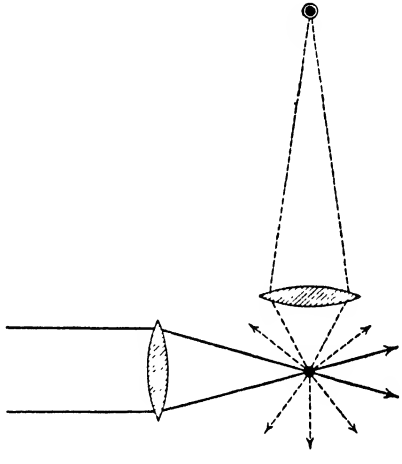


FIG. 10. The diffraction of light impinging upon an ultramicroscopic particle, the axes of the illuminating and diffracted rays being at right angles to each other. (Solid lines, illuminating rays; dotted lines, diffracted rays.)

Utilizing the general principles of the Tyndall phenomenon, Siedentopf and Zsigmondy [54] devised an instrument known as the *ultramicroscope*. If a powerful beam of light is focused to a point within a colloid sol and if the rays are refracted from the surface of the disperse phase, one may view the position and the Brownian movement of individual particles through a microscope. Figure 10 shows diagrammatically the path of such light rays from a light source to a colloid particle through a lens to the observer's eye. The observer, looking into an ultramicroscope, does not see the colloid particle. The particle is too small to be

[53] The Svedberg and A. Tiselius, *J. Am. Chem. Soc.*, **48**, 2272 (1926).

[54] H. Siedentopf and R. Zsigmondy, *Ann. Physik*, [4] **10**, 1 (1903).

seen with the magnification employed. Consequently the individual bright particles which are viewed in the ultramicroscope do not indicate either the form or the size of the ultramicroscopic particle; instead they act merely as mirrors, and the observer sees reflected from the surface of such a mirror the source of light which is used for illumination.

The limit of the size of the particle which becomes discernible in the ultramicroscope depends upon two factors, the intensity of illumination and the difference in refractive index existing between the disperse phase and the dispersions medium. A gold hydrosol is particularly adapted to ultramicroscopic studies because of the completeness with which light is reflected from a gold surface; with an egg-albumin sol the field is practically empty. Using direct sunlight and extreme precautions, Zsigmondy was able to differentiate particles of colloidal gold as small as 1.7 mμ from the water in which they were dispersed. The usual lower limit of resolution of the Siedentopf-Zsigmondy ultramicroscope lies at about 5 mμ under ordinary conditions of illumination.

There are several variations of the ultramicroscope. The illuminating beam may lie in the same axis as that of the observation microscope. This is accomplished through the use of a properly designed condenser for the microscope. A somewhat different lens system is that of Spierer, which is another example of a dark-field microscope. Seifriz [55] and Thiessen [56] have applied the Spierer lens to an examination of biological structures, with excellent results.

[55] Wm. Seifriz, *J. Phys. Chem.*, **35**, 118 (1931).

[56] R. Thiessen, *Ind. Eng. Chem.*, **24**, 1032 (1932).

CHAPTER 4

Hydrogen-Ion Concentration

In order to discuss adequately certain properties of colloid systems and to understand certain biological reactions, it is necessary to diverge at this point and consider briefly the subject of hydrogen-ion concentration.

It is impossible within the space at our disposal to cover the details regarding hydrogen-ion concentration, including the theoretical background and all the technic involved in the various methods used for measurement. We are particularly fortunate in having available such excellent treatises as those of Clark [1] and Michaelis,[2,3] which are indispensable adjuncts to the library of anyone working in this field.

Water is in many respects a unique substance. It is liquid at ordinary temperatures, whereas compounds more or less similar in structure, such as H_2S, SO_2, NH_3, are gaseous; in the liquid state it has a very high surface tension differentiating it from other liquids; it has a minimum volume at +4°C., expanding on solidifying; and, last but not least, solutions of many substances in water as a solvent possess the ability to conduct an electric current. If acids or bases or salts are dissolved in water, the solution becomes a conductor for an electric current. Accordingly, such substances are known as electrolytes.

Arrhenius, in 1887, postulated that when electrolytes are dissolved in water they are dissociated into their corresponding ions and that the flow of electric current takes place through these charged ions. The separation of an electrolyte into its component ions is known as *dissociation.*

Arrhenius noted that electrolytic conductance in solution is not strictly proportional to the amount of electrolyte dissolved. He accordingly suggested that at infinite dilution complete dissociation takes place, whereas in more concentrated solutions the dissociation is not entirely

[1] W. Mansfield Clark, *The Determination of Hydrogen Ions*, 3rd ed., Williams and Wilkins Co., Baltimore, 1928.

[2] L. Michaelis, *Hydrogen Ion Concentration*, Vol. I. Principles of the Theory, translated by W. A. Perlzweig, Williams and Wilkins Co., Baltimore, 1926.

[3] L. Michaelis, *Die Wasserstoffionenkonzentration*, Julius Springer, Berlin, 1914.

complete, part of the original solute remaining in solution in an undissociated condition. In accordance with this view, if we dissolve sodium chloride in water, the following equilibrium is set up:

$$NaCl \rightleftharpoons Na^+ \text{ and } Cl^-$$

In a dilute solution the equilibrium is shifted toward the right until at infinite dilution all the sodium chloride is dissociated. In a concentrated solution the equilibrium is shifted more and more toward the left.

The views of Arrhenius in regard to the dissociation of electrolytes have been altered somewhat by researches of physicists and physical chemists, which have thrown doubt on the existence of individual molecules, such as is represented by NaCl. The x-ray crystal structure of sodium chloride does not indicate the presence of a definite molecule of NaCl, but rather of ions of sodium and chlorine spaced at equal distances from each other throughout the crystal structure. Accordingly, it has been suggested that an electrolyte such as sodium chloride is completely dissociated even in the solid state. If this is so, it must be obvious that it is also completely dissociated in water. It is, therefore, theoretically impossible to have a solution in which the sodium chloride is 98 per cent dissociated. Solutions of sodium chloride can, however, be obtained which have only 98 per cent of the electrolytic conductance theoretically possible. Therefore, the term *activity* has largely displaced the term dissociation, when solutions of strong electrolytes are under consideration. On the assumption that dilute solutions of strong electrolytes are completely dissociated but that, as the concentration increases, the anions are more or less decreased in their activity by the adjacent cations, and correspondingly the activity of the cations may be altered by the adjacent anions, a theory has been built up which substitutes changes in activity for the older conception of changes in dissociation.

The above discussion is inserted at this point merely to indicate some of the more recent trends. It is more accurate to speak of *hydrogen-ion activity* than of *hydrogen-ion concentration*, and throughout the following discussion it will be well to bear in mind that the term hydrogen-ion concentration refers to the "apparent" hydrogen-ion concentration, *i.e.*, the hydrogen-ion activity, rather than to the actual normality of ionized hydrogen present in the solution, for the methods employed for the measurement of hydrogen-ion concentration in reality measure the so-called hydrogen-ion activity.[4]

[4] For those who wish to go further into the theories underlying the activity of ions, the following books are recommended: H. S. Taylor, Editor, *A Treatise on Physical Chemistry*, 2nd ed., D. Van Nostrand Co., New York, 1931; D. A. MacInnes, *The*

Conventionally we write the hydrogen ion by the symbol H^+, and in the remainder of this book we shall continue that symbolism. It must be emphasized, however, that this is only a symbol and does not represent the actual state of affairs, for in reality H^+ is a naked proton and as such is almost infinitesimal in size and accordingly is able to penetrate within the structure of the first molecule with which it comes in contact. In an aqueous system this would probably be a water molecule, so that the actual structure might be represented by $(H_3O)^+$, the *hydronium* (or oxonium) *ion.*[5]

Bernal and Fowler[6] point out that the mobilities of the "hydrogen" and hydroxyl ions are 32.5×10^{-4} and 17.8×10^{-4} cm./sec./volt/cm., respectively, and that all the other ions have mobilities in the neighborhood of 6.7×10^{-4} cm./sec./volt/cm. They account for this by suggesting that, if the ion OH_3^+ comes near an H_2O molecule, there is the possibility that the extra proton in the OH_3^+ can jump over to the other molecule. The net effect is as though the molecules had changed positions, but because of the small mass of the proton its transfer occurs at a higher rate than would be possible if the ions had the same mass. When an external electric field is set up for measuring mobility, it causes the protons to "hop" along from molecule to molecule in the direction of the electric field. In this way the body of the liquid acts as a "conductor of protons." A similar argument can be applied to explain the mobility of OH^- which in this case is simply a water molecule which has lost a proton.

Huggins[7,8] adds to this picture by pointing out that the hydronium ion must be regarded as not merely OH_3^+, because higher polymers, such as $O_2H_5^+$, $O_3H_7^+$, $\cdots$, are equally probable inasmuch as water exists in an indefinite number of polymers of H_2O.

If we have an acid of the type HA, its dissociation into H^+ and A^- is reversible and the dissociation may be expressed as

$$HA \rightleftharpoons H^+ \text{ and } A^-$$

Applying the equation for mass action and expressing concentration by

Principles of Electrochemistry, Reinhold Publishing Co., New York, 1939; H. S. Harned and B. B. Owen, *The Physical Chemistry of Electrolytic Solutions*, Reinhold Publishing Co., New York, 1943; S. Glasstone, *Textbook of Physical Chemistry*, 2nd ed., D. Van Nostrand Co., New York, 1946.

[5] I. M. Kolthoff, *Rec. trav. chim.*, **49**, 401 (1930).

[6] J. D. Bernal and R. H. Fowler, *J. Chem. Phys.*, **1**, 515 (1933).

[7] M. L. Huggins, *J. Phys. Chem.*, **40**, 723 (1936).

[8] M. L. Huggins, *J. Org. Chem.*, **1**, 407 (1936).

complete, part of the original solute remaining in solution in an undissociated condition. In accordance with this view, if we dissolve sodium chloride in water, the following equilibrium is set up:

$$NaCl \rightleftharpoons Na^+ \text{ and } Cl^-$$

In a dilute solution the equilibrium is shifted toward the right until at infinite dilution all the sodium chloride is dissociated. In a concentrated solution the equilibrium is shifted more and more toward the left.

The views of Arrhenius in regard to the dissociation of electrolytes have been altered somewhat by researches of physicists and physical chemists, which have thrown doubt on the existence of individual molecules, such as is represented by NaCl. The x-ray crystal structure of sodium chloride does not indicate the presence of a definite molecule of NaCl, but rather of ions of sodium and chlorine spaced at equal distances from each other throughout the crystal structure. Accordingly, it has been suggested that an electrolyte such as sodium chloride is completely dissociated even in the solid state. If this is so, it must be obvious that it is also completely dissociated in water. It is, therefore, theoretically impossible to have a solution in which the sodium chloride is 98 per cent dissociated. Solutions of sodium chloride can, however, be obtained which have only 98 per cent of the electrolytic conductance theoretically possible. Therefore, the term *activity* has largely displaced the term dissociation, when solutions of strong electrolytes are under consideration. On the assumption that dilute solutions of strong electrolytes are completely dissociated but that, as the concentration increases, the anions are more or less decreased in their activity by the adjacent cations, and correspondingly the activity of the cations may be altered by the adjacent anions, a theory has been built up which substitutes changes in activity for the older conception of changes in dissociation.

The above discussion is inserted at this point merely to indicate some of the more recent trends. It is more accurate to speak of *hydrogen-ion activity* than of *hydrogen-ion concentration*, and throughout the following discussion it will be well to bear in mind that the term hydrogen-ion concentration refers to the "apparent" hydrogen-ion concentration, *i.e.*, the hydrogen-ion activity, rather than to the actual normality of ionized hydrogen present in the solution, for the methods employed for the measurement of hydrogen-ion concentration in reality measure the so-called hydrogen-ion activity.[4]

[4] For those who wish to go further into the theories underlying the activity of ions, the following books are recommended: H. S. Taylor, Editor, *A Treatise on Physical Chemistry*, 2nd ed., D. Van Nostrand Co., New York, 1931; D. A. MacInnes, *The*

Conventionally we write the hydrogen ion by the symbol H^+, and in the remainder of this book we shall continue that symbolism. It must be emphasized, however, that this is only a symbol and does not represent the actual state of affairs, for in reality H^+ is a naked proton and as such is almost infinitesimal in size and accordingly is able to penetrate within the structure of the first molecule with which it comes in contact. In an aqueous system this would probably be a water molecule, so that the actual structure might be represented by $(H_3O)^+$, the *hydronium* (or oxonium) *ion*.[5]

Bernal and Fowler[6] point out that the mobilities of the "hydrogen" and hydroxyl ions are 32.5×10^{-4} and 17.8×10^{-4} cm./sec./volt/cm., respectively, and that all the other ions have mobilities in the neighborhood of 6.7×10^{-4} cm./scc./volt/cm. They account for this by suggesting that, if the ion OH_3^+ comes near an H_2O molecule, there is the possibility that the extra proton in the OH_3^+ can jump over to the other molecule. The net effect is as though the molecules had changed positions, but because of the small mass of the proton its transfer occurs at a higher rate than would be possible if the ions had the same mass. When an external electric field is set up for measuring mobility, it causes the protons to "hop" along from molecule to molecule in the direction of the electric field. In this way the body of the liquid acts as a "conductor of protons." A similar argument can be applied to explain the mobility of OH^- which in this case is simply a water molecule which has lost a proton.

Huggins[7,8] adds to this picture by pointing out that the hydronium ion must be regarded as not merely OH_3^+, because higher polymers, such as $O_2H_5^+$, $O_3H_7^+$, $\cdots$, are equally probable inasmuch as water exists in an indefinite number of polymers of H_2O.

If we have an acid of the type HA, its dissociation into H^+ and A^- is reversible and the dissociation may be expressed as

$$HA \rightleftharpoons H^+ \text{ and } A^-$$

Applying the equation for mass action and expressing concentration by

Principles of Electrochemistry, Reinhold Publishing Co., New York, 1939; H. S. Harned and B. B. Owen, *The Physical Chemistry of Electrolytic Solutions*, Reinhold Publishing Co., New York, 1943; S. Glasstone, *Textbook of Physical Chemistry*, 2nd ed., D. Van Nostrand Co., New York, 1946.

[5] I. M. Kolthoff, *Rec. trav. chim.*, **49**, 401 (1930).

[6] J. D. Bernal and R. H. Fowler, *J. Chem. Phys.*, **1**, 515 (1933).

[7] M. L. Huggins, *J. Phys. Chem.*, **40**, 723 (1936).

[8] M. L. Huggins, *J. Org. Chem.*, **1**, 407 (1936).

inclosing the quantities in brackets, we have the equation

$$\frac{[H^+] \times [A^-]}{[HA]} = K_a \tag{26}$$

where K_a = the ionization or dissociation constant.

Equation (26) states that for any given acid the product of the concentration of the anion, and the cation divided by the concentration of the remaining undissociated acid is a constant. If K_a is large, the greater part of the acid is dissociated into hydrogen ions and anions, whereas, if K_a is small, the greater part of the acid is present in the form of undissociated molecules. Accordingly the dissociation constant K_a is a measure of the relative strength of an acid.

In a similar way the dissociation of a base may be represented by

$$\frac{[B^+] \times [OH^-]}{[BOH]} = K_b \tag{27}$$

where K_b = the dissociation constant of a base and bears the same relation in regard to the alkalinity of a solution that K_a has with respect to the acidity.

Inasmuch as water is the common solvent and inasmuch as it dissociates somewhat into hydrogen and hydroxyl ions, one of which is characteristic of the dissociation of acids, the other of bases, the dissociation constant of water becomes a very important quantity in the calculation of hydrogen-ion concentration. The dissociation of water may be represented by

$$\frac{[H^+] \times [OH^-]}{[HOH]} = K_w \tag{28}$$

This equation states that the product of the concentration of hydrogen and hydroxyl ions divided by the concentration of the undissociated water is a constant. It is usually considered safe to ignore the denominator in this equation, inasmuch as it appears to be approximately constant. The equation accordingly is usually expressed as

$$[H^+] \times [OH^-] = K_w \tag{29}$$

where K_w = the dissociation constant of water.

From equation (29) we read that, no matter how great the concentration of hydrogen ions, there must always remain sufficient hydroxyl ions in the solution to satisfy the equation, and conversely, no matter how great the concentration of hydroxyl ions, there must always be present a residue of hydrogen ions to satisfy this equation. The value

for the dissociation constant of water is, therefore, a very important constant, and its value is approximately 10^{-14}. It is usually expressed as log $(1/K_w)$ which has a value of 13.995 at 22°C.[9] If K_w at 22°C. has a value of 1×10^{-14} and the hydrogen- and hydroxyl-ion concentrations are equal, pure water at this temperature is 1×10^{-7} (or one-ten-millionth) normal with respect to hydrogen and hydroxyl ions. One gram molecule of any substance contains approximately 6.061×10^{23} molecules. Accordingly in pure water there will be $6.061 \times 10^{23} \times 10^{-7}$ hydrogen ions per liter. One liter of water contains 55.56 gram molecules or $55.56 \times 6.061 \times 10^{23}$ water molecules. This same liter, on the other hand, contains only 6.061×10^{16} hydrogen ions. Accordingly only one molecule of water in every 555,000,000 molecules is dissociated into hydrogen and hydroxyl ions.

From equations (26), (27), and (29) it is possible to present an equation which applies equally well to all solutions of acids and bases. Equation (26) may be rewritten

$$\frac{1}{[H^+]} = \frac{[A^-]}{K_a[HA]} \tag{30}$$

or it may be expressed, by using the reciprocal of $[H^+]$ and taking the logarithm of each side of the equation, as

$$\log \frac{1}{[H^+]} = \log \frac{1}{K_a} + \log \frac{[A^-]}{[HA]} \tag{31}$$

The logarithm of the reciprocal of the hydrogen-ion concentration is expressed in the term *p*H. This term, which is commonly met in biological and biochemical studies where hydrogen-ion concentration is referred to, was first introduced by Sörensen [10,11] for reasons which will be referred to later.

By referring to equation (29) it will be noted that the term *p*H can be used equally well to express either the degree of acidity or the degree of alkalinity of a solution, inasmuch as that equation can be rewritten

$$\frac{1}{[H^+]} \times \frac{1}{[OH^-]} = \frac{1}{K_w} \tag{32}$$

or

$$p\mathrm{H} + p\mathrm{OH} = pK_w \tag{33}$$

[9] Clark, on p. 45 of *The Determination of Hydrogen Ions*, gives a table for the variation of log $(1/K_w)$ with temperature, together with the *p*H of the neutral point at the various temperatures.

[10] S. P. L. Sörensen, *Biochem. Z.*, **21**, 131 (1909).

[11] S. P. L. Sörensen, *Compt. rend. trav. lab. Carlsberg*, **8**, 1 (1909).

When a substance yielding H^+ or OH^- ions is added to water, the ionization of the water is repressed so that the hydrogen-ion concentration decreases as the hydroxyl-ion concentration increases, and *vice versa.* Therefore *a decrease in hydrogen-ion concentration may be used to express an increase in hydroxyl-ion concentration, thus permitting one scale to be used for the measurement of both acidity and alkalinity.*

The fact that the symbol *p*H can be used to designate either acidity or alkalinity makes it especially useful in biochemical studies where the degree of acidity or alkalinity ranges around the neutral point. The relationship between *p*H and the concentration of hydrogen or hydroxyl ions which may be present in solution is:

*p*H 1 = 0.1 *N*	Hydrogen Ions and 10^{-13} *N*	Hydroxyl Ions = *p*OH 13
*p*H 2 = 0.01 *N*	Hydrogen Ions and 10^{-12} *N*	Hydroxyl Ions = *p*OH 12
*p*H 3 = 0.001 *N*	Hydrogen Ions and 10^{-11} *N*	Hydroxyl Ions = *p*OH 11
*p*H 4 = 0.0001 *N*	Hydrogen Ions and 10^{-10} *N*	Hydroxyl Ions = *p*OH 10
*p*H 5 = 10^{-5} *N*	Hydrogen Ions and 10^{-9} *N*	Hydroxyl Ions = *p*OH 9
*p*H 6 = 10^{-6} *N*	Hydrogen Ions and 10^{-8} *N*	Hydroxyl Ions = *p*OH 8
*p*H 7 = 10^{-7} *N*	Hydrogen Ions and 10^{-7} *N*	Hydroxyl Ions = *p*OH 7
*p*H 8 = 10^{-8} *N*	Hydrogen Ions and 10^{-6} *N*	Hydroxyl Ions = *p*OH 6
*p*H 9 = 10^{-9} *N*	Hydrogen Ions and 10^{-5} *N*	Hydroxyl Ions = *p*OH 5
*p*H 10 = 10^{-10} *N*	Hydrogen Ions and 0.0001 *N*	Hydroxyl Ions = *p*OH 4
*p*H 11 = 10^{-11} *N*	Hydrogen Ions and 0.001 *N*	Hydroxyl Ions = *p*OH 3
*p*H 12 = 10^{-12} *N*	Hydrogen Ions and 0.01 *N*	Hydroxyl Ions = *p*OH 2
*p*H 13 = 10^{-13} *N*	Hydrogen Ions and 0.1 *N*	Hydroxyl Ions = *p*OH 1
*p*H 14 = 10^{-14} *N*	Hydrogen Ions and 1.0 *N*	Hydroxyl Ions = *p*OH 0

An additional reason for using the symbol *p*H is that we can express, in 14 units, acidities which range from a solution which is 0.1 normal in terms of hydrogen ions to a solution which is one one-hundred trillionth normal (10^{-14}) in terms of hydrogen ions. If we desired to plot the actual hydrogen-ion concentration, C_H, or the hydrogen-ion activity, C_{Ha}, allowing 1 mm. for the unit between 10^{-13} normal and 10^{-14} normal, we should need a piece of paper 111,111 kilometers or approximately 69,444 miles long in order to include the whole range of C_H between *p*H 1 and *p*H 14. In terms of *p*H the entire graph could be placed on an ordinary sheet of coordinate paper. *It is essential to remember that the pH scale is a logarithmic scale.* Accordingly *p*H 6 represents a solution containing 10 times the concentration of hydrogen ions that are present in a solution having a *p*H of 7. A solution with a *p*H of 5 is 100 times as acid as a solution with a *p*H of 7. A solution with a *p*H of 4 is 1,000 times as acid as a solution with a *p*H of 7. Likewise *the fractions in a pH scale do not represent arithmetical values but rather logarithmic values.* Thus, the difference in acidity between *p*H 5.0 and 5.1 is many times greater than that between *p*H 5.9 and *p*H 6.0.

Expressing acidity or alkalinity in terms of the concentration of the

hydrogen or hydroxyl ions or in terms of *p*H permits us to distinguish between the strong acids or bases and the so-called weak acids or bases. Hydrochloric acid and acetic acid ionize respectively as follows:

$$HCl \rightleftharpoons H^+ \text{ and } Cl^-$$

$$HC_2H_3O_2 \rightleftharpoons H^+ \text{ and } C_2H_3O_2^-$$

Hydrochloric acid, however, is a strong acid, and ion concentration measurements indicate that all, or practically all, the hydrogen ions in a 0.1 normal solution of hydrochloric acid are active. On the other hand, only a small fraction of the available hydrogen ions in acetic acid are active (less than 1 per cent in an 0.1 normal solution). We accordingly say that only a small fraction of the acetic acid is dissociated, the major portion remaining as undissociated acetic acid molecules.

If we were to titrate tenth-normal solutions of acetic and hydrochloric acids with standard sodium hydroxide, using a suitable indicator, we should find that both acids required the same amount of sodium hydroxide for neutralization. The amount of standard sodium hydroxide which would be required represents the *titratable acidity*, but this determination tells us nothing in regard to the *actual acidity* of the solution at any particular period of time. *The titratable acidity is a measure of the potential acidity, i.e., the quantity of hydrogen ions which can be made to combine with a base. Hydrogen-ion concentration measurement, on the other hand, is a measure of the actual concentration (or activity) of hydrogen ions present in a given system at a given time, and is therefore the only true measure of how such a solution will affect another system which is sensitive to hydrogen or hydroxyl ions.* Hydrochloric acid is a strong acid, and its hydrogen ions in solution approach ideal activity. A normal solution of hydrochloric acid is a poison, whereas a normal solution of acetic acid is a fairly weak artificial vinegar. Both solutions have the same potential (*i.e.*, titratable) acidity, but their hydrogen-ion activities are widely different.

The reactions of the biological organism toward acids and bases are not controlled by changes in potential acidity but rather are affected by changes in the actual activity of hydrogen ions. Table 8 shows the relationship between the potential acidity and the actual acidity of solutions of certain of the common acids.

Figures 11 and 12 [12] illustrate the differences which may be observed in biochemical systems which are sensitive to the action of acids. In Fig. 11 the viscosity of a 20 per cent wheat flour-in-water suspension is plotted against the normality of the various acids which were added to

[12] R. A. Gortner and P. F. Sharp, *J. Phys. Chem.*, **27**, 481 (1923).

the suspension. It will be noted that there is a great difference in the form of the viscosity curves with acid concentration. However, when the above data were plotted not against the potential acidity of the solution but against the actual acidity, *i.e.*, the hydrogen-ion concentration, Fig. 12 resulted. Instead of a maximum being reached at various points on the curve, as in Fig. 11, the maximum viscosity is in all instances

TABLE 8. THE APPROXIMATE RELATIONSHIP BETWEEN THE NORMALITY OF ACID SOLUTIONS (POTENTIAL ACIDITY) AND THE *p*H (THE NEGATIVE LOG OF THE ACTUAL ACIDITY) AT 25°C.*

Normality of Solution	*p*H of Acid Solution at 25°C.				
	Lactic	Acetic	Ortho-phosphoric	Hydro-chloric	Oxalic
0.500	1.89	2.50	1.33	0.46	1.03
0.200	2.11	2.69	1.59	0.81	1.29
0.100	2.27	2.85	1.79	1.09	1.52
0.040	2.49	3.08	2.05	1.46	1.82
0.020	2.67	3.24	2.26	1.77	2.10
0.010	2.82	3.39	2.49	2.06	2.36
0.005	3.00	3.54	2.73	2.35	2.62
0.002	3.23	3.76	3.08	2.73	3.00

* The values of *p*H are probably not exact for use as physicochemical constants. They represent, however, actual values obtained on the solutions in question by using electrometric methods of determining hydrogen-ion concentration. These values should not be used except as first approximations and are inserted here principally for the purpose of illustrating the variation which may be obtained in actual acidity when solutions of the same potential acidity are compared.

reached at an acidity of approximately *p*H 3.0, and the acids instead of behaving in entirely dissimilar manners yield curves of essentially the same general shape. It is obvious that the normality of the acid which was added was not the true variable affecting the system, but that the hydrogen-ion concentration was in reality the variable concerned.

Another illustration of the importance of *p*H in biological problems is shown in Fig. 13, taken from the paper of Fife and Frampton,[13] who found that there is a hydrogen-ion concentration gradient in the leaf cells of the sugar beet and that the cells of the phloem are decidedly more basic than the parenchyma tissue. Leaf hoppers which transmit

[13] J. M. Fife and V. L. Frampton, *J. Agr. Research*, **53**, 581 (1936).

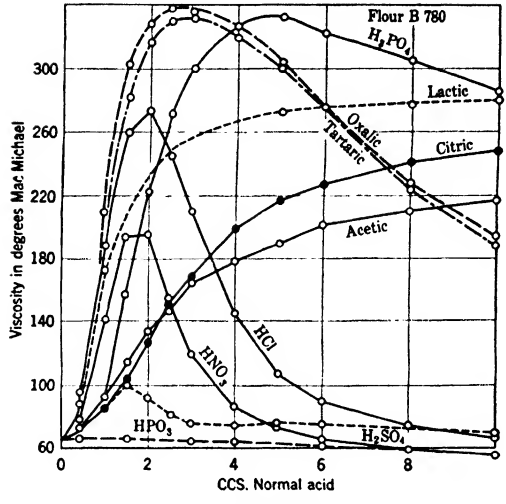


FIG. 11. Showing the viscosity of flour-in-water systems plotted against the normality of solutions of the various acids. (Data of Gortner and Sharp.)

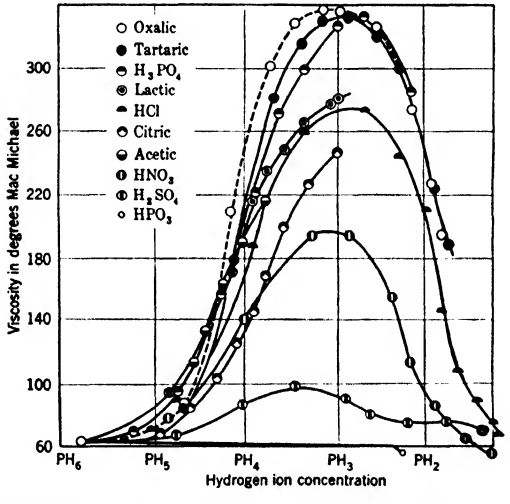


FIG. 12. Showing the same data as Fig. 11, plotted in terms of actual hydrogen-ion concentration.

the virus of the curly-top disease normally feed on the sap of the phloem, and their saliva has approximately the same *p*H as normal phloem sap. When the *p*H of the phloem sap is altered by treatment of the leaves with carbon dioxide, the leaf hoppers lose their sense of direction, their mouth parts do not penetrate into the phloem, and little or no infection with curly-top virus occurs.

It is beyond the scope of this book to more than mention the methods by which hydrogen-ion concentration can be measured. As noted at the

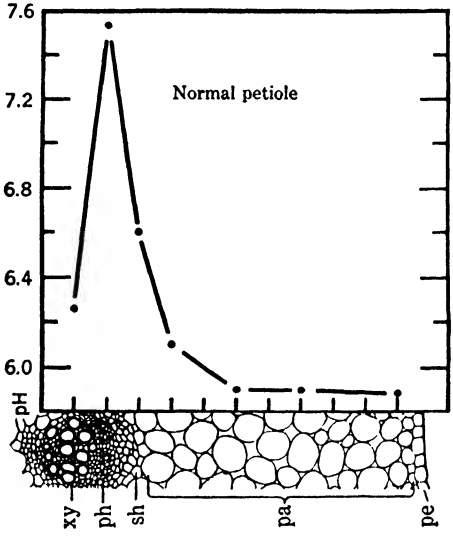


FIG. 13. Showing curve of *p*H gradient in the normal petiole cells of the sugar beet. (Data of Fife and Frampton.)

beginning of this chapter, the excellent manuals of Clark and Michaelis are adequate in all respects. The two methods generally employed are colorimetric measurements and electrometric measurements.

Colorimetric Measurements. In the measurement of hydrogen-ion concentration by colorimetric methods one is concerned with the change in color of an indicator as a measurement of the change in hydrogen-ion concentration. Indicators are in general compounds which form salts with either acids or bases and yield on dissociation at least one colored ion. Accordingly the degree of dissociation of the indicator is altered by changes in the hydrogen- or hydroxyl-ion concentration. The concentration of the colored ion is accordingly altered, thus causing a change in the depth of color of the indicator as measured in a colorimeter. Various indicators have been suggested for the various *p*H ranges. Table 9 shows the indicators which are most suitable, together with their color changes from acid to base and the *p*H range in which they can be used.

TABLE 9. INDICATORS RECOMMENDED FOR THE COLORIMETRIC DETERMINATION OF HYDROGEN-ION CONCENTRATION TOGETHER WITH THEIR USABLE *pH* RANGE AND CERTAIN OTHER PHYSICOCHEMICAL CONSTANTS *

Chemical Name	Common Name	Molecular Weight	*A*	*pK*	Range, *pH*	Color Change		*B*, *pH*	*C*, *pH*	Absorption Max	
						Acid	Alkaline			Acid, mμ	Alkaline, mμ
3,3′-Dimethyl phenolsulfonphthalein	Meta cresol purple	382	26.2	1.51	1.2–2.8	Red	Yellow	conc. HCl	6	533	
2,2′-Diisopropyl-5,5′-dimethyl phenolsulfonphthalein	Thymol blue	466	21.5	1.5	1.2–2.8	Red	Yellow	conc. HCl	6	544	
2,2′,6,6′-Tetrabromophenolsulfonphthalein	Bromophenol blue	669	14.9	3.98	3.0–4.6	Yellow	Blue	0	7	...	592
2,2′,6,6′-Tetrabromo-3,3′-dimethyl phenolsulfonphthalein	Bromocresol green	698	14.3	4.67	3.8–5.4	Yellow	Blue	1	8	..	617
2,2′-Dichlorophenolsulfonphthalein	Chlorophenol red	423	23.6	5.98	4.8–6.4	Yellow	Red	2	9	...	573
o-Carboxy benzene azo dimethylaniline	Methyl red	279	35.6	....	4.4–6.0	Red	Yellow				
2,2′-Dibromophenolsulfonphthalein	Bromophenol red	512	19.5	6.16	5.2–6.8	Yellow	Red	3	10	...	574
6,6′-Dibromo-2,2′-dimethyl phenolsulfonphthalein	Bromocresol purple	540	18.5	6.3	5.2–6.8	Yellow	Purple	3	10	...	591
6,6′-Dibromo-2,2′-diisopropyl-5,5′-dimethyl phenolsulfonphthalein	Bromothymol blue	624	16.0	7.0	6.0–7.6	Yellow	Blue	4	10	...	617
Phenolsulfonphthalein	Phenol red	354	28.2	7.9	6.8–8.4	Yellow	Red	5	11	...	558
2,2′-Dimethyl phenolsulfonphthalein	Cresol red	382	26.2	8.3	7.2–8.8	Yellow	Red	5	11	...	572
3,3′-Dimethyl phenolsulfonphthalein	Meta cresol purple	382	26.2	8.32	7.4–9.0	Yellow	Purple	5	11	...	580
2,2′-Diisopropyl-5,5′-dimethyl phenolsulfonphthalein	Thymol blue	466	21.5	8.9	8.0–9.6	Yellow	Blue	6	12	...	596
o-Cresol phthalein	Cresol phthalein	...	. .	(9.4)	8.2–9.8	Colorless	Red	6	12		

A = milliliters of 0.01 *N* NaOH required per 0.1 gram indicator to form mono sodium salt. Dilute to 250 ml. for 0.04 per cent reagent.
B = approximate *pH* value of solution required for full "acid color" appertaining to *pH* range indicated.
C = approximate *pH* value of solution required for full "alkaline color" appertaining to *pH* range indicated.
* From W. M. Clark, *The Determination of Hydrogen Ions*, 3rd ed., Williams and Wilkins Co., Baltimore, 1928. By permission.

the virus of the curly-top disease normally feed on the sap of the phloem, and their saliva has approximately the same *p*H as normal phloem sap. When the *p*H of the phloem sap is altered by treatment of the leaves with carbon dioxide, the leaf hoppers lose their sense of direction, their mouth parts do not penetrate into the phloem, and little or no infection with curly-top virus occurs.

It is beyond the scope of this book to more than mention the methods by which hydrogen-ion concentration can be measured. As noted at the

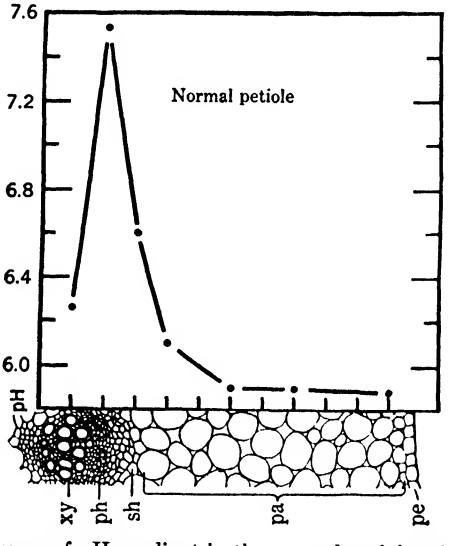


FIG. 13. Showing curve of *p*H gradient in the normal petiole cells of the sugar beet. (Data of Fife and Frampton.)

beginning of this chapter, the excellent manuals of Clark and Michaelis are adequate in all respects. The two methods generally employed are colorimetric measurements and electrometric measurements.

Colorimetric Measurements. In the measurement of hydrogen-ion concentration by colorimetric methods one is concerned with the change in color of an indicator as a measurement of the change in hydrogen-ion concentration. Indicators are in general compounds which form salts with either acids or bases and yield on dissociation at least one colored ion. Accordingly the degree of dissociation of the indicator is altered by changes in the hydrogen- or hydroxyl-ion concentration. The concentration of the colored ion is accordingly altered, thus causing a change in the depth of color of the indicator as measured in a colorimeter. Various indicators have been suggested for the various *p*H ranges. Table 9 shows the indicators which are most suitable, together with their color changes from acid to base and the *p*H range in which they can be used.

TABLE 9. INDICATORS RECOMMENDED FOR THE COLORIMETRIC DETERMINATION OF HYDROGEN-ION CONCENTRATION TOGETHER WITH THEIR USABLE *p*H RANGE AND CERTAIN OTHER PHYSICOCHEMICAL CONSTANTS *

Chemical Name	Common Name	Molecular Weight	*A*	*pK*	Range, *p*H	Color Change		*B*, *p*H	*C*, *p*H	Absorption Max	
						Acid	Alkaline			Acid, mμ	Alkaline, mμ
3,3′-Dimethyl phenolsulfonphthalein	Meta cresol purple	382	26.2	1.51	1.2–2.8	Red	Yellow	conc. HCl	6	533	
2,2′-Diisopropyl-5,5′-dimethyl phenolsulfonphthalein	Thymol blue	466	21.5	1.5	1.2–2.8	Red	Yellow	conc. HCl	6	544	
2,2′,6,6′-Tetrabromophenolsulfonphthalein	Bromophenol blue	669	14.9	3.98	3.0–4.6	Yellow	Blue	0	7	. . .	592
2,2′,6,6′-Tetrabromo-3,3′-dimethyl phenolsulfonphthalein	Bromocresol green	698	14.3	4.67	3.8–5.4	Yellow	Blue	1	8	. .	617
2,2′-Dichlorophenolsulfonphthalein	Chlorophenol red	423	23.6	5.98	4.8–6.4	Yellow	Red	2	9	. . .	573
o-Carboxy benzene azo dimethylaniline	Methyl red	279	35.6	. . .	4.4–6.0	Red	Yellow				
2,2′-Dibromophenolsulfonphthalein	Bromophenol red	512	19.5	6.16	5.2–6.8	Yellow	Red	3	10	. . .	574
6,6′-Dibromo-2,2′-dimethyl phenolsulfonphthalein	Bromocresol purple	540	18.5	6.3	5.2–6.8	Yellow	Purple	3	10	. . .	591
6,6′-Dibromo-2,2′-diisopropyl-5,5′-dimethyl phenolsulfonphthalein	Bromothymol blue	624	16.0	7.0	6.0–7.6	Yellow	Blue	4	10	. . .	617
Phenolsulfonphthalein	Phenol red	354	28.2	7.9	6.8–8.4	Yellow	Red	5	11	. .	558
2,2′-Dimethyl phenolsulfonphthalein	Cresol red	382	26.2	8.3	7.2–8.8	Yellow	Red	5	11	. . .	572
3,3′-Dimethyl phenolsulfonphthalein	Meta cresol purple	382	26.2	8.32	7.4–9.0	Yellow	Purple	5	11	. . .	580
2,2′-Diisopropyl-5,5′-dimethyl phenolsulfonphthalein	Thymol blue	466	21.5	8.9	8.0–9.6	Yellow	Blue	6	12	. . .	596
o-Cresol phthalein	Cresol phthalein	.	. . .	(9.4)	8.2–9.8	Colorless	Red	6	12		

A = milliliters of 0.01 *N* NaOH required per 0.1 gram indicator to form mono sodium salt. Dilute to 250 ml. for 0.04 per cent reagent.
B = approximate *p*H value of solution required for full "acid color" appertaining to *p*H range indicated.
C = approximate *p*H value of solution required for full "alkaline color" appertaining to *p*H range indicated.
* From W. M. Clark, *The Determination of Hydrogen Ions*, 3rd ed., Williams and Wilkins Co., Baltimore, 1928. By permission.

The phenolsulfonphthalein group of indicators are triphenylmethane derivatives containing a sulfophenyl and two phenol radicals, the latter being attached in their para positions to the methane carbon.

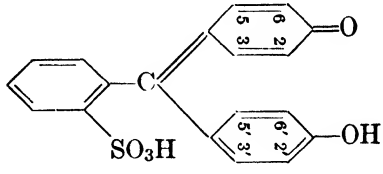


Figure 14 shows the dissociation curves of some of the indicators noted in Table 9. In general an indicator is most useful when it is approximately 50 per cent dissociated.

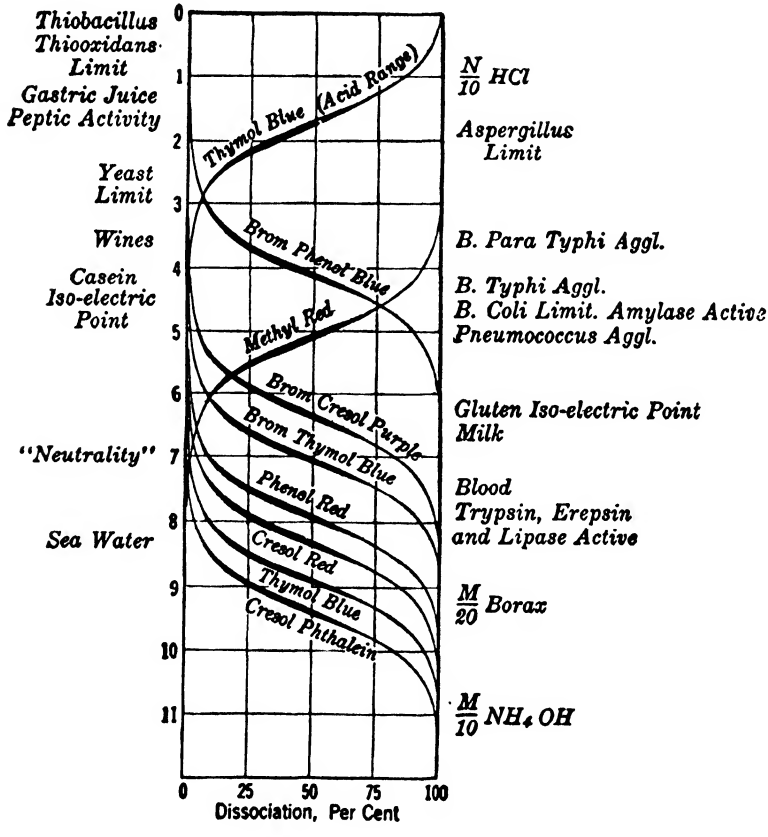


FIG. 14. The dissociation curves for the most useful indicators for the colorimetric estimation of hydrogen-ion concentration. The shaded areas indicate the useful range. (From Clark's *Determination of Hydrogen Ions*. By permission.)

Electrometric Measurements. Electrometric determinations are based on the assumption that ions in solution obey the gas laws. If we allow a metal electrode to dip into a solution containing ions of that metal, the partial pressure of the ions in solution will be proportional to their ionic concentration, and the work of transferring one mole of ions to the electrode will be

$$\text{Work} = D - RT \log_e C \tag{34}$$

where D = a constant involving the nature of the metal electrode
R = the gas constant
T = the absolute temperature
C = the concentration of the ions.

The work of transfer is equal to the quantity of electricity transferred multiplied by the potential at which the transfer is made. Substituting, we have

$$nFE = D - RT \log_e C \tag{35}$$

where n = the number of equivalents of ions
F = the quantity of electricity in one equivalent, which is 96,500 coulombs
E = the potential of the electrode.

Then

$$E = E_0 - \frac{RT}{nF} \log_e C \tag{36}$$

where E_0 = a new constant which is equal to D/nF. When the ions are at unit activity, E is equal to E_0.

The hydrogen electrode is nothing more than a metallic electrode which has been coated with a thin layer of platinum black deposited on an inactive metal and exposed to an atmosphere of hydrogen sufficiently long for the platinum black to become completely saturated with hydrogen. When such an electrode is placed in a solution containing hydrogen ions, it reaches an equilibrium with the solution and behaves as if it were an electrode composed of metallic hydrogen.

It is evident that, in order to realize a potential which can be measured, two electrodes must be connected together. Each of these electrodes is called a half cell. The two solutions are joined together by a salt bridge which is usually a glass tube containing an agar gel saturated with potassium chloride. The difference in electrical potential between the metal electrodes is then measured with a potentiometer.

The hydrogen electrode has been selected as the standard against which all other electrodes are compared, and the E_0 (electromotive force at unit activity and at 760 mm. pressure of hydrogen) has been set equal to zero.

To measure the hydrogen-ion concentration with a hydrogen electrode, the hydrogen electrode is used in conjunction with a calomel half cell. This half cell consists of mercury in contact with a potassium chloride solution saturated with mercurous chloride. The electromotive force observed is then

$$E = E_1 - E_2 \tag{37}$$

where E_1 is the potential due to the calomel half cell and E_2 is that due to the hydrogen half cell. Substituting equation (36) into equation (37), we have

$$E - E_1 = \frac{RT}{F} \log_e \frac{1}{H^+} \tag{38}$$

or, at 25°C.,

$$p\mathrm{H} = \frac{E - E_1}{0.059} \tag{39}$$

The value of E_1 depends on the concentration of potassium chloride in the calomel electrode. For a 0.10 N solution, E_1 is +0.3355 volt whereas, if the solution is saturated with potassium chloride, E_1 is +0.2443 volt.

For various reasons, the hydrogen electrode for the determination of pH is rather inconvenient, and at the present time pH is measured by other electrodes. For a while the quinhydrone electrode was extensively employed. This electrode consists of an inert electrode, such as a bright platinum wire, immersed in a solution of quinhydrone, an equimolar mixture of quinone and of hydroquinone. Quinone is the oxidized form of hydroquinone, and the two together form a reversible oxidation-reduction system the potential of which is a function of the pH of the solution. The glass electrode has proved to have so many advantages over all other types of electrodes that it is now used almost exclusively for the purpose of determining pH.

The glass electrode consists of a glass membrane of suitable composition, on one side of which is a non-polarizable electrode, such as silver-silver chloride. On the other side is the unknown solution, which is connected by means of a salt bridge to another non-polarizable electrode. The arrangement is shown in Fig. 15.

The voltage of the glass electrode is a straight-line function of the pH. The slope of the line varies with the temperature, and at 25°C.

the voltage changes 59 mv. for a change of one *p*H unit. It is necessary to calibrate the glass electrode with a standard buffer. The glass membrane functions as though it were a membrane permeable only to hydrogen ions. The membrane offers a tremendous resistance, and a vacuum tube potentiometer is used to measure the electromotive force produced.

The advantages of the glass electrode are: (1) the glass electrode is independent of oxidation-reduction potentials; (2) it is not necessary to

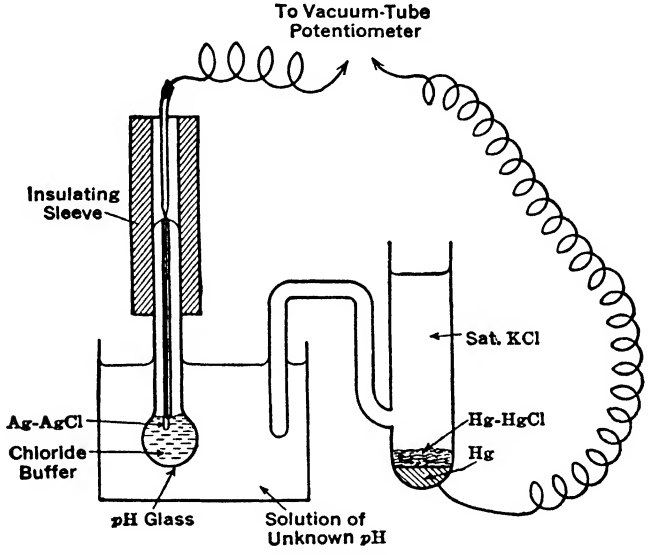


FIG. 15. Glass electrode for the measurement of *p*H.

pass a gas through the solution or to add any material to it; (3) it is possible to use very small quantities of solution; (4) the electrode can be used in colored or turbid solutions; (5) the electrode gives accurate values in unbuffered solutions; (6) equilibrium is rapidly reached.

Dole [14] has written a very complete and useful book dealing with the glass electrode. The glass electrode, complete with compact vacuum tube potentiometer, can be purchased from a number of supply houses.

Buffers. Van Slyke [15] has defined buffers as "*substances which by their presence in solution increase the amount of acid or alkali that must be added to cause unit change in pH.*" The most efficient buffers are mixtures of weak acids or weak bases with their corresponding salts.

[14] M. Dole, *The Glass Electrode*, John Wiley & Sons, New York, 1941.
[15] D. D. Van Slyke, *J. Biol. Chem.*, **52**, 525 (1922).

As noted above, a weak acid or a weak base is characterized by the fact that a large proportion of the material in solution is present in the form of undissociated molecules, and accordingly the dissociation constant is low. In acetic acid only a small fraction of the potential acidity is present at any one time, the dissociation constant having a value of approximately 1.8×10^{-5}. The equilibrium equation can therefore be written

$$\frac{[H^+] \times [C_2H_3O_2^-]}{[HC_2H_3O_2]} = 1.8 \times 10^{-5}$$

The value 1.8×10^{-5} is constant regardless of the concentration of acetic acid in the solution. If more acetic acid is added to the solution, the concentration of undissociated molecules is increased, and a small part of the added acetic acid dissociates to maintain the above equilibrium. If another substance with a common ion, for example, sodium acetate, is added, the concentration of the added acetate ions must be taken into consideration in the above equation, and, since this would cause an increase in the anion portion of the numerator, there would be a corresponding decrease in the hydrogen-ion concentration.

Buffers may also be defined as *those substances which prevent sudden or great changes in hydrogen-ion concentration when strong acids or bases are added to a system.* If one adds hydrochloric acid to a solution of trisodium phosphate, the following reactions take place.

$$Na_3PO_4 + HCl \rightarrow NaCl + Na_2HPO_4$$

$$Na_2HPO_4 + HCl \rightarrow NaCl + NaH_2PO_4$$

$$NaH_2PO_4 + HCl \rightarrow NaCl + H_3PO_4$$

If it is assumed that the original solution of trisodium phosphate is a 0.3 *N* solution and that one equivalent of hydrochloric acid is added to it, the solution of disodium phosphate which results from the adding of the strongly acid hydrochloric acid (pH = 1) would not only be not acid but would actually be alkaline and would have a pH of approximately 8.5 to 9.

If, to this slightly alkaline solution, we again add an equivalent of hydrochloric acid, the resulting monosodium phosphate solution would have a pH of approximately 4.5, and if to this solution of monosodium phosphate a third equivalent of hydrochloric acid is added, the resulting solution of phosphoric acid would have a pH of approximately 2.5. Three equivalents of hydrochloric acid, having a pH of 1.0, have thus been needed to shift the hydrogen-ion concentration of trisodium phosphate from approximately pH = 10.5 to approximately pH = 2.5.

In a similar way the reaction between sodium acetate and hydrochloric acid may be written

$$NaC_2H_3O_2 + HCl \rightarrow NaCl + H \cdot C_2H_3O_2$$

and in this case the hydrochloric acid solution having a *p*H of 1.0 has been "buffered" by the sodium acetate, so that the resulting mixture has a *p*H of approximately 2.5. Sodium carbonate, likewise, acts as a buffer according to the reaction

$$Na_2CO_3 + HCl \rightarrow NaCl + NaHCO_3$$

and, on the addition of a second equivalent of hydrochloric acid,

$$NaHCO_3 + HCl \rightarrow NaCl + H_2CO_3$$

Carbonic acid is such a weak acid that it decomposes according to the equation

$$H_2CO_3 \rightarrow CO_2 + H_2O$$

and the carbon dioxide may be given off from the solution in the form of gas bubbles. Thus, we have here conditions where two equivalents of a strong acid can be added to a slightly alkaline solution, with a resulting solution which is essentially neutral. *This is one of the reactions by which the animal body is protected against sudden changes in hydrogen-ion concentration.* The blood contains a considerable amount of bicarbonate. If diluted hydrochloric acid is injected intravenously, the acid is buffered by the bicarbonate of the blood stream; the carbon dioxide which is formed is eliminated by the lungs; the hydrochloric acid is converted into sodium chloride, a normal constituent of the blood; a part of the remaining bicarbonate dissociates so as to keep a constant K_b; and there is no appreciable change in the hydrogen-ion concentration of the blood stream. Very considerable amounts of strong acids can accordingly be neutralized by the body through the action of the normal buffers of the blood and tissues. If, however, we were to continue the injection of dilute hydrochloric acid until the buffers in the blood stream had been exhausted, there would suddenly occur a marked rise in the hydrogen-ion concentration of the blood, a violent physiological reaction, and if sufficient hydrochloric acid were injected, death would ensue. *All reactions of living protoplasm take place in buffered media.* Carbonates, bicarbonates, and phosphates are the principal buffers encountered in biological processes, although proteins may under certain conditions act as relatively efficient buffers.

The same sort of reactions which take place in the buffering of a strong acid are involved when strong bases are employed. Thus, for example, ammonium acetate will act as a buffer for sodium hydroxide according to the reaction

$$NH_4C_2H_3O_2 + NaOH \rightarrow NaC_2H_3O_2 + NH_4OH$$

where the strong base, sodium hydroxide, has been buffered and replaced by the weak ammonium hydroxide, or, using a phosphate,

$$NaH_2PO_4 + NaOH \rightarrow Na_2HPO_4 + H_2O$$

Most biological reactions take place in an essentially neutral or slightly acid medium, and as a rule biological organisms have a greater capacity for the buffering of acid solutions than for the buffering of bases.

An apparent exception to the general rule that the tissues of living organisms are characterized by hydrogen-ion concentrations in the region of neutrality is afforded by the observations of Kobayashi,[16] who studied the blood plasma and corpuscles of the ascidian, *Chelyosoma siboja*. He found that the body fluid had a *p*H of 1.54, the plasma a *p*H of 1.80, and the corpuscle fluid a *p*H of 0.38. These high acidities were due to the presence of free sulfuric acid which ranged from an $SO_4^=$ concentration in the plasma of 3.95 to 6.53 grams per liter, in the body fluid of 8.33 to 23.50 grams per liter, and in the corpuscle fluid of 42.4 to 55.29 grams per liter, making concentrations of approximately 0.88 *N* in the corpuscle fluid, 0.36 *N* in the body fluid, and 0.027 *N* in the plasma. Further biochemical and physiological studies on such an unusual condition in living organisms are highly to be desired.

Van Slyke [17] has proposed a unit for the measurement of buffering values. The unit adopted is the differential ratio dB/dpH, which expresses the relationship between the increment in gram equivalents per liter of a strong base (B) added to a buffer solution, and the resultant increment in *p*H. Correspondingly for the acid range the increment of strong acid is equivalent to a negative increment of the base ($-dB$). In these terms a solution has a buffer value of 1.0 when a liter will take up a gram equivalent of strong acid or alkali per unit change in *p*H.

As Van Slyke points out, if a base is added to a solution, the *p*H is increased so that both dB and dpH are positive. If an acid is added, dB

[16] S. Kobayashi, *Science Repts. Tôhoku Imp. Univ.* (4th ser.), **8**, 277 (1933); *cf.* also M. Henze, *Z. physiol. Chem.*, **72**, 494 (1911); **79**, 215 (1912); **86**, 340 (1913); S. Hecht, *Am. J. Physiol.*, **45**, 157 (1917).

[17] D. D. Van Slyke, *J. Biol. Chem.*, **52**, 525 (1922).

and dpH both are negative. However, the ratio dB/dpH always has a positive numerical value. If one solution has twice the buffer value of a second solution, it will require twice as much acid or base to change the pH of the former through a unit range. The value of dB/dpH is therefore twice as great for the first solution as for the second. Van Slyke suggests that the symbol β be used to indicate the ratio dB/dpH. Figure 16, taken from the paper by Van Slyke, represents the buffer value of the two solutions referred to in this paragraph.

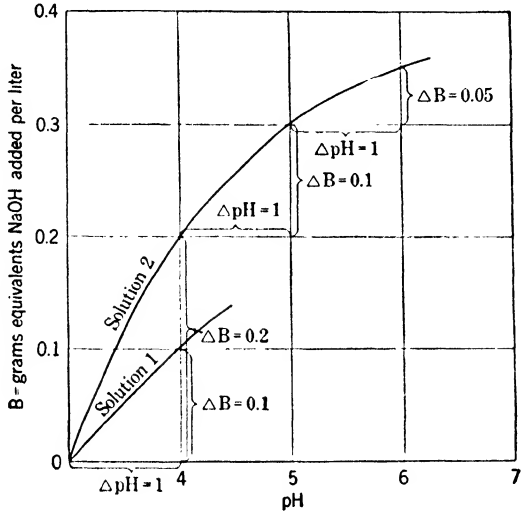


FIG. 16. Showing the relationship between buffer capacity (ΔB) and change in hydrogen-ion concentration. (Data of Van Slyke.)

It would take us too far afield to consider adequately the various factors which must be taken into consideration in a study of buffer action. The manuals by Clark and Michaelis are adequate for a general understanding of the principles involved, and the paper by Van Slyke which is noted above gives an extended discussion of the theories involved. The average biochemist or biologist is usually concerned not so much with the theory underlying the general principle as he is with the practical application of the principle to laboratory problems. Perhaps the greatest application of the principles of buffer action, aside from the interpretation of the resistance of living organisms to changes in acidity or alkalinity, lies in the ability to prepare, from known mixtures of chemicals, solutions which have a definite and stable pH value. Table 10 lists the composition of buffer mixtures having a given pH value at 20°C.

TABLE 10. COMPOSITION OF MIXTURES GIVING *p*H VALUES AT 20°C. AT INTERVALS OF 0.2 *

KCL–HCl Mixtures †

*p*H			
1.2	50 ml. *M*/5 KCl	65.5 ml. *M*/5 HCl	Dilute to 200 ml.
1.4	50 ml. *M*/5 KCl	41.5 ml. *M*/5 HCl	Dilute to 200 ml.
1.6	50 ml. *M*/5 KCl	26.3 ml. *M*/5 HCl	Dilute to 200 ml.
1.8	50 ml. *M*/5 KCl	16.6 ml. *M*/5 HCl	Dilute to 200 ml.
2.0	50 ml. *M*/5 KCl	10.6 ml. *M*/5 HCl	Dilute to 200 ml.
2.2	50 ml. *M*/5 KCl	6.7 ml. *M*/5 HCl	Dilute to 200 ml.

Phthalate–HCl Mixtures

*p*H			
2.2	50 ml. *M*/5 KH Phthalate	46.70 ml. *M*/5 HCl	Dilute to 200 ml.
2.4	50 ml. *M*/5 KH Phthalate	39.60 ml. *M*/5 HCl	Dilute to 200 ml.
2.6	50 ml. *M*/5 KH Phthalate	32.95 ml. *M*/5 HCl	Dilute to 200 ml.
2.8	50 ml. *M*/5 KH Phthalate	26.42 ml. *M*/5 HCl	Dilute to 200 ml.
3.0	50 ml. *M*/5 KH Phthalate	20.32 ml. *M*/5 HCl	Dilute to 200 ml.
3.2	50 ml. *M*/5 KH Phthalate	14.70 ml. *M*/5 HCl	Dilute to 200 ml.
3.4	50 ml. *M*/5 KH Phthalate	9.90 ml. *M*/5 HCl	Dilute to 200 ml.
3.6	50 ml. *M*/5 KH Phthalate	5.97 ml. *M*/5 HCl	Dilute to 200 ml.
3.8	50 ml. *M*/5 KH Phthalate	2.63 ml. *M*/5 HCl	Dilute to 200 ml.

Phthalate–NaOH Mixtures

*p*H			
4.0	50 ml. *M*/5 KH Phthalate	0.40 ml. *M*/5 NaOH	Dilute to 200 ml.
4.2	50 ml. *M*/5 KH Phthalate	3.70 ml. *M*/5 NaOH	Dilute to 200 ml.
4.4	50 ml. *M*/5 KH Phthalate	7.50 ml. *M*/5 NaOH	Dilute to 200 ml.
4.6	50 ml. *M*/5 KH Phthalate	12.15 ml. *M*/5 NaOH	Dilute to 200 ml.
4.8	50 ml. *M*/5 KH Phthalate	17.70 ml. *M*/5 NaOH	Dilute to 200 ml.
5.0	50 ml. *M*/5 KH Phthalate	23.85 ml. *M*/5 NaOH	Dilute to 200 ml.
5.2	50 ml. *M*/5 KH Phthalate	29.95 ml. *M*/5 NaOH	Dilute to 200 ml.
5.4	50 ml. *M*/5 KH Phthalate	35.45 ml. *M*/5 NaOH	Dilute to 200 ml.
5.6	50 ml. *M*/5 KH Phthalate	39.85 ml. *M*/5 NaOH	Dilute to 200 ml.
5.8	50 ml. *M*/5 KH Phthalate	43.00 ml. *M*/5 NaOH	Dilute to 200 ml.
6.0	50 ml. *M*/5 KH Phthalate	45.45 ml. *M*/5 NaOH	Dilute to 200 ml.
6.2	50 ml. *M*/5 KH Phthalate	47.00 ml. *M*/5 NaOH	Dilute to 200 ml.

* From W. M. Clark, *The Determination of Hydrogen Ions*, Williams and Wilkins Co., Baltimore, 1928. By permission.

† The *p*H values of these mixtures are given by Clark and Lubs (1916) as *preliminary* measurements.

TABLE 10. COMPOSITION OF MIXTURES GIVING *p*H VALUES AT 20°C. AT INTERVALS OF 0.2 (*Continued*)

KH_2PO_4–NaOH Mixtures

*p*H			
5.8	50 ml. *M*/5 KH_2PO_4	3.72 ml. *M*/5 NaOH	Dilute to 200 ml.
6.0	50 ml. *M*/5 KH_2PO_4	5.70 ml. *M*/5 NaOH	Dilute to 200 ml.
6.2	50 ml. *M*/5 KH_2PO_4	8.60 ml. *M*/5 NaOH	Dilute to 200 ml.
6.4	50 ml. *M*/5 KH_2PO_4	12.60 ml. *M*/5 NaOH	Dilute to 200 ml.
6.6	50 ml. *M*/5 KH_2PO_4	17.80 ml. *M*/5 NaOH	Dilute to 200 ml.
6.8	50 ml. *M*/5 KH_2PO_4	23.65 ml. *M*/5 NaOH	Dilute to 200 ml.
7.0	50 ml. *M*/5 KH_2PO_4	29.63 ml. *M*/5 NaOH	Dilute to 200 ml.
7.2	50 ml. *M*/5 KH_2PO_4	35.00 ml. *M*/5 NaOH	Dilute to 200 ml.
7.4	50 ml. *M*/5 KH_2PO_4	39.50 ml. *M*/5 NaOH	Dilute to 200 ml.
7.6	50 ml. *M*/5 KH_2PO_4	42.80 ml. *M*/5 NaOH	Dilute to 200 ml.
7.8	50 ml. *M*/5 KH_2PO_4	45.20 ml. *M*/5 NaOH	Dilute to 200 ml.
8.0	50 ml. *M*/5 KH_2PO_4	46.80 ml. *M*/5 NaOH	Dilute to 200 ml.

Boric Acid–KCl–NaOH Mixtures

*p*H			
7.8	50 ml. *M*/5 H_3BO_3, *M*/5 KCl	2.61 ml. *M*/5 NaOH	Dilute to 200 ml.
8.0	50 ml. *M*/5 H_3BO_3, *M*/5 KCl	3.97 ml. *M*/5 NaOH	Dilute to 200 ml.
8.2	50 ml. *M*/5 H_3BO_3, *M*/5 KCl	5.90 ml. *M*/5 NaOH	Dilute to 200 ml.
8.4	50 ml. *M*/5 H_3BO_3, *M*/5 KCl	8.50 ml. *M*/5 NaOH	Dilute to 200 ml.
8.6	50 ml. *M*/5 H_3BO_3, *M*/5 KCl	12.00 ml. *M*/5 NaOH	Dilute to 200 ml.
8.8	50 ml. *M*/5 H_3BO_3, *M*/5 KCl	16.30 ml. *M*/5 NaOH	Dilute to 200 ml.
9.0	50 ml. *M*/5 H_3BO_3, *M*/5 KCl	21.30 ml. *M*/5 NaOH	Dilute to 200 ml.
9.2	50 ml. *M*/5 H_3BO_3, *M*/5 KCl	26.70 ml. *M*/5 NaOH	Dilute to 200 ml.
9.4	50 ml. *M*/5 H_3BO_3, *M*/5 KCl	32.00 ml. *M*/5 NaOH	Dilute to 200 ml.
9.6	50 ml. *M*/5 H_3BO_3, *M*/5 KCl	36.85 ml. *M*/5 NaOH	Dilute to 200 ml.
9.8	50 ml. *M*/5 H_3BO_3, *M*/5 KCl	40.80 ml. *M*/5 NaOH	Dilute to 200 ml.
10.0	50 ml. *M*/5 H_3BO_3, *M*/5 KCl	43.90 ml. *M*/5 NaOH	Dilute to 200 ml.

TABLE 11. COMPOSITION OF SÖRENSEN'S BUFFER MIXTURES

(*a*) Mixtures of Glycine * and Hydrochloric Acid †

Volume of Glycine Solution, ml.	Volume of HCl, ml.	*p*H of Mixture	Volume of Glycine Solution, ml.	Volume of HCl, ml.	*p*H of Mixture
10.00	....	6.106	6.00	4.00	2.279
9.90	0.10	4.411	5.00	5.00	1.932
9.75	0.25	3.991	4.00	6.00	1.645
9.50	0.50	3.679	3.00	7.00	1.419
9.00	1.00	3.341	2.00	8.00	1.251
8.00	2.00	2.922	1.00	9.00	1.146
7.00	3.00	2.607	. . .	10.00	1.038

(*b*) Mixtures of Glycine * and Sodium Hydroxide ‡

Volume of Glycine Solution, ml.	Volume of NaOH, ml.	*p*H of Mixture	Volume of Glycine Solution, ml.	Volume of NaOH, ml.	*p*H of Mixture
10.00	....	6.106	5.10	4.90	11.067
9.90	0.10	7.809	5.00	5.00	11.305
9.75	0.25	8.237	4.90	5.10	11.565
9.50	0.50	8.575	4.50	5.50	12.095
9.00	1.00	8.929	4.00	6.00	12.399
8.00	2.00	9.364	3.00	7.00	12.674
7.00	3.00	9.714	2.00	8.00	12.856
6.00	4.00	10.140	1.00	9.00	12.972
5.50	4.50	10.482	. . .	10.00	13.066

* 7.505 grams of glycine and 5.85 grams NaCl in 1 liter of solution.

† An exactly 0.1 *N* solution of hydrochloric acid.

‡ An exactly 0.1 *N* solution of carbonate-free NaOH.

TABLE 11. COMPOSITION OF SÖRENSEN'S BUFFER MIXTURES (*Continued*)

(*c*) Mixtures of Sodium Monohydrogen Phosphate § and Potassium Dihydrogen Phosphate ‖

Volume of Na_2HPO_4 Solution, ml.	Volume of KH_2PO_4 Solution, ml.	*p*H of Mixture	Volume of Na_2HPO_4 Solution, ml.	Volume of KH_2PO_4 Solution, ml.	*p*H of Mixture
10.00	....	8.302	4.00	6.00	6.643
9.90	0.10	8.171	3.00	7.00	6.468
9.75	0.25	8.038	2.00	8.00	6.239
9.50	0.50	7.863	1.00	9.00	5.910
9.00	1.00	7.648	0.50	9.50	5.600
8.00	2.00	7.347	0.25	9.75	5.305
7.00	3.00	7.146	0.10	9.90	4.976
6.00	4.00	6.976		10.00	4.529
5.00	5.00	6.813			

(*d*) Mixtures of Sodium Borate ¶ and 0.1 *N* Hydrochloric Acid

Volume of Borate Solution, ml.	Volume of HCl, ml.	*p*H of Mixture	Volume of Borate Solution, ml.	Volume of HCl, ml.	*p*H of Mixture
10.00	....	9.241	6.50	3.50	8.506
9.50	0.50	9.168	6.00	4.00	8.289
9.00	1.00	9.087	5.75	4.25	8.137
8.50	1.50	9.007	5.50	4.50	7.939
8.00	2.00	8.908	5.25	4.75	7.621
7.50	2.50	8.799	5.00	5.00	6.548
7.00	3.00	8.678	4.75	5.25	2.371

§ 11.876 grams $Na_2HPO_4 \cdot 2H_2O$ in 1 liter of solution.

‖ 9.078 grams KH_2PO_4 in 1 liter of solution.

¶ 12.404 grams of boric acid dissolved in 100 ml. of exactly 1.0 *N* carbonate-free NaOH and diluting the mixture to 1 liter volume.

TABLE 11. COMPOSITION OF SÖRENSEN'S BUFFER MIXTURES (*Continued*)

(*e*) Mixtures of Sodium Borate ** and 0.1 *N* Sodium Hydroxide

Volume of Borate Solution, ml.	Volume of NaOH, ml.	*p*H of Mixture	Volume of Borate Solution, ml.	Volume of NaOH, ml.	*p*H of Mixture
10.00	....	9.241	6.00	4.00	9.974
9.00	1.00	9.360	5.00	5.00	11.076
8.00	2.00	9.503	4.00	6.00	12.376
7.00	3.00	9.676			

(*f*) Mixture of Sodium Citrate †† and 0.1 *N* Hydrochloric Acid

Volume of Citrate Solution, ml.	Volume of HCl, ml.	*p*H of Mixture	Volume of Citrate Solution, ml.	Volume of HCl, ml.	*p*H of Mixture
10.00	....	4.958	4.75	5.25	3.529
9.50	0.50	4.887	4.50	5.50	3.364
9.00	1.00	4.830	4.00	6.00	2.972
8.00	2.00	4.652	3.33	6.67	2.274
7.00	3.00	4.447	3.00	7.00	1.925
6.00	4.00	4.158	2.00	8.00	1.418
5.50	4.50	3.948	1.00	9.00	1.173
5.00	5.00	3.692	....	10.00	1.038

** Borate solution of same composition as in Section (*d*) of this table.

†† 21.008 grams of citric acid ($C_6H_8O_7 \cdot H_2O$) dissolved in 200 ml. of carbonate-free 1.0 *N* NaOH solution and the mixture diluted to 1 liter.

TABLE 11. COMPOSITION OF SÖRENSEN'S BUFFER MIXTURES (*Continued*)

(*g*) Mixtures of Sodium Citrate ‡‡ and 0.1 *N* Sodium Hydroxide

Volume of Citrate Solution, ml.	Volume of NaOH, ml.	*p*H of Mixture	Volume of Citrate Solution, ml.	Volume of NaOH, ml.	*p*H of Mixture
10.00	....	4.958	5.50	4.50	6.331
9.50	0.50	5.023	5.25	4.75	6.678
9.00	1.00	5.109	5.00	5.00	9.052
8.00	2.00	5.314			10.092
7.00	3.00	5.568	4.50	5.50	12.073
6.00	4.00	5.969	4.00	6.00	12.364

‡‡ Same concentration as solution used in Section (*f*) of this table.

Sörensen [18] gives a series of buffer mixtures different from those suggested by Clark. The compositions of Sörensen's mixtures are shown in Table 11. Temple [19] recommends sodium maleate as a very efficient buffer for the region 5.2–6.8.

As noted above, proteins may act to a certain extent as buffers. Their efficiency as buffers depends to a very great extent on the chemical nature of the amino acids making up the protein molecule. Figure 17 [20] shows the change in hydrogen-ion concentration of two typical proteins when acid or alkali is added to the system. In durumin, the prolamine of durum wheat, there is no evidence of buffer action toward sodium hydroxide, and not much evidence of buffer action toward hydrochloric acid. In casein from milk there is a definite buffering of added sodium hydroxide in the range from approximately *p*H 6.0 to *p*H 7.0, with a slight but less pronounced buffering effect against acid in the range of approximately *p*H 4.0 to *p*H 3.0. On the whole, however, proteins may be regarded as fairly inefficient buffers, and the generalization noted above that carbonates, bicarbonates, and phosphates are the

[18] S. P. L. Sörensen, *Compt. rend. trav. lab. Carlsberg*, **8**, 41, 42, 43 (1909). Compare also I. M. Kolthoff, *J. Biol. Chem.*, **63**, 135 (1925).

[19] J. W. Temple, *J. Am. Chem. Soc.*, **51**, 1754 (1929).

[20] W. F. Hoffman and R. A. Gortner, Colloid Symposium Monograph, Vol. II, p. 209, Chemical Catalog Co., New York, 1925.

principal buffers of biological systems still holds. Dunne [21] discusses in some detail the relation of buffer systems to the problems of plant physiology.

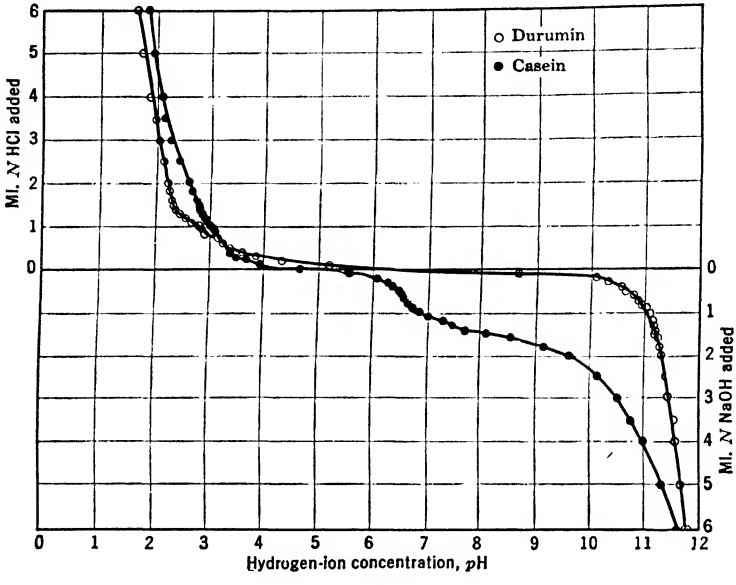


FIG. 17. Buffer curves of durumin and casein in various concentrations of hydrochloric acid and sodium hydroxide. (Data of Hoffman and Gortner.)

Biochemical experiments frequently cover a wide range of *p*H values and extend beyond the ranges of the buffer action of any one system. It

TABLE 12. VERONAL BUFFER SYSTEM

(Data of Michaelis)

a	*p*H	*a*	*p*H
0.0	9.64	6.5	6.12
0.25	9.16	7.0	5.32
0.5	8.90	8.0	4.93
0.75	8.68	9.0	4.66
1.0	8.55	10.0	4.33
2.0	8.18	11.0	4.13
3.0	7.90	12.0	3.88
4.0	7.66	13.0	3.62
5.0	7.42	14.0	3.20
5.5	7.25	15.0	2.62
6.0	6.99		

[21] T. C. Dunne, *Hilgardia*, [5] **7**, 207 (1932).

then becomes necessary to use two or more of the systems to cover the desired range of *p*H. This introduces certain specific ion effects which may assume considerable physiological and biochemical importance, and the effect of the variation of hydrogen ions may be lost. Michaelis [22] proposed a "*universal*" *buffer* which has an adequate buffer capacity over a very wide range of *p*H values. The steps in making up this buffer system are: 9.714 grams of sodium acetate (containing 3 molecules of water of hydration) and 14.714 grams of the sodium salt of veronal are dissolved in water and made up to 500 ml. To each 5 ml. of this solution is added 2 ml. of an 8.5 per cent sodium chloride solution plus a ml. of 0.1 N hydrochloric acid and $18 - a$ ml. of water. Table 12 gives a and the corresponding *p*H. All the *p*H values shown in Table 12 are at constant ionic strength and are isotonic with blood. The veronal system has the advantage that it has no insoluble calcium salt. The phosphate system cannot be used in the presence of appreciable calcium ions because calcium phosphate has a very limited solubility.

[22] L. Michaelis, *Biochem. Z.*, **234**, 139 (1931).

CHAPTER 5

Oxidation-Reduction [1]

We shall limit this short discussion to a consideration of electronic oxidation-reduction and attempt only to outline the general theory of such systems, together with the experimental technic involving the use of indicators and electrometric methods followed by a mention of some of the more important biological systems.[2]

The general term of oxidation-reduction is in some respects an unfortunate one. It is true that many reactions involving oxygen come under this heading, for example,

$$2HCHO + O_2 \rightarrow 2HCOOH$$

but, on the other hand, there are a large number of reactions which are termed oxidations and which do not involve oxygen:

$$FeCl_2 + Cl \rightleftharpoons FeCl_3$$

In the above reaction, the ferrous ion has been said to be oxidized to the ferric state, although no oxygen has been used. Oxidation in this case involves the loss of an electron:

$$Fe^{++} \rightleftharpoons Fe^{+++} + (e)$$

In the same reaction the chlorine was reduced; it gained an electron. We can generalize this experience. If any component of a system is re-

[1] This chapter is contributed by Dr. H. B. Bull, Professor of Chemistry in Northwestern University Medical School; formerly on the staff of the Division of Agricultural Biochemistry, University of Minnesota.

[2] For more extended discussions of the field of oxidation-reduction the reader is referred to W. M. Clark, *et al.*, *Studies on Oxidation-Reduction*, I–X, *U. S. Pub. Health Service, Hyg. Lab. Bull.* 151, Washington (1928); L. Michaelis, *Oxidation-Reduction Potentials* (translated by L. B. Flexner), J. B. Lippincott Co., Philadelphia and London, 1930; I. M. Kolthoff, *The Colorimetric and Potentiometric Determination of pH*, John Wiley & Sons, New York, 1931 (out of print); I. M. Kolthoff, N. H. Furman, and H. H. Laitinen, *Potentiometric Titrations*, 2nd ed., John Wiley & Sons, New York, 1931; L. F. Hewitt, *Oxidation-Reduction Potentials in Bacteriology and Biochemistry*, 4th ed., London County Council, London, England, 1936; S. Glasstone and A. Hickling, *Electrolytic Oxidation and Reduction: Inorganic and Organic*, D. Van Nostrand Co., New York, 1936.

duced, there must be a simultaneous and equivalent oxidation of some other component, and *vice versa*. On the basis of the above example, we may formulate a general definition of an oxidation or a reduction; *an oxidation involves a loss, and a reduction a gain, of electrons.* According to this definition, a metal in the presence of its ions is an oxidation-reduction system.

$$Ag \underset{\text{reduction}}{\overset{\text{oxidation}}{\rightleftharpoons}} Ag^{+} + (e)$$

An example with which all biologists are familiar is the hydrogen electrode. Here the hydrogen gas is the reduced and the hydrogen ions are the oxidized form.

The oxidation-reduction potential is a measure of the tendency for a substance to give up or take up electrons. Stated in a different way, it is a quantitative measure of the free energy of the reaction which is involved in the electronic transfer. Evidently in order to have a potential established, there must be an electronic transfer, although, with some reactions involving no such transfer, there are certain tricks which can sometimes be resorted to and which yield an apparent potential.

Electrode Potentials. If an inert metal such as platinum is immersed in a reversible oxidation-reduction system, and this electrode is connected to a potentiometer together with a calomel half cell in the same way as is done with the hydrogen electrode, a potential difference will be observed. It can be shown theoretically and experimentally that this potential is a function of the ratio of the oxidized to the reduced form; the more of the oxidized component, the higher the potential, and the more of the reduced form, the more negative will be the potential.

In order to indicate how the relation between the measured potential difference and the ratio of the oxidized and reduced forms is derived, the reaction

$$Fe^{++} \rightleftharpoons Fe^{+++} + (e)$$

may be taken as an example. At equilibrium this reaction can be formulated as

$$\frac{[Fe^{+++}][e]}{[Fe^{++}]} = K_1 \tag{40}$$

Now, if a platinum wire is immersed in this system, there will be a tendency for the electrons in solution to flow into the wire. On the other hand, the wire already has a concentration of electrons, and the direction of flow will be governed by these two concentrations. It can be shown that the work in calories done in transferring an equivalent

of electrons from the solution into the wire will be

$$\text{Work} = RT \log_e \frac{[e_m]}{[e_s]} \tag{41}$$

where $[e_s]$ = the concentration of electrons in the solution
$[e_m]$ = the electronic concentration in the metal
R = the gas constant = 1.99 calories.

The work is equal to the quantity of electricity transferred, multiplied by the potential at which the transfer was made. We can substitute these terms in equation (41) and rearrange it to obtain

$$nEF = RT \log_e [e_m] + RT \log_e \frac{1}{[e_s]} \tag{42}$$

where E = the oxidation-reduction potential
F = the faraday
n = the number of equivalents
R = the gas constant, expressed in joules.

We now substitute equation (40) into equation (41), and, since the concentration of the electrons in the metal, $[e_m]$, is a constant, it can be combined with K_1 of equation (40) to form a new constant, K_2; then

$$E = K_2 + \frac{RT}{nF} \log_e \frac{[\text{Fe}^{+++}]}{[\text{Fe}^{++}]} \tag{43}$$

When the concentration of the ferric iron equals that of the ferrous iron, we have the logarithm of 1, which is equal to zero, and then $E = K_2$. E is usually written E_h and K_2 is substituted by E_0. Then we have

$$E_h = E_0 + \frac{RT}{nF} \log_e \frac{[\text{Fe}^{+++}]}{[\text{Fe}^{++}]} \tag{44}$$

This equation can be generalized in the form

$$E_h = E_0 + \frac{RT}{nF} \log_e \frac{\text{Ox}}{\text{Red}} \tag{45}$$

At 30°C. and log to the base 10 and one electron transfer

$$E_h = E_0 + 0.06 \log \frac{\text{Ox}}{\text{Red}} \tag{46}$$

for a reduction involving two electrons at 30°C.

$$E_h = E_0 + 0.03 \log \frac{\text{Ox}}{\text{Red}} \tag{47}$$

The oxidation-reduction potential is always referred to the normal hydrogen electrode which is assumed to have zero potential. In practice, it is usually more convenient to use a calomel half cell and correct for the potential contributed by the calomel half cell; always, however, in reference to the normal hydrogen electrode. E_0 of equation (45) has considerable significance. As pointed out above, if the concentration

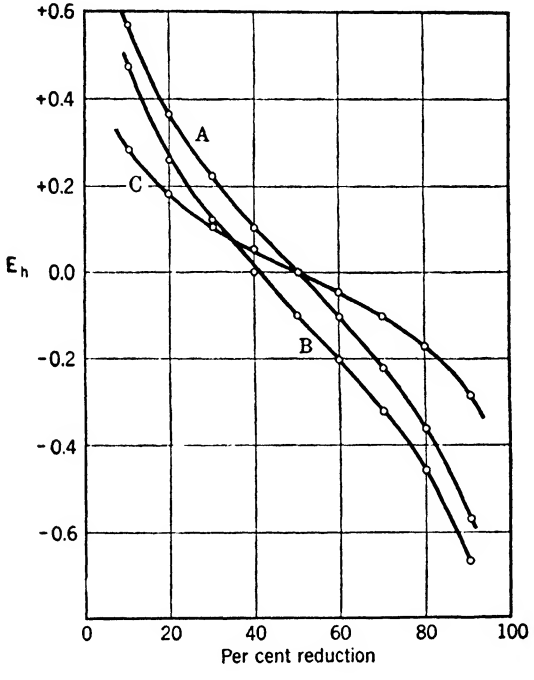


FIG. 18. Redox potential as a function of per cent reduction. (*A*) one electron, (*B*) one electron, (*C*) two electrons.

of the oxidant equals that of the reductant, this term becomes zero and the observed potential is equal to E_0. This is the potential which is used to compare the several oxidation-reduction systems.

It should be added that the oxidation-reduction potential is an *intensity* factor in the same way that temperature is an intensity factor and does not in any way indicate the reducing or oxidizing capacity of a system.

Figure 18 shows the plot of equation (45) for three hypothetical systems involving the transfer of one and two equivalents. We see that, although system *A* has a higher oxidation potential than system *B*, if system *B* contains 90 per cent of the oxidized form, it will tend to oxi-

of electrons from the solution into the wire will be

$$\text{Work} = RT \log_e \frac{[e_m]}{[e_s]} \tag{41}$$

where $[e_s]$ = the concentration of electrons in the solution
$[e_m]$ = the electronic concentration in the metal
R = the gas constant = 1.99 calories.

The work is equal to the quantity of electricity transferred, multiplied by the potential at which the transfer was made. We can substitute these terms in equation (41) and rearrange it to obtain

$$nEF = RT \log_e [e_m] + RT \log_e \frac{1}{[e_s]} \tag{42}$$

where E = the oxidation-reduction potential
F = the faraday
n = the number of equivalents
R = the gas constant, expressed in joules.

We now substitute equation (40) into equation (41), and, since the concentration of the electrons in the metal, $[e_m]$, is a constant, it can be combined with K_1 of equation (40) to form a new constant, K_2; then

$$E = K_2 + \frac{RT}{nF} \log_e \frac{[\text{Fe}^{+++}]}{[\text{Fe}^{++}]} \tag{43}$$

When the concentration of the ferric iron equals that of the ferrous iron, we have the logarithm of 1, which is equal to zero, and then $E = K_2$. E is usually written E_h and K_2 is substituted by E_0. Then we have

$$E_h = E_0 + \frac{RT}{nF} \log_e \frac{[\text{Fe}^{+++}]}{[\text{Fe}^{++}]} \tag{44}$$

This equation can be generalized in the form

$$E_h = E_0 + \frac{RT}{nF} \log_e \frac{\text{Ox}}{\text{Red}} \tag{45}$$

At 30°C. and log to the base 10 and one electron transfer

$$E_h = E_0 + 0.06 \log \frac{\text{Ox}}{\text{Red}} \tag{46}$$

for a reduction involving two electrons at 30°C.

$$E_h = E_0 + 0.03 \log \frac{\text{Ox}}{\text{Red}} \tag{47}$$

The oxidation-reduction potential is always referred to the normal hydrogen electrode which is assumed to have zero potential. In practice, it is usually more convenient to use a calomel half cell and correct for the potential contributed by the calomel half cell; always, however, in reference to the normal hydrogen electrode. E_0 of equation (45) has considerable significance. As pointed out above, if the concentration

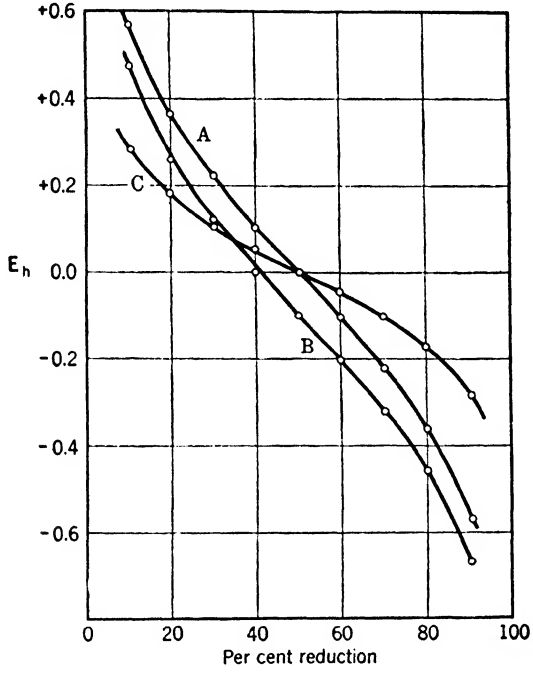


FIG. 18. Redox potential as a function of per cent reduction. (*A*) one electron, (*B*) one electron, (*C*) two electrons.

of the oxidant equals that of the reductant, this term becomes zero and the observed potential is equal to E_0. This is the potential which is used to compare the several oxidation-reduction systems.

It should be added that the oxidation-reduction potential is an *intensity* factor in the same way that temperature is an intensity factor and does not in any way indicate the reducing or oxidizing capacity of a system.

Figure 18 shows the plot of equation (45) for three hypothetical systems involving the transfer of one and two equivalents. We see that, although system *A* has a higher oxidation potential than system *B*, if system *B* contains 90 per cent of the oxidized form, it will tend to oxi-

dize system A, if that system contains 90 per cent of the reduced and 10 per cent of the oxidized form.

The sign and magnitude of the oxidation-reduction potential tells us nothing concerning the relative speeds of oxidation and reduction. Thermodynamics is not capable of giving us any information on this score. Oxidation-reduction potential studies give us a negative answer. We can say definitely that one system with a lower potential can never oxidize a system with a higher potential, but we cannot say definitely that a system at a higher potential will oxidize a system of a lower potential. The reaction may not proceed or, if it does, it may go so slowly that it is of no practical interest. Often, however, an experimental relation is observed between the speed of oxidation and the difference in potential between the two systems. Michaelis and Smythe [3] found that in a series of iron compounds the autoxidizability of the ferro compound at a given pH closely parallels the normal oxidation-reduction potential of the system, ferro compound-ferri compound, at the same pH. The more negative the potential, the greater is the autoxidizability.

Effect of Hydrogen Ions. Hydrogen-ion concentration must usually be taken into consideration when dealing with oxidation-reduction systems, because in many cases the reduced form is an anion which can accept hydrogen ions and so become inoperative in contributing to the potential. Naturally, the potential of such a system is greatly influenced by the hydrogen-ion concentration. Quinhydrone is such a system.

An oxidation-reduction system involving the formation of an anion may be expressed as

$$\mathrm{Red}^- \rightleftharpoons \mathrm{Ox} + (e)$$

and the electrode equation is

$$E_h = E_0 + \frac{RT}{nF} \log_e \frac{\mathrm{Ox}}{\mathrm{Red}^-} \tag{48}$$

but the reduced form ionizes as follows:

$$\mathrm{H\ Red} \rightleftharpoons \mathrm{H}^+ + \mathrm{Red}^-$$

and

$$\frac{[\mathrm{H}^+][\mathrm{Red}^-]}{[\mathrm{H\ Red}]} = K \tag{49}$$

The total reduced form, Red, is equal to the ionized plus the un-ionized forms:

$$\mathrm{Red} = \mathrm{H\ Red} + \mathrm{Red}^- \tag{50}$$

[3] L. Michaelis and C. V. Smythe, *J. Biol. Chem.*, **94**, 329 (1931).

Then, combining equations (49) and (50), we have

$$\frac{[H^+][Red^-]}{[Red] - [Red^-]} = K \tag{51}$$

or

$$[Red^-] = \frac{[Red]K}{[H^+] + K} \tag{52}$$

And substituting equation (52) into the electrode equation (48) we have

$$E_h = E_0 + \frac{RT}{F} \log_e \frac{Ox}{Red} \left[\frac{[H^+] + K}{K}\right] \tag{53}$$

or

$$E_h = E_0 + \frac{RT}{F} \log_e \frac{Ox}{Red} - \frac{RT}{nF} \log_e \frac{K}{[H^+] + K} \tag{54}$$

If the ratio of oxidized to reduced is equal to 1, equation (54) becomes

$$E_h = E_0 - \frac{RT}{F} \log_e \frac{K}{[H^+] + K} \tag{55}$$

If the value of K is so small in comparison with the hydrogen-ion concentration that it can be neglected, we have

$$E_h = E_0 - \frac{RT}{F} \log_e \frac{K}{[H^+]} \tag{56}$$

or

$$E_h = E_1 - \frac{RT}{F} \log_e \frac{1}{H} \tag{57}$$

And at 30°C.

$$E_h = E_1 - 0.06\, pH \tag{58}$$

The system could be used to determine the hydrogen-ion concentration.

The above example involving the formation of an anion with the transfer of only one electron is very unusual for organic compounds. It was selected because it presents the minimum of complexities. As a general rule two electrons are involved in the oxidation-reduction reaction of organic compounds, and it was thought for a number of years that these electrons had to be transferred simultaneously. Michaelis,[4] however, has pointed out that in several organic systems the electrons

[4] L. Michaelis, *Chem. Revs.*, **16**, 243 (1935).

are transferred stepwise and that a definite semi-quinone compound is formed. He has established the criteria for deciding whether the electronic transfer occurs in one or two steps. Michaelis and co-workers have studied a number of systems involving stepwise transfers of electrons, including phenanthrenequinone sulfonate.[5]

Shaffer[6] suggests that the ability of some dyes to catalyze certain oxidation-reduction reactions is related to this property of yielding electrons stepwise. It is known, for example, that the reaction between Ti^{+++} and I_3^- is a slow one; Ti^{+++} can give up one electron but not two, whereas I_3^- can accept two but not one. If a small amount of a dye capable of stepwise oxidation-reduction is added, the reaction between Ti^{+++} and I_3^- is greatly accelerated. The dye has taken electrons one at a time from Ti^{+++} and given them up two at a time to I_3^-.

In addition to the formation of an anion as formulated above, the oxidized form may be a cation which is capable of accepting hydroxyl ions, or we may have the simultaneous destruction of a cation and the creation of an anion. These complications will not be dealt with in detail; suffice it to say that the oxidation-reduction potential usually varies in a highly characteristic manner with the hydrogen-ion concentration.[7] The viologens, however, described by Michaelis and Hill[8] are an exception, as their potentials are not affected by changes in hydrogen-ion concentration.

The dependence of several oxidation-reduction systems on pH is shown in Fig. 19.

The ionization constants of the oxidation-reduction systems are equal to the hydrogen-ion concentration at the inflection points of the curves (pK = pH). The sign in the change in slope of the curve at the inflection point indicates whether the dissociation belongs to the oxidant or to the reductant. When the change in slope $(-dE/dp\text{H})$ is negative, this change is due to the reductant, and when positive to the oxidant. It is also possible to evaluate the dissociation constants by a simple acid-base titration.[9]

A very fundamental analogy exists between acids and bases on one hand and oxidants and reductants on the other. For, according to Brönsted, the relation between the generalized acid A and base B is

$$A \rightleftharpoons B + (+)$$

[5] L. Michaelis and M. P. Schubert, *J. Biol. Chem.*, **119**, 133 (1937).

[6] P. A. Shaffer, *J. Phys. Chem.*, **40**, 1021 (1936).

[7] W. M. Clark and B. Cohen, *U. S. Pub. Health Repts.*, **38**, 666 (1923).

[8] L. Michaelis and E. S. Hill, *J. Gen. Physiol.*, **16**, 859 (1933).

[9] W. L. Hall, P. W. Preisler, and B. Cohen, Supplement 71 to *U. S. Pub. Health Repts.* (1928).

where (+) is a proton. We have already seen how oxidation-reduction can be formulated as

$$\text{Red} \rightleftharpoons \text{Ox} + (e)$$

where (e) is the electron. Thus we see that the creation of an acid involves the addition of the proton to the base, whereas the creation of the reductant involves the addition of an electron to the oxidant.

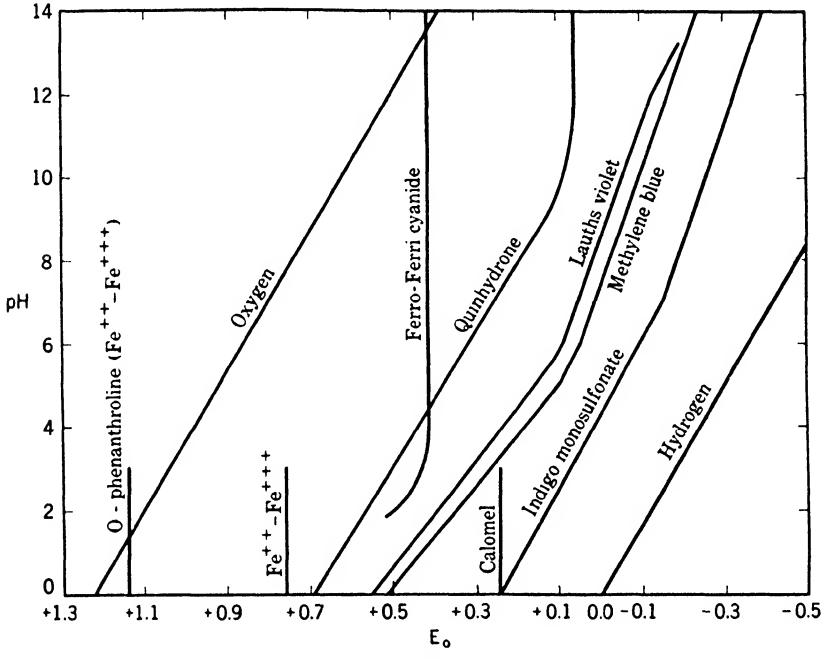


FIG. 19. E_0 values as a function of pH for various redox systems.

To summarize this theoretical discussion, we may state that the oxidation-reduction potential depends on three conditions: (1) the innate tendency of a system to take up or give off electrons, (2) the ratio of the oxidant to the reductant, and (3) the hydrogen-ion concentration.

Measurement of the Oxidation-Reduction Potential. As with the measurement of the hydrogen-ion concentration, two methods are in general use: the indicator and the electrometric methods. For various reasons, however, indicators are less reliable in the measurement of the oxidation-reduction potential than in the measurement of pH. These two methods will now be considered.

Colorimetric Method. It is no accident that a substance which easily undergoes oxidation and reduction is usually colored. Such a substance

TABLE 13. E_0 VALUES FOR OXIDATION-REDUCTION INDICATORS AT 30°C. AS A FUNCTION OF *p*H

Indicator	*p*H										
	5.0	5.4	5.8	6.2	6.6	7.0	7.4	7.8	8.2	8.6	9.0
	E_0	E_0	E_0	E_0	E_0	E_0	E_0	E_0	E_0	E_0	E_0
Phenol-*m*-sulfonate indo-2,6-dibromophenol	+0.390	+0.366	+0.342	+0.319	+0.295	+0.273	+0.251	+0.229	+0.207	+0.187	+0.168
o-Chlorophenol indophenol	.....	.....	.....	+0.288	+0.262	+0.233	+0.203	+0.170	+0.139	+0.109	+0.082
Phenol indophenol	.....	.....	.....	+0.276	+0.254	+0.227	+0.200	+0.170	+0.139	+0.110	+0.083
o-Bromophenol indophenol	.....	.....	+0.308	+0.284	+0.259	+0.230	+0.200	+0.167	+0.137	+0.105	+0.079
2,6-Dichlorophenol indophenol	+0.366	+0.339	+0.310	+0.279	+0.247	+0.217	+0.189	+0.162	+0.137	+0.113	+0.089
Phenol indo-2,6-dibromophenol	.....	.....	.....	.....	.....	+0.218	+0.190	+0.163	+0.138	+0.114	+0.090
m-Cresol indophenol	.....	.....	.....	+0.259	+0.233	+0.208	+0.185	+0.160	+0.134	+0.105	+0.076
o-Cresol indophenol	.....	.....	.....	+0.243	+0.217	+0.191	+0.168	+0.143	+0.116	+0.086	+0.057
2,6-Dichlorophenol indo-*o*-cresol	+0.335	+0.307	+0.277	+0.245	+0.212	+0.181	+0.152	+0.125	+0.099	+0.075	+0.051
Toluylene blue	+0.221	+0.196	+0.173	+0.151	+0.132	+0.115	+0.101	+0.088	+0.075	+0.063	+0.051
Thymol indophenol	.....	.....	+0.244	+0.222	+0.198	+0.174	+0.148	+0.123	+0.097	+0.069	+0.041
1-Naphthol-2-SO_3H-indo-3,5-dichlorophenol	+0.262	+0.236	+0.210	+0.181	+0.150	+0.119	+0.088	+0.060	+0.034	+0.010	−0.012
Thionine	+0.138	+0.112	+0.100	+0.087	+0.074	+0.062	+0.050	+0.037	+0.025	+0.014	−0.001
Methylene blue	+0.101	+0.077	+0.056	+0.039	+0.024	+0.011	−0.002	−0.014	−0.026	−0.038	−0.050
Indigo tetrasulfonate	+0.065	+0.041	+0.017	−0.006	−0.027	−0.046	−0.062	−0.077	−0.090	−0.102	−0.114
Indigo trisulfonate	+0.032	+0.008	−0.016	−0.039	−0.061	−0.081	−0.099	−0.114	−0.127	−0.140	−0.152
Indigo disulfonate	−0.010	−0.034	−0.057	−0.081	−0.104	−0.125	−0.143	−0.160	−0.174	−0.187	−0.199

possesses loosely held electrons which can be transferred without difficulty to another substance. The vibrations of these electrons in the molecule are responsible for the color. Clark and co-workers and others have established a series of oxidation-reduction indicators whose potentials and ionization constants are known. By observing what indicators are oxidized or reduced by the system under investigation, it is possible to assign to the system its approximate oxidation-reduction potential. The indicator method has proved useful in the study of living cells, where it has not as yet been possible to devise an electrode which will not injure the cell. Where attempts at electrode measurement have been made, the cell after a very short time apparently forms a vacuole around the electrode, and the measured potential is not that of the cell proper. Indicators are also of great value in quantitative analysis in determining the end point of an oxidation-reduction reaction. It is also possible for the dye to serve as the oxidizing agent as well as an indicator. For example, Farmer and Abt [10] use 2,6-dichlorobenzenoneindophenol to estimate the ascorbic acid content of small amounts of blood, and this indicator has come into rather general use for the titrametric estimation of ascorbic acid (vitamin C).

Table 13 lists some of the more common indicators which, under certain circumstances, may be used to advantage.

In general there are five considerations which determine the choice of the indicator to be used:

1. The indicator must not act as a hydrogen-ion indicator at the *p*H of the system under investigation, since an alteration in color may be due to a change in *p*H and may thus lead to confusion.

2. The normal oxidation-reduction potential of the indicator must be close to that of the system under study.

3. The indicator must possess a distinctive color, in order not to be confused with the natural color of the system.

4. The indicator should have an intense color, in order that very low concentrations may be used. A number of biological systems have a definite oxidation-reduction potential but possess only a small capacity, *i.e.*, only a small fraction of the system is effective at any given time, although potentially it may have a large reducing or oxidizing capacity. If an excess of dye is used, it may completely oxidize or reduce the active part of the biological system and thus lead to an erroneous idea regarding the actual reducing intensity of the cell. A considerable lapse of time may be necessary to determine the true reducing level of the cell.

[10] C. J. Farmer and A. F. Abt, *Proc. Soc. Exptl. Biol. Med.*, **34**, 146 (1936).

5. The indicator must not enter into the reactions of the system under observation. It must not catalyze reactions which do not ordinarily occur. And, also, it must not have toxic properties which would injure or kill the cells. It must not be adsorbed on or combine in any way with the components of the system.

In determining the reducing intensity of cells with indicators, two methods are used, either (1) the dye is injected into the cell by micro-injection technic, or (2) the dye is allowed to diffuse into the cell. The micro-injection method cannot be used at times because some cells are too sensitive to injury, or the cells may be too small. In using the diffusion method, one has to be sure that the indicator will penetrate the cell. There seems to be a marked correlation between the reducing intensity of individual cells and of a suspension of these same cells.

Electrometric Methods. The electrometric method should be used whenever possible. As indicated above, the use of indicators is rather severely limited. The electrometric method is much broader in its application.

The apparatus used in the measurement is surprisingly simple. The quinhydrone electrode is an oxidation-reduction system which is used so generally that to describe it seems needless. The same set-up is employed to measure the potential of any oxidation-reduction system except that in some cases the reductant is oxidized by the oxygen of the air, which necessitates a closed oxidation-reduction chamber through which an inert gas, such as nitrogen, is bubbled to sweep all oxygen from the system. In such cases it is necessary to purify the nitrogen, and this is done by passing it over heated copper filings or through alkaline pyrogallol solutions. The same technic is used in the study of anaerobic cell suspensions.

Unfortunately, a number of oxidation-reduction reactions do not yield stable, well-defined potentials, and, indeed, some reactions, such as the oxidation of an aliphatic aldehyde to an acid, do not give a potential, and the potential which one obtains in such a system is due to traces of impurities. Other reactions are sluggish and only after several hours yield a stable potential. In some cases, the reason for these difficulties is obscure. In general, the trouble may be due to one of three causes: (1) The material is irreversibly oxidized or reduced, often because the oxidant or reductant is irreversibly altered. (2) The oxidation-reduction is not of an ionic nature and cannot be expected to yield a potential. Perhaps the reason for the delay in the attainment of equilibrium in sluggish systems is that part of the reaction is not ionic but involves some internal change or rearrangement within the molecule. Ionic reactions, as a rule,

proceed with extreme rapidity. (3) Mention has already been made of the suggestion of Shaffer that the difficulty often lies in the reaction between a compound involving a change of one electron with a system involving a change of two electrons.

Conant [11] and others have studied the irreversible oxidation-reduction of organic compounds and have developed methods for determining what they term the apparent potentials. This is done by choosing some easily reversible systems with the oxidant and reductant in equivalent amounts and using them in conjunction with an inert electrode. The substance under investigation is added, and the potential, which is due to the reversible system, is observed. If a 20 to 30 per cent reduction of the reversible system has been accomplished in 30 minutes, the apparent reducing potential is said to be that of the reversible system.

Other methods have been used. Some substances which do not of their own accord exhibit true potentials can be titrated with a reversibly reducible oxidant. For example, potassium ferricyanide has been used to titrate the reduced form of ascorbic acid. The ascorbic acid is oxidized and the ferricyanide is reduced. The observed potential is due to the ferri-ferrocyanide system, but, if sufficient time has been allowed for attainment of equilibrium, this potential must be equal to that of the ascorbic acid system and will continue to be so until all the reduced ascorbic acid has been oxidized.

Some systems are not reversibly reduced until their molecules are activated with an enzyme or, in some cases, by a dyestuff. The reduction of fumaric to succinic acid is reversible only in the presence of a succinic dehydrogenase and oxidation-reduction indicators.

The important sulfhydryl systems have not, as yet, yielded to theoretical treatment. The potential of the cystine-cysteine, as well as of the oxidized-reduced glutathione, is dependent on the concentration of the reductant but not on that of the oxidant.

Biological Systems. A number of systems of biological interest have been studied and, to some extent, characterized. These include systems containing sulfhydryl, hemoglobin, cytochrome, respiratory pigment, ascorbic acid, hermidine and echinochrome, pyrocyanine, dilauric acid, various sugar systems, succinate-fumarate, adrenaline, oxytocic hormone of pituitary gland, as well as tissue and cell suspensions. All together, about 15 biological pigments have been found to consitute oxidation-reduction systems, and respiratory functions have been attributed to some of these pigments. This represents a definite and worth-while achievement; however, the biological systems which yield reversible potentials, and are thus capable of an exact thermodynamic

[11] J. B. Conant, *Chem. Revs.*, **3**, 1 (1926).

treatment, are relatively few, and it appeared a few years ago as if the study of oxidation-reduction must soon be completed with the characterization of these systems. Later work has shown, however, that there is a relation between the reducing potential of cell suspensions and of physiological systems in general and their physiological condition. This realization has added a new interest to the study of oxidation-reduction potentials. The theoretical interpretation of these systems would be extremely difficult, if not impossible. *The derivation of the oxidation-reduction equations rests on the assumption that the system is in equilibrium. It is questionable that any living cell can be said to be in a state of true equilibrium,* and the best that can be achieved is a steady state. Under these circumstances, it is not valid to apply the above equations. The field must be examined experimentally, and the reducing intensity related to known physiological conditions. This method of approach has already proceeded a considerable distance and is proving of especial value in bacteriology.[12]

The assistance which oxidation-reduction potential measurements can yield in the characterizing of a physiological state is strikingly illustrated by the study of Rosner.[13] He studied the oxidation-reduction potentials of lenses of rats before and after the feeding of galactose. It is well known that galactose feeding results in a high incidence of cataract in rats. He found for young rats (6 weeks old, *ca.* 45 grams) average potentials of +19 mv. (range 0 to +39 mv.) for the potential of the cortex and +40 mv. (range −13 to +117 mv.) for the potential of the nucleus, whereas after galactose feeding the corresponding potentials were: cortex, +54 mv. (range +7 to +79 mv.); and nucleus, +150 mv. (range +136 to +160 mv.). For older rats (1 year, *ca.* 190 grams) before galactose feeding the potentials were: cortex, +7 mv. (range −18 to +24 mv.); and nucleus, +132 mv. (range +119 to +150 mv.); after galactose feeding: cortex, +42 mv. (range +32 to +60 mv.); and nucleus, +153 mv. (range +147 to +158 mv.).

The important problem of how the body oxidizes foodstuff to obtain its energy has not as yet shown itself amenable, in all its phases, to quantitative treatment. It is undergoing such active investigation and is of such a complex and diverse nature that it is difficult to summarize.[14]

In its simplest terms physiological oxidation consists of the removal of hydrogen from the substrate and the combining of it with molecular

[12] L. F. Hewitt, *loc. cit.* (reference 2).

[13] L. Rosner, Ph.D. thesis, Northwestern University, July, 1937.

[14] E. S. G. Barron, *Physiol. Revs.*, **19**, 184 (1939); D. E. Green, *Mechanism of Biological Oxidations*, Cambridge University Press, New York, 1941; *A Symposium on Respiratory Enzymes*, University of Wisconsin Press, Madison, 1942.

oxygen to form water or hydrogen peroxide. The living cell seldom, if ever, carries out this process in such a simple, direct fashion. Actually, the hydrogen is taken from the substrate and passed along in successive steps until it is finally combined with oxygen; as many as five steps may be involved. The cell releases the energy which it derives from oxidation somewhat as water is released through a series of locks in a canal. This gradual release of energy is more efficient and more easily controlled.

Four different types of substances are involved in cellular oxidation: dehydrogenases, hydrogen transports, oxidases, and peroxidases. The dehydrogenases act by activating the hydrogen of the substrate. The hydrogen transports convey the activated hydrogen to the oxidases. The oxidases function by activating oxygen so that it will quickly oxidize the hydrogen which is supplied by the hydrogen transports. The function of the peroxidases is to transfer peroxide oxygen to oxidizable substances. Carbon dioxide production is a problem of decarboxylation of the substrate which has been oxidized by the removal of hydrogen and hydrated with the addition of water.

CHAPTER 6

Electrokinetic Phenomena

In the region where two phases come into contact, for example in the interfacial region in a colloid system, there will commonly exist an oriented separation of electrically charged elements derived from one or both of the phases. As a consequence of this phenomenon, we can picture one phase as bearing a positive charge with respect to the other. A potential difference will thus exist between the interfacial regions in which the oppositely charged elements are concentrated. This potential will exist in a direction normal to the plane of the interface and, if an external electrical field is set up in a direction tangential to the plane of the interface, a movement of one phase with respect to the other will result.

Such an interfacial potential has been designated by Freundlich [1,2] as an *electrokinetic potential* (zeta potential) in order to distinguish it from the interphase or *electrochemical potential* (thermodynamic or Nernst potential) commonly considered in physical chemistry. Electrokinetic phenomena are those properties of a system which result from the existence of an electrokinetic potential in the system and are among the most characteristic phenomena of colloid systems.

As we shall see later, the phenomena of mutual precipitation, flocculation, colloid stability, the behavior of a colloid system under the influence of an electric current, and certain phases of the phenomena of adsorption and interfacial tension are all related to the so-called electrokinetic phenomena.

Abramson [3] and Abramson, Moyer and Gorin [4] have discussed in detail the historical background, the derivation of the various mathemat-

[1] H. Freundlich, *Colloid and Capillary Chemistry*, translated by H. S. Hatfield, E. P. Dutton and Co., New York, 1926.

[2] H. Freundlich, *New Conceptions in Colloidal Chemistry*, E. P. Dutton and Co., New York, 1927.

[3] H. A. Abramson, *Electrokinetic Phenomena*, Am. Chem. Soc. Monograph 66, Chemical Catalog Co., New York, 1934; *cf.* also P. H. Prausnitz and J. Reitstötter, *Elektrophorese, Elektro-osmose, Elektrodialyse in Flüssigkeiten*, Theodor Steinkopff, Dresden and Leipzig, 1931.

[4] H. A. Abramson, L. S. Moyer, and M. H. Gorin, *Electrophoresis of Proteins*, Reinhold Publishing Co., New York, 1942.

ical equations, and the application of the electrokinetic technic to a great variety of problems. In view of the ready availability of these excellent books it would be superfluous to repeat here much of the historical background or to enter into the involved mathematical derivations. Suffice it to say that the first observation was made by Reuss, in 1808. He prepared a voltaic pile consisting of 92 silver rubles and 92 zinc plates separated by cloths moistened with salt solution. This battery was attached by wires to a U-tube 0.25 in. in diameter containing powdered quartz in the bottom of the U, the quartz being covered with water. He noted that on applying the current the water level rose approximately 9 in. above the quartz layer on the side in which the negative electrode was inserted and fell correspondingly on the side attached to the positive pole. This observation of Reuss was made only eight years after the experiment of Nicholson and Carlyle (1800), who demonstrated the decomposition of water by the galvanic current, and grew out of a repetition of that experiment. Reuss also demonstrated that a block of moist clay could act as a diaphragm.

In 1816, Porret further studied the phenomenon using both sand membranes and animal (bladder) membranes. He found that water migrated through bladder membrane toward the negative pole. He also coated filter paper with egg white, coagulated the egg white by heat, and then showed a migration with this membrane similar to that which had occurred with the bladder membrane, and in the same paper suggested that minute electric currents may have a great influence in regulating the flow of water through minute pores in living tissues, adding, "Is not this electrofiltration jointly with electrochemical action in constant operation in the minute vessels and pores of the animal system?"

Becquerel, about 1830, observed transport of clay particles in an electric field, and in 1852 Wiedemann showed that the amount of liquid flowing through a porous diaphragm was proportional to the emf. which had been applied. The next major contribution was that of Quincke,[5] who extended the observation of Wiedemann and measured the rate of streaming and the direction of streaming for membranes of various materials. He found that the direction of streaming was determined by the material of the membrane and might be toward either the positive or the negative pole. He also observed that the rate of streaming was greatly influenced by the nature of the material. He furthermore showed that, since an emf. would produce streaming, then conversely streaming would produce an emf. He next turned to the study of suspended particles and showed that most suspended particles were charged negatively in water and positively in turpentine; and lastly, to account

[5] G. Quincke, *Pogg. Ann.*, **113**, 513 (1861); *cf.* also *Pogg. Ann.*, **107**, 1 (1859).

for the electrical effects, he developed a theory of a charge of one sign on the wall of the capillary and of an opposite charge in the liquid bathing the wall of the capillary. This is apparently the first suggestion of an electric double layer.

Here the theory remained until the epoch-making contributions of Helmholtz [6] in 1879. Helmholtz developed the theory of the double layer from the theory of the condenser. He pictured the layers of opposite charge as equivalent to the two plates of a condenser and arrived at the relationship between the density of charge, the distance apart, and the potential existing between the layers, on the basis of the equation describing these relationships for a plane condenser, *i.e.*,

$$\zeta = \frac{4\pi\delta\sigma}{D} \tag{59}$$

where ζ = potential difference existing between the plates (layers)
σ = charge density on unit area of plates
δ = distance between the plates
D = dielectric constant of medium existing between the plates.

He states that, for simplifying the problem of a varying charge or potential, δ (the distance between the layers) may be considered as fixed and equal to the thickness of one molecule, and the dielectric constant may be taken as unity. Perrin,[7] in 1904, introduced the idea that the value of the dielectric constant should be taken as equal to that of the liquid medium across which the double layer could exist. Gouy,[8] in 1910, on the basis of kinetic considerations and the now recognized fact that the layers involved must arise from an oriented arrangement of ions in the region of the interface, pointed out that δ must be variable with electrolyte content of the system and that the outer sheet of the double layer must be regarded as a diffuse layer. δ must thus be interpreted as the distance between the "fixed" layer and a plane containing the "electrical center of gravity" of the diffuse layer. Gouy calculates that, at a surface in contact with 0.10 N NaCl solution, δ approximates 0.96 mμ; for a 0.001 N NaCl solution, δ approximates 9.6 mμ; and, for pure water, δ approximates 1010 mμ, assuming that the dielectric constant of 80 for water exists unchanged in the three solutions.

Practically all the existing textbooks in colloid chemistry contain the statement that electrolytes added to the colloid system reduce the electric charge and that, when the charge is reduced to approximately zero,

[6] H. Helmholtz, *Wied. Ann.*, **7**, 337 (1879); *Wiss. Abhandl.*, **1**, 855 (1882).
[7] J. Perrin, *J. chim. phys.*, **2**, 601 (1904); **3**, 50 (1905).
[8] M. Gouy, *J. phys.*, [4] **9**, 457 (1910).

the system becomes unstable and the colloid particles flocculate. This statement is based upon the theory of a rigid Helmholtz double layer where δ is a fixed quantity. It is obvious from a consideration of the Gouy diffuse layer that the energy of the electric field can be altered by a change in either the density of the charge or in the distances which separate charges of opposite sign, and, as we shall see later, the addition of electrolytes more often than not causes an *increase* in charge but causes δ to decrease so greatly that the electrokinetic potential decreases. At a critically low value of this potential the colloid flocculates.

We have already noted that an electrical charge on a surface may arise from any of several possible causes. In a material in contact with an ionic conducting medium, the charge may arise either from *direct ionization* of the material itself or by the *capture* (*adsorption*) *of an ion*, by the material, from the medium. Where neither of the materials in contact is a conductor of electricity, *electrification by contact* may result. Michaelis [9] classifies these causes, in ionic conducting systems, as due to (1) the forces of residual valences which cause oriented adsorption, (2) the forces of dissociation which cause exchange adsorption, and (3) the spontaneous distribution of ions at a free surface. Cases (1) and (3) differ only in that those classified under (1) involve the lattice forces of crystals acting on ions in solution which are isomorphous to the crystal, whereas those classified under (3) involve weaker or less specific adsorption forces.

In the first instance, silver bromide crystals, for example, contain in their crystal faces both silver and bromine atoms, the charges of which are not completely neutralized. Accordingly, if there is an excess of bromine ions in the surrounding liquid, they will be attracted to the vicinity of the unsaturated silver atoms in the crystal face; if there is an excess of silver ions in the surrounding solution, they will be attracted to the unsaturated bromine atoms in the crystal face. The result has already been shown diagrammatically in Fig. 4. In the event that the bromine ions are in excess, the colloid micelle will be negatively charged, whereas it will be positively charged if the silver ions are in excess. The electric double layer is, therefore, reversed as we pass the isoelectric point. No difference in potential in the double layer will exist at the isoelectric point.

It is easy to picture how dissociation brings about an electric double layer. The dissociation of a ferric hydroxide sol would result in the negative hydroxyl ion being given off into the solution with a corresponding residual positive charge remaining on the micelle. The nega-

[9] L. Michaelis, *The Effects of Ions in Colloidal Systems*, Williams and Wilkins Co., Baltimore, 1925.

tive charge cannot be separated by any great distance from the residual positive charge. Accordingly the double layer is set up.

The third class of substances noted by Michaelis comprises such substances as cellulose, collodion, air bubbles, colloidal carbon, hydrocarbons, where there is probably no great tendency to dissociate and a relatively slight tendency toward oriented adsorption. In this instance Michaelis assumes a selective adsorption of either hydrogen or hydroxyl ions from the aqueous phase (dispersions medium). Assuming that OH^- is more capillary active than H^+, certain of the hydroxyl ions of water would enter more closely to the surface of the disperse phase, so that there would be near the surface of the disperse phase a greater concentration of hydroxyl ions than of hydrogen ions. Accordingly, a double layer would be set up, the micelle acquiring a negative charge. The capillary activity of ions at such surfaces appears to be related to the degree of hydration of the ions. Negatively charged ions are, in general, less hydrated than positive charged ions. Less work is required for them to move into the region of the interface—hence their tendency to reach a position closer to the surface, on the average, than is characteristic of positive ions. Non-ionogenic surfaces are, therefore, usually negatively charged against water. The so-called lyotropic effect of ions may best be explained by the same relationship.

Figure 20 is a diagrammatic representation of the charge relationships in the region of an interface of an ionogenic solid S in contact with a liquid such as water. It is assumed that there will exist a thin layer of liquid adhering to the solid such that it would move with the solid (region A–B). Beyond this layer of liquid, immovable with respect to the solid, the remainder of the liquid can move with respect to the solid. The ionogenic groups of the solid will ionize in contact with the water, and the ionic groups which are chemically a part of the solid (negative in this case) will remain fixed on the surface, while the counter ions (sometimes called *gegen ions*) will pass kinetically into the aqueous phase until this tendency is overcome by electrostatic forces acting between them and the fixed ionic groups of opposite signs. Some counter ions, on the average, may remain within the immovable layer (A–B) while others move on into the movable layer. When equilibrium is attained between the kinetic and electrostatic forces, the counter ions will form a diffuse layer the electrical center of gravity of which will exist in a plane C, at a distance δ from the surface of the immovable layer. If we assume that at a considerable distance D from the interface a difference no longer exists in the volume average content of positive and negative ions, in this region no further change in potential may occur (in direction X perpendicular to the interface plane). But in the region of the double layer a

potential must exist with respect to point D, the course of change of which will be described diagrammatically by the graph of ϵ *vs.* X (where, at any point on X less than point D, ϵ would be negative with respect to D).

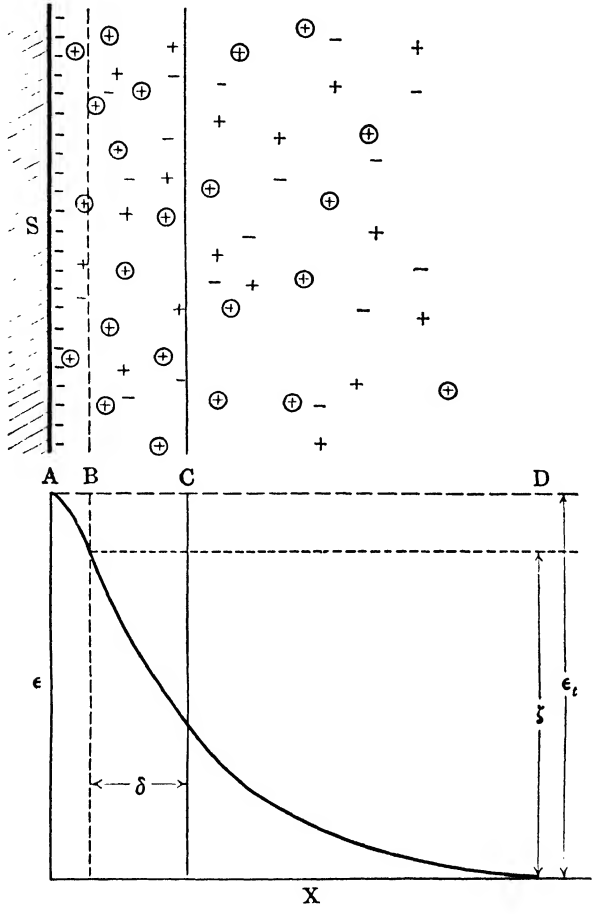


FIG. 20. Diagrammatic representation of charge and potential relationships in the region of an ionogenic solid surface (S) in contact with a liquid such as water.

If we let the entire potential change from the surface (A) to the point D be designated by ϵ_t, it is seen that a fraction of the potential change may occur within the immovable layer and a fraction outside this layer. The latter is identified as the electrokinetic potential (zeta potential) and is responsible for the electrophoretic or electroendosmotic displacement of liquid with respect to the solid when an emf. is imposed in the system in a direction parallel to the plane of the interface. It may be

readily understood that the magnitude of ζ will be a function of the density of net charge per unit area of the solid (plus its immovable layer) and of the thickness δ of the double layer. It is also evident that, as the concentration of other electrolytes is increased in the aqueous phase, the value of δ will be decreased, *i.e.*, planes D and C are forced back toward B. Indeed at a sufficiently high electrolyte concentration the value of δ may be suppressed to zero, wherein all the potential change occurs within the immovable layer of liquid, and ζ will also become zero.

Helmholtz, Perrin, Lamb,[10] and Smoluchowski [11] have expressed the theory of electroendosmosis in mathematical form. In the following discussion we shall make use of the mathematical treatment as presented by Briggs.[12]

Let us assume the existence of an electrical double layer on the wall of a capillary, one part of the electrical double layer being fixed to the solid phase, the other portion, located at distance δ from the opposite charge, being in the liquid phase and free to move with the liquid. If, now, a potential difference is applied to the ends of the capillary, one charge tends to move toward the anode, the other toward the cathode, but, since the charge on the solid is fixed, the charged liquid layer will move along the surface of the capillary at such a rate that the frictional forces R will be at equilibrium with the electrical forces F. The electrical force F, acting on a unit area of surface, is equal to the product of the charge e on the double layer per unit of surface and the potential difference E which is applied tangentially to a unit length along this surface:

$$F = eE \tag{60}$$

At equilibrium this must be equal to the frictional force which must be overcome or balanced. The frictional force R is related to the coefficient of viscosity of the liquid and to the average velocity of flow of the liquid layer between the point of maximum velocity u and the point at the wall where no flow takes place, the distance between these two points being designated as d. The frictional force can therefore be expressed, by assuming the velocity of flow to vary linearly through d, as

$$R = \frac{\eta u}{d} \tag{61}$$

In a capillary of radius r, and where $d \ll r$, the velocity of flow of the body of the liquid in the tube will be equal to u, which may be evaluated

[10] H. Lamb, *Brit. Assoc. Advancement Sci. Repts.*, 495 (1887).
[11] M. von Smoluchowski, *Bull. Acad. Sci. Cracovie*, 182 (1903).
[12] D. R. Briggs, *J. Phys. Chem.*, **32**, 641 (1928).

by measuring the volume V of liquid flowing in a given time through the area of cross section of the tube, according to the equation

$$V = \pi r^2 u \tag{62}$$

Accordingly,

$$R = \frac{\eta V}{d\pi r^2} \tag{63}$$

and from equation (60)

$$eE = \frac{\eta V}{d\pi r^2} \tag{64}$$

On the assumption that the properties of the double layer are those of two plates of a flat condenser, we can state that the capacity C of the double layer is directly proportional to the quantity Q of electricity which it holds and inversely proportional to the potential difference ζ across the plates,

$$C = \frac{Q}{\zeta} \tag{65}$$

The capacity is likewise proportional to the dielectric constant ϵ of the material separating the plates, to the area A of the plates, and inversely proportional to their distance apart δ and to a constant, 4π. Accordingly,

$$C = \frac{\epsilon A}{4\pi\delta} \tag{66}$$

therefore,

$$\frac{Q}{\zeta} = \frac{\epsilon A}{4\pi\delta} \tag{67}$$

and

$$\zeta = \frac{4\pi\delta Q}{\epsilon A} \tag{68}$$

Substituting e, the charge per unit area, for the quantity Q/A, we have

$$\zeta = \frac{4\pi\delta e}{\epsilon} \qquad \text{or} \qquad \pi\delta = \frac{\zeta\epsilon}{4e} \tag{69}$$

By assuming d to be equal to δ and substituting (64) into (69), we have

$$eE = \frac{\eta V 4e}{\zeta\epsilon r^2} \tag{70}$$

Accordingly,

$$V = \frac{r^2 \zeta E \epsilon}{4\eta} \tag{71}$$

and

$$\zeta = \frac{4V\eta}{r^2 E\epsilon} \tag{72}$$

However, in a diaphragm the πr^2 of a capillary tube must be replaced by the area of cross section q, r^2 equaling q/π. Accordingly,

$$V = \frac{q\zeta E\epsilon}{4\pi\eta} \tag{73}$$

and

$$\zeta = \frac{4\pi V\eta}{qE\epsilon} \tag{74}$$

E is equal to H, the emf. applied across the diaphragm, divided by l, the distance between the ends of the diaphragm. Likewise, H, the applied emf., is equal to i, the current, multiplied by w, the resistance, where w is equal to the distance l between the ends of the diaphragm, divided by the product of q, the cross-section area, and κ, the specific conductivity of the liquid. Therefore,

$$V = \frac{\zeta i\epsilon}{4\pi\eta\kappa} \tag{75}$$

and

$$\zeta = \frac{4\pi\eta\kappa V}{i\epsilon} \tag{76}$$

This equation states that, where the electric current i is kept constant, the volume of liquid V which will flow in a given time through a diaphragm is directly proportional to the zeta-potential across the interface, to the dielectric constant ϵ of the liquid, and inversely proportional to the viscosity η, and the specific conductivity of the liquid κ and is independent of the area and length dimensions of the diaphragm.

The derivation of this equation has been given at length in order to illustrate the method of reasoning upon which the various electrokinetic equations, which have been of great service in colloid studies involving biochemical materials, are based. It is apparent that the validity of the values of ζ obtained from these equations must rest upon the validity of the many assumptions upon which the equations are derived. For example, there exists the possibility that the dielectric constant of the medium lying between the two charged layers, owing to its proximity

to a surface and the accompanying probability that orienting forces may be acting upon it in this region, may be different from the bulk dielectric constant of the medium. Likewise, it must be recognized that limiting conditions exist under which the assumption that the double layer acts as the plates of a *plane* condenser is invalid. In very small capillaries and with colloid particles of very small radii, the curvature of the double layer will be so great that the layers must be assumed to approximate the charge-potential relationships of cylindrical and spherical condensers, respectively.

Because of the uncertainty regarding the value of the dielectric constant within the double layer, Bull and Gortner [13] suggested the use of the term $e\delta$, where e is the density of charge per square centimeter of the surface and δ is the thickness of the double layer, as a term possessing a more definite physical meaning than the calculated value of ζ. The quantity $e\delta$ may be thought of as the electric moment of the double layer and can be evaluated by making use only of quantities which can be measured experimentally. Equations involving $e\delta$ will be cited later under the discussion of the streaming potential.

Guggenheim [14] has presented reasons why this term, which he designates by the symbol τ, should better be used as the fundamental electrokinetic unit in place of ζ. As a name for the unit of τ ($= e\delta = \zeta\epsilon/4\pi$), he suggests the term "the Helmholtz."

1 Helmholtz = 1 Debye/A^2 = 1 esu. charge per meter.

Various methods have been proposed for the determination of the sign and magnitude of the electrokinetic potential on colloidal micelles or surfaces. The methods which have been proposed may be classified as *electrophoresis, electroendosmosis, streaming potential,* and *sedimentation potential.* All these methods are interrelated and should be capable of yielding similar values.

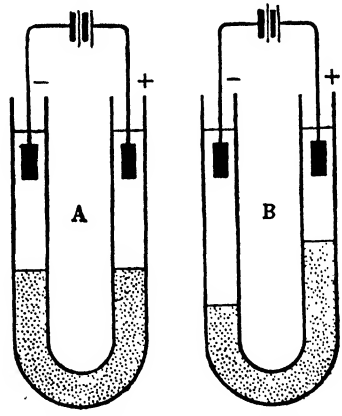


FIG. 21. A diagrammatic representation of electrophoresis.

Electrophoresis. Electrophoresis (formerly termed cataphoresis) may be defined as *the migration of the colloidal particles through the dispersions medium, due to an imposed emf.* Figure 21 shows diagrammatically a U-tube in which a colloid sol has been layered underneath a layer of its dispersions medium. Electrodes have been

[13] H. B. Bull and R. A. Gortner, *Physics,* **2,** 21, (1932).

[14] E. A. Guggenheim, *Trans. Faraday Soc.,* **36,** 139 (1940).

inserted in the overlying medium and connected to a source of direct current. Tube *A* shows the position of the boundary between the sol and the solution before any electric current has passed. Tube *B* shows diagrammatically the displacement of the boundaries, due to the passage of the electric current, on the assumption that the sol particles are negatively charged. It will be noted that the colloidal micelles have migrated toward the anode in the same manner as they would have migrated if they behaved as ions. Various forms of apparatus have been devised for studying not only the direction of migration but also the rate of migration. The direction of migration indicates the sign of the charge on the colloid micelle, the colloid micelle migrating through the liquid toward the pole of the opposite sign. *The velocity of migration is proportional to the electrokinetic potential existing across the double layer.*

The equations which can be applied in electrophoresis studies are

$$v = \frac{\zeta E \epsilon}{4\pi\eta} \tag{77}$$

or

$$\zeta = \frac{4\pi v \eta}{E\epsilon} \tag{78}$$

where v = velocity of migration
E = the applied emf. per unit length in the cell
ϵ = the dielectric constant
η = the viscosity.

This method has proved the most applicable of the four mentioned above. There are two general modifications of the electrophoresis method for determining the migration velocity of particles in an electric field, namely, the microelectrophoresis method, and the macro or U-tube method. Both have reached a high degree of technical perfection.

The microelectrophoresis method for determination of electrophoretic mobility (velocity per second per volt per centimeter) depends on the visual measurement under a microscope of the velocity and direction of displacement of microscopically visible particles when subjected to an electrical field. Many modifications of micro cells have been devised for this purpose.

Mattson,[15] Northrop,[16] Kunitz,[17] Northrop and Kunitz,[18] Szent-

[15] S. E. Mattson, *Kolloidchem. Beihefte*, **14**, 309 (1922).
[16] J. H. Northrop, *J. Gen. Physiol.*, **4**, 629 (1921–22).
[17] M. Kunitz, *J. Gen. Physiol.*, **6**, 413 (1923–24).
[18] J. H. Northrop and M. Kunitz, *J. Gen. Physiol.*, **7**, 729 (1924–25).

Gyorgyi,[19] Bull,[20] Abramson, Moyer, and Voet,[21] Abramson,[22] and Briggs [23] have described apparatus suitable for measuring electrophoretic velocity with small quantities of materials and with microscopic particles. In these micro cells precautions must be taken to make the measurement of electrophoretic velocity at the proper depth in the cell, since there is a varying velocity of displacement of the particle at varying depths in the cell. This variation occurs because, in addition to the electrophoretic migration of the particle toward the pole of the opposite electric sign, there is an electro-endosmotic streaming of the liquid along the walls of the cell. Since the cell is closed at one end, this liquid must return through the center of the cell, thus producing a retardation and sometimes an actual reversal in sign of the direction of migration of the particle close to the wall while intensifying the rate of migration in the center of the cell. There are only two levels at which true electrophoretic velocity is observable. In a flat cell these two levels lie at 21 per cent of the cross-sectional distance from the top or from the bottom of the cell. For a microcylindrical cell the levels lie at 14 per cent of the diameter from the wall of the capillary. Figure 22 shows a displacement *vs.* depth curve for particles in a flat cell. Abramson's monograph should be referred to for a discussion of the theory of this behavior. Non-gassing electrodes must be employed and current densities must be maintained low enough to avoid the occurrence of convection currents within the cell due to heating. Detailed directions for setting up the apparatus for making the measurements with a micro cell have been described by Moyer.[24]

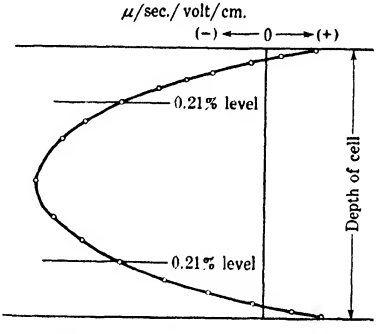


FIG. 22. Showing a displacement *vs.* depth curve for particles at different levels in a flat microelectrophoretic cell. Note that correct velocity values are obtained only at the 0.21 per cent and 0.79 per cent levels

The use of the microelectrophoresis technic is not limited to microscopically visible sol particles but may be applied in the study of any surface-active material (such as protein) which will be adsorbed on the

[19] A. von Szent-Gyorgyi, *Biochem. Z.*, **139**, 74 (1923).

[20] H. B. Bull, *J. Phys. Chem.*, **39**, 577 (1935).

[21] H. A. Abramson, L. S. Moyer, and A. Voet, *J. Am. Chem. Soc.*, **58**, 2362 (1936).

[22] H. A. Abramson, *Electrokinetic Phenomena*, Am. Chem. Soc. Monograph 66, p. 70, Chemical Catalog Co., New York, 1934.

[23] D. R. Briggs, *Ind. Eng. Chem., Anal. Ed.*, **12**, 703 (1940).

[24] L. S. Moyer, *J. Bact.*, **31**, 531 (1936).

surface of microscopically visible particles of substances such as quartz, glass, or collodion, thereby imparting to the surfaces of these particles the surface properties of the material adsorbed. The preparation of collodion particles for this purpose is described by Cannon and Marshall.[25]

The macroelectrophoresis, or moving boundary, method depends on the detection and following of the displacement of a boundary between a sol and its dispersions medium when an electric field is imposed across the boundary. All macroelectrophoresis apparatus consist of some form of U-tube in the bend of which the colloid sol is layered beneath its dispersions medium. The boundary to be observed, then, is that between the sol and its medium in either leg of the U-tube. Accuracy in measurement of the electrophoretic mobility by the moving boundary method is dependent on the establishment of sharp initial boundaries and the accurate detection of the position of the boundaries in the cell before and after they are displaced by the imposed electric field. Displacement or disturbance of the boundaries by gas formation at the electrodes and by convection currents due to heating during passage of the electric current must be avoided. Electrochemical changes which may occur in the region of the electrodes must be prevented from spreading to the region of the boundaries in the U-tube. Accurate determination of the field strength under which migration of the sol occurs and elimination of appreciable field strength changes in the region of the boundary must be attained. Variations in the many macroelectrophoresis apparatus which have been devised [26-28] are primarily progressive improvements in the degree of perfection to which the above requirements have been met.

In the apparatus developed by Tiselius [29] the macroelectrophoresis technic has been carried to a high degree of perfection. By its use a protein preparation, for example, can be analyzed not only to yield the mobility of the colloid component, but also to determine the number, mobilities, and relative amounts of the electrophoretically different components which may be present in the preparation. The electrophoretic homogeneity or degree of heterogeneity of the components can be estimated. It thus has become a method for the analysis of a mixture and for the study of the *purity* of a particular protein preparation.

[25] P. R. Cannon and C. E. Marshall, *J. Immunol.*, **38**, 365 (1940).

[26] E. F. Burton, *Phil. Mag.*, **11**, 440 (1906); A. Coehn, *Z. Elektrochem.*, **15**, 653 (1909).

[27] I. Engel and W. Pauli, *Z. physik. Chem.*, **126**, 247 (1927); H. R. Kruyt and C. P. van der Willigen, *Kolloid-Z.*, **44**, 22 (1928).

[28] The Svedberg and A. Tiselius, *J. Am. Chem. Soc.*, **48**, 2272 (1926); H. Theorell, *Biochem. Z.*, **275**, 1 (1934).

[29] A. Tiselius, *Trans. Faraday Soc.*, **33**, 524 (1937).

Purity, in this instance, is with respect to surface composition which is reflected in terms of the uniformity of charge per unit of surface of the molecules or particles involved. Owing to widespread use in this connection it seems desirable to give here a somewhat more elaborate description of the apparatus than is ordinarily allowed.

Figure 23 shows the electrophoresis cell of Tiselius as modified by Longsworth and MacInnes.[30] The U-tube is made in sections which fit together and with the electrode system through flat ground-glass plates. The whole assembly is held in place by a frame. The bottom and center sections of the U-tube may be moved to the right or left, thus making or breaking the continuity of the lumen of the U-tube. This arrangement makes possible the easy formation of sharp boundaries between the sol and its medium. The cell segments consist of flat pieces of glass sealed or fused together and to the ground-glass plates to give a rectangular cross-sectional lumen with flat faces which eliminate distortion in the optical system which is employed in the detection of the boundary position. The large electrode vessels are filled with the same medium as will form the liquid overlying the sol in the U-tube. Large non-polarizing electrodes, such as silver-silver chloride electrodes in contact with potassium chloride, make possible the carrying of considerable amounts of current without gas formation, and the long path from electrode to boundary eliminates danger of contamination of the boundary region with electrolytes from the electrode region. Convection currents in the U-tube are minimized by keeping the current density low, and by the shape of the cell which gives a large cooling surface per unit volume of cell contents. By working in a constant temperature bath below 4°C., density differences in the boundary region, due to heat arising from wattage dissipation, are minimized in their effects.

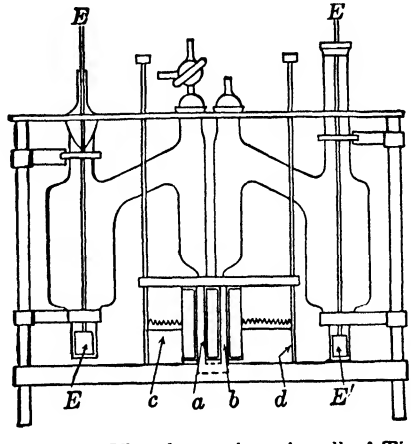


FIG. 23. The electrophoresis cell of Tiselius for quantitative study of the moving boundary and isolation of protein components. E and E' are silver-silver chloride electrodes. a and b are sections of the electrophoresis U-tube. c and d are mechanical devices for moving sections a and b. (From Bull, *Physical Biochemistry*.)

The method which Tiselius introduced for observing the boundaries

[30] L. G. Longsworth and D. A. MacInnes, *Chem. Revs.*, **24**, 271 (1939).

is the most unique feature of this apparatus. Although with colored sols a boundary can be detected visually, there are many sols, such as those of most proteins, which are colorless. By using quartz U-tubes and ultraviolet light, the position of protein-buffer boundaries can be detected by absorption or fluorescent methods, since proteins both absorb and fluoresce in ultraviolet light. None of these methods, however, has proved as versatile in detecting and characterizing boundaries as have those which are based upon the refractive index changes which occur in the region of the boundaries. Tiselius applied the so-called "schlieren" (shadow) method to the detection of boundaries existing in the U-tube. When light passes through a region of changing refractive index, its path will be bent toward the region of the higher refraction.

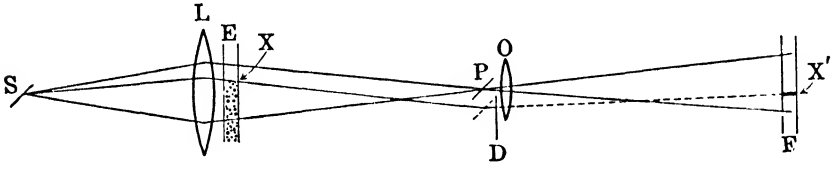


FIG. 24. Diagrammatic representation of the optics involved in the schlieren method for detection of the position of a boundary in the U-tube of the Tiselius apparatus.

The boundary between a protein solution and its equilibrated medium (buffer) constitutes such a region of changing refractive index. The index of refraction of the protein solution will be greater than that of the buffer by an amount which is proportional to the concentration of the protein and to its refractive increment.

Figure 24 shows diagrammatically the manner in which this principle is applied to the detection of the position of a boundary in the Tiselius cell. A slit source of light S is allowed to pass through a long focal length (schlieren) lens L, which forms an image of the slit at point P. If the light in passing from L to P encounters no region of refractive index change (existing in a direction perpendicular to its direction of propagation), all the light will be brought to focus as a simple image at P. If, however, an electrophoresis cell E is placed in the path of the beam from L to P and this cell contains a boundary at X across which there is a refractive index change in a vertical direction, that fraction of the light passing through this region of refractive index change will be bent down and will be brought to focus in the same plane as P but at a point vertically below P. If a camera of long focal length is focused on the electrophoresis cell with the camera lens O just beyond P in the path of the light beam, all the light passing the cell will be brought to focus on the camera plate F. If a diaphragm D is raised in the plane of P to a

point where it will intercept the light thrown down, but not the main beam, there will appear on the camera plate a shadow X' in the image of the cell corresponding to that region containing the refractive index gradient, *i.e.*, the region of the boundary. Thus the position of the boundary in the cell can be determined from the position of the shadow in its image.

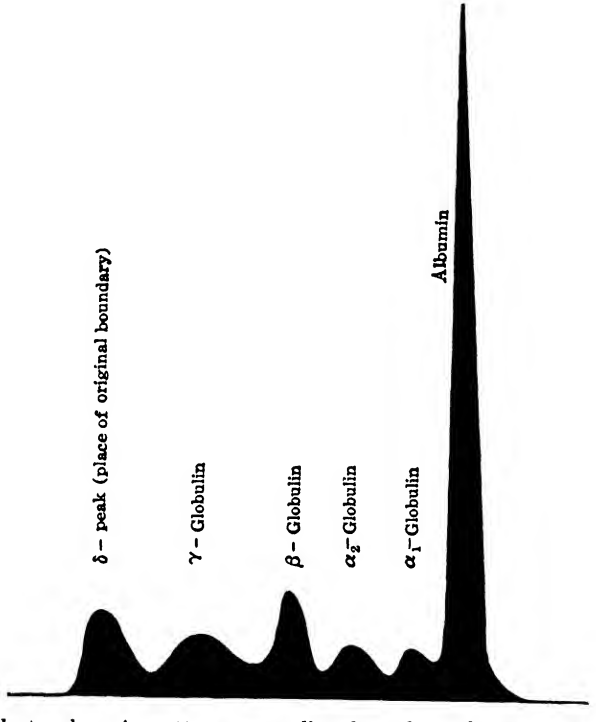


FIG. 25. Electrophoresis pattern, ascending boundary, for normal human blood serum in veronal buffer, *p*H 8.6, ionic strength 0.1. Pattern scanned after 7,000 seconds of electrophoresis at field strength of 7.52 volts per centimeter in cell.

The degree to which the light beam passing through the boundary region is thrown down at point P will be proportional to the rate of change of the refractive index n with distance dn/dx within the cell. This rate of change of the refractive index will be proportional to the rate of change of protein concentration in the region of the boundary, varying, as we pass from pure buffer through the boundary into uniform protein-containing solution, from zero up to a maximum and down to zero again. Thus, as the diaphragm is raised, the width of the shadow will increase. If we plot the rate of change of n with distance in the cell, dn/dx, against its position in the cell, x, a simple probability curve will be obtained for a symmetrical boundary. Such a figure is obtained

automatically for each such boundary present in the cell if the photographic plate of the camera is moved uniformly across the cell image as the schlieren diaphragm is raised. This scanning method for obtaining dn/dx *vs.* x patterns was introduced by Longsworth.[31] Such curves describe the total refractive index change across the boundary and thus the total concentration change in protein across the boundary. The area under the curve is, therefore, proportional to the protein concentration change at the boundary. If more than one electrophoretically different protein is present, a scanning picture taken after the electric current has been allowed to pass for a time will indicate, as separate peaks in the pattern, the presence of more than one boundary. The number of boundaries observed corresponds to the number of electrophoretically different components present in the preparation, and the area under the curve for each boundary will be proportional to the relative amount of each such component in the mixture. Thus, this apparatus can be used to analyze a protein preparation as to the number and relative amounts of electrophoretically different components present and as to the electrical mobility of each. Figure 25 shows the scanned pattern obtained with a normal blood serum.

Electroendosmosis. The theory of electroendosmosis has already been discussed in the early part of this chapter. *Electroendosmosis can be defined as the passage of liquid through a membrane or colloidal gel under the force of an applied electric current.* The direction of migration of the water through the membrane or gel is toward the pole which has the same sign as the membrane or gel. Figure 26 shows diagrammatically in *A* a membrane inserted in a tube, together with the liquid levels in the arms of the tube, before the electric current has been applied. In *B* is represented diagrammatically the relative position of liquid in the two arms of the tube after an electric current has been applied to the electrodes, if the gel was negatively charged. The direction of flow accordingly indicates the sign of the charge on the interface. *The volume of flow is proportional to the electrokinetic potential existing across*

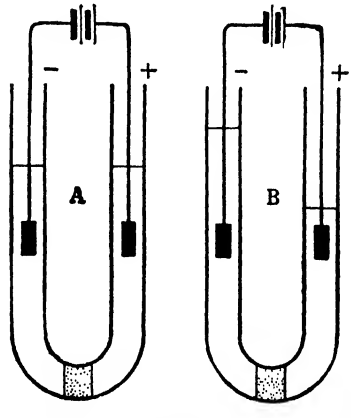


FIG. 26. A diagrammatic representation of electroendosmosis.

[31] L. G. Longsworth, *J. Am. Chem. Soc.*, **61**, 529 (1939). For other scanning devices see J. S. L. Philpot, *Nature*, **141**, 283 (1938), and H. Svensson, *Kolloid-Z.*, **90**, 141 (1940).

the double layer. The equations which apply in electroendosmotic measurements are, for a single capillary tube,

$$v = \frac{r^2 \zeta H \epsilon}{4 \eta l} \tag{79}$$

and

$$\zeta = \frac{4 v \eta l}{r^2 H \epsilon} \tag{80}$$

where v = volume of electroendosmotic flow per unit time
l = the length of the capillary
H = the emf. applied across the ends of the capillary, $H/l = E$, the emf. drop per unit length, as in equation (78).

If the liquid is not allowed to flow out but accumulates so as to produce a hydrostatic pressure P, the equations then become, for a single capillary

$$P = \frac{2 \zeta H \epsilon}{\pi r^2} \tag{81}$$

and

$$\zeta = \frac{P \pi r^2}{2 H \epsilon} \tag{82}$$

The dimensions of the capillary (r and l) can be eliminated from the endosmotic equations provided that the current i and the specific conductivity κ are introduced. In such a case equations (75) and (76) apply, either for a single capillary or for a diaphragm.

Numerous papers have been published which deal with measurements involving electroendosmotic technic. Among the more interesting to the biochemist are those of Stamm,[32] where sections of wood have been used as the membranes under investigation.

Most investigations on electroendosmosis have been conducted with water as the liquid medium. However, Strickler and Mathews [33] have investigated the behavior of various organic liquids, using membranes of filter paper.

Briggs [34] has compiled an excellent review on electroendosmosis, together with its application to industrial processes. Some of the industrial applications suggest interesting possibilities. It is pointed out

[32] A. J. Stamm, Colloid Symposium Monograph, Vol. IV, p. 246, 1926; Vol. V, p. 361, 1928, Chemical Catalog Co., New York.

[33] A. Strickler and H. Mathews, *J. Am. Chem. Soc.*, **44**, 1647 (1922).

[34] T. R. Briggs, Second Report on Colloid Chemistry, *Brit. Assoc. Advancement Sci.*, p. 26 (1918).

that a considerable quantity of water can be removed from colloidal gels by passing an electric current through the gel. For instance, peat can be compressed into blocks, electrodes applied at the two ends of the block, and, when a current is turned on, water will flow from the moist peat to the cathode.

In a similar manner, clay can be collected and partially dewatered. If one has a thin suspension of clay, the clay particles can be made to collect on the anode, owing to electrophoretic migration, and later the mass of moist clay can be partly dewatered by subjecting the mass to conditions which will cause electroendosmotic flow. Briggs points out that Dawkins, in 1913, proposed a novel application of electroendosmosis to the process of brick manufacture. Wet clay has a marked tendency to adhere to a smooth metallic surface. In making wire-cut bricks, it is usual to employ a lubricant of some sort in order to prevent the clay from adhering to the cutting wires. It was found, however, that if the wires were connected with a source of direct current, the wire being made the cathode, the anode being inserted in the clay block, the clay would no longer stick to the wire, but that the wire would cut the clay cleanly, and with this electrical "lubrication" the power consumption was reduced by 25 to 30 per cent. The action is due to water collecting in a film on the surface of the cutting wires due to electroendosmotic flow, this water film then acting as a lubricant for the wire, the mass of clay itself never coming in contact with the metal.

Crowther and Haines [35] applied this method to a study of plowing. If the plowshare were made the cathode, being connected with an overhead power line, and the anode were imbedded in the soil of the field, it would be theoretically possible to lubricate the plowshare with a film of water so that the soil would never come in contact with the metal. They accordingly tested this theory, using both laboratory-scale experiments and actual plowing tests.

A slider, consisting of a weighted steel slab, was so arranged as to be drawn by weights across the surface of a plane of moist soil. The weight necessary to keep the iron block in steady motion was considered to be a measure of the friction of the block upon the soil surface. Figure 27 shows the results obtained. In the absence of electrification of the slab a weight of 0.6 kg. was necessary to keep the slab in motion. When the slab was made the anode, there was a sharp increase in the friction until at point B 1,500 grams were necessary to keep the slab in motion. At point B the current was shut off, the friction falling to point C. The higher friction at point C over that at point A was due to the drying of the surface, the moisture being drawn into the interior of the soil. At

[35] E. M. Crowther and W. B. Haines, *J. Agr. Sci.*, **14**, 221 (1924).

point *C* the iron slider was made the cathode. Moisture, now, was drawn from the soil to the surface of the moving iron plate, thus forming a

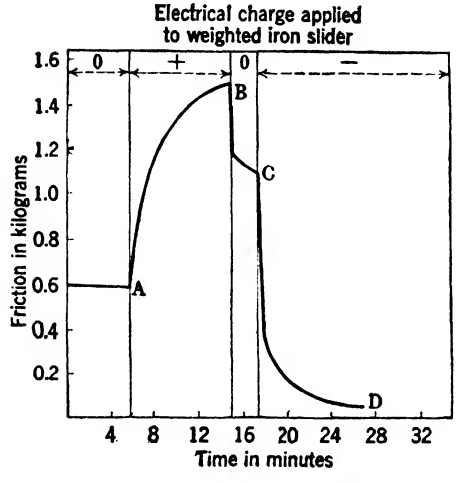


FIG. 27. Showing how the friction between an iron slider and soil may be increased or decreased by electroendosmotic water flow.

lubricating film between the iron and the soil, the weight necessary to move the iron plate dropping rapidly to point *D*.

Figure 28 shows a similar experiment in which the soil moisture content was varied and variations in electrical potential applied to the iron

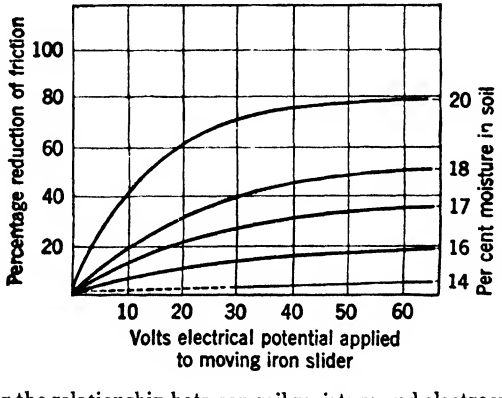


FIG. 28. Showing the relationship between soil moisture and electroendosmotic water flow through the soil. (Data of Crowther and Haines.)

plate were introduced. It will be noted that the frictional force was reduced to approximately 20 per cent of its original value when soil con-

taining 20 per cent moisture was used. When the soil contained only approximately 14 per cent of moisture, very little reduction in friction took place. Accordingly the moisture content of the material which is subjected to electroendosmosis becomes an important factor.

In the actual plowing experiments where the drawbar pull was recorded by means of a dynamometer, there was a reduction in the energy necessary for plowing, but the magnitude of the reduction was not sufficient to compensate for the current used.

Streaming Potential. It is obvious that, if an electrical current induces a flow of liquid through a capillary, a condition which we have already discussed under electroendosmosis, the reverse phenomenon must take place, *i.e., forcing a liquid through a capillary will produce a difference in electrical potential between the two ends of the capillary.* The emf. so induced is known as the *streaming potential.*

Kruyt,[36] Freundlich and Rona,[37] and Kruyt and Willigen [38] developed technics for the quantitative estimation of the streaming potential and equations for the calculation of the zeta potential from the streaming potential. They used a single glass capillary and studied changes in the streaming potential produced by varying the kind of glass making up the walls of the capillary tube, as well as the effect of the various ions which are present in the liquid being streamed through the capillary. Using a glass capillary, Kruyt found that the emf. set up was directly proportional to the hydrostatic pressure and that the quantity H/P was a constant for a given capillary as long as the same solution was being forced through the capillary. Table 14, taken from Kruyt's paper, illustrates this point.

TABLE 14. THE EMF. SET UP BETWEEN THE TWO ENDS OF A GLASS CAPILLARY WHEN A MILLIMOLAR SOLUTION OF POTASSIUM CHLORIDE IS FORCED THROUGH THE CAPILLARY UNDER VARIOUS PRESSURES

P, cm. Hg	*H, mv.*	*H/P*
61.2	271	4.4
86.4	367	4.2
54.8	236	4.3
70.8	315	4.4

Briggs [39] in an attempt to apply streaming potential methods to the determination of the electrokinetic charge on cellulose fibers observed that H/P was not a constant for different diaphragms of the same sample

[36] H. R. Kruyt, *Kolloid-Z.*, **22**, 81 (1918).
[37] H. Freundlich and P. Rona, *Sitz. preuss. Akad. Wiss.*, **20**, 397 (1920).
[38] H. R. Kruyt and P. C. van der Willigen, *Kolloid-Z.*, **45**, 307 (1928).
[39] D. R. Briggs, *J. Phys. Chem.*, **32**, 641 (1928).

of cellulose. However, by introducing κ_s, the specific electrical conductivity of the liquid as it exists inside of the diaphragm and which is being streamed through the diaphragm, into the streaming potential equation in place of κ, the "bulk" specific conductivity of the liquid which was being streamed through the diaphragm, he [40] was able to obtain a constant ratio of potential difference to the pressure used to induce streaming. Briggs accordingly proposed the modified equation

$$\zeta = \frac{4\pi\eta\kappa_s H}{P\epsilon} \tag{83}$$

where η = the viscosity of the liquid being streamed through the diaphragm
P = hydrostatic pressure under which the liquid flows
ϵ = the dielectric constant of the liquid
κ_s = the specific electrical conductivity of the system, *i.e.*, the liquid as it exists in the pores of the diaphragm material
H = the potential difference existing across the diaphragm.

The values that must be determined are P, H, and κ_s, and from these the electrokinetic potential can be calculated. The coefficient of viscosity for dilute aqueous solutions can be taken as 0.01. ϵ, the dielectric constant of water, is considered to have a value of 80. P, which is observed in centimeters of mercury, must be converted into dynes in order to be expressed in absolute units. Accordingly the height of the mercury column in centimeters must be multiplied by the specific gravity of mercury, 13.6, and the gravity constant, 981 dynes. H, which is read in millivolts, must be divided by 1,000 to be reduced to volts, and again divided by 299.86 to reduce volts to cgs. electrostatic units. κ_s observed in reciprocal ohms must be multiplied by 9×10^{11} to convert it to cgs. electrostatic units. The value of ζ so obtained will be in electrostatic units. In order to obtain this value in volts it must be multiplied by 299.86. Then

$$\zeta = \frac{H\kappa_s}{P} \times \frac{9 \times 10^{11} \times 4 \times 3.1416 \times 0.01 \times 299.86}{13.6 \times 981 \times 10^3 \times 299.86 \times 80}$$

$$= 1.0596 \times 10^2 \times \frac{H\kappa_s}{P}$$

where ζ is expressed in volts, H in millivolts, κ_s in reciprocal ohms, and P in centimeters of mercury. Similar methods of calculation, employ-

[40] *Cf.* also D. R. Briggs, Colloid Symposium Monograph, Vol. 6, p. 41, Chemical Catalog Co., New York, 1928.

ing electrostatic units of conversion, must, of course, be employed in the application of any of the electrokinetic equations.

We have already noted that the term $e\delta$ may be introduced into electrokinetic calculations. Substituting equation (69) into equation (83) gives

$$\delta e = \frac{\eta \kappa_s H}{P} \tag{84}$$

All the terms on the right-hand side of equation (84) are measurable experimentally. The quantity δe, which may be regarded as *the electric moment per square centimeter of the double layer*, is an expression for the determination of the symmetry of the double layer in much the same way as the electric moment of a molecule is the expression of the symmetry of a molecule. Furthermore, this equation does not involve any factors which cannot be measured experimentally, and it is believed that it will be a much more valuable measure of the intensity of electrical forces at interfaces than is the ζ potential as usually calculated.

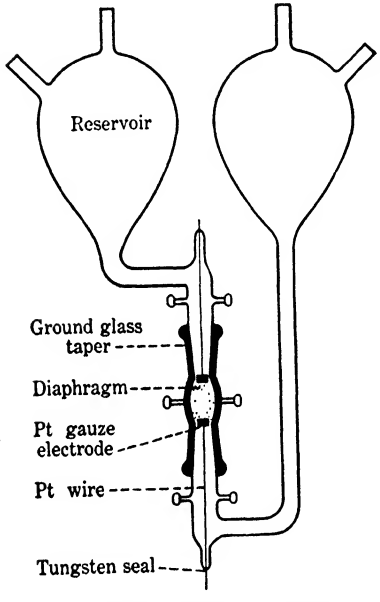


FIG. 29. Solid-liquid streaming potential cell.

The apparatus devised by Briggs was later modified by Martin and Gortner [41] and still later by Lauffer.[42] Figure 29 shows the latest modification of the streaming potential cell. In this cell the electrodes are made of 80-mesh platinum gauze welded onto a platinum grid and are permanently mounted on the ends of the inside parts of the two standard taper joints. Each electrode is held in place by the tension on the platinum wire which connects it to the tungsten seal at the end of the cell. By use of standard ground-glass joints mechanical clamps and gaskets have been eliminated. This cell can be inserted in the wiring diagram as described by either Briggs or Martin and Gortner. It is essential, however, that the potential which is set up by the liquid streaming between the two electrodes be measured by means of a quadrant electro-

[41] W. McK. Martin and R. A. Gortner, *J. Phys. Chem.*, **34**, 1509 (1930).
[42] M. A. Lauffer and R. A. Gortner, *J. Phys. Chem.*, **42**, 641 (1938).

meter. Use of a galvanometer causes polarization of the electrodes and may give rise to erroneous potentials.

Bull [43] has shown that gold or platinum electrodes used in this way in streaming potential cells give results identical with those obtainable when calomel half cells are used as the electrodes and that the gold or platinum ones have the advantage of avoiding a dangerous source of contamination with electrolytes.

Another type of streaming potential cell which is designed for the study of electrical forces at liquid-liquid interfaces was designed by Martin and used by Bull and Gortner.[44] This cell is shown in Fig. 30.

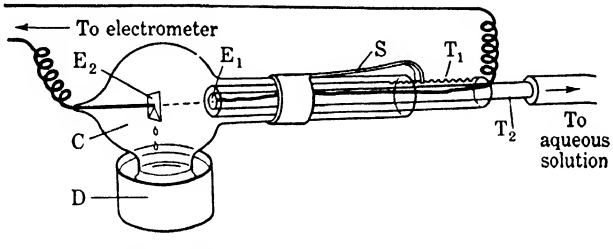


FIG. 30. Oil-water streaming potential cell.

The operation of the cell depends on the streaming of a liquid through a small orifice in one electrode under sufficient pressure so that a "solid rod" of liquid (water or an aqueous solution) streams from that electrode and impinges upon the opposite electrode, the two electrodes being separated by the second immiscible liquid (oil) under investigation. Using this cell Bull and Gortner found that aqueous solutions of electrolytes streamed through paraffin oil yielded curves similar in shape to other curves resulting from similar studies, in which the Briggs or Martin and Gortner type of apparatus was used, of the effect of electrolyte solutions streamed through cellulose or quartz membranes. This liquid-liquid streaming potential apparatus should be of value in studying emulsification and the physical and chemical behaviors of different samples of oils where slight impurities, added components, or changes in structure make great differences in surface behavior.

Sedimentation Potential. Dorn,[45] in 1878, discovered that when particles fall through a liquid, a difference of electrical potential is established between the top and the bottom of the liquid. This is the converse of electrophoresis in the same way as the streaming potential is the converse of electroendosmosis. In electrophoresis the particles are

[43] H. B. Bull, *J. Am. Chem. Soc.*, **57**, 259 (1935).
[44] H. B. Bull and R. A. Gortner, *Physics*, **2**, 21 (1932).
[45] E. Dorn, *Ann. phys.*, **5**, 20 (1878); **9**, 513; **10**, 46 (1880).

drawn through the liquid under the influence of an electric current. Accordingly *an emf. is set up when particles are allowed to fall through a liquid.* Similar observations were made by Billitzer,[46] Freundlich and Mäkelt,[47] and Stock.[48] Stock allowed fine quartz powder to fall through nitrobenzene, ether, and toluene, and measured the magnitudes of the emf. which was produced.

Bull [49] has applied this technic to the determination of the sign of the electrical charge on lead sulfide and has attempted to explain the action of certain toxic ions in the industrial process known as ore flotation. Although he did not attempt to determine the absolute value of the electrokinetic potential on the galena particles, he was nevertheless able to show that galena possesses a positive charge when in contact with pure water, and that certain ions may greatly increase whereas other ions cause a pronounced decrease in the magnitude of the potential.

In later studies Bull, Ellefson, and Taylor,[50] using electrophoretic technic, could not confirm the form of the curves which had been earlier secured by the sedimentation potential technic, and efforts in the author's laboratories to use the sedimentation potential as a quantitative method have so far resulted in failure to secure constant and reproducible data. Failure probably results from the fact that it has so far been impossible to control exactly the path of the fall of the individual particles, so that the particles take a more or less random and irregular path in falling through the liquid, and the side motion of the particles, together with convection currents in the liquid, produce disturbing effects which interfere with a strict reproducibility of conditions.

For non-conducting liquids (or gases) the potentials developed through the Dorn effect may be of rather large magnitude. Stock measured potentials in excess of 80 volts for powders falling through organic liquids. Probably electrical effects in dust storms, in which radio aerials often become highly charged and emit sparks, and instances in which lightning bolts fall from a clear sky may be due to such origins.[51, 52]

Determination of the Charge Density on the Surface of Particles. We have already noted that the zeta potential involves not only the magnitude of the charge per unit area of the double layer but likewise the distance which separates the positive and negative charges in the double layer. Nearly all colloid textbooks carry the statement that

[46] J. Billitzer, *Ann. physik*, [4] **11,** 937 (1903).
[47] H. Freundlich and E. Mäkelt, *Z. Elektrochem.*, **15,** 161 (1909).
[48] J. Stock, *Anzeiger Akad. Wiss. Krakau* [A] 131 (1913); [A] 95 (1914).
[49] H. B. Bull, *J. Phys. Chem.*, **33,** 656 (1929).
[50] H. B. Bull, B. S. Ellefson, and N. W. Taylor, *J. Phys. Chem.*, **38,** 401 (1934).
[51] R. A. Gortner, *Science*, **70,** 118 (1929).
[52] P. E. Shaw, *Proc. Roy. Soc.* (*London*), (A) **122,** 49 (1929).

"electrolytes may reduce the *charge* to zero or may reverse it." What this statement really means is that electrolytes may reduce the zeta potential to zero or may reverse it and that this may be brought about by an alteration in either charge or distance. As a matter of fact, the distance which separates the positive and negative charges in the double layer is usually the factor which is most affected. This was demonstrated in studies [53] of the behavior of electrolytes at a cellulose-water interface, in which streaming potential technic was used. Figures 31, 32, and 33 illustrate the effect of KCl, $CaCl_2$, and $ThCl_4$ on the zeta potential, the charge, and the thickness of the double layer at a water-cellulose interface. It is evident from these curves that the distance which separates the charges in the double layer is the factor which causes the decrease in the zeta potential and that e (charge density per square centimeter of the double layer) actually is increasing with increasing salt concentration, while the double layer is collapsing. Apparently at a large value of e and a negligible value of δ the layer collapses and reverses. This is particularly shown in the curves for thorium chloride.

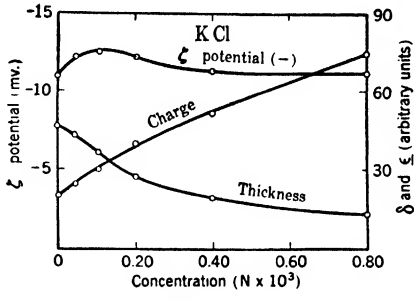


FIG. 31. The effect of increasing concentrations of KCl on the zeta potential, the thickness of the Helmholtz double layer, and the density of charge in the double layer at a cellulose-water interface. (Data of Bull and Gortner.)

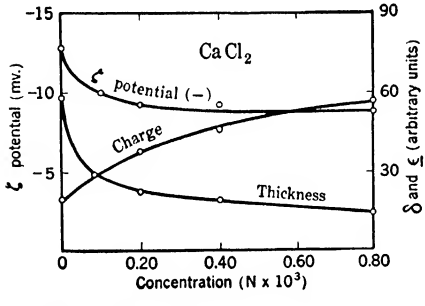


FIG. 32. The effect of increasing concentrations of $CaCl_2$ on the zeta potential, the thickness of the Helmholtz double layer, and the density of charge in the double layer at a cellulose-water interface. (Data of Bull and Gortner.)

It has been recognized for a long time that the flocculation of lyophobic sols by electrolytes may take place when there is still an appreciable zeta potential. There is accordingly a range about the isoelectric point within which lies the so-called critical zone [54] and a level of potential which has been called the critical potential.

[53] H. B. Bull and R. A. Gortner, *J. Phys. Chem.*, **35**, 309 (1931).

[54] H. R. Kruyt, *Proc. Acad. Sci. Amsterdam*, **17**, 623 (1914); *Rec. trav. chim.*, **39**, 618 (1920).

This is shown in Fig. 34. The critical potential is probably determined not so much by the charge at the interface as by the thickness of the double layer, and when this has been reduced to a given point, the system becomes unstable and flocculation begins. Here again the factor of *time* should be emphasized, since, in a short time period, flocculation may not be perceptible, whereas, in a much longer time period, flocculation may be complete. Thus the critical zone narrows with increasing time.

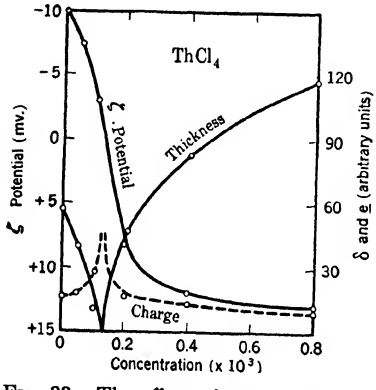


FIG. 33. The effect of increasing concentrations of $ThCl_4$ on the zeta potential, the thickness of the Helmholtz double layer, and the density of charge in the double layer at a cellulose-water interface. (Data of Bull and Gortner.)

The observation that charge density in the double layer increases with increasing electrolyte concentration is confirmed by Abramson [55] where he has calculated charge density on typhoid bacteria as a function of increasing salt concentration. Certain of his data are given in Table 15.

The form of the curves for NaCl, KCl, LiCl, etc., in electrokinetic studies at a cellulose-salt solution interface indicates that in all probability both anions and cations are being adsorbed into the double layer.

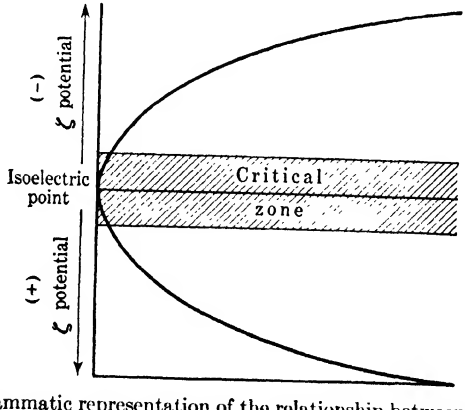


FIG. 34. A diagrammatic representation of the relationship between colloid stability, the zeta potential, and the "critical zone."

For example, the lower concentrations of NaCl cause an increase in the zeta potential from approximately −10 mv. to −14.10 mv. at a salt con-

[55] H. A. Abramson, *J. Bact.*, **27**, 89 (1934).

centration of 1×10^{-4}. The zeta potential then decreases to -10.90 mv. at a salt concentration of 1.6×10^{-3}. The initial increased negativity, in all probability, is due to a differential adsorption of anions by the surface, which causes an increase in charge e, whereas the decrease in the negativity which occurs later is in all probability due to a decrease in δ, the thickness of the double layer.

TABLE 15. CHANGES IN CHARGE DENSITY ON BACTERIA WITH INCREASING SALT CONCENTRATION

(Data of Abramson)

NaCl Concentration	*Zeta Potential, volts*	*Charge Density, esu.*
0.001 *M*	0.30	224
0.004	0.25	417
0.01	0.20	530
0.02	0.10	700
0.04	0.006	924

At the time Bull and Gortner made their studies no method was available for evaluating the absolute surface charge. More recently equations have been devised which make such evaluation possible. The generalized equation is [56]

$$\sigma = \sqrt{\frac{NDkT}{2{,}000\pi}} \sqrt{\Sigma C_i(e^{-Z_i(e\zeta/kT)} - 1) + \Sigma C_j(e^{+Z_j(e\zeta/kT)} - 1)} \quad (85)$$

where σ = the charge per unit area on the particle
N = Avogadro's number
D = the dielectric constant
k = the Boltzmann constant
T = the absolute temperature
C = the concentration of electrolytes in moles per liter
e = the elementary charge
Z = the valence of the ions

and i and j indicate the anion and cation, respectively.

When the valences of the ions making up the electrolyte are the same, equation (85) reduces to

$$\sigma = 2\sqrt{\frac{NDkT}{2{,}000\pi}} \sqrt{C} \sinh Z \frac{e\zeta}{2kT} \quad (86)$$

[56] H. A. Abramson, *Electrokinetic Phenomena*, American Chemical Society Monograph 66, p. 110, Chemical Catalog Co., New York, 1934.

The importance of determining the charge density in electrokinetic studies is illustrated by the fact that an investigation [57] of the electrokinetic behavior of NaCl, KCl, $CaCl_2$, $MgCl_2$, and mixtures of NaCl:KCl, NaCl:$CaCl_2$, NaCl:$MgCl_2$, KCl:$CaCl_2$, KCl:$MgCl_2$, and $CaCl_2$:$MgCl_2$, at cellulose-aqueous solution interfaces, in an attempt to determine whether or not electrokinetic phenomena are involved in the problem of ion antagonism, failed to demonstrate ion antagonism as reflected by the zeta potential. Later a recalculation [58] of the data demonstrated the marked ion antagonism shown in Fig. 35.

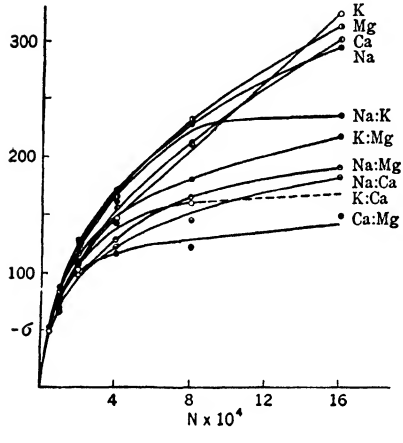


FIG. 35. The relationships between charge density and salt concentration at a cellulose-water interface. Note ion antagonism exhibited by mixed salt solutions. (All salts as chlorides. Data of Bull and Gortner recalculated by Moyer and Bull.)

Comparison of Various Electrokinetic Technics. It is perhaps pertinent at this point to inquire how electrokinetic measurements conducted by the various methods compare with each other. Briggs [59] in a study of the electrokinetic potential utilized the streaming potential method on egg albumin adsorbed on a quartz-water interface and compared his data with data which had been reported by Abramson [60] for the electrokinetic potential of egg albumin adsorbed on quartz determined by electrophoresis. The comparison was carried out through a *p*H range of 3.38–7.50. The two series of experiments lay essentially on the same curve, although less scattering was shown by the data obtained by the streaming potential method than by those obtained by electrophoresis.

Bull [61] compared the streaming potential technic, electrophoretic technic, and electroendosmotic technic on a Pyrex capillary and Pyrex particles coated with electrodialyzed gelatin or egg albumin. He found average ratios of 1.01 for ζ determined electroendosmotically and electrophoretically; 0.97 for ζ determined by the streaming potential technic

[57] H. B. Bull and R. A. Gortner, *J. Phys. Chem.*, **35**, 700 (1931).
[58] L. S. Moyer and H. B. Bull, *J. Gen. Physiol.*, **19**, 239 (1935).
[59] D. R. Briggs, *J. Am. Chem. Soc.*, **50**, 2358 (1928).
[60] H. A. Abramson, *J. Am. Chem. Soc.*, **50**, 390 (1928).
[61] H. B. Bull, *J. Phys. Chem.*, **39**, 577 (1935).

and electrophoretically, and 0.99 for ζ determined by the streaming potential technic and electroendosmotically. These ratios indicate that, at least for protein-covered surfaces and over the range of *p*H which was studied (*p*H 3.62–4.49), the three electrokinetic methods yield values which are within experimental error of one another.

He did not find a ratio of one to hold for bare quartz surfaces or for cellulose fibers, and he suggested that studies on surfaces which derive their charge by adsorption rather than by ionization are needed. A somewhat similar disagreement is noted by White, Monaghan, and Urban,[62] although they likewise report that when Pyrex particles are coated with gelatin the three methods yield comparable data. These authors report that the same values are given by electrophoretic and electroendosmotic technics with somewhat different values by the streaming potential technics where bare glass surfaces are studied, and they suggest that in the electrophoretic and electroendosmotic technics we are dealing with an electrical pull, whereas in streaming potentials we have a pressure force which they believe does not deform the fixed double layer; that is, they suggest that only a diffuse portion of the double layer moves under hydrostatic forces, whereas both the diffuse and outer portions of the fixed double layer move in an electrical field.

Capillary Size and Electrokinetic Behavior. Komagata [63] considered the influence of capillary size on the streaming potential and concluded that there is a limiting radius of the capillary below which the true streaming potential is not observed. This is due in part to the change in the thickness of the double layer with change in ionic concentration and in part to endosmotic effects which result from the streaming potential. Thus, if a given pressure is applied across a diaphragm, a streaming potential is developed, and this developed streaming potential may induce an endosmotic flow countercurrent to the original direction of streaming, so that the effective pressure is less than the hydrostatic pressure which is applied.

Bull and Moyer [64] have considered Komagata's equations and the electrical back pressure and concluded that the radii of the capillaries in the cellulose and quartz membranes studied by Bull and Gortner lie outside the critical radius. Bull [65] developed equation (87) for describing electrical back pressure.

$$\frac{P}{P_e} = \frac{\eta\kappa_s r^2}{8(\mu)^2} + 1 \tag{87}$$

[62] H. L. White, B. Monaghan, and F. Urban, *J. Phys. Chem.*, **39**, 585, 611 (1935).
[63] S. Komagata, *Researches Electrotech. Lab.* (*Tokyo*), (1934).
[64] H. B. Bull and L. S. Moyer, *J. Phys. Chem.*, **40**, 9 (1936).
[65] H. B. Bull, *Kolloid-Z.*, **60**, 130 (1932).

where P = the applied hydrostatic pressure
P_e = the electroendosmotic back pressure
r = the radius of the capillary
η = the viscosity of the liquid
μ = the electric moment (δe) of the double layer
κ_s = the specific conductivity of the liquid as it exists in the capillary.

He points out that the difficulties which some workers have experienced are unquestionably due to their working with capillaries which are below the critical radius. In electrolyte solutions the double layer is reduced in thickness, and accordingly valid results may be obtained at electrolyte concentrations of the order of 10^{-4}. When working with more dilute aqueous solutions or with organic liquids, particularly if the systems possess a high electrokinetic potential or are characterized by low specific conductances, special precautions must be taken to see that the results obtained are not affected by too small capillary radii. The cellulose diaphragms which are above the critical radius for dilute solutions of electrolytes have been found to be below the critical radius for certain pure organic liquids.[66]

The Electroviscous Effect. From a consideration of the foregoing section it is evident that the rate of flow of a liquid through a capillary tube would be influenced by the presence of an electrokinetic potential at the wall of the tube. The existence of such a potential would tend to reduce the rate of flow, for an observed pressure gradient, and the viscosity of the liquid calculated on this basis would be increased above the true viscosity. The equation describing this effect would take the form

$$\frac{\eta_e}{\eta_0} = 1 + \frac{8\mu^2}{\eta_0 \kappa_s r^2} \tag{88}$$

where η_0 = true viscosity of the fluid
η_e = apparent viscosity of the fluid

and $\eta_e - \eta_0$ would be the viscosity increment due to the electroviscous effect in the capillary. This effect would be greatest at high values of μ and low values of κ_s and r. It would attain significant values only under those conditions outlined above where the critical radius is approached.

Smoluchowski,[67] in 1916, predicted that the viscosity of a suspension of charged particles should differ from that of a similar suspension in which the particles were uncharged. This effect he pictured as being due

[66] M. A. Lauffer and R. A. Gortner, *J. Phys. Chem.*, **43**, 721 (1939).
[67] M. von Smoluchowski, *Kolloid-Z.*, **18**, 190 (1916).

to the setting up of a "cataphoretic potential" (Dorn potential) in the region of the charged particle caused by a displacement of the particle with respect to the liquid, with the accompanying effect that a greater expenditure of energy would be required to cause a given shear displacement in the system. He considered this effect to be additive to the volume effect described by the Einstein equation.[68] Smoluchowski failed to publish the derivation of his equation but essentially the same relationship has been derived by Krasny-Ergen;[69] it takes the form

$$\eta_{sp} = 2.5\phi\left[1 + \frac{3}{2\kappa\eta_0 r^2}\left(\frac{\zeta\epsilon}{2\pi}\right)^2\right] \tag{89}$$

where κ = specific conductivity of the sol
r = radius of the suspended particle (assumed spherical)
η_{sp} and η_0 = the specific viscosity of the sol and the viscosity of the solvent, respectively
ϕ = the percentage volume occupied by the disperse phase in the sol

and the other terms are the same as already defined in the electrokinetic equations. That this effect may be very important in influencing the observed viscosity of sols of colloid electrolytes at low ionic strength is indicated by the fact that small amounts of added electrolytes will often cause a very marked decrease in the viscosity of such sols.[70]

Attempts to test the veracity of the above equation of Smoluchowski, carried out by Bull [71] and by Briggs and co-workers, [72-74] have indicated that the phenomenon is, at best, only partially described by this equation.

Applications of Electrokinetic Technics. Abramson's book lists many applications of electrokinetic technics in the fields of biology and medicine. Only a few examples will be noted here.

Electrokinetic technics have been employed in the attempt to learn something of the nature of the surface composition of microorganisms and other particulate materials of biological origin. For example, Moyer [75] studied the species relationships in *Euphorbia*, as shown by the

[68] A. Einstein, *Ann. Physik*, **34**, 591 (1911).
[69] W. Krasny-Ergen, *Kolloid-Z.*, **74**, 172 (1936).
[70] H. R. Kruyt and H. G. Bungenberg de Jong, *Kolloid Biehefte*, **28**, 1 (1928).
[71] H. B. Bull, *Trans. Faraday Soc.*, **36**, 80 (1940).
[72] D. R. Briggs, *J. Phys. Chem.*, **45**, 866 (1941).
[73] C. L. Hankinson and D. R. Briggs, *J. Phys. Chem.*, **45**, 943 (1941).
[74] D. R. Briggs and M. Hanig, *J. Phys. Chem.*, **48**, 1 (1944).
[75] L. S. Moyer, *Am. J. Botany*, **21**, 293 (1934); *Botan. Gaz.*, **95**, 678 (1934).

electrophoresis of latex particles derived from the plant sap, and showed that the various taxonomic groups in twenty-one species of *Euphorbia* can be differentiated by the form of the electrophoretic curves of the latex particles. In some species the isoelectric point of the latex particles lies close to *p*H 3.0. In others it is near *p*H 4.7. Apparently in the former the latex possesses a sterol surface and in the latter a protein surface. In one taxonomic group (the poinsettias) he found a marked difference in the form of the curves and in the isoelectric point for the latex particles of *E. heterophylla*. Figure 36, taken from Moyer's data, shows this divergent curve. On investigation of the nature of the nucleus, *E. heterophylla* was found to be a tetraploid form possessing 56 chromosomes, whereas all the others in this botanical group possessed the diploid number of 28 chromosomes. Later Moyer [76] studied the constancy of the latex isoelectric points and compared data on latex from plants grown from seed in Minnesota in 1933–1934 with similar data from plants grown from seed in Pennsylvania in 1932–1933, and in all instances found that the maximum deviation of the isoelectric point did not exceed 0.1 *p*H. In a study of the electrophoretic behavior of the latex of *Asclepias*, he [77] could not distinguish between forms having 22 chromosomes and those having 24. However, he did find constant species differences in the isoelectric points and also constant species differences in the form of the electrophoretic curves. In all cases the surface of the latex in *Asclepias* appeared to be protein coated. He accordingly suggested that electrokinetic studies may provide a useful tool for the taxonomist.

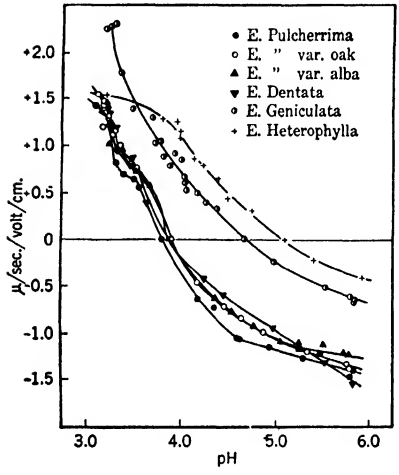


FIG. 36. Electrokinetic curves *vs.* *p*H for the latex particles of various species of *Euphorbia*. (Data of Moyer.)

Sen [78] inserted microelectrodes into the root hairs of *Azolla pinnata* and the petiole hairs of *Urtica dioica* and found that the protoplasmic granules were negatively charged and that the velocity under which they migrated was independent of their size and shape.

[76] L. S. Moyer, *Protoplasma*, **21**, 588 (1934).
[77] L. S. Moyer, *Botan. Gaz.*, **97**, 860 (1936).
[78] B. Sen, *Ann. Botany*, **48**, 143 (1934).

In a study of bacteria at various stages in the growth cycle,[79] young cells of *Escherichia coli* up to seven hours of age had a higher mobility than had older cells, and the "rough" and "smooth" cultures likewise differed in mobility.

Martin [80] used the streaming potential technic to study the identity or non-identity of various protein fractions extracted from wheat flours and found the method to yield information which could not be obtained by the usual technics of the cereal chemist.

An application in the field of industry is the observation [81] that latex particles could be "electroplated" by electrophoresis upon metal, cloth, or wooden surfaces. This observation of the electrodeposition of rubber has created a large industry.

The perfecting of the U-tube method for electrophoresis by Tiselius, Longsworth, and others has enormously increased the usefulness of this electrokinetic technic in the study of materials of biological origin. Since this method constitutes a means for analyzing proteins and other colloid electrolytes for electrophoretic homogeneity of components as well as for the number and relative amounts of such components present in a given preparation, it can be used to characterize naturally occurring mixtures such as the proteins of the blood where variations from normal have proved of diagnostic value in some cases.[82] Proteins previously considered as single entities, such as casein, have been shown to be mixtures.[83] Even crystalline egg albumin has been shown to consist of two components slightly different electrophoretically.[84] The interaction of proteins and detergents can be followed by this method.[85] Some phases of the denaturation reactions of proteins are being elucidated by this means.[86,87] The usefulness of this method in helping to characterize biocolloids better becomes increasingly evident with each new system on which it is applied.

[79] L. S. Moyer, *J. Bact.*, **32,** 433 (1936).

[80] W. McK. Martin, *J. Phys. Chem.*, **35,** 2065 (1931); **38,** 213 (1934).

[81] S. E. Sheppard, *Trans. Am. Electrochem. Soc.*, **52,** 47 (1927).

[82] See H. A. Abramson, L. S. Moyer, and M. H. Gorin, *Electrophoresis of Proteins*, p. 186 *et seq.*, Reinhold Pub. Corp., New York, 1942.

[83] R. C. Warner, *J. Am. Chem. Soc.*, **66,** 1725 (1944).

[84] L. G. Longsworth, R. K. Cannan, and D. A. MacInnes, *J. Am. Chem. Soc.*, **62,** 2580 (1940).

[85] F. W. Putnam and H. Neurath, *J. Biol. Chem.*, **159,** 195 (1945).

[86] F. W. Putnam, J. O. Erickson, E. Volkin, and H. Neurath, *J. Gen. Physiol.*, **26,** 513 (1943).

[87] D. R. Briggs and R. Hull, *J. Am. Chem. Soc.*, **67,** 2007 (1945).

CHAPTER 7

Surface Tension, Interfacial Tension, Surface Energy, and Adsorption

We have already discussed under emulsions certain phenomena which are dependent on surface tension or interfacial tension. We have noted that an efficient emulsifying agent is one which lowers the interfacial tension between the two mutually insoluble phases. In order, however, to discuss adequately certain fundamental properties of colloid systems, it is necessary to consider in somewhat more detail the forces which operate to produce the phenomena which are designated as surface tension and interfacial tension. Such a discussion will necessarily be limited to the barest outline which will be adequate to lay the foundation for the discussions which are to follow.

Willows and Hatschek [1] have presented a most excellent discussion of surface tension and surface energy and their influence on chemical phenomena. The reader is referred to their book for a more complete elaboration of the theory. Likewise, Rideal [2] and Adam [3] have devoted a considerable part of their books to these questions.

SURFACE TENSION

The energy of an ideal gas is due solely to the kinetic energy of the gas molecules. Owing to kinetic energy, the gas molecules are in constant motion, and an incessant bombardment of an individual gas molecule by the surrounding molecules takes place.

A liquid differs from a gas in that its molecules are closer together and, owing to high intermolecular attractions, it is capable of assuming a definite form, *i.e.*, it has a boundary surface. A gas always completely fills the container in which it is placed, no matter how large the container

[1] R. S. Willows and E. Hatschek, *Surface Tension and Surface Energy*, 3rd ed., J. and A. Churchill, London, 1923.

[2] E. K. Rideal, *An Introduction to Surface Chemistry*, 2nd ed., University Press, Cambridge, England, 1930.

[3] N. K. Adam, *The Physics and Chemistry of Surfaces*, 3rd ed., Oxford University Press, 1941.

may be, whereas a liquid occupies a more or less fixed volume. A liquid possesses kinetic energy similar to a gas, but, since it has a surface, it likewise possesses surface energy, and this surface energy confers upon liquids many of the properties which are not possessed by gases.

The surface of a liquid differs from the body of the liquid, in that the molecules making up the surface are at a higher energy content and may be oriented in some particular direction. The molecules in the body of the liquid are largely distributed at random, but the work of Hardy, Harkins, Langmuir, Adam, and others has shown that the molecules in the surface film of liquids are in general arranged in an orderly fashion. This question of molecule orientation will shortly be discussed at length. Suffice it to say here that the surface of a liquid behaves as if there were a "skin" drawn over the bulk of the liquid, differing in physical properties and in molecular arrangement from the bulk of the liquid beneath the surface layer. The concentration of molecules per unit area in this surface layer is usually greater than in an equivalent volume within the bulk of the liquid. The layer of molecules on the surface is more or less rigid and gives rise to the phenomenon which we call surface tension.

Surface tension is due to molecular cohesion. According to Laplace, molecules in a liquid have a pronounced attraction for one another. This attractive force, however, operates over only a short distance and is greatest at not to exceed one or two molecular diameters. It is probably negligible at a distance of 5 mμ but is quite large as we approach one molecular diameter. Figure 37 shows diagrammatically the force which

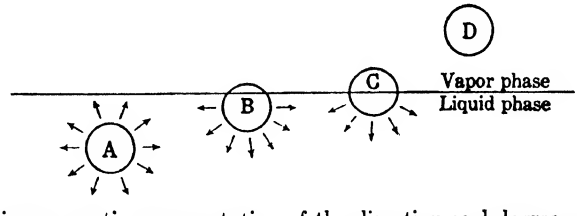


FIG. 37. A diagrammatic representation of the direction and degree of the forces of molecular attraction acting upon a molecule, A in the body of the liquid, B touching the surface, C in the interface, and D in the vapor phase.

must be overcome in bringing a molecule from the interior of the liquid phase into the surface of the liquid and from the surface of the liquid into the vapor phase. We have at A within the body of the liquid a molecule which is attracted equally from all sides by other molecules. However, as this molecule approaches and moves into the surface at B and C, it must absorb energy sufficient to overcome the unequal attraction which acts upon it as it assumes this position. The arrows in the diagram show the direction of this cohesional attraction. At A there is an equal pull

in all directions; at B there is an excess of downward and lateral pull. In order for the molecule to reach the interfacial film at C, energy must be expended to overcome a part of the downward and lateral attractions, and a further amount of energy must be expended to overcome the downward pull when the molecule passes from C into the vapor phase at D. This downward and lateral pull on the molecules which are in the surface film, or which are entering the surface film, is the tension which we measure and call surface tension. We can define surface tension in a purely experimental way as the force (dynes) necessary to equilibrate the pull exerted by 1 cm. cross section of the surface (γ = dynes per centimeter) or as the free surface energy which will be measured in terms of the work (ergs) necessary to extend that surface isothermally by 1 sq. cm. (γ = ergs per sq. cm.). An excellent example to illustrate surface tension is diagrammatically represented by the Dupré frame shown in Fig. 38. A wire loop $ABCD$ is constructed, and a movable crossbar EF of length a is placed on this loop. The area $BCFE$ is covered with a liquid film. The crossbar, which is free to move without friction, is held in equilibrium by force G (dynes). This force acts parallel to the surface of the film and perpendicular to an edge of the surface equal to $2a$ (film has two surfaces). The force per centimeter is $(G/2a)$. If bar EF is moved to position $E'F'$, through a distance d cm., the work done will be equal to Gd, and the resulting increase in surface will be $2ad$. The work per unit of surface increase (square centimeters) will be equal to $Gd/2ad = G/2a = \gamma$, which, thus, can be measured either as dynes per centimeter $(G/2a)$ or as ergs per square centimeter $(Gd/2ad)$.

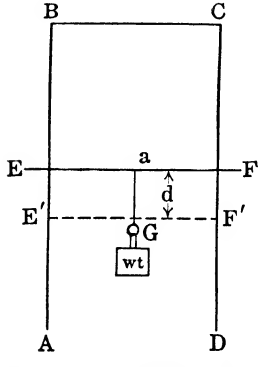


FIG. 38. A diagrammatic representation of the fact that work must be done to increase surface area.

Table 16 shows the surface tension of certain common liquids. *Surface tension decreases with an increase in temperature and becomes zero at the critical temperature.* The decrease in surface tension is directly proportional to the temperature except possibly for a short range close to the critical temperature.

Referring again to Fig. 38, we have noted that energy is expended in bringing molecules into the surface film and in removing a molecule from the surface film into the vapor phase. Stefan,[4] in 1886, proposed a generalization known as Stefan's law, in which he stated that one half of the latent heat of vaporization of a liquid is expended in pulling a

[4] J. Stefan, *Wied. Ann.*, **29**, 655 (1886).

molecule of the liquid into the surface film, the remainder being expended in pulling it out of the surface film into the vapor phase.

TABLE 16. SURFACE TENSION (γ) IN CGS. UNITS FOR CERTAIN LIQUID-AIR INTERFACES

	Temperature, °C.	γ
Chloroform	25	26.2
Methyl alcohol	20	22.61
Ethyl alcohol	20	22.27
Ethyl ether	20	17.10
Acetic acid	20	27.63
Benzene	20	28.86
Benzaldehyde	20	40.04
Aniline	20	42.58
Glycerin	20	63.4
Mercury	0	480.3
Gallium	30	735
Hydrogen peroxide (56.06%)	0	77.46
Water	0	75.64 ± 0.1
	10	74.22 ± 0.05
	20	72.75 ± 0.05
	25	71.97 ± 0.05
	30	71.18 ± 0.05

Harkins and Roberts [5] pointed out that *Stefan's law is only a rough approximation*, and that associated liquids and non-associated liquids differ markedly in their surface energy relations. Table 17 shows certain of their data. It will be noted that, instead of half of the energy of vaporization being expended in pulling the molecule out of the surface film into the vapor phase as required by Stefan's generalization, the energy expenditure for this process ranges from 51.60 per cent for mercury to 84.91 per cent for water.

Harkins and Roberts give the total energy necessary for the vaporization of a molecule of water at 10°C. as 69.6×10^{-14} erg. Of this amount, 10.5×10^{-4} erg was expended in drawing the molecule from the interior of the liquid into the surface film, and 59.1×10^{-14} erg was expended in moving the molecule from the surface film into the vapor phase. These values are widely divergent from the 50:50 ratio suggested by the generalization of Stefan.

The energy required to remove a molecule from a surface film into the vapor phase is to some extent dependent on the form of the surface. In

[5] W. D. Harkins and L. E. Roberts, *J. Am. Chem. Soc.*, **44**, 653 (1922).

TABLE 17. SHOWING FOR SOME COMMON LIQUIDS THE INTERNAL LATENT HEAT OF VAPORIZATION PER MOLECULE (λ) AND THE WORK EXPENDED (e) IN BRINGING A MOLECULE FROM THE INTERIOR OF A LIQUID INTO THE SURFACE FILM AND (j) IN CAUSING THE MOLECULE TO "JUMP" FROM THE SURFACE FILM INTO THE VAPOR PHASE

(The energy values are expressed in 10^{-14} erg; data of Harkins and Roberts.)

Liquid	Absolute Temperature, degrees	(λ)	(e)	(j)	(e/λ) Per Cent	(j/λ) Per Cent
Carbon tetrachloride	298	50.72	17.1	33.6	33.70	66.30
Ethyl ether	283.9	41.71	12.7	29.0	30.46	69.54
Benzene	298.1	52.8	19.5	33.3	36.93	63.07
Chlorobenzene	298	65.1	21.9	43.2	33.63	66.37
Ethyl acetate	303	53.1	16.0	37.1	30.13	69.87
Nitrogen	70	8.67	3.84	4.83	44.29	55.71
Oxygen	70	10.81	4.50	6.31	41.63	58.37
Methyl alcohol	413	43.2	10.1	33.1	23.38	76.62
Ethyl alcohol	383	55.6	11.7	43.8	21.04	78.77
Acetic acid	423	34.1	11.8	22.3	34.60	65.40
Water	283	69.6	10.5	59.1	15.09	84.91
Mercury	313	96.3	46.6	49.7	48.40	51.60

Fig. 39 we have 3 molecules, molecule D in a concave surface, molecule E in a plane surface, and molecule F in a convex surface, the center of the molecule being spaced at an equal distance from the surface. More of the molecule protrudes from the surface at F than at E, and more at E than at D. Accordingly there is the least cohesional downward and

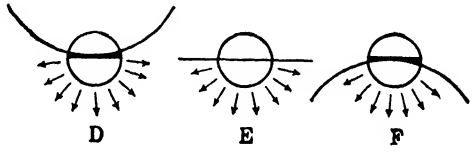


FIG. 39. A diagrammatic representation of the direction and the intensity of the forces of adhesion acting upon a molecule in concave, convex, and plane surfaces.

lateral pull of adjacent molecules on the molecule at F, and the greatest downward and lateral pull on the molecule at D. The molecule at F can escape from the surface with less expenditure of energy than molecule E, and molecule E can escape with less energy expenditure than molecule D. Accordingly molecules will spontaneously evaporate from sur-

face F to surface E, and from surface E to surface D. This difference in surface energy relationships becomes important when we are dealing with very minute particles or droplets. The small droplets with a highly curved surface, as in F, will spontaneously evaporate to larger droplets with more nearly plane surfaces. Correspondingly, a small particle with a highly curved surface will have a greater solubility in a given liquid than will a large crystal of the same material. As we shall see later, this is an important consideration when we are dealing with particles of colloidal dimensions.

INTERFACIAL TENSION

Interfacial tension differs from surface tension only in that surface tension refers to a liquid-vapor interface, whereas interfacial tension refers to liquid-liquid or solid-liquid interfaces. No satisfactory method has yet been devised for measuring accurately interfacial tensions existing at solid-liquid interfaces. The interfacial tension of liquid-liquid interfaces can, however, be very conveniently measured by technic similar to that which is used for measuring the tension existing at liquid-vapor boundaries.

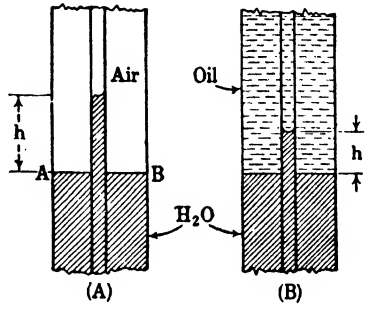


FIG. 40. A diagrammatic representation of the relationship between surface tension and interfacial tension as measured by the capillary tube method.

One of the earliest methods for the measurement of surface tension involved the determination of the height to which a liquid would rise in a capillary tube inserted into a plane surface in the liquid. Diagram A, Fig. 40, shows such a measurement. If we have a plane surface AB of a liquid in an open vessel, and into this plane surface a capillary tube is inserted and the liquid wets the wall of the capillary tube, the liquid will rise in the capillary because of the pull exerted by the adhesional forces existing between the liquid and the wall of the tube. Surface tension of the liquid tends to maintain the surface area within the tube at a minimum (*i.e.*, straight across the tube where area = πr^2). The result is that equilibrium is attained when the surface tension force is equal to the pressure (weight) of the column of liquid in the tube. The force of surface tension equals $2\pi r\gamma'$, and the weight of the liquid column equals $\pi r^2 hD$. At equilibrium

$$\gamma' 2\pi r = \pi r^2 hD \qquad \text{and} \qquad \gamma' = \tfrac{1}{2} hrD \tag{90}$$

where γ' = surface tension in grams per square centimeter
h = height, in centimeters, of the rise of the liquid in the capillary
r = the radius of the capillary tube, in centimeters
D = the density of the liquid.

In order to convert γ' into γ, the surface tension, *i.e.*, into dynes or ergs per square centimeter, it is necessary to multiply by 980.1, the gravity constant.

Diagram *B*, Fig. 40, indicates how a similar arrangement can be used to measure interfacial tension (*cf.* Pound [6]). The interfacial tension at the interface of two immiscible liquids can be calculated by the equation

$$\gamma = \frac{1}{2} Gr \left(h + \frac{r}{3} \right) (D - D') \qquad (91)$$

where γ = interfacial tension in ergs per square centimeter
G = 980.1 dynes
r = radius of the capillary tube in centimeters
h = difference in level of the two interfaces (the highest to the lowest points of the meniscus) inside and outside the capillary tube in centimeters
D and D' = density of the two liquids, respectively.

The factor $r/3$ added to h in this equation corrects for the volume of the liquid in the meniscus; it should also be included in the equation for calculation of surface tension where highest accuracy is to be attained.

Table 18 shows certain values of the interfacial tension of pure liquids against a water surface, taken from the data of Pound. In a later paper, Pound [7] determined the interfacial tension between 42 organic liquids and water or various aqueous solutions. He points out that the greater the mutual solubility between the two liquids, the lower the interfacial tension between them. He also notes that the interfacial tension between certain esters in contact with dilute acid or alkali solutions in water decreases with time, and ascribes this decrease to the hydrolysis of the esters and the formation of products of the hydrolysis which cause a decreased interfacial tension.

Bartell and Miller [8] and Mack and Bartell [9] have devised improved apparatus for the measurement of interfacial tension which is based on the double capillary principle and which can be used with liquids that are either heavier or lighter than water. The later forms of their ap-

[6] J. R. Pound, *J. Chem. Soc.*, **123**, 578 (1923).
[7] J. R. Pound, *J. Phys. Chem.*, **30**, 791 (1926).
[8] F. E. Bartell and F. L. Miller, *J. Am. Chem. Soc.*, **50**, 1961 (1928).
[9] G. L. Mack and F. E. Bartell, *J. Am. Chem. Soc.*, **54**, 936 (1932).

paratus have the advantage that very small quantities of organic liquid, not to exceed 2 ml., are necessary, and even with such small quantities of liquid precision data can be obtained.

TABLE 18. INTERFACIAL TENSION BETWEEN CERTAIN PURE LIQUIDS AND WATER

(Data of Pound)

	Interfacial Tension, 30°C.	*Interfacial Tension, 10°C.*
Benzene	32.5	34.10
Ether	11.13	10.19
Chloroform	31.41	32.17
Carbon tetrachloride	42.75	.....
Carbon disulfide	46.31	.....
Toluene	34.6	.....
Aniline	6.00	.....
Nitrobenzene	24.1	.....
Turpentine	23.0	.....
Paraldehyde	9.6	.....
Amyl alcohol	4.89	.....
Amyl acetate	10.88	.....
Ethyl acetate	6.27	.....
Cresol	4.28	.....

Interfacial tension can be obtained from surface tension data using Antonoff's theorem, that the interfacial tension is equal to the surface tension of substance A saturated with substance B, minus the surface tension of substance B saturated with substance A, or

$$\gamma_{AB} = \gamma_A - \gamma_B \tag{92}$$

Harkins [10] states that this rule does not hold rigidly since it ignores molecular orientation and therefore the composition at the interface differs from the composition of the bulk of the liquid. He suggests that the adhesional forces residing in the different ends of the molecule at the two interfaces of the oriented molecular layer have to be taken into consideration.

Harkins and Humphrey,[11] Reynolds,[12] Antonoff,[13-16] and Harkins and

[10] W. D. Harkins, *Proc. Natl. Acad. Sci. U. S.*, **5**, 569 (1919).
[11] W. D. Harkins and E. C. Humphrey, *J. Am. Chem. Soc.*, **38**, 228, 242 (1916).
[12] W. C. Reynolds, *J. Chem. Soc.*, **119**, 460 (1921).
[13] G. N. Antonoff, *J. Russian Phys. Chem. Soc.*, **39**, 342 (1907). Cited in *Chem. Z.*, [II] 1295 (1907).
[14] G. N. Antonoff, *J. chim. phys.*, **5**, 372 (1907).
[15] G. N. Antonoff, *Phil. Mag.*, [VI] **36**, 377 (1918).
[16] G. N. Antonoff, *J. Phys. Chem.*, **46**, 497 (1942).

Zollman [17] likewise studied the question of the accurate measurement of interfacial tension and have presented data on various systems. Certain data of Harkins and Zollman have already been recorded in Table 3. A striking result of their work is the extremely low interfacial tension which was secured between olive oil and a 0.15 *N* solution of sodium chloride in 0.001 *N* sodium hydroxide.

Various other methods for the measurement of surface tension and interfacial tension have been used. Thus, Morgan and Stevenson [18] showed that surface tension could be calculated from the weight of a drop falling from the tip of a capillary. Harkins' values for interfacial tension were determined by a modification of this drop-weight method. Until recently it has been believed that drop-weight technics yield the most accurate value for surface tension. However, Hauser and coworkers,[19] using the high-speed motion-picture camera, have shown that the form of the drop falling from a capillary tip is not so regular as has been believed and that secondary drops are formed which complicate the theoretical treatment of the phenomena. Viewed in the light of these high-speed motion pictures, the task of deriving appropriate mathematical corrections for the drop-weight surface tension equations appears to be formidable.

Du Noüy [20] devised a torsion balance for measuring surface tension by recording the pull required to overcome the adherence of a ring of platinum wire to the surface of a liquid. This apparatus has been modified to measure interfacial tensions. Green [21] combined the torsion balance principle with the drop-weight principle and devised an instrument to measure surface tension by recording the weight of a given number of drops of liquid. MacDougall [22] pointed out certain limitations of the du Noüy tensiometer. Hauser, *et al.*, likewise brought into question the accuracy of surface tensions and interfacial tensions determined with the du Noüy apparatus, since the high-speed motion camera shows that a complete film of liquid is drawn out from the surface when the ring leaves the surface. A valuable description and evaluation of the common methods for measurement of surface and interfacial tensions is to be found in a recent article by Harkins.[23]

Changes in interfacial tension are responsible for many of the phenom-

[17] W. D. Harkins and H. Zollman, *J. Am. Chem. Soc.*, **48**, 69 (1926).

[18] L. R. Morgan and R. Stevenson, *J. Am. Chem. Soc.*, **30**, 360 (1908).

[19] E. A. Hauser, H. E. Edgerton, B. M. Holt and J. T. Cox, Jr., *J. Phys. Chem.*, **40**, 973 (1936).

[20] P. L. du Noüy, *J. Gen. Physiol.*, **1**, 521 (1919); **7**, 625 (1925).

[21] R. G. Green, *Ind. Eng. Chem.*, **15**, 1024 (1923).

[22] F. H. MacDougall, *Science*, **62**, 290 (1925).

[23] W. D. Harkins, p. 141, *Physical Methods of Organic Chemistry*, Vol. I, edited by A. Weissberger, Interscience Publishers Inc., New York, 1945.

ena characteristic of colloid systems. Emulsification appears to be primarily dependent on interfacial tension changes. Interfacial tension changes are likewise involved in cell division. In 1876, Bütschli stated that, if a drop of oil was suspended in an aqueous solution, the droplet would break into two drops provided that a high surface tension could be produced locally at the equator of the drop. Robertson [24] contended that the droplet would pull apart at the point where the surface tension was the lowest. McClendon [25] contended that Bütschli's view was correct and carried out experiments to test the hypothesis. A chloroform-rancid olive oil droplet was suspended in a sodium chloride solution of specific gravity such that the oil droplet floated in the liquid. A dilute solution of sodium hydroxide was then allowed to flow from the tips of two pipets against the opposite poles of the oil droplet. The alkali uniting with the acid in the droplet formed a soap which decreased the surface tension at the poles of the droplet and caused the droplet to elongate into an hour-glass form or to break in two completely, owing to constriction by the higher interfacial tension at the equator.

Bancroft and Gurchot [26] repeated these experiments under somewhat more exact control and confirmed Bütschli's view. Undoubtedly interfacial tension is not the only factor operating in cell division. It does, however, provide a mechanical force which would be capable of producing the observed effects.

SURFACE ENERGY

As already noted, a liquid differs from a gas in that it has a bounding surface. The energy of a gas is solely kinetic energy. The energy of a liquid, on the other hand, is composed of two factors, the internal or kinetic energy and the surface energy. The amount of kinetic energy available depends on the initial value of the kinetic energy per unit of mass multiplied by the weight of the material. *The surface energy of a system is the product of two quantities, the intensity factor and the capacity factor.* The *intensity factor* is the *surface tension* or the *interfacial tension.* The *capacity factor* is the *extent of surface area.* The surface energy of a system can accordingly be expressed by the equation

$$S = \gamma s \tag{93}$$

where S = the surface energy
γ = the surface tension or interfacial tension in ergs per square centimeter
s = the surface area in square centimeters.

[24] T. B. Robertson, *Arch. Entwicklungsmech. Organ.*, **27,** 29 (1909).
[25] J. F. McClendon, *Am. J. Physiol.*, **27,** 240 (1910).
[26] W. D. Bancroft and C. Gurchot, *J. Phys. Chem.*, **31,** 430 (1927).

Surface energy is a very active form of energy and is readily converted either into work or into other forms of energy. Accordingly the free surface energy of a given system is decreased under many conditions. *A decrease in the free surface energy can be accomplished by (a) a reduction in area (reducing the capacity factor) or (b) a reduction of interfacial tension or surface tension (reducing the intensity factor).*

An example of the reduction of the surface energy by reducing the surface area is the coalescing of two mercury droplets into a single larger droplet. Such a coalescence takes place spontaneously when two clean mercury surfaces touch each other. The only way in which a pure liquid can decrease its surface energy is to decrease its surface area. Solutions, on the other hand, may have a higher or a lower surface tension than the pure solvent. In most instances, the surface tension of a solution is lower than that of the solvent. A solution may decrease its surface energy by bringing into the surface area an excess of the solvent or of the solute, depending on which one lowers the surface tension to the greatest extent.

The surface energy of a pure liquid or of a crystalloidal solution in bulk is insignificant in comparison with the kinetic energy of the system, inasmuch as a relatively insignificant amount of surface is present. On the other hand, the surface energy in a colloid system may be relatively enormous, by reason of a very large absolute surface. The ratio of surface energy to kinetic energy is expressed by the term specific surface, which is the ratio of the surface area of the disperse phase to the volume of the disperse phase,

$$\text{Specific surface} = \frac{\text{absolute surface}}{\text{volume}} \tag{94}$$

Table 19 shows the changes in total surface and specific surface of a hypothetical cube, 1 cm. on an edge, which has been progressively subdivided until the particles have reached colloidal dimensions. It will be noted that the surface area has increased from 6 sq. cm. to 60 sq. m. at the upper limit of the colloidal realm, and to 6,000 sq. m. for particles 1 $m\mu$ on an edge. The specific surface has increased from 6 to 60,000,000. During this great increase in specific surface, the internal energy or kinetic energy has remained constant. It is usually agreed that if the specific surface s/V is less than 10,000, the internal energy of a system predominates. On the other hand, if s/V is greater than 10,000, the surface energy becomes the predominating form of energy, and, when the specific surface reaches such enormous values as are reached within the colloid realm, the reactions which take place are to a large measure re-

actions involving surface energy changes. For this reason many substances which are totally different in their chemical composition acquire properties, when in the colloid state, which are similar or identical.

Every colloidal particle in suspension is surrounded by a film of the dispersions medium. Accordingly we have operating in these films an interfacial tension which controls, at least in a large measure, the energy relationships of a liquid-solid or a liquid-liquid interface in the same way as the surface tension expresses the energy relationships at a gas-liquid interface. Although we can measure interfacial tension at a liquid-liquid interface, we are as yet unable to measure directly the forces of

TABLE 19. SURFACE AREA POSSIBLE BY THE SUBDIVISION TO COLLOIDAL DIMENSIONS OF A CUBE 1 CM. ON AN EDGE

Edge of Cube	Number of Cubes	Total Surface	Relative Surface Energy,* Assuming that Water Is Being Subdivided, ergs
1 cm.	1	6 sq. cm.	438
0.1 μ	1,000,000,000,000,000	60 sq. m. (645.84 sq. ft.)	43,800,000
0.01 μ	1,000,000,000,000,000,000	600 sq.m. (6,458.4 sq. ft.)	438,000,000
1 mμ	1,000,000,000,000,000,000,000	6,000 sq.m. (64,584 sq. ft.)	4,380,000,000 (105 cal.)

* The energy in this column is the calculated surface energy due to increased surface. No account has been taken of other possible forms of energy, such as electrical.

interfacial tension at a solid-liquid or a solid-gas interface. It is highly probable, however, that these forces are as great as, if not greater than, the known forces operating at liquid-liquid or liquid-gas interfaces. Even if we postulate forces no greater than those of liquid-gas interfaces, we have forces of sufficient magnitude to explain the surface reactions characteristic of colloid systems.

Wo. Ostwald [27] ascribes to Wenzel the law that "*the amount of chemical change in a unit of time is proportional to the absolute surface.*" Since colloid systems show a great absolute surface, they may also show marked chemical activity.

[27] Wo. Ostwald, *A Handbook of Colloid-Chemistry*, translated by Martin H. Fischer, p. 93, P. Blakiston's Son and Co., Philadelphia, 1915.

Powdered sulfur is practically without action on a silver surface; colloidal sulfur, on the other hand, has an energetic action on a silver surface, causing the formation of the black silver sulfide. A smooth polished platinum sheet can be inserted into a solution of hydrogen peroxide without causing any appreciable amount of decomposition of the peroxide; if the platinum foil is roughened, a slow evolution of oxygen takes place; if powdered platinum is added to the hydrogen peroxide, the oxygen is formed rather rapidly; platinum black causes an energetic decomposition; and a colloidal platinum sol added to hydrogen peroxide may cause a violent explosion. Colloidal platinum in a dilution as great as 1 gram atom of platinum in 70,000,000 liters of water can be detected by its decomposing action on hydrogen peroxide. No change in the chemical nature of the platinum has taken place, but the change in specific surface has shifted the chemical reaction rate in the same way as a rise in temperature would shift it. Bancroft and Magoffin [28] note that a number of reactions are affected by the nature of the cathode surface. Thus, at 25° there is practically no reduction of chlorate to chloride when a smooth platinum cathode is used, but reduction takes place readily when a pulverulent platinum or copper cathode is used. Similarly NO_3 and NO_2 are reduced about equally readily with a smooth platinum cathode, but NO_2 is reduced more readily than NO_3 by a platinized platinum cathode, and NO_3 is reduced more readily than NO_2 by pulverulent copper or silver cathodes. Bancroft suggests that these results may be caused by an adsorption and activation which take place prior to reduction. It is entirely probable that many of the reactions which we call catalytic are in reality due to a change in the specific area of the substance used as the catalyst. In fact, this may well be the predominating factor.

Powdering a substance before dissolving it replaces, by mechanical work, work which must be expended in the solution process. Consequently more energy is available for the work of solution of a small particle than of a large particle. Accordingly, finely powdered substances show a greater solubility and a greater heat of solution than do coarsely powdered substances or large crystals. Stas, in 1870, obtained the following values as the solubility of silver chloride: granular AgCl, 0.001 gram per liter at 15°; powdered AgCl, 0.0060 gram per liter at 17°; flocculent AgCl, 0.0140 gram per liter at 20°.

Hulett,[29,30] in a study of the solubility of gypsum, found that particles

[28] W. D. Bancroft and J. E. Magoffin, *J. Am. Chem. Soc.*, **57**, 2561 (1935).

[29] G. A. Hulett, *Z. physik. Chem.*, **37**, 385 (1901); **47**, 357 (1904).

[30] G. A. Hulett, Chap. 36, *Colloid Chemistry*, Vol. I, edited by J. Alexander, Chemical Catalog Co., New York, 1926.

0.3 μ in diameter had a solubility of 18.2 millimoles per liter, whereas particles 2 μ in diameter showed the normal solubility of 15.3 millimoles per liter. Similarly, he found that small particles of barium sulfate of 0.1 μ radius had a solubility of 4.15 millimoles per liter, whereas particles of 1.8 μ radius had the normal solubility of 2.29 millimoles. From his data he calculated the interfacial tension of the solid-liquid. However, Freundlich [31] pointed out that the equation which Hulett used was incorrect. In its place, he proposed the equation

$$\frac{RT}{M} \log_e \frac{C_2}{C_1} = \frac{2\gamma_{sl}}{rD} \tag{95}$$

or

$$\gamma_{sl} = \frac{RTDr}{2M} \log_e \frac{C_2}{C_1} \tag{96}$$

where R = the gas constant = 8.31×10^7 ergs
T = the absolute temperature
D = the density of the solid phase
r = the radius of the smaller particles
C_2 = the concentration of the solution in equilibrium with the particles of radius r
C_1 = the concentration of the solution in equilibrium with the larger particles, *i.e.*, the concentration of a solution in equilibrium with the massive solid phase
γ_{sl} = the interfacial tension of the solid-liquid
M = the molecular weight of the substance the solubility of which is being determined.

By use of equation (96) and Hulett's data, the interfacial tension of gypsum-water is found to be 1,140 ergs per square centimeter, and of the barium sulfate-water system, 1,420 ergs per square centimeter.

Jones [32] later proposed a more complicated equation, taking into consideration the dissociation of the salts in the solution. His equation yields 1,048 ergs per square centimeter for the gypsum system and 1,332 ergs per square centimeter for the barium sulfate system.

Dundon and Mack [33] and Dundon [34] repeated Hulett's experiments and extended them to other salts. They pointed out that Hulett's values

[31] H. Freundlich, *Kapillarchemie*, p. 144, Akademische Verlagsgesellschaft, Leipzig, 1909.

[32] W. J. Jones, *Z. physik. Chem.*, **82**, 448 (1913).

[33] M. L. Dundon and E. Mack, Jr., *J. Am. Chem. Soc.*, **45**, 2479 (1923).

[34] M. L. Dundon, *J. Am. Chem. Soc.*, **45**, 2658 (1923).

for gypsum are too high, probably because during the powdering of the gypsum a part of the crystal water is lost and $CaSO_4 \cdot H_2O$ has a much greater solubility than gypsum.

Table 20 gives their data. The last column in the table represents, as nearly as can be determined with our present technic, the probable interfacial tensions in ergs per square centimeter of the various substances against a water interface. Glasstone [35] from similar studies calculated an interfacial energy between lead oxide and sodium hydroxide of 1,860 ergs per square centimeter at 20°.

TABLE 20. INTERFACIAL TENSIONS AT SOLID-WATER INTERFACES

Substance	*Diameter of Particle, μ*	*Increase in Solubility, per cent*	γ
PbI_2	0.4	2.0	130
$CaSO_4 \cdot 2H_2O$	0.5	4.8	356
$CaSO_4 \cdot 2H_2O$	0.2 0.3	12.3	385
Ag_2CrO_4	0.3	10.0	575
PbF_2	0.3	9.0	900
$SrSO_4$	0.25	26.0	1,400
$BaSO_4$	0.1	80.0	1,250

Lipsett, Johnson, and Maass [36] found a difference of 24.0–40.7 calories per mole in the heat of solution of finely divided sodium chloride, where the particles had an average diameter of 1.2–1.4 μ, over that for the heat of solution of the coarsely crystalline salt. They point out that the difference in heat of solution is the heat which is bound up in the excess surface energy of the fine salt, and from their experimental values they find the surface energy of sodium chloride to range from 356 to 406 ergs per square centimeter.

We have already noted that colloidal particles may increase in size, as the result of crystal growth, and that the rate of crystal growth depends in part on the supersaturation of the solution. It must be obvious that, if small particles have a greater solubility than large particles and if a substance dispersed in small particles has a greater specific surface and consequently a greater surface energy per unit of mass, small particles will dissolve in a solution which has not yet reached the saturation value for such particles but which is already supersaturated with

[35] S. Glasstone, *J. Chem. Soc.*, **119**, 1689 (1921).

[36] S. G. Lipsett, F. M. G. Johnson, and O. Maass, *J. Am. Chem. Soc.*, **49**, 925; 1940 (1927); **50**, 2701 (1928).

respect to large crystals or a plane surface. Accordingly in a system containing particles of various sizes, equilibrium will eventually be established when all small particles have disappeared and only relatively large crystals are present.

Many papers have been written to present experimental work involving these considerations of crystal growth. Lipsett, Johnson, and Maass noted that, when their finely divided salt was placed in a moist atmosphere, water was rapidly absorbed, 1 gram of the salt absorbing 1 mg. of water in 20 minutes, 3 mg. in 2 hours, and 6.9 mg. in 5 hours. The 1 gram of salt originally occupied a volume of 8 cc. At the end of 5 hours the volume had contracted to approximately one fourth of its original bulk. A microscopical examination showed that, whereas the original dry salt had contained innumerable particles approximately 1 μ in diameter, after 30 minutes the main bulk of the salt was made up of particles ranging from 4 μ to 10 μ in diameter. They ascribed this change to the greater solubility of the small particles which dissolved in the water that was absorbed and then recrystallized as larger particles, and noted that a few milligrams of water act catalytically in transforming a large mass of finely divided salt into salt of much coarser degree of subdivision.

Kolthoff [37] considered the solubility of large and small crystals as a factor in analytical procedures, pointing out that with hard crystals, such as $BaSO_4$, the *relative* supersaturation may be enormous ($\pm$1,000) when large crystals and small crystals are compared, whereas with soft crystals, such as Ag_2CrO_4 (hardness *ca.* 2), the ratio value is only about 4.0 and with PbI_2 (hardness <2) the ratio is only 1.38. In the latter substances the ratio is not great enough to make appreciable differences in analytical determinations, although for nephelometric analyses it would still be a factor. Grosvenor [38] observed air droplets in Nujol (γ = 44 dynes). The air droplets disappeared rapidly after they reached a diameter of about 5 μ. At about 3 μ the shrinkage was rapid enough to be followed continuously, and at 1.5 μ they disappeared in less than 0.5 sec. At 5 μ the pressure within the droplet, due to surface tension, was about 8 lb. per sq. in. above atmospheric pressure, and at 1.5 μ the pressure had risen to about 2 atmospheres. The collapse was due to the solution of the small bubble in the oil, larger bubbles growing at the expense of the smaller ones. Grosvenor states that bubbles less than 3 μ in diameter cannot be stable in the liquid and they will vanish

[37] I. M. Kolthoff, *Z. anal. Chem.*, **86**, 34 (1931).
[38] W. M. Grosvenor, *Science*, **72**, 244 (1930).

by loss of gas either to the larger bubbles or to the air-liquid surface. Many highly viscous liquids, such as syrups and lacquers, become freed from air bubbles by this differential solution-gas liberation process.

If finely divided materials possess an excess surface energy, they should melt at a lower temperature than larger particles. Here, again, we have the principle of mechanical work expended in subdividing the particle replacing thermal work. Palow [39] observed that increasing specific surface of salol (phenyl salicylate) one hundred times depresses the melting point 2.8°C. This may explain why chemists do not always agree on the exact melting point of a chemical compound.

This phenomenon may also explain in part why certain biological organisms are not frozen when the temperature falls to a few tenths of a degree below the freezing point, or why water in fine capillary spaces,

TABLE 21. CHANGE IN THE FREEZING POINT OF LIQUIDS WHEN IN THE FORM OF FILMS ON FINELY DIVIDED POWDERS

(Data of Parker)

Substance	Liquid Added, per cent by weight	Depression of Freezing Point over that of Pure Liquid, °C.	Substance	Liquid Added, per cent by weight	Depression of Freezing Point over that of Pure Liquid, °C.
Water			Benzene		
Al_2O_3	25.0	2.118	SiO_2	3.3	0.670
Al_2O_3	30.0	1.227	SiO_2	5.0	0.490
Al_2O_3	35.0	0.650	SiO_2	10.0	0.225
Al_2O_3	40.0	0.370	SiO_2	20.0	0.110
Benzene			Nitrobenzene		
Al_2O_3	30.0	1.337	Al_2O_3	50.0	1.720
Al_2O_3	35.0	0.682	Al_2O_3	60.0	1.175
Al_2O_3	40.0	0.492	Al_2O_3	70.0	0.810
Al_2O_3	50.0	0.212	Al_2O_3	80.0	0.580
			Al_2O_3	100.0	0.200

[39] P. Palow, *Z. physik. Chem.*, **65**, 1, 545 (1908–9); **74**, 562 (1910).

such as the interstices of a clay soil, has a freezing point below 0°C. If water is finely divided, it should cause a depression of the temperature at which ice forms. Parker [40] investigated the effect of mixing finely divided materials with water, benzene, and nitrobenzene. In each instance, he used materials which are insoluble in the liquid phase, so that the depression of the freezing point which was observed cannot be ascribed to the presence of a solution. It must be obvious from Table 21 that it is impossible to calculate the concentration of a solute in a solvent from freezing point data, if inert materials possessing large specific surface are present in the system at the time of making the freezing point measurements.

Reinders' Theorem. Reinders [41] investigated the distribution of a suspended powder or of a colloidally dispersed material between two mutually insoluble liquids. Let us assume, for example, that we have a red gold hydrosol and that it is mixed with petroleum ether. Will the disperse phase remain in the water or will it migrate to the petroleum ether, or will it form a layer at the interface between the water and the petroleum ether? The distribution will depend on three interfacial tensions, the interfacial tension between solid and water, γ_{sw}; the interfacial tension between water and petroleum ether, γ_{wo}; and the interfacial tension between solid and petroleum ether, γ_{so}.

If $\gamma_{so} > \gamma_{wo} + \gamma_{sw}$, the solid will remain suspended in the water.

If $\gamma_{sw} > \gamma_{wo} + \gamma_{so}$, the solid will leave the water and go into the oil phase.

If $\gamma_{wo} > \gamma_{sw} + \gamma_{so}$, or if none of the three interfacial tensions is greater than the sum of the other two, the solid particles will collect at the boundary between the water and the oil. Strictly speaking, for a three-dimensional particle, if the particle can arrange itself in any position where the sum of the energies of its interface left in contact with the water and the oil, $A_{sw}\gamma_{sw} + A_{so}\gamma_{so}$, is less than the surface energy of the oil-water interface which it displaces, $A_{ow}\gamma_{ow}$, it will come to equilibrium in that position. A distinctly heteropolar solid particle would have the highest probability of coming to rest in the interface. These conditions will also hold for a particle which comes in contact with a film of oil instead of a layer of oil or a globule of oil, and will determine whether or not the particle is wetted by the oil film. These principles apply to practical problems and are particularly important in dealing with the adherence of insecticidal and fungicidal dusts and oil sprays.

[40] F. W. Parker, *J. Am. Chem. Soc.*, **43**, 1011 (1921).

[41] W. von Reinders, *Kolloid-Z.*, **13**, 235 (1913).

Contact Angles, Adhesion Tension, and Degree of Wetting. Bartell [42-55] and his students have contributed greatly to our knowledge of adhesion tension and technics by which we can express quantitatively the degree of wetting. His entire series of papers should be consulted by the student who is interested in the energy relationships of solid-liquid interfaces.

Contact Angles. If a drop of liquid is placed in contact with a solid, the resulting equilibrium will be determined by three factors—the surface tension of the liquid, the surface tension of the solid, and the solid-liquid interfacial tension. We can see from Reinders' theorem that, if the surface tension of the solid is greater than the sum of the two other tensions, the drop will spread and become a film upon the solid. If, however, the surface tension of the solid is less than the sum of the other two tensions, the liquid drop will remain upon the solid, and there will be a definite angle of contact between the liquid and the solid which will be related to the adhesion tension between the liquid and the solid and also to the degree that the liquid wets the solid. If the angle of contact is zero, there will be complete wettability and a high adhesion tension. If the angle of contact is 180 degrees, no wetting of the solid by the liquid occurs. Figure 41 shows contact angles of different magnitudes. Where the liquid actually contacts the solid,

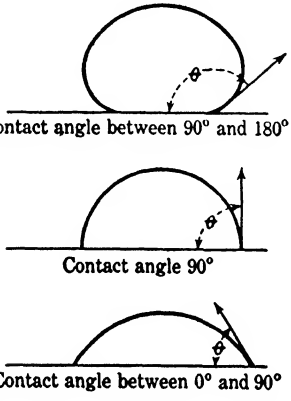


FIG. 41. Types of liquid-solid contact angles as indicative of different degrees of wettability.

[42] F. E. Bartell and H. J. Osterhof, *Ind. Eng. Chem.*, **19**, 1277 (1927).

[43] F. E. Bartell and H. J. Osterhof, Colloid Symposium Monograph, Vol. V, p. 113, Chemical Catalog Co., New York, 1928.

[44] F. E. Bartell and F. L. Miller, *Ind. Eng. Chem.*, **20**, 738 (1928).

[45] F. E. Bartell and C. K. Sloan, *J. Am. Chem. Soc.*, **51**, 1637 (1929); 1643 (1929).

[46] F. E. Bartell and C. N. Smith, *Ind. Eng. Chem.*, **21**, 1102 (1929).

[47] H. J. Osterhof and F. E. Bartell, *J. Phys. Chem.*, **34**, 1399 (1930).

[48] F. E. Bartell, p. 41, in Alexander's *Colloid Chemistry*, Vol. III, Chemical Catalog Co., New York, 1931.

[49] N. S. Davis, Jr., and H. A. Curtis, *Ind. Eng. Chem.*, **24**, 1137 (1932).

[50] F. E. Bartell and H. Y. Jennings, *J. Phys. Chem.*, **38**, 495 (1934).

[51] F. E. Bartell and C. W. Walton, Jr., *J. Phys. Chem.*, **38**, 503 (1934).

[52] F. E. Bartell and G. B. Hatch, *J. Phys. Chem.*, **39**, 11 (1935).

[53] F. E. Bartell, J. L. Culbertson, and M. A. Miller, *J. Phys. Chem.*, **40**, 881 (1936).

[54] F. E. Bartell and M. A. Miller, *J. Phys. Chem.*, **40**, 889 (1936); 895 (1936).

[55] F. E. Bartell and F. C. Benner, *J. Phys. Chem.*, **46**, 847 (1942).

there is a decrease in free surface energy, and the following relationship must hold:

$$\gamma_s + \gamma_l = \gamma_{sl} + \Delta F \tag{97}$$

where ΔF = the decrease in free surface energy.

Young, in 1805, derived the equation for the equilibrium conditions existing in a solid-liquid system in terms of the contact angle and the interfacial tension. His equation is

$$\gamma_s - \gamma_{sl} = \gamma_l \cos \theta \tag{98}$$

where θ = the contact angle.

Contact angles can be measured by placing a drop of a liquid on a plane plate of the solid material and viewing the drop through a microscope with the objective horizontal to the plane of contact, or the drop on the plane surface can be placed in the beam of a projection lens (the optical system of the Zsigmondy slit ultramicroscope is well adapted to such use), and the enlarged image of the drop in contact with the plane surface can be thrown upon a ground-glass screen where the outline of the contact angle can be drawn and later measured.

Bartell has developed methods for the determination of the contact angle by the pressure necessary to displace one liquid by another, and this method has been found of great value in certain industrial problems. Unfortunately it is necessary to know the average pore radius in the diaphragm in which the liquid displacement occurs, and no method has yet been devised for using this method in diaphragms of hydrophilic (or lyophilic) colloids where the pore radius is different in one liquid from that in another liquid.

Bartell has likewise devised methods of measuring the contact angle by dipping a vertical rod of the solid to be tested into a liquid and throwing the image of the angle either on the screen where it can be measured or on a photographic plate where it can be permanently recorded. With the vertical rod method the advancing contact angle can be measured by slowly lowering the rod into the liquid, or the receding angle can be measured by slowly raising the rod from the liquid.

Adhesion Tension. Freundlich states that the adhesion tension between a solid and a liquid is a fraction of the surface tension of the liquid and equals the term $\gamma_l \cos \theta$ in equation (98).

$$A_{1,2} = \gamma_l \cos \theta \tag{99}$$

where $A_{1,2}$ = the adhesion tension of solid-liquid.

Then from equation (98), we have

$$A_{1,2} = \gamma_s - \gamma_{sl} \tag{100}$$

The adhesion tension, therefore, is the difference in dynes per centimeter or ergs per square centimeter between the surface tension of the solid against air and its interfacial tension against the liquid. It is the decrease in free surface energy which occurs when 1 sq. cm. of solid-liquid interface is substituted for 1 sq. cm. of solid-air interface, and accordingly it is the energy factor on the magnitude of which depends the wetting or non-wetting of a solid by a liquid.

Using the displacement method, Bartell and Osterhof determined adhesion tensions at silica-liquid and carbonblack-liquid interfaces. Table 22 shows certain of their data. It is obvious that a suspension

TABLE 22. SHOWING ADHESION TENSIONS OF SILICA AND CARBONBLACK FOR VARIOUS LIQUIDS

(Data of Bartell and Osterhof)

Liquid	Silica-Liquid Interface		Carbonblack-Liquid Interface	
	Displacement pressure, dynes/cm.	Adhesion tension, dynes/cm.	Displacement pressure, dynes/cm.	Adhesion tension, dynes/cm.
Water	...	82.82	...	54.74
Aniline	8	82.00	1,266	60.51
Carbon tetrachloride	409	40.69	6,935	89.45
Hexane	395	42.13	3,330	69.93
Benzene	295	52.43	5,775	81.08
α-Bromonaphthalene	397	41.92		

will be most stable in that liquid which has the highest adhesion tension for the solid. Thus, from Table 22, it is evident that silica will form the most stable dispersion in water and carbonblack will be most stable in carbon tetrachloride. These discussions of adhesion tensions are simply restating in other terms the factors already noted under Reinders' theorem.

Since immersion in a particular liquid involves wettability, solid-liquid adhesion tensions become of great importance in the adherence of insecticides, dusts, oil sprays, paints, varnishes, etc. Bartell and Walton [51] essentially propose adhesion tensions as a quantitative measure of the hydrophilic or hydrophobic properties of surfaces, *i.e.*, wettability by water or wettability by oils with Al_2O_3 and carbon as solid extremes, and water and *n*-heptane as extreme liquids. They were able

to prepare by heat treatment samples of stibnite (Sb_2S_3) varying from extreme wettability by water to extreme wettability by oil. Table 23 shows certain of their adhesion tension data. Figure 42 shows the adhesion tensions of the various stibnites noted in Table 23 and of various other solids as measured against a variety of liquids. Rather interestingly, *n*-butyl acetate appears to show no preferential wetting in favor of any of these solids. Liquids more polar than *n*-butyl acetate tend to preferentially wet alumina and silica. Liquids less polar tend to preferentially wet carbon. As Bartell suggests, studies of this sort may give means of quantitatively evaluating degrees of hydrophilic or hydrophobic characteristics.

Harkins and Livingston [56] have pointed out that the adhesion tension as here defined is not a true measure of the affinity of the liquid for the solid. This affinity can better be estimated in terms of the energy or work of adhesion, W_a, earlier defined by Harkins [57] as the work required to separate, at the solid-liquid interface, a unit area of solid from a unit area of liquid. In this process a new unit area of the liquid and one of the clean solid will be formed.

$$W_a = \gamma_{so} + \gamma_l - \gamma_{sl} \tag{101}$$

In this equation γ_{so} refers to the surface tension of the clean solid surface in contrast to the value γ_s of equation (98) which must refer to the surface tension of the solid surface saturated with the vapor of the liquid. These values, γ_{so} and γ_s, are not equal but will differ by an energy factor F which is the energy change occurring when the vapor is adsorbed to saturation by a unit area of the solid surface. Thus

$$\gamma_s = \gamma_{so} - F \tag{102}$$

and therefore

$$W_a = \gamma_s - \gamma_{sl} + \gamma_l + F \tag{103}$$

and the value of the adhesion tension $A_{1,2}$ is less than the energy of adhesion W_a by a factor equal to $\gamma_l + F$. Because of the tendency of many workers to confuse the terms γ_s and γ_{so}, Harkins and Livingston suggest that use of the term "adhesion tension" be dropped and be replaced by the term *spreading pressure* which is defined by them as

$$\phi_{l/s} = \gamma_s - \gamma_{sl} = \gamma_l \cos\theta \tag{104}$$

The value of $F = \gamma_{so} - \gamma_s$, for many solids, will be much greater in magnitude than the value of $\gamma_s - \gamma_{sl}$. Boyd and Livingston [58] have

[56] W. D. Harkins and H. K. Livingston, *J. Chem. Phys.*, **10**, 342 (1942).
[57] W. D. Harkins, G. L. Clark, and L. E. Roberts, *J. Am. Chem. Soc.*, **42**, 700 (1920).
[58] G. E. Boyd and H. K. Livingston, *J. Am. Chem. Soc.*, **64**, 2383 (1942).

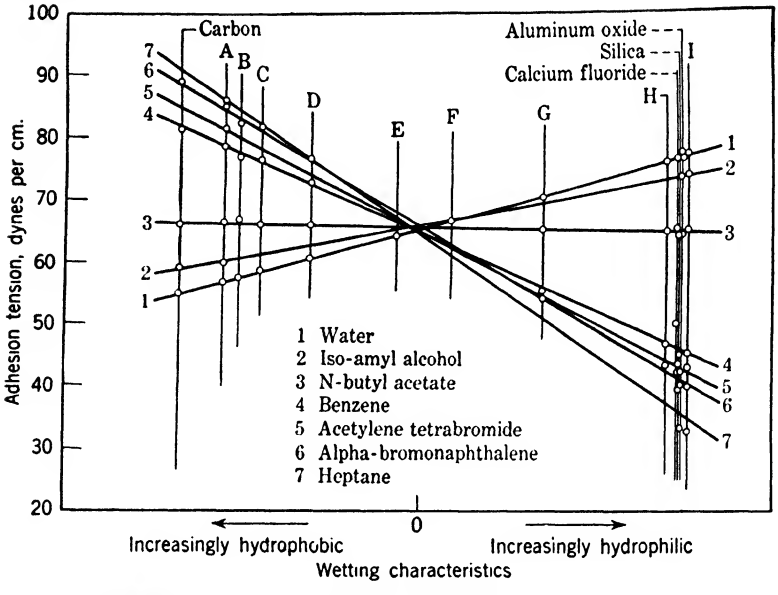


FIG. 42. Showing the relationships between adhesion tension, the polar or non-polar nature of the liquid, and the hydrophilic or hydrophobic nature of the solid. (Data of Bartell and Walton. By permission, *Journal of Physical Chemistry.*)

TABLE 23. ADHESION TENSIONS OF HEAT-TREATED STIBNITE AGAINST WATER AND BENZENE

(Data of Bartell and Walton)

Stibnite	Treatment	A_{sl} (H_2O), dynes/cm.	A_{sl} (C_6H_6), dynes/cm.
A	Ground only	56.5	78.4
B	Heat 1 hr. at 170°	57.2	76.6
C	Heat 2 hr. at 170°	58.1	76.0
D	Heat 3 hr. at 170°	60.3	72.6
E	Heat 4 hr. at 170°	64.2	....
F	Heat 5 hr. at 170°	66.5	....
G	Heat 6 hr. at 170°	70.5	55.0
H	Heat 8 hr. at 170°	76.0	47.0
I	Heat 8 hr. at 400–420° (in nitrogen)	77.1	45.4

been able to calculate this quantity for several solids and liquid pairs by use of the equation

$$\gamma_{so} - \gamma_s = F = \frac{RT}{M\Sigma}\int_0^{p_0} \frac{q}{p}\,dp \tag{105}$$

where M = molecular weight of the vapor molecules adsorbed in quantity q (in milligrams) per gram of solid of total surface Σ per gram, at the relative vapor pressure p of the vapor. Integration of the equation from $p = 0$ to $p = p_0$, the saturation pressure, is obtained by a graphical method.

If a solid is brought into contact with the interface of contact between two liquids, liquid 1 will tend to displace liquid 2 in its contact with the solid if $\gamma_{sl_1} < \gamma_{sl_2}$, where γ_{sl_1} is the interfacial tension at the interface of the solid, saturated with respect to the vapor of both liquids, against liquid 1, itself saturated with respect to liquid 2, and γ_{sl_2} is the corresponding interfacial tension for the interface solid-liquid 2. The displacement pressure $D_{l_1l_2}$ will be

$$D_{l_1l_2} = \gamma_{sl_2} - \gamma_{sl_1} = \gamma_{l_1l_2}\cos\theta_{sl_1l_2} \tag{106}$$

and can be calculated from the interfacial tension between the two liquids (each saturated with respect to the other) and the angle of contact made by the interface of the two liquids with the solid.

Mudd and Mudd [59] suggested a technic whereby one could study the relative wettability of particles by oil or by water. The particles suspended in water were placed on a microscope slide and a drop of oil was added adjacent to the water drop. When a cover slip was placed over this preparation, an oil-water boundary was formed; and, when the slide was observed under a microscope (preferably with dark-field illumination), it was observed that the oil boundary advanced across the field. When the oil boundary reached a particle which was easily wetted by oil, the particle progressed through the oil-water interface with little or no distortion of the interface. If, however, the particle was not wetted by oil or was very difficultly wetted, the interface was compressed and distorted, although eventually the particle might break through and be engulfed by the oil. In extreme cases such engulfment left the particle with a surrounding water film so that in reality there was no wetting of the particle by the oil.

Nugent [60] diagrammed the appearance of the oil-water interface of the Mudd and Mudd phenomenon in eight degrees of wettability from an extreme case where a particle "dissolves" in the oil to one where a par-

[59] S. Mudd and E. B. H. Mudd, *J. Exptl. Med.*, **40**, 633 (1924); 647 (1924).

[60] R. L. Nugent, *J. Phys. Chem.*, **36**, 449 (1932).

ticle passes the interface only when surrounded by a water film or where the moving boundary simply pushes the particle ahead of it. His diagram is shown in Fig. 43. Moyer [61] used this technic to determine the physical state of the surface of latex particles from various species of *Euphorbia* and found that species with a low isoelectric point of *p*H ±3.0 showed preferential wettability by oil, whereas those with a higher isoelectric point of *p*H ±4.5–5.0 possessed surfaces which were resistant to oil wetting and were preferentially wetted by water. These observations confirmed the electrokinetic studies where the assumption had been made that the former probably had a sterol-like surface, whereas the latter possessed a protein protective surface.

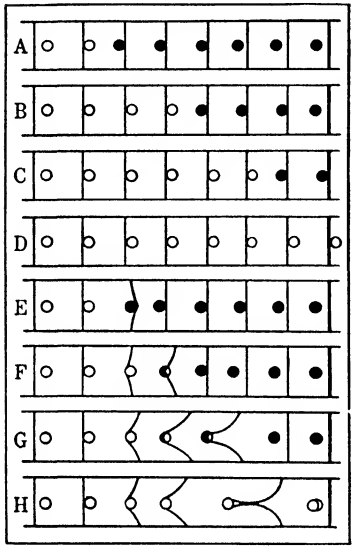


FIG. 43. A diagrammatic representation of various degrees of wettability, as shown by the Mudd and Mudd technic. (After Nugent.)

Orientation of Molecules in Surface or Interfacial Films. We have already referred to the work of Hardy, Harkins, Langmuir, and Adam as having shown that, whereas the molecules in the body of a liquid are apparently distributed at random, those molecules which enter into the surface film are in general arranged in a more or less orderly fashion. Although undoubtedly the germ of the idea of molecular orientation originated with Hardy,[62] nevertheless Harkins,[63-66] and Langmuir,[67] independently developed this idea along somewhat different lines. The work of Harkins originated from his interest in the subject of energy at inter-

[61] L. S. Moyer, *Am. J. Botany*, **22**, 609 (1935).

[62] W. B. Hardy, *Proc. Roy. Soc.* (*London*), **A86**, 610 (1912); **88**, 303, 313 (1913).

[63] W. D. Harkins, F. E. Brown and E. C. H. Davies, *J. Am. Chem. Soc.*, **39**, 354 (1917).

[64] W. D. Harkins, E. C. H. Davies, and G. L. Clark, *J. Am. Chem. Soc.*, **39**, 541 (1917).

[65] W. D. Harkins and Y. C. Cheng, *J. Am. Chem. Soc.*, **43**, 35 (1921).

[66] W. D. Harkins and N. Beeman, *J. Am. Chem. Soc.*, **51**, 1674 (1929). (*Cf.* also Harkins' papers in Alexander's *Colloid Chemistry*, p. 192, and Colloid Symposium Monograph, Vol. II, p. 141.)

[67] I. Langmuir, *J. Am. Chem. Soc.*, **39**, 1848 (1917). (*Cf.* also Colloid Symposium Monograph, Vol. III, p. 48.)

faces between liquids; that of Langmuir originated from his work on liquid films. Both came to the general conclusion that surface tension or interfacial tension phenomena are, in general, characterized by the orientation and the packing of molecules in surface layers and that the forces involved in this action are the forces of solution and the forces of interfacial tension. Hardy [68] extended his work in a different direction in studying the problems of friction and lubrication.

In reality, monomolecular films existing at interfaces constitute a different state of matter from that which we are accustomed to consider, for they act as *two-dimensional gases, liquids, and solids.*[69] They are free to move only in two directions and not in a third direction. In a highly expanded form they behave as two-dimensional gases; when the molecules come closer together so as to approximate the distances between molecules in the liquid state, they behave as two-dimensional liquids; and when still further compressed they become two-dimensional solids. The equations of state for gases, liquids, and solids accordingly have to be modified to account for the kinetic behavior of the molecules in these two-dimensional films.

Concepts introduced by studies on molecular orientation may be summarized as follows.[70]

Liquids may be divided, in general, into two great classes, those in which the molecule is essentially symmetrical and those in which the molecule is more or less unsymmetrical. Pentane may be taken as an example of a symmetrical molecule, and acetic acid as an example of an unsymmetrical molecule. In the pentane molecule the two ends of the hydrocarbon chain are identical. Accordingly we would expect the two ends to behave identically toward an interface. In acetic acid the two ends of the molecule are very dissimilar, one being of hydrocarbon nature, the other (—OH) very similar to water. Accordingly we would expect the two ends of the acetic acid molecule to behave differently toward an interface. To express these differences, the term *polar group* has been introduced, and the radicals —OH, —COOH, —CHO, —CN, —$CONH_2$, —SH, —NH_2, —$NHCH_3$, —NCS, —COR, —COOM, —COOR, —NO_2, —CH=CH_2, —C≡CH, and groups which contain oxygen, nitrogen, sulfur, iodine, bromine, and chlorine, and double and triple bonds have

[68] W. B. Hardy, Chap. XIII of *Colloid Chemistry, Theoretical and Applied*, Vol. I, edited by Jerome Alexander, Chemical Catalog Co., New York (1926); see also *Phil. Trans. Roy. Soc. (London)*, **A230**, 1 (1931).

[69] I. Langmuir, *Science*, **84**, 379 (1936).

[70] In addition to cited reference see E. K. Rideal, *An Introduction to Surface Chemistry*, 2nd ed., University Press, Cambridge, England, 1930; N. K. Adam, *The Physics and Chemistry of Surfaces*, 3rd ed., Oxford University Press, 1941.

been called polar groups, and compounds containing these groups have been designated *polar compounds.*

A polar group confers upon an organic compound a certain solubility tendency in water. Thus, we have in methane a gas which is relatively non-water-soluble. When we introduce a polar group, as in methyl alcohol and methyl amine, we form a compound which is water-soluble. One portion (CH_3—) of a molecule of methyl alcohol is of hydrocarbon nature; the other portion (—OH) is closely allied to water, and as such shows an affinity for water. If then a small amount of methyl alcohol is dissolved in water, it will concentrate in the surface film, because methyl alcohol lowers the surface energy at an air-water interface. The fact that the CH_3— group retains a part of the properties of a hydrocarbon and the —OH group retains a part of the properties of water causes the methyl alcohol molecules to orient themselves in the interfacial film, with the hydrocarbon chain toward the vapor phase and the —OH group toward the water phase. As we lengthen the hydrocarbon chain in passing to ethyl alcohol, propyl alcohol, butyl alcohol, amyl alcohol, etc., there is a progressive intensification of the hydrocarbon properties residing in the molecule and a corresponding lessening in the similarity of the molecule to a water molecule. Accordingly the solubility of the alcohols in water decreases as the carbon chain is lengthened, and correspondingly their solubility in organic solvents increases.

In acids, when we approach lauric acid, we reach a point at which the interface between an aqueous solution of lauric acid and air is essentially a hydrocarbon interface.[71] The effect on surface tension of the sodium salts of the saturated fatty acids of carbon chains of varying lengths becomes clear when the theories of molecular orientation and polar groups are understood.

Harkins likens a polar group to a metallic weight attached to logs of wood. If the metallic weight is kept constant and exceeds the buoyant capacity of the log, the log will be immersed in water. This is what happens when methyl alcohol is dissolved in water. The affinity of the —OH group for water is so great as to overcome the effect of the hydrocarbon chain. If, however, the log of wood is increased in size or in length, we eventually reach a point where the weight no longer will submerge it and

[71] The point at which the liquid-air interface becomes essentially a hydrocarbon-air interface depends to some extent on the concentration of the acid in the aqueous layer. At low acid concentrations a pure hydrocarbon surface probably occurs above C_{12}. At high acid concentrations the point may be lowered below C_{12}. The point likewise depends on the temperature, for, as the temperature increases, the area occupied by a single molecule becomes greater and the film expands like a two-dimensional gas.

the wooden end of the log will project above the surface of the water, the weight at the bottom tending to hold the log in an erect position.

If we imagine a bar of liquid which we divide into two parts by an imaginary plane, when the imaginary plane is lifted the upper layer rises with it and two surfaces appear where there was no surface previously. In these two surfaces we have a rearrangement of the molecules of the liquid from a random distribution, such as is characteristic of the interior

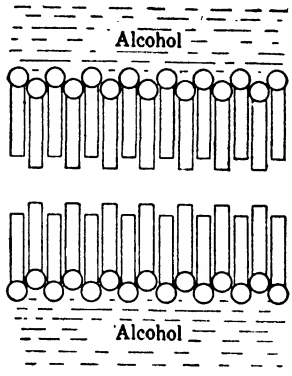


FIG. 44. A diagrammatic representation of the orientation of molecules on surfaces when a column of alcohol is broken apart to give two surfaces. (After Harkins.)

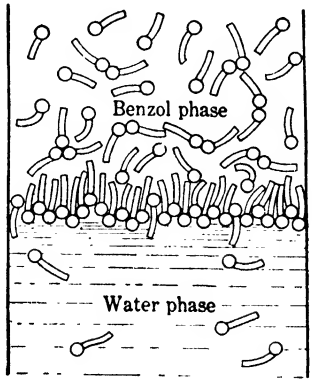


FIG. 45. A diagrammatic representation of butyric acid distributed between water, benzol, and the interface between the two. (After Harkins.)

of liquids, to one where the molecules are oriented in a more or less orderly fashion. Figure 44 is a diagrammatic example of such surfaces.

Similarly, if we dissolve butyric acid in benzene and pour this solution upon a water surface, the water will have a greater affinity for the carboxyl group of the butyric acid than for the hydrocarbon chain, whereas, on the other hand, benzene will have a greater affinity for the hydrocarbon chain than for the carboxyl group, and these two opposing forces will cause an orientation of the butyric acid at the benzene-water interface, similar to the orientation diagrammatically shown in Fig. 45.

Polar liquids are, in general, mutually soluble, and slightly polar liquids, such as hexane or octane, are soluble in other slightly polar liquids but are relatively insoluble in very polar liquids. It is the old theorem that "like attracts like." In general, the introduction of polar groups causes hydrocarbons to become soluble in water when the hydrocarbon chains are short. As the length of the hydrocarbon chain of the solute

is increased, a decrease in the surface tension of the solution occurs up to a point where a hydrocarbon interface is formed, after which the free energy of the interface remains essentially constant. As Langmuir has shown, the surface energy of the paraffin hydrocarbons from hexane to molten paraffin is essentially the same (46–48 ergs per square centimeter), although the molecular weights differ very greatly.

From the above hypotheses, it is evident that the solubility of one substance in another may be the resultant of several opposing forces, or, in other words, that *different parts of a molecule may possess different affinities toward the solvent.*

For a liquid which is totally miscible with water the interfacial tension between it and water will be zero, *i.e.*, no interface will exist. A liquid which is immiscible with water is so because its molecules show a low or negligible overall attraction for water molecules. Individual groups in the molecules of the immiscible liquid, however, may show a high attraction for water. When a small quantity of such a liquid is brought into contact with a clean water surface the resultant of the attracting and repelling forces acting at the interface and within the body of the liquid will determine whether or not its molecules will spread as a monomolecular film on the water, will spread as a film thicker than a monolayer, or will remain as a lens on the water.

Langmuir's [72] theory with regard to the probability of a non-aqueous, non-miscible liquid spreading on water is to the effect that totally nonpolar liquids do not spread at all but remain as a lens on the water surface, the thickness of which is determined primarily by gravitational forces. But if the molecules of the liquid contain polar groups (COOH, OH, etc.) they will spread to form a monomolecular layer on the water surface.

The spreading of films of one substance upon the surface of another and the conditions determining the manner and degree of spreading have been studied particularly by Harkins and his students. They have developed "a general thermodynamic theory of spreading which applies to all liquids, to spreading of solids on liquids and the spreading of mobile films, *i.e.*, volatile or liquid substances, upon solids." [73] The theory is based upon the concept that there are two types of spreading, namely, duplex spreading and non-duplex spreading. A duplex film is defined as a film of sufficient thickness (greater than monomolecular) to produce complete independence between the energy (surface tension) of its outer surface and that of the interface with the phase upon which it is spread.

[72] I. Langmuir, *Cold Spring Harbor Symposia Quant. Biol.*, **6**, 173 (1938).

[73] W. D. Harkins, *Chem. Revs.*, **29**, 385 (1941); see also p. 12 in J. Alexander, *Colloid Chemistry*, Vol. V, Reinhold Pub. Corp., New York, 1944.

Duplex films, where formed, tend to be unstable and to separate at equilibrium in the system into lens and monolayer films.

If a drop of insoluble oil b is placed on the surface of clean water a, it will fail to spread if the oil has a higher affinity for itself than for the water, or it will spread if its affinity for the water surface is greater than its affinity for itself. Harkins expressed this condition quantitatively, for duplex spreading, in terms of the spreading coefficient, $S_{b/a}$, which is defined in terms of the so-called work of adhesion (oil for water) W_a and the work of cohesion W_c (oil for oil).

$$S_{b/a} = W_a - W_c \tag{107}$$

The work of adhesion W_a is equal to the work required to separate a unit area of the oil-water interface and form the two new unit surfaces, oil-air and water-air in its place.

$$W_a = \gamma_a + \gamma_b - \gamma_{ab} \tag{108}$$

The work of cohesion can similarly be defined as the work necessary to separate a similar area of cross section within the pure liquid.

$$W_c = 2\gamma_b \tag{109}$$

Thus

$$S_{b/a} = \gamma_a + \gamma_b - \gamma_{ab} - 2\gamma_b = \gamma_a - (\gamma_b + \gamma_{ab}) \tag{110}$$

The oil will spread as a duplex film on the water if the surface energy of the disappearing surface γ_a is greater than the sum of the energies of the two new surfaces formed, γ_b and γ_{ab}, as the oil spreads. In any case where $S_{b/a}$ is positive, duplex spreading will occur, whereas, where $S_{b/a}$ is negative, duplex spreading will not take place.

Monolayer spreading will occur in any case where the surface tension of the water plus monolayer, γ_a', is less than that of the water alone, γ_a. If $\gamma_a' > \gamma_b + \gamma_{ab}$ the liquid b will spread as a duplex layer over its monolayer. In general, such is not the case, and usually, even when liquid b spreads initially as a duplex layer over liquid a it will change to a monolayer and a lens because $\gamma_a' < \gamma_b + \gamma_{ab}$. No liquid will spread to form a duplex film on a surface with which it makes a contact angle greater than zero. A lens at equilibrium in contact with its monolayer will always show a positive contact angle.

The film density (number of molecules per unit area) which a given spreading material will reach as it spreads as a monolayer will be determined by the extent of the free energy change experienced when a molecule passes from the bulk phase into the surface layer. The limit will be reached when the escaping tendency of the limiting molecules in the film becomes equal to that of the molecules of like nature in the bulk

phase with which the surface film is in equilibrium. The amount which will pass into and accumulate in the film before this equilibrium has been reached will be determined by the extent to which the spreading molecules can lower the surface tension of the phase upon which spreading occurs. If this is high, many molecules will be able to collect before equilibrium is attained; if this is low, few will accumulate. There is thus a parallelism between the surface tension lowering capacity of a substance and the equilibrium film density (*i.e.*, degree of adsorption).

The value of $\gamma_a - \gamma_a'$ (water surface) may be taken as a measure of the degree of heteropolarity of the spreading compound. If the compound contains a strongly polar group and a weak non-polar group (*e.g.*, CH_3COOH) it will go into solution in water, and little will tend to collect in the surface; $\gamma_a - \gamma_a'$ will be small. If the compound contains only strongly non-polar groups or these plus weakly polar groups (hexane, CS_2), the substance will not go into solution and will tend to spread as a very disperse monolayer, if at all; $\gamma_a - \gamma_a'$ will again be small. If, however, it contains a strong non-polar and a strong polar group it will tend to spread as a monolayer of high film density; $\gamma_a - \gamma_a'$ will be large.

When oleic acid is placed on water, the —COOH groups are immersed in the water phase. The long hydrocarbon chains, however, have too much attraction for each other and too little for water to be drawn underneath the surface merely by the affinity of the —COOH group for water. Accordingly the oil spreads on the surface as a monomolecular layer. The spreading of an oil is thus due to the presence of a polar group in the molecule. Hardy found that many oils, such as the pure saturated aliphatic hydrocarbons, did not spread on a water surface, and explains this by stating that the great chemical stability of the paraffins makes chemical interaction with water impossible.

Harkins and McLaughlin [74] used values derived from a study of adhesional and cohesional energies and found 2.78×10^{14} molecules of butyric acid per square centimeter of the interfacial film between hexane and water. This corresponds to a molecular area of 32 Å.^{2}.

Langmuir devised a so-called "surface tension balance" which was later modified by Adam and others. Figure 46 shows a commercially available model. This instrument is essentially a shallow trough with smooth, parallel sides. Suspended in the trough is a "float" which extends almost across the water surface and which is connected with the sides of the trough by extremely thin platinum or gold foil. This float acts as a movable barrier to prevent the escape of the film past it to the other end of the trough. The surface of the water in the trough is swept clean of contaminants, and a small amount of surface-active substance is

[74] W. D. Harkins and H. M. McLaughlin, *J. Am. Chem. Soc.*, **47**, 1610 (1925).

then placed upon the water surface. Many strongly heteropolar substances, virtually insoluble in water, will be found to spread readily on the surface of the water to form monomolecular films. A movable barrier is now brought toward the film, compressing the film against the float.

As long as the film is expanded, *i.e.*, the molecules in the film are not in contact with one another, the movable float will not change position. The movable barrier is advanced until a deflection of the float takes

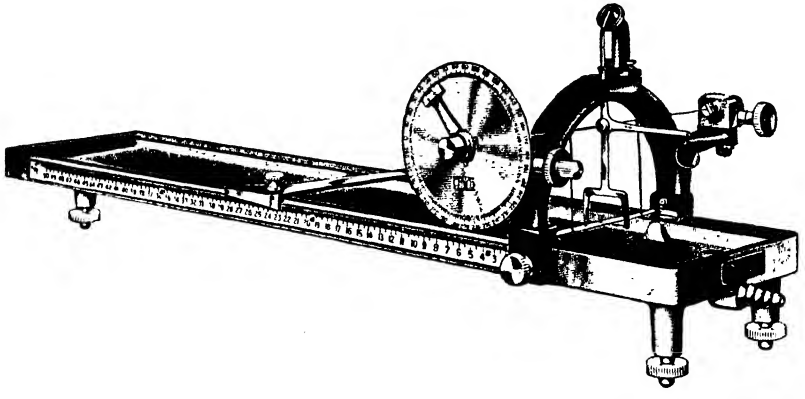


FIG. 46. A hydrophile balance for measuring molecular areas and spreading characteristics of surface films.

place. The instrument shown in Fig. 46 is designed to measure the pressure which the film exerts against the float, and the data so obtained can be converted into dynes per centimeter. This pressure, pictured as being exerted against the barrier, is in reality the difference between the surface tension of the pure water on one side of the barrier and the surface tension of the water covered by the film on the other side of the barrier. The film pressure π is equal to $\gamma_a - \gamma_a'$. As the movable barrier is brought closer to the floating barrier, higher force must be applied to the float to maintain it unmoved. As the area occupied by the molecules of the film is decreased the value of γ_a' is diminished (π is increased) until a sufficient force is applied to remove the film from contact with the water. This is the collapse pressure. π *vs.* area curves will show how the value of $\gamma_a - \gamma_a'$ varies with film density. Figure 47 shows such a curve for a palmitic acid film when spread on hydrochloric acid (0.01 *N*) at 25°C.

Discontinuities in these pressure-area curves are taken to indicate phase changes in the two-dimensional system of which the film consists. Regions may be noted where the pressure *vs.* area relations are very

analogous to the pressure *vs.* volume relationship of three-dimensional gases, liquids and solids, together with several intermediate phase conditions of the films. Harkins identifies seven distinct surface film phases as recognizable, not all for one substance at a given temperature but for various substances and at various temperatures. The occurrence and meaning of such phase relationship in these two-dimensional systems are treated also by Adam [75] and Dervician.[76] Although these relationships have been considered to apply primarily to monofilms of insoluble

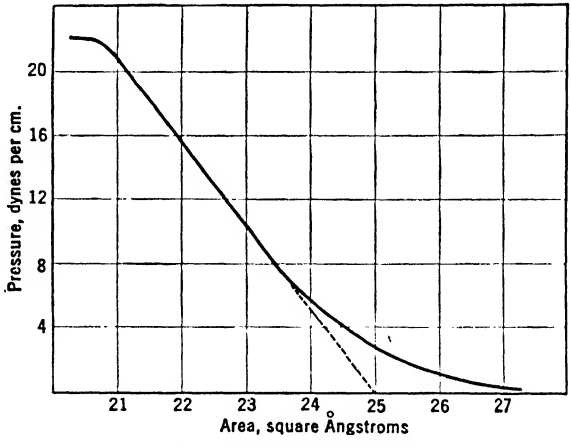


FIG. 47. Pressure-area curve for a palmitic acid film spread on 0.01*N* HCl at 25°C.

substances spread on a liquid subphase, Jura and Harkins [77] have recently indicated that identical phase changes occur in monolayers spread on solid subphases.

Since the area of the film can be exactly measured and since we can also measure with exactness the weight of matter which was added to produce the film, we can calculate the number of molecules occupied by each square centimeter of the monomolecular film at any observed value of π.

Table 24 shows certain of the data for the surface area occupied by a single molecule containing various groups as determined by Adam,[78] who used an apparatus similar to that shown in Fig. 46. It will be noted that the first eleven substances in the table show some differences in cross-sectional area. Nevertheless they are relatively alike. When we

[75] N. K. Adam, *Physics and Chemistry of Surfaces*, 3rd ed., Oxford University Press, New York, 1941.

[76] D. G. Dervician, *J. Chem. Phys.*, **7**, 931 (1939).

[77] G. Jura and W. D. Harkins, *J. Am. Chem. Soc.*, **68**, 1941 (1946).

[78] N. K. Adam, *J. Phys. Chem.*, **29**, 87 (1925).

pass, however, to a triglyceride, it is apparent that the glycerol portion of the molecule is lying flat in the interface with the three hydrocarbon radicals of the fatty acids projecting into the surface, inasmuch as the cross section of the triglyceride is approximately three times that of the hydrocarbon chains. Similarly, in the dipalmitate we have two hydrocarbon chains projecting into the surface, with the glycol lying horizontally beneath.

TABLE 24. CROSS-SECTIONAL DIMENSIONS OF THE AREA OCCUPIED BY A SINGLE MOLECULE IN AN INTERFACIAL FILM

(Data of Adam)

Group	Cross Section, Å.	Group	Cross Section, Å.
Hydrocarbon chain	20.7	$-CH_2NH_2CONH_2$	25.5
$-CH_2CH_2COOH$	25.1	$-C_6H_4OH$	23.8
$-CH{=}CHCOOH$	28.7	$-C_6H_4NHCOCH_3$	28.2 or 25.8 †
$-CH_2CH_2COOC_2H_5$ *	22	Triglycerides	63
$-CH{=}CHCOOC_3H_5$	28.7	Glycol dipalmitate	42
$-CH_2OH$	21.7	Cholesterol	39
$-CONH_2$	<21	Hydrolecithin	53
$-CN$	27.5		

* Ethyl, methyl, and allyl esters pack into the same area.

† According to temperature.

Harkins and McLaughlin [79] and Harkins and Gilbert [80] studied the structure of films on aqueous salt solutions. They found the surface of a salt solution to be covered with a monomolecular water film which ranges in thickness from 4 Ångström units on a 0.1 *M* NaCl solution to 2.3 Ångström units on a 5.0 *M* NaCl solution. As we have already noted (Table 20), the surface tension of solutions of inorganic salts is relatively high; consequently an inorganic salt solution can decrease its free energy by bringing pure water molecules into the surface. Apparently, however, from these studies, the water film on the surface is monomolecular in thickness, the difference in cross section being explained by the intensity of molecular "packing."

Many proteins have been shown to spread as monolayers on water. Deveaux, in 1903, applied a small amount of dry egg white to the sur-

[79] W. D. Harkins and H. M. McLaughlin, *J. Am. Chem. Soc.*, **47**, 2083 (1925).

[80] W. D. Harkins and E. C. Gilbert, *J. Am. Chem. Soc.*, **48**, 604 (1926).

face of water and noted that it spread as a coherent film which he considered one molecule thick. It was not, however, until Gorter [81] began his studies on the spreading of proteins that much interest was developed in this field. His work has been followed by work in many other laboratories. A good discussion on the subject is that of Langmuir and Schaefer.[82] In general, it is found that many proteins, some of which are initially water-soluble, will spread, under proper conditions, on the surface of water to form films which are no longer water-soluble and which, in all cases at low film pressures, show about the same film thickness. Thus, proteins which are initially quite diverse as to molecular shape and solubility properties are rendered very similar in nature when they unroll into surface films. The native protein molecules, rolled up in a definite and characteristic arrangement in three dimensions, rearrange into two-dimensional films under the influence of surface forces to form films in which the hydrophilic groups of the molecule are faced toward the water surface and the hydrophobic side chains face away from that surface. The configuration or special arrangement of the protein molecules is radically changed (denatured) from its native form to one which is nearly the same for all proteins. In this process they have lost their characteristic physical configurations although they still retain their original chemical configurations.

Mixed films of proteins and other surface-active substances have been studied by use of the film balance and film potentials.[83] Film potentials can be measured by placing a reversible electrode in contact with the water and suspending a flat metallic electrode above the surface of the film with a small amount of radioactive substance placed on its lower surface to cause ionization of the air and permit conductance between the upper electrode and the water. The potential measured will vary from that of pure water with the number and kind of molecules oriented in the film and below it. It is observed, for example, that, when a surface-active substance is introduced below a protein or lipid film, it may do any of three things. (*a*) It may cause no effect on either the film pressure or the film potential, in which case it must not have been in any way concentrated in the surface region. (*b*) It may change the film potential but not the pressure, in which case it is pictured as having moved into the film region and become firmly oriented and anchored to the lower side of the film. (*c*) It may cause a change in both film pressure and film potential, in which case it is pictured as penetrating into the film,

[81] E. Gorter and F. Grendel, *Trans. Faraday Soc.*, **22**, 477 (1926).

[82] I. Langmuir and V. J. Schaefer, *Chem. Revs.*, **24**, 181 (1939).

[83] J. H. Schulman and E. K. Rideal, *Proc. Roy. Soc.* (*London*), **B122**, 29, 46 (1937); J. H. Schulman and E. Stenhagen, *ibid.*, **126**, 356 (1938).

thus forcing all film molecules closer together. Schulman [84] found that, if the protein of red blood cells was spread, all substances which would penetrate the film were also hemolytic in their action on intact red cells. Similarly if the red cell lipid constituents were spread as a film only cell hemolysers would penetrate the film (different hemolytic agents in the two cases).

Most physicochemical studies of molecular orientation have dealt with the formation of monomolecular films. Blodgett,[85] working in Langmuir's laboratory, devised a technic for the production of polymolecular films and built up polymolecular layers more than two hundred molecules deep. Blodgett uses a Langmuir surface tension balance upon which a surface tension depressant (*e.g.*, the calcium or barium salts of palmitic, stearic, or other long-chain fatty acids) has been allowed to spread. This film is then compressed on the surface of the water by the spreading of another surface tension depressant (*e.g.*, castor oil), the two films being kept separate by a waxed silk thread. In this way definite pressure can be brought to bear on the calcium palmitate film by the spreading of the so-called "piston oil" (*e.g.*, the castor oil noted above). The pressure so produced ($\pm$16.5 dynes per centimeter) is not sufficient to crumple the film but is sufficient to cause it to move upon another suitable surface upon which it can be spread. If now a clean glass or metal slide is *raised* through the film of calcium palmitate, the area occupied by the calcium palmitate on the water surface decreases by exactly the surface area of the slide, and a monomolecular layer of calcium palmitate is deposited over the surface of the slide with the carboxyl groups of the palmitic acid oriented toward the slide and the hydrocarbon "tails" extending outward from the slide. If now the slide is *lowered* through the calcium palmitate film, a second layer of calcium palmitate is deposited on the slide, in this instance the hydrocarbon portion of the molecule being attracted to the hydrocarbon surface already on the slide and the —COOH portion of the molecule extending in the surfaces. Thus films may be built up, one film at a time, by successive raising and lowering of the slide through the surface film, the 1-3-5-7 $\cdots$ films being hydrophobic and being wetted by oil and not wetted by water, and the 2-4-6-8 $\cdots$ films being hydrophilic and being wetted by water and not wetted by oil. Such polymolecular films when sufficiently thick show a beautiful series of interference colors. They may be dried even by baking in an oven without destruction of the film structure. Blodgett notes that Langmuir has suggested that "films

[84] J. H. Schulman, *Trans. Faraday Soc.*, **33**, 1116 (1937).

[85] Katharine B. Blodgett, *J. Am. Chem. Soc.*, **57**, 1007 (1935). (*Cf.* also I. Langmuir and V. J. Schaefer, *J. Am. Chem. Soc.*, **58**, 284 (1936).)

could be built for use as diffraction gratings for soft x-rays by depositing $(2n + 1)$ layers of barium stearate, then $2n$ layers of stearic acid, then $2n$ layers of barium stearate, and so on in alternating succession. The stearic acid would be more transparent to radiations of short wave length than the barium stearate and would therefore serve to space the series of layers of barium stearate at known intervals apart."

Later studies [86] have dealt with built-up films of proteins and the properties of such protein films. By use of the Blodgett technic it has been found possible to build up multiple films of proteins of almost any thickness that may be desired. If the interference colors which such films show are used, and if the number of layers of protein in the film are known, it is possible to measure with rather high precision the cross-sectional area of the protein molecule. Egg albumin gives a film about 20 Å. in thickness, whereas a film of zein is only about 10 Å. thick. When the film is deposited on a chromium-plated surface, egg albumin, pepsin, insulin, and the tobacco mosaic virus protein form monolayers, the "outside" of which is hydrophilic, *i.e.*, easily wetted by water but not easily wetted by oil, but in casein and zein the monolayer is hydrophobic.

These studies, demonstrating beyond peradventure of a doubt the possibility of polymolecular films of oriented molecules, are regarded by the writer as of great importance in extending the theories of membrane formation, etc., in biological systems, and we shall have occasion to refer to Blodgett's studies again.

Another method of attacking the problem of molecular orientation is by studying the electrical behavior at interfaces. Compounds containing polar groupings are all electrical dipoles, and an oriented layer of dipoles should induce an oriented layer of any ionic species present and give rise to an electric double layer. Thus, in such systems we can study two problems at once, the electrical behavior of the double layer and the degree of specific orientation in the oriented molecular layer.

According to Gibbs' theorem (p. 187), surface tension forces need not draw the molecule into the interface with any particular degree of orientation. However, electrical studies made on such surfaces may form a means of determining the extent to which the molecules in the surface layer have assumed any specific orientation. Accordingly the electrokinetic potential at the interface between cellulose or Al_2O_3 and a number of pure organic liquids has been studied [87-89] in an attempt to relate the structure of organic compounds and the electrical interfacial energy

[86] I. Langmuir, V. J. Schaefer, and D. M. Wrinch, *Science*, **85**, 76 (1937).
[87] W. McK. Martin and R. A. Gortner, *J. Phys. Chem.*, **34**, 1509 (1930).
[88] O. G. Jensen and R. A. Gortner, *J. Phys. Chem.*, **36**, 3138 (1932).
[89] M. A. Lauffer and R. A. Gortner, *J. Phys. Chem.*, **42**, 641 (1938).

of such systems. It was found that the structure of the molecule apparently determined not only the sign but also the magnitude of the electrokinetic forces at the interface. Table 25 shows certain of the data obtained. In the *n*-aliphatic alcohols, for example, a definite relation-

TABLE 25. ZETA POTENTIAL AND ELECTRIC MOMENT (δe) PER UNIT AREA OF THE DOUBLE LAYER AT INTERFACES BETWEEN CELLULOSE OR Al_2O_3 AND CERTAIN PURE ORGANIC LIQUIDS

	Cellulose-Liquid Interface		Al_2O_3-Liquid Interface		
	Zeta potential, mv.	(δe), esu. $\times 10^6$	Zeta potential, mv.	(δe), esu. $\times 10^6$	Unbalanced orientation of molecules in interface, per cent
Methyl alcohol	− 55.3	−464.1	− 26.7	−218.0	.....
Ethyl alcohol	− 19.9	−136.3	− 14.7	− 94.2	.....
n-Propyl alcohol	+ 17.1	+101.3	− 11.7	− 64.7	.....
Isopropyl alcohol	− 16.2	−112.8	+ 11.5	+ 73.6	.....
n-Butyl alcohol	+ 51.7	+265.2	− 3.7	− 17.8	.....
Isobutyl alcohol	+ 12.4	+ 66.4	+ 2.0	+ 10.2	.....
n-Amyl alcohol	.....	.....	+ 8.9	+ 36.8	.....
n-Hexyl alcohol	+ 33.6	+118.4	+ 35.3	+122.4	.....
n-Heptyl alcohol	− 6.6	− 18.7	+ 8.9	+ 26.0	.....
Acetic acid	+ 5.27	+ 8.74	+ 39.4	+ 65.2	11.40
n-Propionic acid	+ 1.31	+ 1.12	+ 18.7	+ 16.0	2.34
n-Butyric acid	− 0.483	− 0.35	+ 17.4	+ 13.0	3.52
n-Valeric acid	− 1.11	− 0.77	+ 12.1	+ 8.4	.....
n-Caproic acid	− 1.67	− 1.40	+ 29.0	+ 24.8	.....
Ethyl formate	− 6.31	− 11.80	+113.0	+212.0	24.10
Ethyl acetate	− 3.39	− 5.72	− 24.8	− 41.4	5.05
Ethyl *n*-propionate	− 1.96	− 3.02	− 19.1	− 29.0	3.56
Ethyl *n*-butyrate	− 2.04	− 2.85	− 32.1	− 43.6	.....
Benzene	0.00	.....	0.0	.....	.....
Methyl benzene	− 0.2	.....	.....	.....	.....
Chlorobenzene	− 1.0	.....	.....	.....	.....
Bromobenzene	− 6.7	.....	.....	.....	.....
Aminobenzene	− 49.7	.....	.....	.....	.....
Nitrobenzene	−142.0	.....	.....	.....	.....

ship was found to exist between the length of the hydrocarbon chain and the zeta potential at the alcohol-cellulose interface. The introduction of a $—CH_2$ group into the main carbon chain changed the zeta potential

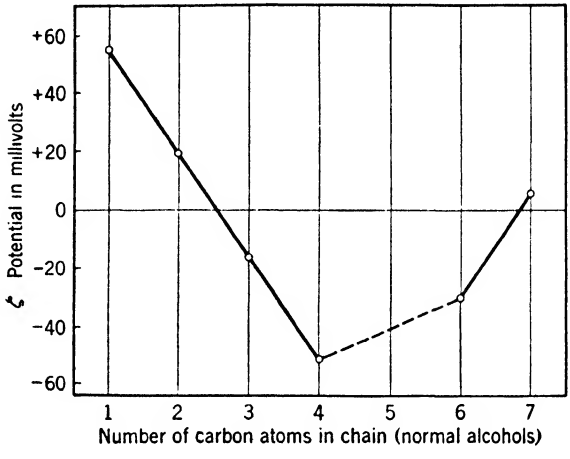


FIG. 48. Showing relationship between the molecular structure of certain *normal*-aliphatic alcohols and the zeta potential at a cellulose-alcohol interface.

by approximately 35 mv. (*cf.* Fig. 48), whereas the introduction of a $—CH_2—$ into a side chain (*e.g.*, replacement of a hydrogen by $—CH_3$) changed the zeta potential by only approximately 4 mv.

In the last column of Table 25 are certain calculations directed at ascertaining the degree of unbalanced orientation of the molecules in the interface. These calculations were based on the following assumptions. If organic dipoles are oriented at an interface, we might expect an arrangement more or less like that shown diagrammatically in Fig. 49.

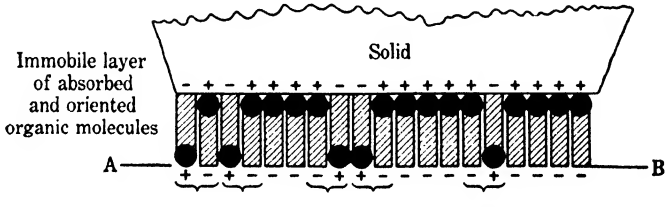


FIG. 49. A diagrammatic representation of oriented dipoles at an interface postulated as the source of the zeta potential.

If it is assumed that the boundary between the movable and immovable liquid is at AB, the bracketed pairs of dipoles oriented in opposite directions might be expected to neutralize each other, whereas the "unbal-

anced orientation" of the remaining molecules should give rise to a net negative charge on the immovable layer side of the interface with a corresponding positive charge in the streaming liquid. On this hypothesis, it should be possible to calculate the percentage of "unbalanced orientation" of the organic molecules in the immovable layer, assuming (1) that the layer is monomolecular, close-packed, and oriented, and (2) that the electric moment per unit area of the double layer is the product of the dipole moment of the organic molecule and the number of "unbalanced" molecules oriented per unit area.

In these calculations, the values for the cross-sectional area of the molecules (A) are those given by Rideal for the limiting areas per molecule in the liquid condensed form. For esters $A = 22.0$ Å.^{2}, and for acids $A = 24.4$ Å.^{2}.

The percentage of the total surface occupied by oriented but "unbalanced" molecules is given by the expression $\delta e A/\mu \times 10^{-1}$, where μ is the molecular dipole moment, and δe and A have the meanings noted.

It will be seen that a relatively small percentage of unbalanced orientation is sufficient to account for the electric moment of the double layer.

If these studies are confirmed by further investigations and if the electrokinetic forces at a solid-organic liquid interface are due to the electrical dissymmetry of oriented organic molecules, the molecular orientation theory of Hardy, Harkins, Langmuir, Adam, *et al.*, can well be extended to include surface electrical forces as well as surface tension and interfacial tension.

It has been possible to indicate only in barest outline some of the theories and applications involved in the problem of molecular orientation. It is hoped, however, that sufficient has been given to stimulate the interest of the student and direct him to literature sources, for undoubtedly in this field of molecular orientation lies an explanation for many biochemical and biological phenomena.

ADSORPTION

As noted above, a solution may decrease its surface energy by bringing into the surface either an excess of the solvent or an excess of the solute, depending on which one causes a lowering of the surface tension. In the same way a colloid system can decrease its surface energy by bringing into the interfacial film substances which lower the interfacial tension. *Those substances which decrease surface energy tend to concentrate at a liquid-vapor interface, and those substances which decrease interfacial energy tend to concentrate at a liquid-solid or liquid-liquid interface. This phenomenon of concentrating at the interface is called* ADSORPTION.

Sodium oleate decreases the surface tension of water. If air is bubbled through a solution of sodium oleate, the foam which is carried over contains a higher percentage of sodium oleate than did the original sol, and the residual sol left behind is correspondingly more dilute. A dilute acetic acid solution or a solution of certain dyestuffs may be used equally well to demonstrate the same phenomenon.

The phenomenon of adsorption, as a rule, does not bring into play the forces of primary valence, *i.e.*, it is not possible to write a stoichiometrical chemical equation to represent the process, for, after all, adsorption is a process which depends on the concentration of the material which is being adsorbed and the extent of surface upon which adsorption can take place.

Many equations have been suggested as mathematical representations of the characteristics of an adsorption reaction.

Freundlich [90] proposed an empirical equation from a fit of experimental data. His equation is

$$\frac{x}{m} = aC^b \tag{111}$$

where x = the weight of substance adsorbed in grams
m = the weight of adsorbent in grams
C = the concentration of the solution at equilibrium
a and b = constants depending on the nature of the adsorbent and the substance which is adsorbed.

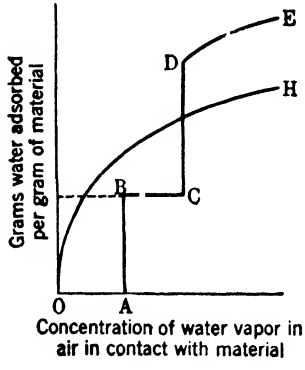
Equation (111) is the mathematical expression of a parabola and may be written

$$\log \frac{x}{m} = \log a + b \log C \tag{112}$$

The characteristic features of the curve are that there is no single point where the process appears to be completed and there are no regions of discontinuity. Figure 50 is a diagrammatic representation of an adsorption isotherm *OH*, which describes the reaction occurring when water vapor is adsorbed on charcoal, as contrasted with the discontinuous isotherm *ABCDE*, which describes the reaction of water vapor with anhydrous sodium sulfate. Both are reversible. In curve *ABCDE* fixation of water vapor occurs in definite steps, each completed at a definite relative vapor pressure. Such relationships are characteristic of reactions (chemical or physical) where the energy of reaction for each unit

[90] H. Freundlich, *Kapillarchemie*, Akademische Verlagsgesellschaft, Leipzig, 1909.

of reactants involved is equal to that of every other unit undergoing the same reaction, *i.e.*, the whole reaction takes place at the same chemical potential of the reactants and is stoichiometric in character. In the adsorption reaction the potential at which the reaction occurs is continuously shifting. The adsorption curve can be considered to represent a process which occurs in a very large number of steps, each of which occurs at a definite potential but is extremely small in extent or capacity.



Surface phenomenon. Curve *O–H:* the adsorption of water vapor by dry charcoal.
"Chemical" reactions. Curve *ABCDE:* Behavior of Na_2SO_4 to water vapor.
OA = anhydrous Na_2SO_4
AB = anhydrous Na_2SO_4 in contact with $Na_2SO_4 \cdot 10H_2O$
Point *B* = only $Na_2SO_4 \cdot 10H_2O$
CD = $Na_2SO_4 \cdot 10H_2O$ in contact with a saturated solution of Na_2SO_4
Point *D* = a saturated solution of Na_2SO_4
DE = varying concentrations (less than saturated) of Na_2SO_4

FIG. 50. Contrasting the form of curves characteristic of surface phenomena with that characteristic of a chemical reaction. (After Bancroft.)

The constants a and b of the Freundlich adsorption isotherm can be determined by plotting, in terms of logarithms, the data of an adsorption process. The value of a is the value on the ordinate axis, where the straight line cuts the axis of ordinates at unit concentration, as shown in Fig. 51. The constant b is an expression denoting the slope of the curve and is the tangent of the angle θ which the line makes with a line drawn parallel to the axis of abscissa. Accordingly it is not necessary to plot the logarithmic curve in expressing graphically an adsorption reaction. Very extensive series of data can be concisely presented by tabulating the values for these two constants.

The graphic method yields at the best only a close approximation for the value of these fundamental constants. Their absolute value may be determined by the method of least squares. Denoting values on the axis of ordinates as (y) and values on the axis of abscissa as (x),

$$a = \frac{\Sigma(x) \cdot \Sigma(xy) - \Sigma(x^2) \cdot \Sigma(y)}{[\Sigma(x)]^2 - n\Sigma(x^2)} \tag{113}$$

$$b = \frac{\Sigma(x) \cdot \Sigma(y) - n\Sigma(xy)}{[\Sigma(x)]^2 - n\Sigma(x^2)} \tag{114}$$

where Σ = the sum of the numbers involved
n = the number of individual items which were summed.

The Freundlich equation has been found to describe, fairly generally, the process of adsorption of dissolved materials by solids in contact with the solution.

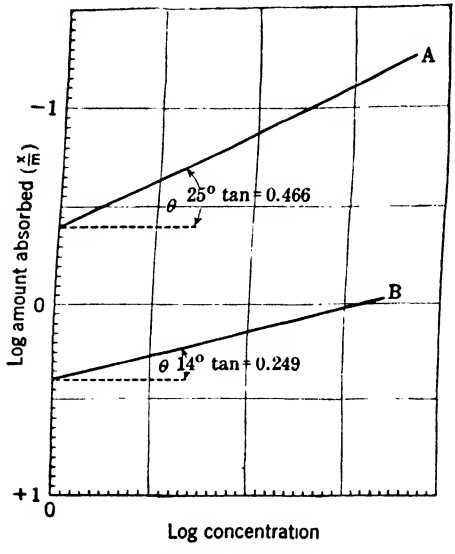


FIG. 51. Showing the significance of the constants in the expression

$$\frac{x}{m} = aC^b \quad \text{or} \quad \log \frac{x}{m} = \log a + b \log C$$

Curve A,

$$\log \frac{x}{m} = -0.4 + 0.466 \log C$$

Curve B,

$$\log \frac{x}{m} = 0.4 + 0.249 \log C$$

Langmuir [91] proposed an equation which was set up, primarily, to describe the adsorption of a gas on a solid. In certain cases, this equation appears to describe adsorption from solution as well. It may be stated as

$$x = \frac{\alpha\beta C}{1 + \alpha C} \tag{115}$$

where x = the amount of material adsorbed
C = the equilibrium concentration
α and β = constants.

[91] I. Langmuir, *J. Am. Chem. Soc.*, **40**, 1361 (1918).

Equation (115) is the equation for a hyperbola, and it indicates that the interface will become saturated at some definite value of C. This value of C is determined by constant β which becomes equal to x when C is a maximum. This point is usually reached [92] when the equilibrium concentration is the concentration of a saturated solution.

In certain instances β may exceed the free solubility of a material. In discussing this point Brintzinger [93] presents the data shown in Table 26

TABLE 26. ADSORPTION OF SALICYLIC ACID ON CHARCOAL AT VARIOUS TEMPERATURES

Temperature (°C.)	25	30	35	42	49	53
Solubility of acid (grams)	0.221	0.270	0.320	0.411	0.541	0.651
Acid adsorbed per gram charcoal	0.49	0.50	0.52	0.54	0.57	0.60

for the maximal adsorption of salicylic acid. He explains the excess adsorption over the true solubility as due to a surface "pull" which is greater than the water "pull" on crystals of salicylic acid.

Langmuir's equation was derived on the assumption that adsorption occurs at definite active points on the surface of the adsorbent and that, when these are all occupied by adsorbate molecules, no more will be adsorbed. Thus, the process of adsorption is considered to be limited to a monolayer of adsorbate. Freundlich's equation recognizes no such limit and must assume that adsorption can and does exceed that of a monolayer.

Brunauer, Emmett, and Teller [94] have elaborated on the picture set up by Langmuir for the adsorption of gases and have arrived at an equation which includes the possibility of polylayer adsorption. They have been able to fit their theory, in cases of gases adsorbed on solids, to the experimental observations even where these deviate very greatly from the types of curves expected from either the Langmuir or the Freundlich equations.

From a consideration of the rates of evaporation of the adsorbate molecules from, and their condensation onto, the surface in both the unimolecular layer and possible polymolecular layers, they arrive at the following equation to describe the equilibrium conditions.

$$\frac{p}{V(p_0 - p)} = \frac{1}{sV_m} + \frac{(s-1)p}{sV_m p_0} \tag{116}$$

[92] E. R. Linner and R. A. Gortner, *J. Phys. Chem.*, **39**, 35 (1935).
[93] H. Brintzinger, A. Schall, and H. G. Beier, *Kolloid-Z.*, **74**, 29 (1936).
[94] S. Brunauer, P. H. Emmett, and E. Teller, *J. Am. Chem. Soc.*, **60**, 309 (1938).

where V = volume of gas adsorbed at pressure p
V_m = volume of gas which will just complete a unimolecular layer on the adsorbent surface
p_0 = the liquefaction pressure of the gas (at the temperature of observation)
s = a constant given by the relationship $s = e^{-(E_a - E_l)/RT}$, where
e = base of natural logarithms
E_a = heat of adsorption of first layer
E_l = heat of adsorption of second and succeeding layers and is considered to approximate the heat of liquefaction of the gas.

This equation, for the conditions under which the Langmuir equation would hold, *i.e.*, at low gas pressures where the number of layers would be less than or equal to 1, reduces to

$$V = \frac{sV_m \dfrac{p}{p_0}}{1 + s\dfrac{p}{p_0}} \tag{117}$$

which is equivalent to equation (115). Here V is equivalent to x, p/p_0 to C, s to α, and V_m to β.

By plotting values of $p/[V(p_0 - p)]$ *vs.* p/p_0 the values of s and V_m can be determined from the graph. By determining the value of V_m from the adsorption isotherm for a given gas on a solid, and assuming that the gas molecules are close-packed with known dimensions (as liquid or solid) in the monolayer, Emmett [95] has arrived at a means for determining the surface area of the solid which has proved of value in characterizing finely divided substances such as iron oxide catalysts, silica gels, and carbonblacks.

Harkins and Jura [96] have also devised a method for the determination of the area of a finely divided solid which they designate an "absolute" method (as distinguished from the "relative" method mentioned below). This method differs from the Emmett method in that no assumption has to be made as to the area occupied by the adsorbed gas molecule or as to the accuracy with which the value of V_m can be determined.

They distinguish between the free energy of a surface and the total energy of a surface. For water, for example, the free surface energy is equal to the surface tension γ (equal to 72.8 ergs per sq. cm. at 20°C.).

[95] P. H. Emmett, *Advances in Colloid Science*, Vol. I, p. 1, Interscience Publishers, Inc., New York, 1942.

[96] W. D. Harkins and G. Jura, *J. Am. Chem. Soc.*, **66**, 1362, 1366 (1944).

The total energy h of the water surface is, however, defined by the relationship

$$h = \gamma - T\left(\frac{\partial\gamma}{\partial T}\right)_{P_1\Sigma_1} \tag{118}$$

or

$$h = \gamma + l \tag{119}$$

where l = the latent heat of the surface. The value of l for water at 20°C. is 45.7 ergs per sq. cm. Thus, the total energy of the surface of water at 20°C. is 118.5 ergs per sq. cm.

If a finely divided powder is placed in a saturated vapor (*e.g.*, water vapor) it becomes covered by a film adsorbed from the vapor. If the liquid film makes zero contact angle with the solid a duplex film will form on the solid surface. Every surface of the solid will thus be covered by a duplex film of the liquid. The outer surface of the duplex film will show a total surface energy which is identical with that of any free liquid surface. For water this is 118.5 ergs per sq. cm. If, now, the solid with its film of liquid is allowed to fall into a volume of the liquid in a sensitive calorimeter, heat will be liberated when the surface of the film on the solid disappears. This heat, in ergs, divided by the value of h for the liquid will give directly the area of the solid.

Harkins and Jura offer a new linear form of the adsorption isotherm which they claim covers accurately a greater range of conditions than those covered by any other isotherm (equation) which has been proposed. This is

$$\log\frac{p}{p_0} = B - \frac{A}{V^2} \tag{120}$$

where p/p_0 = the relative vapor pressure of the gas
V = the volume of gas adsorbed
A and B = constants.

A "relative" method which they propose for the estimation of the area of a solid is based upon an empirically found relationship between the constant A in this equation and the surface area Σ of the solid; *i.e.*,

$$\Sigma = K\sqrt{A} \tag{121}$$

where K is a constant varying only with the nature of the gas being adsorbed and has a value for water of 3.83 at 25°C., for N_2 gas of 4.06 at −195.8°C., a value of 13.6 for *n*-butane at 0°C., etc.

The methods of Emmett and of Harkins and Jura yield equal values for the surface areas of identical samples of finely divided crystalline (non-porous) materials such as anatase. They would not be expected

to do so for finely porous materials. It is probable that neither method would yield significant values for the surface areas of biocolloids which, in the dispersed or semi-dispersed state in aqueous media, form the surfaces at which many important biological processes are catalyzed. Although the theory and technics for gaseous adsorption have reached a high degree of perfection, it must be admitted that adsorption from solution is not yet so well characterized or understood. Finely divided crystalline (and insoluble) materials, whose surface areas have been previously determined by the above methods, should prove extremely useful in the study of the mechanism of adsorption by these materials from solution.

The Mechanism of Adsorption. There are, in reality, two schools of thought with respect to the phenomenon which we are discussing under the heading adsorption. To one of these schools, the ideas which will be propounded in this chapter, and even the term adsorption, are anathema. This school has considered primarily the behavior of electrolytes toward surfaces and is interested in such phenomena as acid-base exchange in soils and minerals, in acid and alkali binding of proteins, and in relationships which exist between the biocolloids and the electrolytes in biological systems and organisms. This school insists that stoichiometrical relationships account for the phenomena which we are calling adsorption and that the union between the electrolyte and the substrate is a true "salt" rather than an "adsorption complex."

Probably they are correct to a degree. It is well known that, in order to replace one equivalent of calcium or magnesium in a zeolite, it is necessary to add approximately two equivalents of an alkali metal. However, the various alkali metals differ among themselves in their replacement ability. Thus, Jenny [97] points out that there is a 240 per cent difference between the replacement ability of the lithium ion and the cesium ion on ammonium permutite and a 700 per cent difference in the replacement ability between the lithium ion and the potassium ion on hydrogen permutite, and that the monovalent ions arrange themselves in the lyotropic series $Li < Na < K < Rb < Cs < H$ with respect to permutite systems.

Apparently the binding takes place on *surfaces*, and the *extent of surface area* available to the adsorbate determines not necessarily the type of reaction that the system will undergo but rather the magnitude of the reaction which will take place. Unquestionably stoichiometrical chemical combination and surface adsorption in many instances involve the same chemical forces. This can be illustrated diagrammatically, if we take a hypothetical mass of carbon and project a plane through it. As-

[97] Hans Jenny, *J. Phys. Chem.*, **36**, 2217 (1932).

suming that the carbon atoms are arranged with a definite space relationship to each other, we might postulate an arrangement similar to that shown in Fig. 52. Those carbon atoms which are imbedded in the body of the carbon mass and are not exposed at any surface will obviously not take part in an adsorption or a stoichiometrical chemical reaction. Those carbon atoms which are exposed in the surface of the plane have one free valence bond and accordingly may enter into stoichiometrical reactions to a limited extent. Those carbon atoms which are exposed on edges are

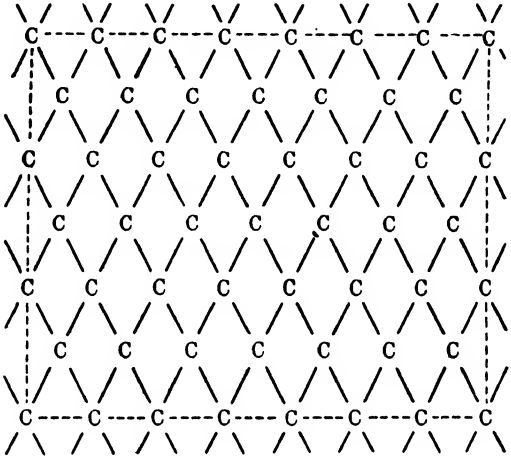


FIG. 52. A diagrammatic representation of the unsatisfied valences at the sides and corners of a plane surface of space-oriented carbon atoms.

represented as having two free valence bonds and accordingly may be expected to be more reactive than those atoms which are exposed only in surfaces. Those carbon atoms which are exposed at corners are represented as having three free valence bonds, and accordingly such atoms should be more reactive than those which are exposed in edges and still more reactive than those exposed in surfaces. Thus, in order to prove the stoichiometrical nature of the binding, it would be necessary to know not only the surface area of the solid phase but also the ratios which exist between the "atoms" which are involved in the binding and which are exposed at the "corners," at the "edges," and at the "faces" of the exposed planes, since those portions which constitute the "corners" will be theoretically more reactive than those which form the "edges," and these in turn will have a higher binding capacity than those in the "faces."

Taylor [98] has suggested that catalytic surface activity is due to sur-

[98] Hugh S. Taylor, *J. Phys. Chem.*, **30**, 145 (1926), and Georg-Maria Schwab, *Catalysis*, translated by H. S. Taylor and R. Spence, D. Van Nostrand Co., New York, 1937.

face atoms which have a high degree of valence freedom, thus approximating more nearly the gaseous state than the solid state, and that less than 1 per cent of the surface may be catalytically active. In addition, at surfaces we have those attractions which have been designated as the van der Waals forces. These forces differ only in degree from primary valence forces, since both are probably electrical in origin.[99]

Huggins [100] attempted to give a physical picture of the affinities causing adsorption and listed four energy sources: (*a*) an unpaired electron in the valence shell of an atom, (*b*) a positive atomic kernel (H, Na, Cu, etc.) not surrounded by electron pairs, (*c*) an electronegative atom or a lone electron pair in the valence shell of a negative atom, and (*d*) double or triple bonds in which one of the bounding electron pairs is not near the line joining the centers of the two atoms which it holds together. Adsorption may result from (*ab*), (*ad*), and (*ac*) forces in carbon and fresh metal surfaces. Attractions (*bc*), (*bd*), (*dd*), and (*cd*) may take place in organic compounds. He noted that (*bd*) adsorption will be "polar" and that (*dd*) will probably be "apolar," and that adsorption in any specific case may involve all the above types.

Inasmuch as the rare gases are positively adsorbed by charcoal and show fairly high heats of adsorption (argon, 3,450 calories per mole at 0°C. as compared with 3,654 calories per mole for nitrogen and 1,870 calories per mole for hydrogen at the same temperature), it is evident that surface fixation and heats of adsorption do not necessarily reflect compound formation.

Probably the truth actually lies somewhere in the intermediate ground between strictly stoichiometrical combination and purely physical surface adsorption. In the rare gases we are dealing wholly with surface behavior. In relatively inert organic molecules we are probably dealing largely with surface energies, and in electrolytes we have probably in part stoichiometrical unions with a residuum of surface energy attraction. The proteins have been singled out as a great class of the biocolloids in which the assumption has been made that only stoichiometrical reactions occur. Thus, Loeb[101] and his school interpret the reactions of proteins with acids and bases wholly on the basis of stoichiometrical salt formation. However, there is evidence [102] that adsorption also accounts for at least a part of the apparent acid and base binding in protein sys-

[99] K. F. Herzfeld and H. M. Smallwood, Chap. IV in *Treatise on Physical Chemistry*, edited by H. S. Taylor, 2nd ed., D. Van Nostrand Co., New York, 1931.

[100] M. L. Huggins, *Science*, **63**, 454 (1926).

[101] J. Loeb, *Proteins and the Theory of Colloid Behavior*, McGraw-Hill Book Co., New York, 1922.

[102] W. F. Hoffman and R. A. Gortner, Colloid Symposium Monograph, Vol. II, p. 209, Chemical Catalog Co., New York, 1925.

tems, and it has been shown [103] that, in contrast to succinic acid, casein does not possess a "maximum base binding capacity," but that together with the stoichiometrical binding of the base by the acid casein there is a second reaction which apparently yields an adsorption isotherm.

In order to account for the osmotic behavior of the red blood cells, Peters [104] has suggested that the potassium "salt" of hemoglobin in the red blood corpuscles probably exists largely in an un-ionized form. It is difficult to picture a true potassium salt which fails to ionize. It is easy to picture an alternative potassium ion-hemoglobin adsorption complex which osmotically would behave as a unit, and that picture is suggested as an alternative to explain the osmotic behavior of the red cells.

Positive and Negative Adsorption. Adsorption may be either positive or negative. *Positive adsorption from solution is the concentration of the solute in the interfacial film.* This concentration may reach high values and cause a very marked difference in the concentration of the original solution. *Negative adsorption is the concentration of the solvent in the surface film,* leaving an excess of the solute in the body of the liquid. Negative adsorption is always small in amount. Apparently in most cases only a monomolecular film of solvent is adsorbed at the interface.

The nature of the solvent is frequently of very great importance in determining not only whether or not adsorption will take place, but also the extent of the adsorption reaction. As a general rule, *adsorption is greatest from solvents which have a high surface tension and least from solvents which have a low surface tension.* This is what one would expect, since surface energy interchanges are involved.

Willard Gibbs [105] predicted, long before experimental evidence was available, that substances which lower the surface tension or the interfacial tension would tend to concentrate in the surface film, because of the decrease in surface energy which would result. Gibbs' equation may be written

$$C_2 = -\frac{C}{RT} \cdot \frac{d\gamma}{dC} \tag{122}$$

where C_2 = the excess concentration in the interface
C = the equilibrium concentration in the liquid phase
dC = the increment of change in concentration of the solution for an increment of change ($d\gamma$) in the interfacial tension.

[103] A. D. Robinson, R. A. Gortner, and L. S. Palmer, *J. Phys. Chem.*, **36**, 1857 (1932).

[104] J. P. Peters, *Body Water—The Exchange of Fluids in Man*, Charles C. Thomas, Springfield, Illinois, and Baltimore, Maryland, 1935.

[105] *The Scientific Papers of Willard Gibbs*, Vol. I, Thermodynamics, Longmans Green and Co., New York, 1906.

C_2 is positive when $d\gamma/dC$ is negative, *i.e.*, when the interfacial tension is decreased. Under such conditions we have positive adsorption. C_2 is negative when $d\gamma/dC$ is positive or when there is an increase in the interfacial tension of the system. Under such conditions we have negative adsorption. Bartell and Sloan (*loc. cit.*) restated Gibbs' theorem in terms of adhesion tensions as, "*There will be an excess in the concentration of the solute in the interfacial layer, if the ratio of change of the adhesion tension of the solution against the adsorbent is positive.*"

Characteristics of the Adsorption Process. It will be seen from equations (111) and (115) that the amount of material adsorbed is in direct proportion to the weight of the adsorbent, on the assumption that a portion of the same adsorbent in the same physical state is used. It would be more accurate to say that *the amount of material adsorbed is directly proportional to the specific surface of a given adsorbent.*

The nature of the solvent is of very great importance in adsorption reactions. As already noted, adsorption is greatest from solvents which have a high surface tension and least from solvents which have a low surface tension, because a greater change ($d\gamma/dC$) can be brought about in a dispersions medium with high surface energy than in one with low surface energy. In many instances a substance will be positively adsorbed from water and be only slightly or not at all adsorbed from an organic solvent. Picric acid dissolved in water is very readily adsorbed on a charcoal surface and can be removed completely from the aqueous phase by filtering off the charcoal. This charcoal-picric acid adsorption complex can be washed with large volumes of water without any appreciable quantity of picric acid appearing in the wash water. If, however, the charcoal-picric acid adsorption complex is washed with alcohol or with ether, the lower surface tension of the solvent causes the alcohol or the ether to replace the picric acid in the charcoal-water interfacial film, and the displaced picric acid moves out of the interface, giving a yellow solution in the alcohol or ether.

This reaction is of particular importance in the purification of organic compounds. Decolorizing a solution with boneblack or an activated vegetable carbon is a very common example of selective adsorption. After satisfactory decolorization has been secured, it often happens that the student washes, with an organic solvent, the carbon which has been filtered off, in order to be sure that he is not losing any appreciable amount of the valuable constituent which he wishes to conserve. In many instances these wash liquors are highly colored and contain the impurities which he originally wished to remove. Thus, the washing of the carbon adsorption complex with an organic solvent may entirely defeat the original purification program.

On the other hand, it may be possible to concentrate on an interface by the process of adsorption all the valuable constituent which is present in a solution, and then by later treating the adsorption complex with the proper solvent to remove the valuable constituent completely and obtain it in a highly concentrated and a relatively pure condition. Alkaloids can be quantitatively removed from an acid solution by "Lloyd's reagent" [106] (hydrous aluminum silicate). The affinity of Lloyd's reagent for alkaloids is so great and adsorption is so complete that no bitter taste can be detected either in the mother liquor or in the Lloyd's reagent-alkaloid adsorption complex, when such substances as quinine or strychnine are adsorbed. The adsorption complex can be filtered off, washed thoroughly with water to get rid of all contaminating sugars, salts, etc., and the alkaloid which was adsorbed upon the surface can be quantitatively liberated by dilute alkali and can then be extracted from the solution by shaking out with chloroform, ether, etc. Similar examples of selective adsorption are numerous.

The adsorptive process is so general that care should be taken at every step in the preparation or the quantitative estimation of biochemical compounds in order to prevent adsorption reactions from interfering with the desired results. Accordingly *in quantitative studies it is never justifiable to decolorize a solution with carbon or with any other adsorbent unless preliminary experiments have definitely proved that such decolorization does not remove that constituent which is being quantitatively estimated.* This is especially important in experiments where the material is estimated by colorimetric methods. Although it is essential that the original solution be colorless and clear, many of the methods which are used to produce clear, colorless solutions of biological fluids introduce errors due to forces of adsorption.[107]

The presence or absence of polar groups in the molecules of the solvent, of the solute, or of the adsorbent may affect an adsorption reaction, probably by reason of their influence on the ease with which the surface of a solid adsorbent is wetted. Freundlich [108] extended Traube's rule, that surface tension depression increases as we ascend a homologous series, to adsorption from aqueous solution and stated that *the adsorption of organic substances from aqueous solutions increases strongly and regularly as we ascend the homologous series.* Accordingly, in general, butyric acid is adsorbed more strongly than propionic, and propionic more

[106] J. U. Lloyd, *J. Am. Pharm. Assoc.*, **5**, 381, 490 (1916).

[107] *Cf.* J. C. Bock, *J. Am. Chem. Soc.*, **42**, 1564 (1920); R. A. Gortner and G. E. Holm, *J. Am. Chem. Soc.*, **42**, 1678 (1920).

[108] Herbert Freundlich, *Colloid and Capillary Chemistry*, p. 195, translated by H. S. Hatfield, E. P. Dutton and Co., New York, 1926.

strongly than acetic. However, *these generalizations are not invariably true.* Holmes and McKelvey [109] pointed out that Freundlich's adsorption studies dealt with carbon, a non-polar solid, in water, a polar liquid. They reversed the condition, using silica, a polar solid, and the various fatty acids dissolved in toluene, a non-polar liquid. They suggested that in Freundlich's studies the polar end (—COOH) of the fatty acid molecule was oriented toward the water phase and the hydrocarbon chain toward the carbon surface, whereas, in their own studies, the highly polar carboxyl end was oriented toward the silica surface and the non-polar alkyl group toward the hydrocarbon solvent. In addition to a reversal of Traube's rule they found that acetic acid was very strongly adsorbed from toluene on silica gel, propionic acid less strongly adsorbed, butyric acid still less adsorbed, and caprylic acid relatively slightly adsorbed. These observations afford another example of the role that molecular orientation plays at interfaces.

The adsorption process is an equilibrium, i.e., if an adsorbing surface is brought into contact with a solution which has an equilibrium concentration of 0.01 N, and the adsorbing substance is then removed from that solution and placed in a more dilute solution which has an equilibrium concentration of 0.001 N, the amount of material adsorbed will be identical with the amount that would have been adsorbed at an equilibrium concentration 0.001 N. Accordingly one cannot filter off an adsorbent which has reached equilibrium in a chemical solution and wash it with a solution which has a concentration different from the equilibrium solution without changing the amount of material adsorbed on the adsorbing surface.

Adsorption reactions are, in general, characterized by a positive heat of adsorption. In some instances the amount of heat liberated is large and indicates the great affinity existing between the surface of the adsorbent and the material being adsorbed. Harkins and Ewing [110] found that 1 gram of bone-charcoal gave a maximum heat of adsorption for water of 18.5 calories, and 1 gram of fuller's earth gave 32.0 calories. Lamb and Coolidge,[111] in an extensive study of the adsorption of various organic vapors on charcoal, found that the heat of adsorption could be expressed as

$$\log h = \log a + b \log x \qquad (123)$$

where h = total heat evolved during adsorption
x = cubic centimeters of gas adsorbed per gram of charcoal
a and b = constants.

[109] H. N. Holmes and J. B. McKelvey, *J. Phys. Chem.*, **32**, 1522 (1928).
[110] W. D. Harkins and D. T. Ewing, *J. Am. Chem. Soc.*, **43**, 1787 (1921).
[111] A. B. Lamb and A. S. Coolidge, *J. Am. Chem. Soc.*, **42**, 1146 (1920).

When they plotted the logarithms of h and x, series of straight lines were obtained, all having the same slope (constant b) and differing only slightly in the position at which they intersect the ordinate axis (constant a). The fact that all the vapors gave parallel lines indicates that the same fundamental mechanism is operating in all cases. Over the range studied the heat of adsorption which was evolved for each successive increment decreased but slightly and was found to be expressed by the equation

$$\frac{dh}{dx} = \frac{ab}{x^{1-b}} \tag{124}$$

It is evident that, when $1 - b = 0$, dh/dx becomes a constant, *i.e.*, the heat evolved becomes independent of the amount of vapor already adsorbed. For the systems studied $1 - b$ ranged from 0.044 to 0.100, indicating only a slight change in heat increment. The film was postulated to be polymolecular and for CS_2 was calculated to be approximately 40 molecules deep. The attractive force holding the molecules of the adsorbed vapor on the charcoal was calculated to be in the neighborhood of 37,000 atmospheres, when 1 ml. of liquid was adsorbed on 10 grams of charcoal. The heat of adsorption is therefore in reality the heat of compression.

Keyes and Marshall [112] studied the heat of adsorption of various gases on charcoal and found the following values: ether 7,250, ammonia 6,456, carbon dioxide 5,450, methane 4,600 calories per mole. They also noted that in the adsorption of gases one is not limited to a monomolecular layer. They reasoned that the initial reaction forms a monomolecular layer of adsorbed gases upon the surface of the adsorbent, and that "the first and succeeding layers, because of their special state, constitute new adsorptive surfaces which may adsorb molecules of the same species as the first layer or molecules of a different species. Each succeeding layer (the adsorbing molecules all of the same species) partakes of the special state of the first layer in lessened degree until finally a layer is reached wherein the molecular state differs little from what may be imagined as a molecular 'contact' arrangement."

Nutting [113] calculated the differential curve dc/dz, where c is the energy in calories evolved by the adsorption of water on a square centimeter $\times 10^4$ of a silica surface, and z equals the thickness of the adsorbed water film in centimeters $\times 10^{-6}$.

Figure 53 shows that the differential curve turns sharply at a thickness of water film between 4 and 5×10^{-6} cm., representing a layer of

[112] F. G. Keyes and M. J. Marshall, *J. Am. Chem. Soc.*, **49**, 156 (1927).

[113] P. G. Nutting, *J. Phys. Chem.*, **31**, 531 (1927).

water 100–120 molecular diameters deep, and that the energy is not by any means confined to the first molecular layer, nor is there any apparent break after the first molecular layer has been formed. Nutting calculates the maximum pressure at the surface to be of the order of 17,410 atmospheres and adds that, if the solid pulls on the adsorbed liquid with a pressure equal to 17,000 atmospheres, the liquid must also pull on the solid to the same degree, and accordingly *the pull of water on a quartz surface approximates the tensile strength of quartz!* This leads to a hydration and dispersion of quartz by water which is adsorbed onto the surface.

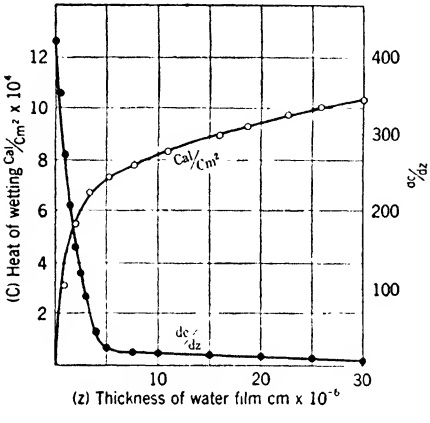


FIG. 53. Thermal changes as related to the thickness of the water film on silica gel. (Data of Nutting.)

Bartell and Almy [114] measured the heat of wetting of activated silica gel and found that it was markedly decreased if the silica gel was dehydrated, falling from 18.55 calories per gram for a gel containing 3.18 per cent water to 9.97 calories per gram when the water content was reduced to 0.83 per cent.

If heat is evolved by the adsorption of water upon a surface, at least an equivalent amount of heat must be added before the adsorbed water is released from the surface. From the heat of adsorption of liquids Patrick and Grimm [115] calculated a surface area of 6.9×10^6 sq. cm. per gram of silica gel. Neuhausen and Patrick [116] heated a silica gel to 300°C. *in vacuo* for a period of 6 hours without reducing the water content below 4.8 per cent, and Bartell and Almy stated, "Water persists within the gel at temperatures well above the critical temperature of water." The adsorptive force of silica gel for water is so great that even at high temperatures water would be positively adsorbed over the interface. Such water, at least up to 4.36 per cent water in the gel, is under a high state of compression and has a density greater than liquid water.[117]

Adsorption reactions are characterized by a negative temperature coefficient. This is to be expected, provided that there is a positive heat of

[114] F. E. Bartell and E. G. Almy, *J. Phys. Chem.*, **36**, 475 (1932).
[115] W. A. Patrick and F. V. Grimm, *J. Am. Chem. Soc.*, **43**, 2144 (1921).
[116] B. S. Neuhausen and W. A. Patrick, *J. Am. Chem. Soc.*, **43**, 1844 (1921).
[117] D. T. Ewing and C. H. Spurway, *J. Am. Chem. Soc.*, **52**, 4635 (1930).

adsorption. The fact that adsorption becomes less as the temperature increases is probably due to the increased kinetic energy of molecules with an increase in temperature. An adsorption reaction can accordingly, in some instances at least, be distinguished from a chemical reaction, in that a decrease in rate of adsorption and a decrease in the amount of material adsorbed occurs at a higher temperature, whereas a classical chemical reaction is characterized by an increase in rate with an increase in temperature, although in many instances the same equilibrium is reached. Accordingly the constant of the adsorption equations will be different at different temperatures.

Pseudo-Adsorption. True adsorption is an equilibrium, and diluting the system shifts the equilibrium concentration to the same point that would have been reached had the reaction been carried out originally at the lower concentration. It occasionally happens, however, that what appears to be an adsorption reaction is not a true equilibrium and is not reversible. For example, if porous porcelain is placed in a solution of auric chloride, gold which cannot be subsequently dissolved out of the adsorbing surfaces is adsorbed from the solution. The reaction in this case is probably similar to the reaction which Latshaw and Reyerson [118] utilize to prepare metallized silica gels. When silica gel is placed in solutions of salts of certain metals, including the noble metals, the salts are decomposed and the surface of the silica gel becomes covered with a thin layer of metal. Apparently the initial reaction is the adsorption of the metallic salt at the interface, and this adsorption is followed by a reduction of the salt by hydrogen (or other reducing substances) adsorbed on the surface of the silica gel. Accordingly the concentration of the metallic salt in the interface is altered, and additional adsorption takes place from the solution until finally by repetition of these processes the solution is almost entirely depleted of metallic ions. Graham, in 1830, noted that similar reductions occur on the surface of charcoal.

When an aqueous solution of egg albumin is beaten to a foam, the protein is positively adsorbed at the gas-liquid interface and *the surface energy coagulates the adsorbed egg albumin in the same way as egg albumin can be coagulated by heat.* Such denatured albumin will no longer dissolve in water, and accordingly we have an insoluble film of albumin which persists and prevents a reversal of the adsorption equilibrium. By stirring or bubbling air through a dilute solution of egg albumin, the protein can be removed practically quantitatively, and, if the foam is allowed to stand until it breaks and is then filtered, it is possible to collect nearly all the original egg albumin in the form of insoluble films.

[118] M. Latshaw and L. H. Reyerson, *J. Am. Chem. Soc.*, **47**, 610 (1925).

Pseudo-adsorption can, therefore, be defined as adsorption followed by a chemical change in the material which has been adsorbed, the chemical change being of such a nature as to inhibit the reversal of the adsorption process.

Pseudo-adsorption probably plays a role in the "poisoning" of hydrogen electrodes by Hg, Cu, As, etc. The ions are probably reduced by the hydrogen, so that we are no longer dealing with a platinum surface but rather with a mercury, copper, or arsenic surface which gives an erroneous or an unstable potential. Arsenic is notoriously poisonous to the platinum catalyst used in the contact process for the manufacture of sulfuric acid. Here again the explanation probably lies in the adsorption upon the platinum surface of arsenic or arsenic compounds, which causes the surface to become catalytically inert. Biological catalysts may be "poisoned" by a similar process.

Catalysis by Adsorption. We have already noted Taylor's view in regard to the nature of an active surface of nickel used in hydrogenation, *i.e.*, that we have certain free valences at isolated points in the catalytic surface, where atoms of nickel behave more or less as gaseous nickel. Almquist [119] takes a somewhat similar view of the nature of the surface of iron catalysts used in the synthesis of ammonia from nitrogen and hydrogen.

Boswell and Dilworth,[120] in a study of the role which aluminum oxide plays in the catalysis of certain organic reactions, such as the formation of ethylene and water from ethyl alcohol, point out that all the reactions which are catalyzed by aluminum oxide involve either hydrogen or hydroxyl ions, or the addition of water to, or the removal of water from, the reacting compounds, and reach the conclusion that aluminum oxide functions by means of a surface film of water which is the real seat of the catalysis. "There is still a water film on the surface, even after heating at atmospheric pressure at 500°C. for twenty hours, followed by two days' heating with a Meker burner." In regard to catalysis they conclude that "*The catalyst from this point of view does not accelerate a reaction already in progress . . . but actually initiates the change from one to the other.*" Taylor has made a similar statement that a catalyst may initiate a reaction because of the energy residing in the catalytic surface, and when we discuss enzyme reactions we shall see that a combination (adsorption) between the enzyme and substrate usually is the first reaction to occur.

Some Applications of Adsorption. The adsorption of gases on surfaces became of extreme importance when gas warfare was introduced.

[119] J. A. Almquist, *J. Am. Chem. Soc.*, **48**, 2820 (1926).

[120] M. C. Boswell and H. M. Dilworth, *J. Phys. Chem.*, **29**, 1489 (1925).

The problem of offense was to use poison gases which were not readily adsorbed upon the charcoal or other adsorbents used in the canisters of the gas masks, whereas the problem of defense was to prepare quickly suitable adsorbents for the gases which were introduced. In many instances, special adsorbents were found, and throughout the entire period of World War I there was a rapid advance in the development of new adsorbents and improving the adsorbing power of the common materials such as charcoal and silica gel. Since that time the knowledge gained in this way has been applied to industrial problems in removing vapors of solvents, etc., from the atmosphere of factories, for the purpose either of reducing the health hazards or of recovering valuable solvents.

In the manufacture of sulfuric acid by the contact process, adsorption plays a very important role. The oxygen and sulfur dioxide are adsorbed upon the contact surface so strongly that they are brought within the radius of chemical attraction and unite to form sulfur trioxide which is less strongly adsorbed than the initial reacting materials. The surface accordingly becomes supersaturated with respect to the sulfur trioxide. Sulfur trioxide leaves the surface, allowing room for additional adsorption of oxygen and sulfur dioxide. Such a cycle will continue indefinitely as long as the surface remains in its original condition. The poisoning of such a surface is undoubtedly due to the irreversible adsorption of some foreign material on the catalyst where it is held with a greater tenacity than the reagents of the reaction which it is proposed to catalyze, that portion of the surface which is covered by the adsorbed poison film becoming catalytically inert.

Adsorption is used to de-ink magazines and old newspapers, making the paper available for re-use as newsprint. A small amount of bentonite mixed with pulped newspapers or magazines selectively adsorbs the printer's ink from the fiber, and when the mixture is run over wire screens and washed with water the bentonite and the adsorbed printer's ink are readily washed from the fiber, yielding a reclaimed pulp which is practically ink-free and which can be made into a satisfactory sheet of paper with very little loss of the original fiber content.

The phenomenon of selective adsorption or rather of the wettability of surfaces by liquids is strikingly illustrated in the ore-flotation process. Sulfide ores, such as copper sulfide and lead sulfide, have a non-polar surface and are not easily wetted by polar liquids like water. On the other hand, they are rather easily wetted by non-polar liquids such as oils. Silica and siliceous minerals, on the other hand, having a polar surface, are readily wetted by water but are not readily wetted by oils. If a rock, containing small amounts of valuable sulfides intermixed with large amounts of silicates, is finely ground and the ground mixture

agitated with water to which a certain amount of oil has been added, the oil is selectively adsorbed by the sulfide particles, so that they become coated with a film of oil, or a number of the sulfide particles adhere to a single oil droplet. If air is then blown through this mixture, the finely ground sulfides adhering to the oil particles rise to the surface in a foam which can be skimmed off, the silica and silicates of the gangue settling to the bottom of the agitator tank. In this way a relatively small amount of a valuable mineral may be separated from a relatively large amount of inert material. A concentrated high-grade ore is thus obtained. The ore-flotation process has made it possible to utilize ores of much lower grade than was previously feasible. As a matter of fact, it has made possible the working of large masses of residues which had been earlier discarded. More than 60,000,000 tons of ore are concentrated in the United States each year by the flotation process.

De Groote [121] uses the theories of molecular orientation to break natural crude oil emulsions on the theory that, if an emulsion of water-in-oil is stabilized by oleophilic compounds, such an emulsion may be broken by adding hydrophilic colloids. He has been very successful in this attempt by adding various types of organic acids (used as their ammonium salt) to crude oil emulsions, tank bottoms, etc.

An important application of adsorption, and apparently molecular orientation, is afforded by the protection which small amounts of amines confer upon an iron surface in inhibiting corrosion.[122] Pieces of steel were immersed in N sulfuric acid containing various amines and the rate of corrosion measured by weighing the steel sheet from time to time. Ammonium sulfate was relatively ineffective, and the inhibition of corrosion increased progressively from methylamine through ethyl-, *n*-propyl-, *n*-butyl- to *n*-amylamine, which was the most effective of the primary amines. Secondary amines were more effective than primary amines, and, with the exception of trimethylamine, tertiary amines were more effective than secondary amines. In some instances corrosion was inhibited 99.9 per cent in solutions containing as little as 0.005 per cent amine nitrogen.

The Adsorption of Water on Biological Products. The hydrophilic colloids, characteristic of biological systems, possess a great affinity for water. The major phenomena involved will be discussed later when we consider the problem of gels. At this point it will be sufficient to consider only the problems pertaining to the drying of biological materials.

[121] Melvin De Groote, p. 616 in *The Science of Petroleum*, Vol. I, Oxford University Press, England, 1937.

[122] C. A. Mann, B. E. Lauer, and C. T. Hultin, *Ind. Eng. Chem.*, **28**, 159, 1048 (1936)

The determination of the moisture content of a biological material is a purely empirical procedure determined by the three variables, temperature, pressure, and time. In order to make a definite statement that such and such a biological material has such and such a moisture content, it is necessary to define the conditions in regard to temperature, pressure, and time of drying under which the moisture determination was carried out. *The removal of water from a biocolloid is merely shifting one equilibrium between a colloid surface and water to a new equilibrium,* and the extent to which the equilibrium is shifted is determined by these three variables. This has been emphasized by Nelson and Hulett,[123] who in a study of this problem conclude that moisture is retained by biological products at least to temperatures as high as 365°C., the critical temperature of water.

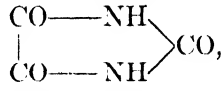
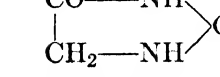
Enzymatic Activity. Apparently the first reaction between an enzyme and its substrate is an adsorption of the enzyme by the substrate. This is followed by the chemical action promoted by the enzyme. The initial adsorption reaction can be readily demonstrated. If pepsin is added to a suspension of fibrin in 0.2 per cent hydrochloric acid, and the mixture is vigorously shaken and immediately filtered, it will be found that the pepsin has been quantitatively adsorbed by the fibrin. This can be proved by the addition of fresh fibrin to the acid filtrate, under which conditions no appreciable digestion of the fibrin will be observed. If fresh acid is added to the fibrin granules which were filtered off, hydrolysis will proceed, and the fibrin will become completely digested. The initial combination of the enzyme with the substrate usually takes place almost instantaneously, the rapidity of reaction and other characteristics indicating rather conclusively that it is purely an adsorption phenomenon.

Quastel [124] emphasizes adsorption as an essential feature of enzyme action and accounts for specificity on this basis. He studied the oxidation $\rightleftharpoons$ reduction reactions of bacteria on 108 organic compounds. Of these, 56 were "activated" so that they acted as hydrogen donators, *i.e.*, the hydrogen atoms in the molecule were transferable to a suitable "acceptor" such as oxygen or methylene blue. Strict specificity of enzyme action would require 56 dehydrogenases. Quastel found that the 56 compounds could be grouped into classes, *e.g.*, a formic acid class, a lactic acid class, a succinic acid class. He shows that *the rate of oxidation* $\rightleftharpoons$ *reduction of any compound in a given class is altered (lessened) by the presence in the system of any other compound of the same class grouping but is unchanged in the presence of a compound belonging to another class.*

[123] O. A. Nelson and G. A. Hulett, *Ind. Eng. Chem.*, **12**, 40 (1920).

[124] J. H. Quastel, *Trans. Faraday Soc.*, **26**, 853 (1930).

Quastel assumes adsorption at "active centers" on the bacterial membrane surface, these "centers" being *specific* for certain *molecular groupings*. Thus the "lactic acid center" adsorbs the grouping —CO—COH*— or —CHOH—COH*—, where H* is mobile. Glycollic, oxalic, glyoxylic, hydroxymalonic, glyceric, α-hydroxybutyric, mandelic, and pyruvic acids are all "adsorbed" and "activated" at the "lactic acid center," and in a mixture of two or more of these compounds there is competition for the area available for adsorption.

The "succinic acid center" is apparently specific for the grouping —C—CH_2COOH or —C—CH—COOH and in addition to succinic acid is the locus for adsorption of malonic, glutaric, phenylpropionic, tricarballylic, and pyrotartaric acids, all of which compete with one another for adsorption and for "activation" but none of which interferes noticeably with the "lactic acid group" of compounds. The "formic acid center" appears to be specific for formic acid. Even acetic acid does not interfere with reactions occurring at this center. An excellent demonstration of the specificity of adsorption as determined by chemical configuration is afforded by the fact that malonic acid is adsorbed and activated at the succinic acid center, hydroxy malonic acid at the lactic acid center, but ethyl malonic acid is not adsorbed at any of the centers. Furthermore, parabanic acid,  , was strongly adsorbed at the lactic acid center, whereas hydantoin,  , was not adsorbed, the hydrogenation of one carbonyl group destroying the specificity.

In summing up the work Quastel states, "Substances appear to act as 'poisons' simply by competing with the substrate for the space available for adsorption at the centers. The 'poison' and the substrate apparently compete with each other for adsorption on fairly equal terms. . . . A relatively large number of substances can be adsorbed in this specific manner, but out of this large number only a few can be *activated* to function as donators of hydrogen. . . . Specificity of enzyme action is seen to depend upon three factors:

"(1) Specificity of adsorption at the active centre.

"(2) The nature and strength of the polarizing field at the active centre.

"(3) The constitution of the adsorbed molecule.

" . . . The reason therefore why an enzyme is so specific in its action is, in the first place, because only a limited number of substances—con-

taining a certain type of structure—is accessible to or adsorbed by the enzyme, and in the second place, because out of this limited number of substances specifically adsorbed only a few are capable of being turned into the 'active' molecules capable of the reactions under investigation."

Glick and King [125] have extended Quastel's studies in an investigation of the effect of the structure of organic compounds on the inhibition of the action of liver esterase. They found inhibition to increase rapidly as the length of the carbon chain of the *n*-aliphatic alcohols increased. One part of nonyl alcohol caused the same inhibition as 840 parts of methyl alcohol. When various groups were attached to the *n*-amyl alkyl radical, a lyotropic series of $CN > I > NO_3 > SH > Br > OH > Cl > CO > CONH_2 > NH_2$ was obtained from the inhibiting effect. One part of amyl cyanide was as effective as 270 parts of amylamine. When the alkyl radical was replaced by C_6H_5—, the inhibition followed a series which was only slightly altered, *i.e.*, $I > Br > OH > Cl > NO_2 > CN > CH_3 > CONH_2 > NH_2$. One part of phenyl iodide was as effective as 58 parts of aniline.

Quastel's view of active centers is still further extended by the observation that coupled reactions occur in biological systems.[126] On the surface of *Escherichia coli*, pyruvate is reduced to lactate by means of the energy of the anaerobic oxidation of formate to bicarbonate, and fumarate is reduced to succinate through the anaerobic oxidation of lactate to pyruvate. In both instances an intermediary energy carrier is necessary. This carrier must be reducible at the locus where one of the compounds is oxidized and reoxidized at the locus where the other compound is reduced. For the lactate-pyruvate-formate-bicarbonate system methylene violet serves as the intermediary, and methylene blue serves for the succinate-fumarate-lactate-pyruvate system. Without an intermediary no reaction occurs.

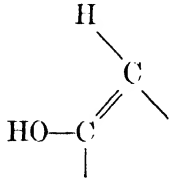
The bearing of the above experiments on various physiological and medical problems is obvious. We have as yet no specific information about the types of surfaces which are exposed at the various centers of the bacterial mosaic surface to account for the areas characteristic of specific adsorption. It was in the hope of securing such specific information that the molecular orientation studies conducted in the author's laboratories were undertaken.

One other example from the field of medicine may suffice to illustrate the importance of specific adsorption and molecular orientation for biological studies. Morphine is known to possess the desirable property of deadening pain but to have associated with this property the un-

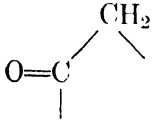
[125] D. Glick and C. G. King, *J. Biol. Chem.*, **94**, 497 (1931); **95**, 477 (1932).

[126] H. F. Schott and H. Borsook, *Science*, **77**, 589 (1933).

desirable properties of inducing habit formation, producing nausea, and favoring constipation. The morphine molecule contains the grouping



Alvarez [127] notes that when the hydroxyl group is converted to a carbonyl group and the double bond is hydrogenated to yield a structure containing the grouping



the pain-deadening properties of the drug are increased about fivefold, the constipation and psychic effects are markedly decreased, and the habit-forming properties essentially disappear. In this case we are presumably dealing with adsorption and specific orientation of the original morphine on at least two brain centers, one which has to do with habit formation and the other with pain. Apparently a slight change in chemical configuration of the molecule intensifies adsorption and orientation on the pain center and destroys adsorption affinities (or alters the specific orientation) of the drug on the habit-forming center. Admittedly the above explanation is a hypothesis, but in the light of Quastel's observations it appears to be an extremely logical explanation. We need specific adsorption and molecular orientation studies on all the physiologically active drugs and those compounds which are structurally closely related to them. It is believed that, when such data are available, the pharmacological and physiological behavior of drugs can be much more rationally interpreted.

[127] W. C. Alvarez, *Proc. Staff Meetings Mayo Clinic*, **7**, 480 (1932).

CHAPTER 8

Electrolytes and Colloid Systems

We have already noted that a lyophobic colloid micelle is stabilized by an electric double layer existing in the region of its interface with the medium of dispersion, and that, when this double layer is modified to decrease the potential (electrokinetic) existing across it, the micelle becomes unstable and flocculation results. Small amounts of added electrolytes cause strong modifications in the magnitude of this electrokinetic potential. Solutions of electrolytes accordingly have a very marked effect upon the properties of colloidal systems. Colloids which are *negatively charged* are flocculated by the *cations* of an added electrolyte, whereas *positively charged* micelles are sensitive to the added *anions*.

Referring to Fig. 4, the addition of Br^- causes the flocculation of positively charged silver bromide micelles until the isoelectric point is reached; then a further addition of Br^- imparts a negative charge to the isoelectric silver bromide with the formation of negative silver bromide micelles. Accordingly, as an electrolyte is added to a colloidal sol, the electrokinetic potential is progressively decreased to zero. Further addition of electrolyte may cause the formation of a new sol possessing a charge opposite to that possessed by the original sol.

In general, the flocculating power of an ion follows the Hardy-Schulze rule that "*the precipitating power of an electrolyte depends upon the valence of the ion whose charge is opposite to that on the colloidal particle.*"

Although the Hardy-Schulze rule is not a hard-and-fast one, it is in general true, and differences between ions of the same valence can probably be explained by differences in ionic radii, in degree of hydration of the ions, in ionic mobility, and in the degree to which they are adsorbed on the surface of the colloid micelle.

Negatively charged sols are coagulated by the cations, Na^+, Ca^{++}, Al^{+++}, and are relatively slightly affected by the anions, Cl^-, $SO_4^=$, $PO_4^{\equiv}$, the reverse being true for the positively charged sols. *The influence of valence is not an arithmetical 1:2:3 ratio but more nearly a geometrical progression* $1:x:x^2$, where, in some instances at least, x has a value lying somewhere between 16 and 32. If x should have a value of

32, Ca^{++} would be 32 times as efficient a flocculating agent as Na^{+} and Al^{+++} would be 1,024 times as effective as Na^{+}.

The data in Table 27 explain the use of alum or iron sulfate in water purification. In large cities the water is frequently taken from a muddy river, pumped through a station where chlorine is added to destroy microorganisms and where iron or aluminum sulfate is added to flocculate the suspended particles. The water then goes on to the filter beds

TABLE 27. EFFECT OF ELECTROLYTES IN FLOCCULATING FERRIC HYDROXIDE SOLS (DATA OF HARDY) AND ARSENOUS SULFIDE SOLS (DATA OF FREUNDLICH)

(Positive) Ferric Oxide Sols		(Negative) Arsenous Sulfide Sols	
Coagulating ion and salt used	Coagulation concentration, gram equivalents per liter	Coagulating ion and salt used	Coagulation concentration, millimoles per liter
Cl^{-} (NaCl)	0.5	Na^{+} (NaCl)	51.0
Cl^{-} (HCl)	0.5	Mg^{++} ($MgCl_2$)	0.717
$SO_4^{=}$ (H_2SO_4)	0.002	Mg^{++} ($MgSO_4$)	0.810
$SO_4^{=}$ (K_2SO_4)	0.0006	Ca^{++} ($CaCl_2$)	0.649
		Ba^{++} ($BaCl_2$)	0.691
		Al^{+++} ($AlCl_3$)	0.093

and passes through the filters into the water mains. Relatively small amounts of Al^{+++} or Fe^{+++} are required to clarify such waters. The clarification could be conducted by the addition of sodium chloride, but the amount of Na^{+} that would have to be added to produce the same degree of flocculation would be so great that the water passing into the city main would probably taste salty, owing to the excess of sodium chloride remaining in the water.

Deltas are formed where rivers carrying clay and silt meet the salt water of the ocean, the clay and silt being deposited because of the neutralization of their electric double layer by the electrolytes in the ocean water.

The flocculating effect of electrolytes acting upon lyophobic colloids is due almost entirely to their effect on the stabilizing electrokinetic potential. We have already indicated, Fig. 34, that the electrokinetic potential does not have to drop to zero before the sol becomes unstable, but that instead there is a *critical zone* in the vicinity of the isoelectric point where the magnitude of the potential is not sufficiently great to insure indefinite stability.

That matter dispersed in colloidal dimensions should remain stable for long periods of time must mean that in such cases the forces of repulsion existing between the particles are greater than the forces of attraction, and, as a result, the individual particles never approach each other closely or do not remain together at the point of their closest approach. In lyophobic colloid systems it has long been recognized that the electrical forces resulting from the existence of the double layer of ions in the region of the interface constitute the forces of repulsion which promote their stability. It is generally assumed that, if in spite of all repelling forces two lyophobic particles should be made to approach contact with each other, attracting forces which may be variously described as interfacial tension forces, van der Waals-London forces of attraction, etc., would act to hold them together. As this process continues, flocculation results. It is also generally assumed that the force, which in an unstable sol brings about this required close approach, arises from the kinetic motion of the particles as they exist in solution. If, then, the maximum kinetic energy of approach attainable by any two particles in a sol is less than the electrical energy absorbed as a result of distortion of their double layers during their approach, movement toward each other will stop short of the distance at which the attracting forces are great enough to maintain them in a state of combination strong enough to prevent their subsequent spontaneous peptization.

Hardy,[1] Ellis,[2] and Powis,[3] among others, early considered the magnitude of the stability factor, and hence the electrical energy involved in this process, to be proportional to the electrokinetic potential. If, at a given temperature and with a given sol, the electrokinetic potential were reduced to a critically low value, *e.g.*, by electrolyte addition, such that not all particles approaching each other in the sol were prevented from penetrating to close contact, flocculation would proceed at a measurable and definite rate. Hence the concept of a "critical" potential which would be the minimum potential required to prevent this flocculation process. Experiment has confirmed this theory only partially. Salts of monovalent ions generally cause flocculation while the electrokinetic potential of the sol is significantly higher than that which is "critical" when salts containing polyvalent ions of sign of charge opposite that of the particle are used.[4,5]

[1] W. D. Hardy, *Z. physik. Chem.*, **33,** 385 (1900).

[2] R. Ellis, *Z. physik. Chem.*, **78,** 321 (1912); **84,** 145 (1915).

[3] F. Powis, *J. Chem. Soc.*, **109,** 734 (1916).

[4] H. R. Kruyt, A. C. W. Roodvoets, and P. C. van der Willigen, Colloid Symposium Monograph, p. 4, 1926.

[5] D. R. Briggs, *J. Phys. Chem.*, **34,** 1326 (1930).

Eilers and Korff [6] have concluded that the factor ζ^2/κ (where $\kappa = 1/d$ and d is the thickness of the electrical double layer) is a more accurate measure of the electrical energy of repulsion which determines the stability than is the electrokinetic potential ζ alone. They present evidence that a critical value of ζ^2/κ gives better agreement with experiment than a critical value of ζ. It would, therefore, be more accurate to speak of a critical value of the electrical energy of the interface (proportional to ζ^2/κ) than of a critical potential (proportional to ζ) as the determinant factor in the stability of a lyophobic sol.

Levine and Dube,[7] Hamaker,[8] and Verwey [9] have offered theoretical considerations depicting the manner in which the attracting and repelling forces acting between particles may vary with their distance apart and with changes in their electrolyte environment. These forces acting between colloidal particles are assumed to be due to a superposition of an attracting potential P_A, arising from van der Waals-London forces, and a double layer interaction potential P_R caused by the interpenetration of the diffuse charges of the electrical double layers surrounding the particles in the liquid. The result of the double layer interaction (for particles bearing like sign of charge) is always repulsion. Depending on the magnitudes of these potentials and the manner in which they vary with the distance between the particles, the free energy of the interfaces involved (proportional to $P_R + P_A$, where force of repulsion is given a positive sign and force of attraction a negative sign) can vary with distance in three type manners as shown graphically in Fig. 54. (P_R decays exponentially with d and P_A varies according to d^{-2}.) Curve c in this figure represents the interfacial energy conditions for a sol which is completely devoid of stability. Such a system reaches a state of equilibrium only when all particles reach a distance apart, equal to T in the figure, representative of actual contact or at least very close approach. The particles are flocculated. Any lyophobic sol in an isoelectric condition is an example.

Curve b represents the relationships of a stable lyophobic sol, where the maximum energy of the interface M is greater than the highest possible kinetic energy of approach attainable by the particles. Addition of electrolytes, by decreasing the repulsion potential and thus the magnitude of M (which can be thought of as the activation energy of flocculation) modifies curve b toward curve c and at some intermediate

[6] H. Eilers and J. Korff, *Trans. Faraday Soc.*, **36,** 229 (1940).

[7] S. Levine and G. P. Dube, *Trans. Faraday Soc.*, **35,** 1125 (1939); **36,** 215 (1940)

[8] H. C. Hamaker, *Rec. trav. chim.*, **55,** 1015 (1936); **56,** 3, 727 (1937); **57,** 61 (1938).

[9] E. J. W. Verwey, *Trans. Faraday Soc.*, **36,** 192 (1940); *Chem. Weekblad*, **39,** 563 (1942); *Philips Research Repts.*, **1,** 33 (1945).

value, some critical value of M, allows flocculation to occur. Also, at a given value of M, an increase in temperature increases the kinetic energy of approach to such an extent that the particles can approach past the distance at which M occurs, and the attracting forces then becoming predominant lead to flocculation.

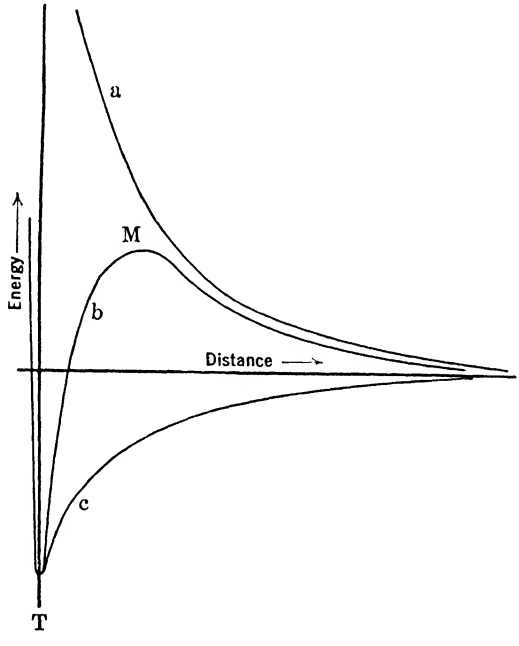


FIG. 54. Schematic representation of the manners in which the free energy of the interface (proportional to $P_R + P_A$) may vary with distance apart of the particles of a sol.

Curve *a* is representative of a sol in which the value of M is so high or the attracting potential of such a low value that the energy of the interface will be relatively increased at all distances closer than infinity. Curve *a* then is representative of a very strongly charged lyophobic sol in very dilute electrolyte solution, or it is representative of a lyophilic sol, in which case, apparently, the attraction potential between particles is zero owing to a preferential attraction of the particle surface for the dispersion medium. Intermediate values between these type curves can be devised to describe the stability relationships for any given sol.

Bungenberg de Jong,[10] working in Kruyt's laboratory, has shown that lyophilic sols are stabilized not only by an electric potential but likewise by solvation. Figure 55 represents his argument schematically. In *A*

[10] H. G. Bungenberg de Jong, *Rec. trav. chim.*, **42**, 437 (1923); **43**, 35 (1924).

we have a neutral particle. This can become negatively charged, as in *B*, by the adsorption of an anion. The addition of more salt or a salt with a polyvalent cation will reduce the potential to zero, and the particle will revert to its former condition as at *C*. The adsorption of a cation will produce a positively charged micelle at *D*, which again can be made isoelectric by the neutralizing effect of an electrolyte producing

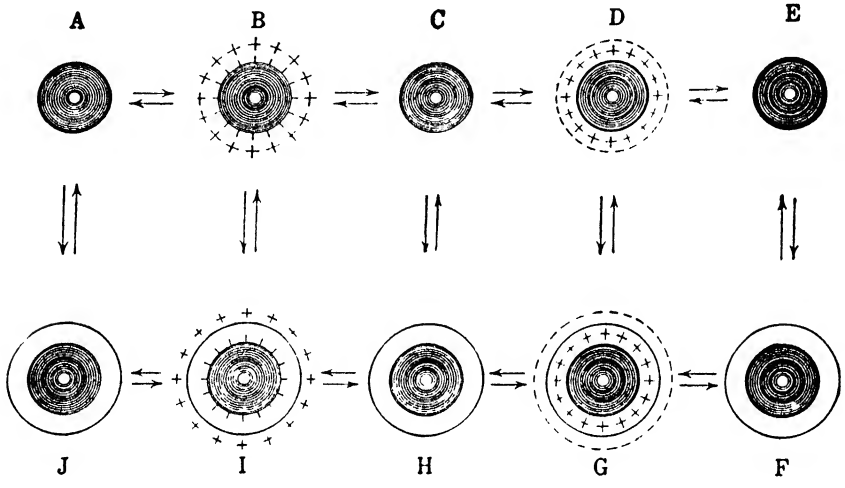


FIG. 55. A diagrammatic representation of the relationships which exist between the solvation of a particle and the electrical charge of a particle, showing that both are important as factors influencing stability. *A*, *B*, *C*, *D*, *E* are lyophobic particles of which only *B* and *D* will exist in stable sols. *F*, *G*, *H*, *I*, *J* are lyophilic particles. All will be relatively stable as contrasted to the lyophobic series, but *G* and *I* have two factors for stability, solvation, and electric charge. Suitable electrolytes in low concentration will neutralize or even reverse the charges. "Desolvating" agents (*e.g.*, alcohol for water systems) will convert the solvated particles to lyophobic systems. (After Bungenberg de Jong.)

a particle, as at *E*; or particle *A* can become solvated (or hydrated if the dispersions medium is water) as at *J*, and this solvated particle can, under the same conditions as above, become negatively charged, be neutralized, become positively charged, and again be neutralized without losing its solvation.

The addition of small amounts of electrolytes influences only the electrokinetic potential on the micelle. The addition of larger quantities of electrolytes influences the degree of solvation and causes flocculation to take place. In Fig. 55 flocculation will occur at *A*, *C*, and *E*. *B* and *D* will be stable lyophobic sols, *J*, *H*, and *F* isoelectric lyophilic sols possessing a considerable degree of stability, and *I* and *G* charged lyophilic sols with a higher degree of stability. Dehydrating agents, such as alcohol,

will cause the lyophilic colloid to lose its solvation and accordingly effect the transformation from *I* to *B* or from *G* to *D*. Because of the influence which solvation has upon the stability of a lyophilic system, the apparent effect of electrolytes on such systems is much less marked than on lyophobic sols.

The "salting out" of lyophilic colloids, such as proteins, from solution, by the addition of ammonium sulfate or by saturation of the solution with magnesium sulfate is not an electrokinetic phenomenon but rather a dehydration of the neutral micelle whose electrokinetic potential has already been reduced to nearly zero by the first addition of the electrolyte but whose affinity for the solvent is overcome only in solutions which contain a large amount of the electrolyte where the ions of the electrolyte compete with the lyophilic colloid for the water which is available.

In an attempt to determine the magnitude of the repulsion force existing between surfaces which are solvated, Derjaguin [11] measured the pressure variation in relation to distance apart of surfaces separated by a liquid film. The thickness of the liquid film on a glass surface (microscope slide) was measured through observation of interference patterns under the microscope when a bubble of air was pressed against the film surface on the glass slide. Films 50 mμ in thickness could withstand very high pressures without being forced off the glass surface, and the pressure decreased to zero only when the thickness of the film was of the order of 200–300 mμ. This points to a comparatively large radius of action of solid surfaces on the molecules of the neighboring liquid layers. The nature of these "solvating" forces is problematic since they, according to Derjaguin, are neither of the nature of van der Waals forces nor of dipole forces. These forces are of major importance in determining the stability of lyophilic colloids, the action and mechanism of the stabilizing action of added surface-active substances, etc., and a study of their nature and action is of the highest importance to those interested in the application of colloid chemistry to biological systems because these systems are principally lyophilic in nature.

Thixotropy. The term thixotropy was introduced in 1927 to describe those colloidal systems in which a gel is converted into a sol isothermally by shaking or stirring the gel, the sol setting to a gel again when left undisturbed. The term is derived from the Greek *thixis* (touching or striking) and *trepo* (to turn or to change). Almost all the literature dates from 1927, although before that time there were a few isolated observations of this peculiar behavior. Freundlich [12] and his students have been especially active in investigating this phenomenon.

[11] B. Derjaguin, *Trans. Faraday Soc.*, **36**, 203 (1940).

[12] H. Freundlich, *Thixotropy*, Hermann et Cie., Paris, 1935.

There is a close relationship between thixotropy and coagulation. *Ions which cause coagulation of a dilute sol bring about solidification and form thixotropic gels in more concentrated sols of substances which are capable of showing thixotropic behavior.* For example, concentrated hydrous ferric oxide sols or concentrated hydrous aluminum oxide sols are converted into thixotropic gels by the proper concentration of anions, whereas concentrated, negatively charged sols of such materials as bentonite (a colloidal clay) are influenced chiefly by cations. Time of solidification of the sol for a given concentration of electrolyte is used to measure thixotropic behavior.

Thixotropic behavior depends on the existence of a marked affinity between the particles and the liquid. The particles are surrounded by rather thick layers of liquid so that there is an interlocking between the particles when the system is at rest, and when the thixotropic gel is in equilibrium, the competition between the repelling forces and attracting forces results in the particles being rendered motionless. The result is a thixotropic gel. When the orientation of the particles with respect to each other is disturbed by shaking, the system loses its rigidity and acquires the properties of a sol. Since there must be a balance between attracting forces and repelling forces, the thixotropic behavior is realized only when the electrokinetic potential has been exactly adjusted to a rather narrow range.

Starch is a biocolloid which shows distinct thixotropic behavior. If a mass of starch is placed in a beaker, if sufficient water is added to cover the starch mass, and if the preparation is allowed to stand for a few minutes, it will be found that considerable energy will be necessary to move a stirring rod through the mass. However, once stirring has been started, the mass becomes decidedly liquefied, and much less energy is necessary to continue the stirring. On standing again the starch-water system solidifies.

The phenomenon of thixotropy is of major importance in the drilling of deep oil wells. A "drilling fluid" is usually injected into the well during the drilling operation to carry rock chips, etc., to the surface. If the drilling operation is suspended for a short period of time, these rock chips settle to the bottom and form a compact mass around the drill bit so that it might be impossible to move the drill bit when operations are resumed. Accordingly, colloidal materials, such as bentonite, are added to the drilling fluid, and as long as the bit is in operation the agitation keeps the drilling fluid in the sol form. If, however, drilling is suspended, the drilling fluid solidifies to a gel, and the rock fragments remain suspended in the gel. Drilling fluids have been devised in which the sol-to-gel transformation occurs within less than 2 seconds after drilling opera-

tions have ceased. Quicksand, in contrast to ordinary sand, is thixotropic. It appears solid but when agitated becomes fluid.[13]

It is highly probable that protoplasm has thixotropic properties which may explain the ready sol ⇌ gel transformations which so many biological workers have observed to take place in it. Chambers [14] comments on the fact that, when protoplasm is stirred with the fine needles of a micromanipulator, it becomes highly mobile but that it again becomes resistant to the movement of the needle if left undisturbed for a short period of time.

A few thixotropic systems exhibit the very surprising effect of solidifying rapidly if the sol is given a slow circular to-and-fro motion. This phenomenon has been called *rheopexy*.[15] Freundlich notes that in such systems the particles are rod- or plate-like, the slow circular motion appears to cause an orientation of the particles, and, when the particles lie parallel to one another, solidification takes place. Such sols solidify on standing but require a much longer time before solidification takes place than when they are gently rotated. Strong agitation again converts the system to a sol.

Protective Colloids. Inasmuch as lyophilic colloids are stabilized not only by an electric double layer but also by the adsorbed dispersions medium, small amounts of lyophilic colloids are, in many instances, sufficient to "protect" lyophobic sols from the flocculating action of electrolytes. This behavior is known as *protective colloid action*.

Schulz and Zsigmondy [16] have given us a measure of the protective efficiency of lyophilic colloids in the "*gold number*." They define the gold number of a colloid as *that weight of the colloid in milligrams which will just fail to prevent a change in color from red to violet when 1 ml. of 10 per cent sodium chloride solution is added to 10 ml. of a Zsigmondy (formaldehyde) red gold sol to which the colloid has been added.* The mechanism of the reaction is an adsorption of the protective colloid on the surface of the gold micelles, so that the interface is no longer a gold-water interface but rather a hydrophilic colloid-water interface, the micelle acquiring the characteristics of the lyophilic colloid which was added. Table 28 shows the gold numbers of various lyophilic colloids.

Colloidal silver and colloidal silver oxide have pronounced bactericidal properties, and the "argyrol" of the physician is colloidal silver protected by protein split-products.

The protective action is characterized by a definite time interval.

[13] H. Freundlich and F. Juliusburger, *Trans. Faraday Soc.*, **31**, 769 (1935).
[14] R. Chambers, *Proc. Soc. Exp. Biol. Med.*, **19**, 87 (1921).
[15] H. Freundlich and F. Juliusburger, *Trans. Faraday Soc.*, **31**, 920 (1935).
[16] F. N. Schulz and R. Zsigmondy, *Beitr. chem. physiol. Path.*, **3**, 137 (1903).

Sufficient time must elapse after the protective colloid has been added to the red gold sol to complete the initial adsorption of the protective colloid at the gold-water interface. Otherwise erroneous values will be obtained. The time interval before the sodium chloride is added, as a rule, need not exceed 10 minutes. If the protective colloid is added to the solution of the electrolyte and then this mixture is added to the colloid system, very much smaller quantities of electrolyte are sufficient to coagulate the lyophobic sol than would be required if the protective colloid was first added to the sol. The explanation probably is that the

TABLE 28. GOLD NUMBERS OF VARIOUS LYOPHILIC COLLOIDS

Substance	Gold Number	Substance	Gold Number
Dextrin (British gum)	125–150	Gum arabic	0.10 –0.125
Soluble starch	10–15	Protalbinic acid	0.15 –0.20
Sodium oleate	2–4	Lysalbinic acid	0.10 –0.125
Egg albumin	0.08–0.10	Gelatin	0.005–0.0125

colloidal micelles adsorb the protective colloid and form a micelle which is not readily flocculated. No such pronounced adsorption takes place between the solution of the electrolyte and the lyophilic colloid, and, when this mixture is added to the lyophobic system, sufficient time does not elapse for the stabilizing adsorption to take place, flocculation occurring at once.

Various other measures of protective value have been suggested. Windisch and Bermann [17] proposed an *iron number* to characterize hydrophilic colloids which prevent the coagulation of Fe_2O_3 sols. They thus identified a gum in the wort as being responsible for stabilizing the foam of beer.

Wo. Ostwald [18] has suggested a *rubine number*, in which Congo rubine sols are used in place of the gold sol, the rubine number being defined as the amount of colloid in grams per 100 ml. of solution which prevents the change of color of a Congo rubine sol from red to blue. The electrolyte used is potassium chloride, 160 millimolar concentration.

Undoubtedly various hydrophilic colloids will give different values for the iron number, the gold number, and the rubine number, and within a group of hydrophilic colloids the various values for these numbers will not necessarily follow the same order. Only a few of the factors involved

[17] W. Windisch and V. Bermann, *Wochschr. Brau.*, **37**, 129 (1920).

[18] Wo. Ostwald, *Kolloidchem. Beihefte*, **10**, 234 (1919).

in protective action have been ascertained in a qualitative way, and, as long as different lyophobic micelles possess different electrokinetic potentials and different interfacial tensions, they must possess different adsorptive capacities, and the molecules which are adsorbed are probably not always oriented in a similar manner. Because of these variations, it is probably too much to expect that gold numbers should exactly parallel the iron numbers, or that either one of these should parallel the rubine numbers.

If the electrokinetic potential on the hydrophobic micelle is opposite in sign to the potential on the hydrophilic colloid which is being adsorbed, protection is not conferred by small additions of the hydrophilic colloid, but rather the system is made very unstable, owing to a neutralization of the charge on the lyophobic micelle by the charge carried by the hydrophilic micelle.

The sequence of reactions is (1) increased sensitivity toward electrolytes, (2) neutralization of the potential, with resulting flocculation, and (3) if sufficient of the hydrophilic colloid is added, the reversal of the charge on the lyophobic micelle, and protection.

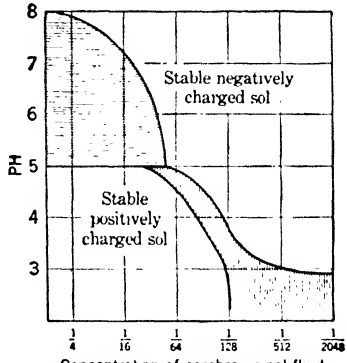


FIG. 56. Showing the areas characteristic of flocculation of gum benzoin sols by cerebrospinal fluid from pathological cases. (Data of Wright and Kermack.)

The gold number of cerebrospinal fluid was introduced into medicine a number of years ago as a diagnostic technic. Since the cerebrospinal fluid of patients suffering from general paresis of syphilitic origin shows a different flocculation area from that characterizing normal cerebrospinal fluid, Wright and Kermack [19] propose the use of gum benzoin sols for the diagnosis of cases of general paresis. Certain of their data are shown in Fig. 56. The area of precipitation shown by the vertical lines is the area characteristic of normal cerebrospinal fluid. The area denoted by the horizontal lines appears to be characteristic of cerebrospinal fluid from cases of general paresis. The constituents which cause flocculation of gum benzoin in this area are abnormal constituents not present in the cerebrospinal fluid of normal persons. Similar technics [20] have been proposed for the diagnosis of cancer.

[19] H. D. Wright and W. O. Kermack, *Biochem. J.*, **17**, 635, 658 (1923).

[20] G. J. Van den Bovenkamp and W. H. Kisjes, *Nederland. Tijdschr. Geneesk.*, **77**, III: 2319 (1933); **78**, II: 2762 (1934).

Mutual Precipitation of Colloids. The preceding paragraphs have indicated that positively charged colloids will cause flocculation when added to systems containing negatively charged micelles, and *vice versa*. This property can often be utilized for the qualitative determination of the sign of the charge on the micelles in a given system. Filter paper (cellulose) is negatively charged against water, and, if a strip of filter paper is dipped into a hydrosol, the liquid will be drawn up by capillary action. If the micelles present in the sol are negatively charged, they will ascend the strip of filter paper, although as a rule at a somewhat slower rate than the water. On the other hand, if the micelles are positively charged, their charge will be neutralized at the point where the filter paper touches the surface of the sol and only the dispersions medium will ascend by capillary action. The reverse behavior will take place if the absorbing column is a positively charged colloid.

This method of determining the sign of the charge on colloid micelles is known as *capillary analysis*, and sometimes it can be advantageously applied to determine whether or not a chemical system contains only one kind of colloidal micelles. Thus, if a purple dye solution is tested, it may be found that the filter paper is colored a uniform purple color, in which case the dyestuff is probably a single chemical compound, or the filter paper may have a zone of red (or blue) advancing ahead of the purple boundary, in which case the purple solution was prepared by mixing red and blue dyestuffs. The purity of a dye can, as a rule, be roughly ascertained by this method, owing to a difference in rate of movement of the different dyestuff molecules along the capillaries of the filter paper.

Mutual precipitation of colloids is of extreme importance in the industry. Dyeing of fabrics is largely dependent on the adsorption of the dye by the fiber. Animal fibers, such as wool, are amphoteric, *i.e.*, the charge on the fiber is rather readily altered so as to be either positive or negative, depending on the hydrogen-ion concentration of the medium. Animal fibers, therefore, can be dyed with either positive or negative dyestuffs by simply shifting the hydrogen-ion concentration of the dye bath.

Cotton possesses a negative charge which is not readily reversed by changes in hydrogen-ion concentration. Substantive dyes color cotton directly. These dyes, derivatives of the benzidine series, possess basic groups, and the micelles carry a positive charge. The attraction between dyestuff and fiber is, therefore, the attraction between positively charged and negatively charged colloids. Mutual precipitation occurs on or in the interstices of the fiber, and the adsorbed dyestuff is firmly fixed on the oppositely charged fiber.

In order to dye cotton or other negatively charged fibers with acid dyestuffs, it is necessary to mordant the fiber. The mordants commonly used are aluminum or chromium salts. The fiber is put through a mordanting bath where Al^{+++} or Cr^{+++} is adsorbed by the negatively charged fiber in such amounts that the original negative charge on the fiber is reversed by the positively charged metallic ions, the mordanted fiber acquiring a strong positive charge which permits the adsorption and mutual precipitation of the negatively charged dyestuff on the surface and in the interstices of the fiber.

Moore [21] utilized the principle of the mutual precipitation of colloids in preparing arsenical insecticides which adhere over long periods to the foliage of plants. Plant tissues in general possess a negative charge. Furthermore, leaf surfaces are covered with an adsorbed water film. Diagram *A*, Fig. 57, shows diagrammatically a positively charged par-

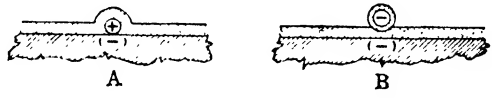


FIG. 57. A diagrammatic representation of the relationships which exist between a leaf surface with its adsorbed water film and a particle of (*A*) a positively charged or (*B*) a negatively charged insecticide.

ticle neutralized by being adsorbed on a negatively charged surface, and the position which the water film would assume. Diagram *B*, Fig. 57, shows the hypothetical position which would be taken by a negatively charged particle sprayed onto a negatively charged surface, with the relative position of the water film. It is easily seen that the particle in *B* would be readily washed off the surface by a stream of water. All the commercially available arsenicals were found to be negatively charged. Positively charged arsenicals were easily prepared by utilizing the principles involved in the complex theory of colloids. Laboratory tests of adherence showed that, if leaves were sprayed with the negatively charged arsenical and, after drying, were placed under a spray of water, practically all the arsenical was removed within a few minutes. However, if leaves were sprayed with the positively charged preparations, and then, while still wet, were immediately placed under a spray, a very considerable amount of the arsenic remained on the leaves after several hours' washing. Hooker [22] found that colloidal copper hydroxide, owing to its positive charge, is an efficient fungicide when sprayed in a concentration of 1 part of the colloid to 5,000 parts of water.

Powdered okra has been used with success in the treatment of gastric

[21] W. Moore, *Ind. Eng. Chem.*, **17**, 465 (1925).
[22] H. D. Hooker, *Ind. Eng. Chem.*, **15**, 1177 (1923).

ulcer.[23] Its effectiveness appears to be due to the phenomenon of mutual precipitation. At the *p*H of gastric digestion (*ca.* *p*H 3) the gastric mucosa is positively charged. Okra contains a plant mucilage which is negatively charged at that *p*H. Apparently therefore the negatively charged carbohydrate slime is precipitated on the surface of the gastric mucosa, covers the ulcer, and prevents further abrasion by the gastric contents, thus relieving pain and affording conditions for rapid healing. Apparently the okra treatment is of no avail in cases of duodenal ulcer, which is in accord with the theory noted above, since at the *p*H of the duodenum the mucosa and the carbohydrate slime would both be negatively charged.

Coacervation. Although the phenomenon now known as coacervation had been observed earlier, it remained for Kruyt [24] and his student, H. G. Bungenberg de Jong, first to describe the fundamental principles underlying the phenomenon. The word is derived from the Latin *acervus* (heap or swarm), combined with the prefix *co* (together), and means, literally, to swarm together. It has been allegorically compared to a swarm of bees which behaves as a unit but in which each bee retains its own individuality.

Coacervation is a phenomenon which involves the lyophilic colloids and particularly the hydrophilic colloids. It has been extensively studied by Bungenberg de Jong [25] and his students. We have already noted that the hydrophilic colloids possess two factors of stability, an electric charge and a hydration shell. Many lyophilic sols may be brought to the isoelectric point and still remain relatively stable, stabilized by the hydration shell. The phenomenon of coacervation is the separation of microscopic liquid droplets when sols of two hydrophilic colloids of opposite electric charge are mixed. After a time these droplets may unite to form a viscous liquid layer at the bottom of the container. They constitute a new phase. Coacervation is a special application of the phenomenon of mutual precipitation. Gelatin and gum acacia sols may be taken as representative systems which form coacervates.[26] Figure 58 shows diagrammatically the conditions which must be met. Gelatin, in common with most proteins, assumes either a positive or a negative charge de-

[23] J. Meyer, E. E. Seidmon, and H. Necheles, *Illinois Med. J.*, p. 339 (Oct., 1933).

[24] H. G. Bungenberg de Jong and H. R. Kruyt, *Proc. Acad. Sci. Amsterdam*, **32**, 849 (1929).

[25] See numerous papers in *Kolloid-Zeitschrift*, *Kolloid-Beihefte*, and *Biochemische Zeitschrift* since 1929. See also H. G. Bungenberg de Jong, Actualités scientifiques et industrielles, Nos. 397 and 398, Hermann et Cie., Paris, 1936; and P. Koets, *J. Phys. Chem.*, **40**, 1191 (1936).

[26] L. W. J. Holleman, H. G. Bungenberg de Jong, and R. S. T. Modderman, *Kolloid-Beihefte*, **39**, 334 (1934).

pending on the hydrogen-ion concentration of the system. At a *p*H greater than 4.7 gelatin is negatively charged. At a *p*H below 4.7 the micelles are positively charged. The reversal of sign of gelatin is shown by curve *A*, Fig. 58. Gum acacia retains its negative charge over a wide range of hydrogen-ion concentration, as indicated by curve *B*, Fig. 58. At the right of the line *xy* both gelatin and gum acacia are negatively charged. The two systems do not interact when mixed, and the relative viscosity of the mixture is the average of the relative viscosities of the sols used (note Fig. 59, system at *p*H 5.06). To the left of line *xy*, Fig. 58, mutual precipitation will occur between the gelatin and gum acacia micelles, but because of the water shell surrounding the individual micelles they cannot coalesce with complete destruction of the double layers but are held apart by the resistance of the water shells. *Electrostatic forces tend*

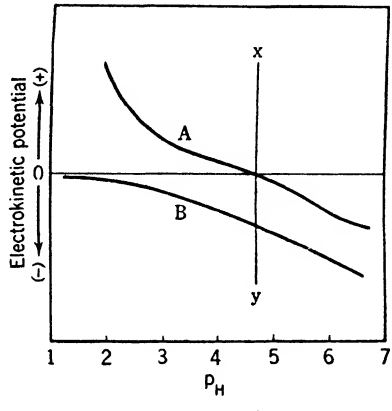


FIG. 58. A diagrammatic representation of the relationship between *p*H and electrokinetic potential for protein (*A*) and non-protein (*B*) sols. Coacervates will form at greater acidities than *p*H 4.7

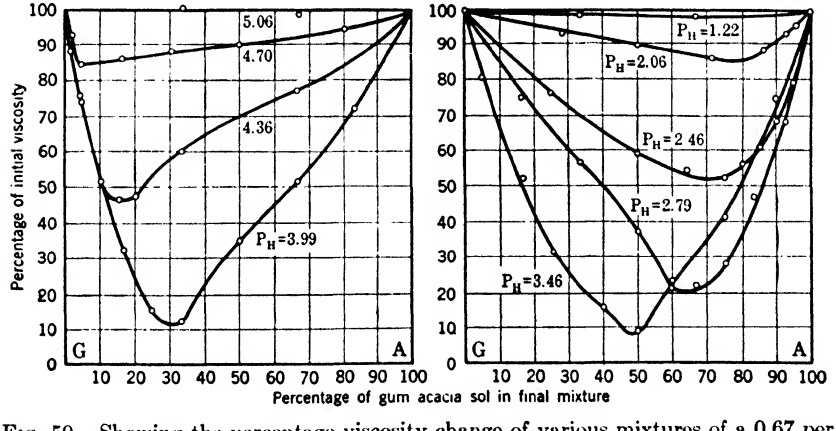


FIG. 59. Showing the percentage viscosity change of various mixtures of a 0.67 per cent gelatin sol (*G*) and a 0.67 per cent gum acacia sol (*A*) at various *p*H values, with particular reference to coacervate formation. (Data of Bungenberg de Jong.)

to cause aggregation, but these forces are resisted by the elasticity of the water shells surrounding the hydrophilic micelles (cf. Fig. 60), so that the

individual charged micelles retain their identity but are held together in a swarm by electrostatic attractions. The net result is that droplets of liquid separate as a new phase.

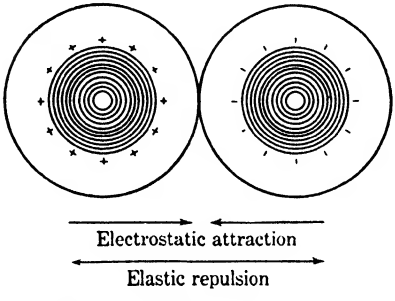


FIG. 60. A diagrammatic representation of electrostatic attraction and elastic repulsion (of water shells) as the factors responsible for the phenomenon of coacervation.

Figure 61 [27] shows a mixture of the coacervate of positively charged gelatin and negatively charged gum acacia together with that of positively charged gelatin and negatively charged nucleic acid. Methyl green colors gelatin-nucleic acid coacervates but does not color gelatin-gum acacia coacervates. Accordingly mixtures of the two complexes can be distinguished as in Fig. 61. Although for any given ratio of the sols essentially isoelectric coacervate droplets may be obtained by the proper adjustment of the hydrogen-ion concentration, the individual micelles of which these droplets are composed retain their characteristic

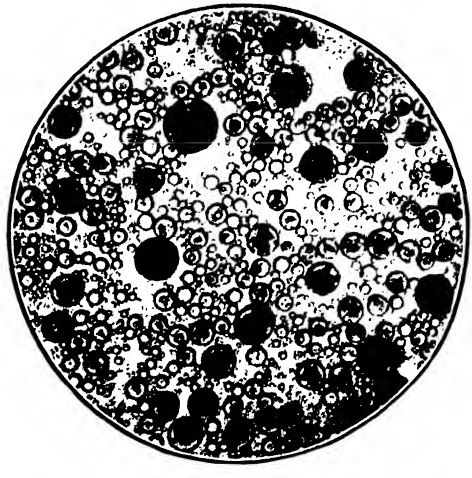


FIG. 61. Showing mixed coacervate droplets formed by mixing gelatin (+), gum acacia (−), and nucleic acid (−) sols. The coacervate of gelatin-nucleic acid stains with methyl green; the coacervate of gelatin-gum acacia does not stain. (Data of Bungenberg de Jong.)

electric charge; consequently a shift in the hydrogen-ion concentration or the addition of electrolytes to the system may cause the coacervate to

[27] H. G. Bungenberg de Jong and A. de Haan, *Biochem. Z.*, **263**, 33 (1933).

redisperse into a sol composed of a mixture of the micelles of the original components of the coacervate.

The real test of whether or not a phase separating from a colloid-containing system is a coacervate depends on the effects that salts may have upon this phase. Since the force of attraction which causes the coacervate to form depends on the presence of electrokinetic potentials of opposite sign on the two reacting colloids, it is evident that the removal of the electrokinetic potential on either colloid destroys the attracting force and likewise the coacervate. On a colloid carrying a negative charge, the zeta potential is most readily reduced by addition of polyvalent cations; a colloid carrying a positive charge is most susceptible to high valent anions. The sensitivity of the potential to salts in either case follows the Schulze-Hardy rule. The destruction of the zeta potential on either component of a coacervate allows the separated phase to redisperse, *i.e.*, the force of mutual attraction disappears. *In a true coacervate the separated phase may be made to redisperse (at constant pH) by adding salts, and the effect will be found to follow a double Schulze-Hardy rule, i.e.*, the effect is produced by *both* cations and anions.

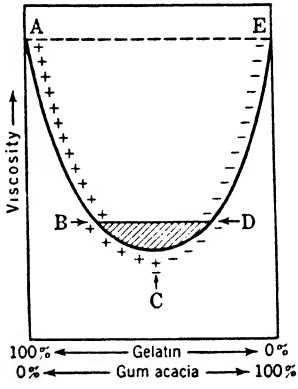


FIG. 62. A diagrammatic representation of the effect of *p*H on the additive viscosity effects and coacervate formation when sols of gelatin and gum acacia are mixed. Dotted line AE = sols mixed at *p*H = 6.0, both sols (−) charged. Solid line = sols mixed at *p*H 3.5, gelatin (+), gum acacia (−). Coacervate droplets separate in area BCD. Coacervate droplets are neutral electrically, *i.e.*, (+) ≅ (−), only at C.

In order for a coacervate to form, it is not necessary that the electrostatic forces of the two components be exactly balanced. Figure 62 illustrates this diagrammatically. As negatively charged gum acacia is added to a positively charged gelatin sol, the viscosity of the mixture progressively decreases until at point B the coacervate begins to separate. Between B and C the coacervate droplets retain a net positive charge, because the zeta potential of the gelatin which they contain is greater than that of the gum acacia component. At C the electrokinetic potentials of the two components are exactly balanced, and the individual droplets are of themselves isoelectric, *although they still contain both positively charged and negatively charged micelles.* From C to D the potential on the gum acacia exceeds the potential of the gelatin, and the coacervate droplets as a whole possess an excess negative potential. In mixtures

represented over the ranges *AB* and *DE* the electrokinetic charges and the kinetic energy of the individual micelles prevent the separation of a coacervate. The cross-hatched area of the diagram is the area within which a coacervate may be expected to separate.

The phenomenon of coacervation appears to be of considerable importance in biological systems. Bungenberg de Jong [28] considers that many of the phenomena of biology can be duplicated by model systems of coacervates. Protoplasm is an aqueous system containing hydrophilic colloids but retains its identity when immersed in water. Its behavior is remarkably paralleled by systems of coacervates. Complex coacervates on aging show the phenomenon of vacuole formation and the extrusion of liquid from the coacervate droplet. Most of the lyophilic biocolloids readily form coacervates either with proteins or with themselves, since many of the biocolloids are zwitterions or contain groups some of which ionize as acids and others as bases (*e.g.*, phosphatides, proteins, nucleic acid). Accordingly, at the proper hydrogen-ion concentration or salt concentration, sols containing such compounds may concentrate into coacervate droplets. In some instances at least these coacervate droplets which separate have been confused with true liquid crystals. When we discuss the nucleoproteins we shall see that they are systems of basic proteins combined with nucleic acid. Here we are possibly dealing with coacervate systems.

The Lyotropic Series. Valency is not the only factor involved in the behavior of ions toward a colloid system. A group of monovalent ions possesses within itself varying degrees of ability to flocculate sols. Hofmeister was the first to investigate this phenomenon. In a series of papers [29, 30] from his laboratory he showed the effects of various anions and cations upon protein systems. Numerous workers have since made similar studies, using both lyophilic and lyophobic systems, and it is generally agreed that the anions can be arranged in a series of citrate > tartarate > SO_4 > acetate > Cl > NO_3 > Br > I > CNS, where, at least for protein systems, citrate shows the greatest precipitating effect, and thiocyanate the least. The cation effects are somewhat less marked but may be expressed approximately as Th > Al > H > Ba > Sr > Ca > K > Na > Li.

[28] H. G. Bungenberg de Jong, *Protoplasma*, **15**, 110 (1932); (with G. C. P. Saubert), *Proc. Acad. Sci. Amsterdam*, **40**, 302 (1937); (with J. Bonner), *Proc. Acad. Sci. Amsterdam*, **38**, 797 (1935); (with G. C. P. Saubert), *Proc. Acad. Sci. Amsterdam*, **40**, 295 (1937).

[29] S. Lewith, *Arch. exptl. Path. Pharmakol.*, **24**, 1 (1888).

[30] F. Hofmeister, *Arch. exptl. Path. Pharmakol.*, **24**, 247 (1888); **25**, 1 (1889); **27**, 395 (1890); **28**, 210 (1891).

These or similar series of ionic effects have come to be known as the lyotropic series, or the Hofmeister, or the irregular series of ions. Various theories have been propounded to account for the observed differences in the behavior of the various salts. The behavior is not limited to colloid systems, since it has been abundantly demonstrated in other connections. Thus, Jaeger [31] found the surface tension of molten alkali salts at 1000°C. to show a series of $F > SO_4 > Cl > Br > NO_3 > I$ and $Li > Na > K > Rb > Cs$, and essentially the same order holds for the effect of these ions in increasing the surface tension of water. Freundlich suggested that in all probability the series correspond to the order of the hydration of the ions, the most hydrated ions being at the sulfate and lithium ends of the lyotropic series. Ionic radius likewise appears to be an important factor.

Frumkin [32] measured the potential difference at an air-solution interface. Table 29 shows certain of his data. The effect of the anion on the

TABLE 29. PHASE BOUNDARY POTENTIALS ϵ BETWEEN AIR AND THE SURFACE LAYER OF CERTAIN INORGANIC SALT SOLUTIONS

(Data of Frumkin)

Electrolyte	Solution $N/1$, ϵ (mv.)	Solution $2\,N$, ϵ (mv.)	Electrolyte	Solution $N/1$, ϵ (mv.)	Solution $2\,N$, ϵ (mv.)
KF		+ 5	Na_2CO_3	+ 3	+ 5
KCl	− 2	− 6	RbCl	.....	− 7
KBr	−10	−16	CsCl	.....	− 6
KI	−34	−52	LiCl	.....	− 9.5
KCNS	−57	−87	$Al_2(SO_4)_3$	0	+ 6
KOH	0	+ 5	HF	−71	−103
K_2CrO_4	....	− 1	HCl	−23	− 55
K_2CO_3	....	+ 3	HBr	−34	− 78
NaI	−39	−55	HI	−61	−112
$NaNO_3$	−17	−23	H_2SO_4	−13.5	− 55(?)
NaCl	− 1	− 4	H_3PO_4	− 0.5	
Na_2SO_4	+ 3				

phase-boundary potential follows the order $F, SO_4 > Cl > Br > NO_3 > I > CNS$, and he suggests that this series is related to the hydration of the anions. A similar view has been expressed by Kruyt and Robinson,[33]

[31] F. M. Jaeger, *Z. anorg. allgem. Chem.*, **101**, 1 (1917).

[32] A. Frumkin, *Kolloid-Z.*, **35**, 340 (1924); *Z. physik. Chem.*, **109**, 34 (1924).

[33] H. R. Kruyt and C. Robinson, *Proc. Koninkl. Akad. Wetenschap. Amsterdam* **29**, 1244 (1926).

who suggest that molecular orientation may well be a factor. They note that the solubility of various materials may be very different in a salt solution from what it is in pure water. In the case of quinone they found a lyotropic series of anions influencing solubility, the solubility of quinone in a 1.5 M solution of KCNS being 170 per cent of its solubility in water, whereas in the case of K_2SO_4 of an equivalent concentration the solubility was only 63.7 per cent of the water solubility. When hydroquinone was used, it was found that the cations showed the more pronounced lyotropic series, ranging from CsCl solutions, where the solubility of the hydroquinone was slightly lower than in pure water, to LiCl, where the solubility was only 56.6 per cent of the water solubility. They suggest that there is an orientation of the dipoles of water in the immediate neighborhood of the ions and that two kinds of orientation are possible. Either the positive end or the negative end of the water molecule may be turned toward the molecules of the solute. Similar orientations of water are postulated to occur on the anions and cations of the electrolytes, and the interaction of these three orientations determines the behavior of a given system.

Büchner [34] and his students, from a study of the effect of ions on a number of systems, measured by a variety of technics, propose an arbitrary "lyotropic number" for the various anions and cations which can be applied to reduce the behavior of the ions to a common denominator. The numbers proposed apparently have no theoretical explanation but are derived empirically from a comparison of the behavior of the ions on experimental systems. For the cations the numbers proposed are Li 115, Na 100, K 75, Rb 69.5, Cs 60, and for anions F 4.8, IO_3 6.25, H_2PO_2 8.3, BrO_3 9.55, Cl 10, NO_2 10.2, ClO_3 10.65, Br 11.3, NO_3 11.6, ClO_4 11.8, I 12.5, and CNS 13.25. A remarkable parallelism is shown between these numbers and the heats of hydration of the ions. For the anion series the relationship is

$$H = 164 - 8N \tag{125}$$

and for the caton series

$$H = 27.25 + 0.73N \tag{126}$$

where H = heat of hydration of the ion
N = the lyotropic number.

Briggs [35] summarized the literature in regard to the effect of ions on electrokinetic potentials and reached the conclusion that the lyotropic

[34] E. H. Büchner, *Kolloid-Z.*, **75**, 1 (1936); (with A. Voet and E. M. Bruins) *Proc. Acad. Sci. Amsterdam*, **35**, 563 (1932); (with C. S. Büchner de Gruiter), *Kolloid-Z.* **76**, 173 (1936). *Cf.* also A. Voet, *Kolloid-Z.*, **78**, 201 (1937).

[35] D. R. Briggs, *J. Phys. Chem.*, **32**, 1646 (1928).

series is dependent not only on ionic mobility and hydration, but also on a number of other factors, only a few of which can be quantitatively evaluated at the present time. Among these factors is the specific capillary activity or adsorbability of ions. Figure 63 shows changes in the electrokinetic potential for a series of chloride solutions against a cellulose membrane, as measured by streaming potential methods. The initial increase in the negative potential of a cellulose surface, when small amounts of the chlorides of Li, Na, K, and Cs are added, is probably due to a greater adsorption of the anion than of the cation. At slightly higher concentrations a relatively greater adsorption of the cation takes place, and the potential is decreased. The divalent ions strontium and barium are more strongly adsorbed throughout the entire range of concentrations than the chloride ion. However, they are not sufficiently adsorbed in the concentrations studied to reduce the potential to zero. The thorium ion was the only ion which reversed the charge on the cellulose interface.

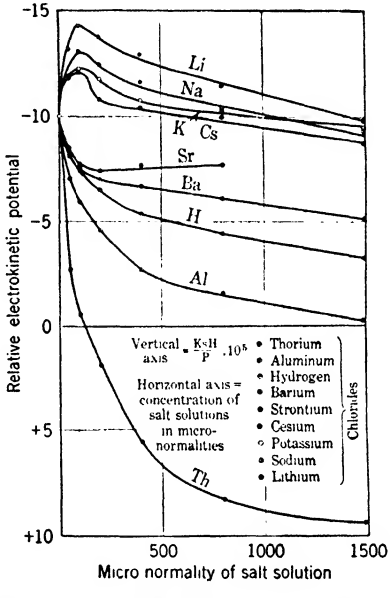


FIG. 63. Showing a lyotropic series of ions determined by measuring the electrokinetic potentials at a cellulose-salt solution interface. (Data of Briggs.)

The behavior of any ion will depend not only on the concentration of that ion but also on the nature of the system and on the type of reaction which is being studied. This explains why one author will give a certain lyotropic series as illustrative of the effect of ions on viscosity, and another worker will give a very similar but somewhat different series for the effect of the same ions in flocculating a sol or in causing peptization. Each worker is securing results which indicate the behavior of ions in the system which he is studying, but such behavior would be slightly altered if other and different factors were introduced.

When organic acids are adsorbed on charcoal [36] a definite lyotropic series can be constructed from the adsorption data. However, the adsorption in this case is apolar,[37] *i.e.*, there is an adsorption of the undis-

[36] E. R. Linner and R. A. Gortner, *J. Phys. Chem.*, **39**, 35 (1935).

[37] V. L. Frampton and R. A. Gortner, *J. Phys. Chem.*, **41**, 567 (1937).

sociated molecule, and the magnitude of the electrical charge at the interface is not altered. Accordingly no lyotropic series is evident from *electrokinetic studies.*

The peptizing behavior of various inorganic salts on wheat flour proteins shows a very pronounced lyotropic series, *e.g.*, 1.0 *N* solution KF peptized on the average 13.07, KCl 22.7, KBr 37.22, and KI 63.89 per cent of the wheat flour proteins.[38] Carpenter [39] found a similar ion series of I > Br > Cl and Li > Cs > Rb > Na for the effect of the alkali halides on the optical rotation of gelatin in the gel state. The cation effect was greater than the anion effect. The bearing of peptization studies on problems of protein chemistry will be discussed later, it being sufficient to point out here that, *within a series of solutions of equal ionic concentration, there will be found very marked differences in the behavior of such solutions toward protein systems, and as a matter of fact toward all biocolloids.*

The removal of an ion from an interface may often be brought about by adding another ion to the system. Seth [40] pointed out that adsorption is the predominant feature in the interaction between the accumulative poisons, such as lead, arsenic, copper, and mercury, and proteins of the body. The heavy metals are selectively adsorbed by the body proteins and tend to alter the nature of the proteins until finally sufficient quantities are accumulated to cause death. We have already indicated that these same ions are the chief metallic poisons of catalytic surfaces.

Carder and Coffindaffer [41] argued that, if mercury were adsorbed by the proteins, it should be possible to replace mercury by using an innocuous electrolyte. Thirty-eight dogs were given a fatal dose of 20 mg. of $HgCl_2$ per kilogram of body weight. Fifteen of these dogs received no further treatment and all died within 1 to 6 days. The 23 remaining dogs received periodically intravenous and intraperitoneal injections of an 0.8 per cent solution of NaCl which were repeated daily until either death or apparent recovery occurred, one month being selected as the time limit, and 19 of the 23 survived.

Colloid Electrolytes. Materials which contain ionogenic groups and which may be dispersed colloidally in solvents which promote ionization, such as water, will act as electrolytes. Examples are proteins at *p*H

[38] R. A. Gortner, W. F. Hoffman, and W. B. Sinclair, Colloid Symposium Monograph, Vol. V, p. 179 (1928); *Kolloid-Z.*, **44**, 97 (1928).

[39] D. C. Carpenter, *J. Phys. Chem.*, **31**, 1873 (1927); (with J. J. Kucera), *J. Phys. Chem.*, **35**, 2619 (1931); (with F. E. Lovelace) *J. Am. Chem. Soc.*, **57**, 2337, 2342 (1935).

[40] Trilok Nath Seth, *Biochem. J.*, **17**, 613 (1923).

[41] J. R. Carder and R. S. Coffindaffer, *J. Am. Med. Assoc.*, **81**, 448 (1923).

values removed from their isoelectric points, many carbohydrates such as pectin and gums, clay soils and permutite, and solutions of soaps or other detergents which are concentrated enough so that a high proportion of the solute molecules are agglomerated to form colloid micelles. Such ionizable colloids are commonly referred to as colloid electrolytes, and in solution they may be considered to consist of a colloid ion, usually multivalent, and an equivalent number of small ions (Na^+, H^+, etc.) designated as *gegen* or counter ions.

The equivalent conductance of colloid electrolytes at varying concentrations usually show a relationship like that illustrated in Fig. 64

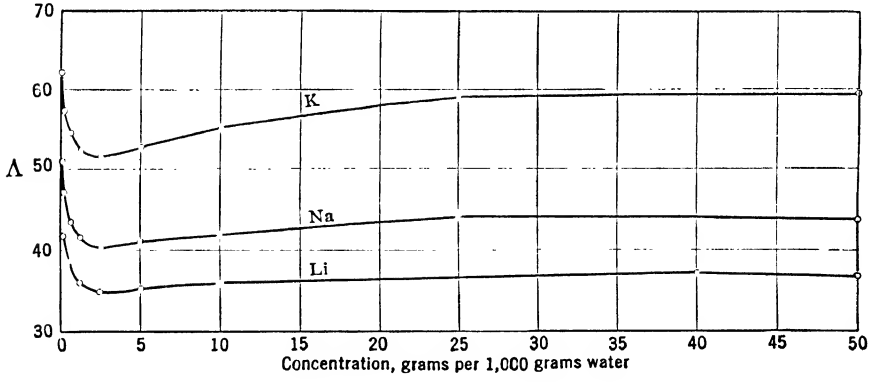


FIG. 64. Equivalent conductivity Λ *vs.* concentration of potassium, sodium, and lithium salts of gum arabic.

for salts of gum arabic, where, as concentration increases, the equivalent conductance falls first to a minimum, then increases to a maximum, and then decreases again. This is in contrast to the continuous decrease in equivalent conductance with concentration increase which is characteristic of ordinary electrolytes (as NaCl, etc.) in solution. The observed maximum for colloid electrolytes possibly can be explained by means of the assumption that surface conductance, *i.e.*, the movement of counter ions in an electric field from point to point along the surface of the colloidal particle, can take place with less energy expenditure than that required for an equal displacement in the bulk of the solvent. Thus the normal decrease in equivalent conductivity due to decrease in ion activity with increase in concentration could be partially nullified or the normal trend even reversed through a part of the concentration range by the increased fraction of total conductance which occurs in the surface layer.

As already noted in the discussion of electrokinetic phenomena, Briggs found that the electrical conductivity of a colloid gel was not necessarily

related to the presence or the concentration of free electrolytes. He found that a membrane of pure cellulose immersed in conductivity water acted as a fairly efficient conductor of an electric current, and was able to demonstrate that the conductance was not due to inorganic constituents present in the system. Thus, if a mixture of cellulose fibers in conductivity water was placed in a conductivity cell, and the fibers were allowed to settle below the level of the electrodes, a specific conductivity

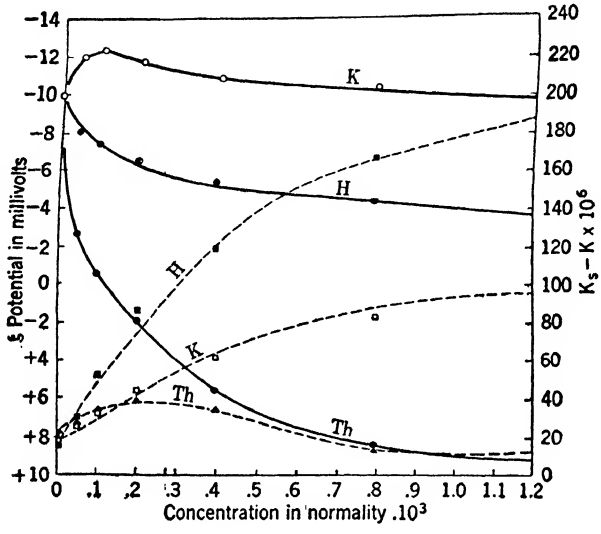


FIG. 65. Showing that surface conductance is not a function of the electrokinetic potential for a cellulose membrane against salt solutions of varying concentrations. (Data of Briggs.)

as low as 4×10^{-6} mho was realized. On shaking the cell to bring the cellulose fibers into suspension between the electrodes, a very marked increase in specific conductivity could be obtained, and, when the fibers again were allowed to settle, the water between the electrodes regained its original conductivity.

Later Briggs [42] discussed the theory of surface conductance, pointing out that Smoluchowski [43] was incorrect in stating that surface conductance was a function of the zeta potential. Briggs found surface conductance to be independent of the zeta potential, as illustrated in Fig. 65, where the solid lines represent zeta potential curves and dotted lines the surface conductance curves on the systems.

[42] D. R. Briggs, Colloid Symposium Monograph, Vol. VI, p. 41, Chemical Catalog Co., New York, 1928.

[43] M. von Smoluchowski, *Physik. Z.*, **6**, 529 (1905).

Certain systems, such as silica gel, show high surface conductance, whereas other systems, such as flowers of sulfur and Al_2O_3, show relatively slight surface conductance.

Bikerman [44] in a theoretical paper points out that the difficulty with Smoluchowski's equation lies in the fact that the thickness of the double layer is assumed to be a constant and that the dielectric constant within the double layer is the same as the dielectric constant of the liquid in bulk. Both these assumptions are unjustified. On the basis of the Gouy diffuse layer, Bikerman concludes that surface conductance is a property of the "ion atmosphere" of the surface. Accordingly for any given colloidal material the amount of electricity which will be conducted along the surfaces will be determined by (*a*) the degree of dispersion, (*b*) the nature of the molecules and ions present in the system, and (*c*) their particular orientation, and probably also (*d*) the affinity of the surface (wettability) for the dispersions medium. There is every indication that the specific conductivity of biological materials is in part due to surface conductance through the lyophilic colloid gel structure, as well as to ionic conductance through the liquid in the cells and intracellular spaces.

Electrical Precipitation. A discussion of the precipitation of suspended particles upon electrified surfaces can be justified at this point because the same basic principles hold for such precipitation as for precipitation by means of electrolytes and the mutual precipitation of colloids.

Cottrell [45] has discussed the general theory which is involved and has devised apparatus for satisfactorily causing the precipitation of suspended particles. He notes that "If we bring a needle point connected to one side of a high potential direct current line opposite to a flat plate connected to the other side of the line we find that the air space between becomes highly charged with electricity of the same sign as the needle point irrespective of whether this is positive or negative, and any insulated body brought into this space instantly receives a charge of the same sign. If this body is free to move, as in the case of a floating particle, it will be attracted to the plate of opposite charge and will move at a rate proportional to its charge and the potential gradient between the point and plate."

In "Cottrell precipitation," a high-tension current is applied to chains or rods hung in the stack, and the particles suspended in the gas, becoming electrified, are later precipitated upon plates having the oppo-

[44] J. J. Bikerman, *Z. physik. Chem.*, Abt. A, **163**, 378 (1933).

[45] F. G. Cottrell, *Ind. Eng. Chem.*, **3**, 542 (1911). *Cf.* also *Ind. Eng. Chem.*, **11**, 147 (1919).

site electrical charge, in the same way as two oppositely charged colloidal particles would attract and neutralize each other. The method has been applied not only to gases from smelters [46] but likewise to the dusts from cement plants.[47] In a single cement plant at Riverside, California, approximately 350,000 tons of dust were collected by means of the Cottrell process during a twelve-year period. In this particular installation the treatment was necessary because the cement dust, settling on the citrus groves, interfered with normal photosynthesis. Similar installation has also been used in the breaking of emulsions, particularly natural emulsions of crude petroleum and salt water.[48,49]

McClendon [50] constructed a laboratory precipitator to avoid the loss of dust when relatively large quantities of grains or other foodstuffs are burned in the process of analyzing them for iodine content. Although the gases of combustion were passed several times through solutions, McClendon observed that solid particles were still present in the gases, and, feeling that iodine might be lost, he passed the gases through a Cottrell precipitator. Even at a high velocity of gas flow, the suspended particles precipitated practically quantitatively. Undoubtedly there are many occasions in the laboratory where similar technic could be employed advantageously. "Precipitrons" built on this principle are now being introduced to eliminate dust and pollen from the atmosphere of home and office with the idea of both decreasing the work required to keep a home clean and eliminating substances from breathed air which are active antigens in various allergies.

[46] At Anaconda, Montana, in a stack 72 ft. in diameter at the base and 572 ft. high, there are four batteries of Cottrell precipitators which require the current from 16 d-c. generators at 62,500 volts. Seventy-six tons of dust with an As_2O_3 content of 42 per cent are collected daily.

[47] W. A. Schmidt, *Ind. Eng. Chem.*, **16**, 1038 (1924).

[48] J. L. Sherrick, *Ind. Eng. Chem.*, **12**, 133 (1920).

[49] B. Speed, *Ind. Eng. Chem.*, **11**, 153 (1919).

[50] J. F. McClendon, *J. Am. Chem. Soc.*, **50**, 1093 (1928).

CHAPTER 9

Gels and the Water Relationships of the Hydrophilic Colloids

A gel may be defined as a colloid system (*i.e.*, a system in which there exists a high degree of interface between the two or more phases present) which, however, possesses many of the properties of a solid system. The most characteristic solid properties of a gel are its rigidity and its elastic response to applied distorting forces. These properties imply the existence of a structure within the system, a continuity of contact or attracting forces between the macromolecules, micelles, or particles of one phase which necessitates that the energy required for disruption of these forces be somewhat greater than the kinetic energy attainable by these components at the temperature at which the gel exists. The second phase present in a gel is liquid, *i.e.*, the dispersions medium of the sol from which the gel may have been formed. The liquid phase in a gel must be regarded also as a continuous phase which fills the interstitial space between and around the network of structural elements. The forces of attraction acting between elements of the structural phase and on which the existence of the gel is dependent may be very weak, in which case the rigidity of the gel is low, the limit of elastic deformation is small, and flow will result upon the application of a small external shearing force, as in the case of a thin jelly; or these forces may be strong, resulting in very rigid structures, such as hair, horn, cellulose fibers, tendon, which can withstand strong forces of distortion without exceeding their elastic limits.

Gels are usually prepared, experimentally, from sols. The process of gel formation, *i.e.*, *gelation* or *gelatinization*, is intimately related to the process of sol destabilization and, therefore, to the factors recognized as governing the stability of a sol, namely, solvation and the electrical energy of the interface. Grossly, gelation differs from flocculation in that gelation results in the solidification of the whole system, disperse phase and dispersions medium, into one apparently homogeneous mass, whereas flocculation shows a visible separation of the two.

Gels may also be prepared without the intermediate of sol formation, or with only a transient sol formation. This may be brought about by

use of many of the methods described earlier which lead to sol formation, such as double decomposition and change of solvent, provided that the conditions are such that the concentration of the phase separating will be high enough to yield a structure capable of incorporating all the liquid phase and that the stabilizing factors are sufficiently ineffective to yield a sol. For example, although dilute solutions of ferric chloride and potassium ferrocyanide upon mixing form a stable Prussian blue sol, concentrated solutions of the same salts will lead to formation of a weak gel. Or, when a molecular solution of dibenzoyl-L-cystine in alcohol is mixed with a large volume of hot water, in which the material is insoluble, a stiff gel will result after a few hours of standing, even when the dibenzoyl-L-cystine content of the gel is very low. The percentage of the structural phase required to form a gel will depend strongly on the degree of asymmetry of the particles of which it is composed. A sol containing only spherical particles of uniform size and unsolvated cannot form a structure at less than about 75 per cent disperse phase. But if the particles formed during phase separation are highly solvated or very asymmetric, very small amounts may form the structure required for gel formation.

Gel Structure. In order to account for the structure of gels in cases where a relatively small amount of the structural phase is present, it is necessary to picture some sort of three-dimensional network which will be characteristic of the solid phase and which will allow for a continuity of this phase. At the same time it is necessary to imagine some limitation to the degree to which the areas of contact can occur, else the solid elements will pull together in such a way as to exclude entirely, or nearly so, the liquid phase from the aggregate. Two general views regarding this structure have been put forward. Bütschli maintained that gels have a more or less honeycomb structure, the disperse phase forming a network of cell-like walls, the interior being filled with the dispersions medium. Nägeli, on the other hand, suggested a brush-heap structure, in which minute ultramicroscopic fibrils are interlaced throughout the system and the dispersions medium is held within the fibrillar mass by capillary forces in the same way as a crystalline mass of tyrosine or caffeine possesses a certain degree of rigidity.

The evidence appears to be entirely in favor of the brush-heap structure. As little as 0.2 per cent of dibenzoyl-L-cystine forms a relatively rigid hydrogel,[1] and even 0.1 per cent forms a soft gel. Ultramicroscopic studies indicate that these gels are of a brush-heap structure, formed by the interlacing of relatively long crystals of dibenzoyl-L-cystine, the

[1] R. A. Gortner and W. F. Hoffman, *J. Am. Chem. Soc.*, **43**, 2199 (1921).

crystals being so thin as to have no apparent cross section. Dibenzoyl-L-cystine is not hydrophilic and does not crystallize with water of crystallization. The extreme minuteness of the crystal fibrils apparently favors an intricate network enmeshing the dispersions medium.

Kraemer [2] used the motion-picture camera as an adjunct to ultramicroscopic studies of gelatin gels and dibenzoyl cystine gel during the gelling process. The gelatin gels were also studied during the melting process. By introducing small mercury particles into the gels, he showed that the resistance of the gel to Brownian movement is not uniform, but that in a dibenzoyl cystine gel the space between the fibrils is occupied by the dispersions medium, which appeared to have essentially the same viscosity as water in bulk. In gelatin gels, channels appeared to exist where the Brownian movement of the mercury particles was slightly impeded, and at other points definite resistance to Brownian movement could be demonstrated. Although no fibers were visible at these points of apparent resistance to the Brownian movement of the mercury particle, it appears as if fibers having approximately the same refractive index as water may well have been present in these areas.

Furthermore, if Bütschli's view were correct, one would expect to find a resistance to the diffusion of a given material or the resistance to the passage of an electrical current through a dilute gel. However, such differences, if they do exist, are extremely slight.

The author has observed many instances in which a plant sap, when expressed from leaves, was, though relatively viscous, definitely not a gel. The specific electrical conductivity of such a system remained constant during and following complete gelation, and in a number of instances the gel structure underwent contraction with the squeezing out of a clear fluid and the separation of a distinct "clot" of organic material, the liquid which was squeezed out of the clot being perfectly limpid and still possessing the same specific conductivity as was possessed by the freshly expressed sap or by the gel. The specific electrical conductivity of blood likewise undergoes practically no change during the clotting process.[3] It is only when a system possesses a relatively high percentage of disperse phase that the specific electrical conductivity decreases even slightly on passing from a sol to a gel.

Many of the earlier papers on gel structure were based on observations of sections of gels which had been studied by the usual microscopic technic, *i.e.*, the gels had been "fixed" and hardened by the usual biological laboratory reagents, after which they were dehydrated, sectioned,

[2] E. O. Kraemer, Colloid Symposium Monograph, Vol. II, p. 57, Chemical Catalog Co., New York, 1925.

[3] S. Gelfan and J. P. Quigley, *Am. J. Physiol.*, **94**, 531 (1930).

and studied under the microscope. The fixing and hardening reagents of the cytological laboratory are in reality reagents which transform the hydrophilic biocolloids into lyophobic systems. Accordingly one of the most important phases of the biocolloid system, namely the water, is eliminated at an early stage in the process, and undoubtedly rearrangements take place between the organic portions when their affinity for water is destroyed. It may well be that the distribution of the organic colloid particles is entirely different in the original gel from what it appears to be in the fixed, dehydrated, and stained sections which are ultimately studied. The honeycomb structures originally observed by Bütschli appear to have been artifacts.

We can accept then, as a working hypothesis, that the structure of a gel is due to a "brush heap" of asymmetrical particles, touching or strongly attracted to each other at widely separated points along the particle surfaces. The intervening areas of the particles of a true gel, however, must have little tendency toward agglomeration but rather a strong attraction for the solvent which results in the tendency of the solvent to remain within the interstices of the structural network. In order for a sol to form a gel, it is evident that, first, the particles of the sol must be more or less asymmetric, and, second, that the particles must be heterophilic in nature. We can understand that, if the particles were lyophilic entirely over their surface, then, even at very high concentration, there would be no opportunity for particle-to-particle contact or attraction, and no structure would develop even though the system might be highly viscous in nature. If, on the other hand, the particles were entirely lyophobic in nature, they would form only a dense flocculum when their stabilizing factors (other than solvation) were destroyed. The intermediate condition that the particles must possess areas of lyophilic and lyophobic character would lead to formation of a structure which, nevertheless, would tend to retain the solvent within itself.

Classification of Gels. Even though individual gels may differ considerably in properties it is possible to make an approximate classification of gels on the basis of either their mode of formation or their reversibility with respect to the removal and addition of their liquid phase. Thus we find for the true or permanent gels:

(1) The group of *temperature-gelating gels*, which are also called swelling or elastic gels because they will change in volume, as they give up or take up their liquid phase, more or less reversibly. Such gels are also reversible with respect to the sol $\rightleftharpoons$ gel transformation, this change being effected by changes in temperature alone. Examples are gelatin and agar gels.

(2) The group of *electrolyte-gelating gels,* which are also called non-swelling or non-elastic gels. These gels will shrink in volume only to a limited extent when their liquid phase is removed. Removal of liquid beyond this point results in replacement of liquid by air in the gel interstices. Replacing the liquid is not accompanied by any appreciable swelling of the gel. These gels, which are formed by addition of electrolyte to a sol, are not reversible, *i.e.*, they are not peptized back to the sol state by the removal of the electrolyte which originally caused the gel to form. Examples are silica gel and aluminum oxide gels.

In addition to the true gels, there exists a group of unstable gels which upon standing for a time revert or break down to form crystals in a supernatant fluid. Such gels are often formed when a new phase is generated in a system in which the rate of growth of crystal or particle is great in one dimension and low in the other two so that fiber crystals result. These fibrils, in forming, result in a network of the separating phase which mechanically enmeshes the liquid phase without actually or necessarily possessing any marked lyophilic attraction for the liquid, as is the case with the true gels. After a variable length of time crystal growth in the other dimensions results in an eventual destruction of the fibrils and of the gel. The time required for this change would be expected to vary inversely with the absolute solubility of the solid phase in the final liquid phase. Such gels may be called *false gels.* Examples are many of the gels formed by change of solvent, such as that of calcium acetate formed when a saturated aqueous solution is mixed with 10 volumes of alcohol, or that of dibenzoyl-L-cystine formed when a small amount in alcohol solution is mixed with several volumes of hot water.

Mechanism of Gel Formation. The electrolyte-gelating gels may be formed from stable sols by the addition of small amounts of electrolytes. The action of electrolytes in this phenomenon follows lyotropic and Schulze-Hardy relationships with respect to the valence of the ion of opposite sign to that carried by the disperse phase of the sol (see Table 30).[4,5] This can be taken to indicate that the gel formation results primarily from a destruction of a stabilizing electrokinetic potential on the sol particles. That a gel is formed instead of a flocculum must be interpreted to mean that the sol was originally stabilized on only limited regions of the particle surface by an electrokinetic potential and that a large part of the sol particle surface was lyophilic and stabilized by its solvation. The area of solvation would not be affected by the electrolyte added and, therefore, the gel particle remains strongly solvated but in contact with other particles at the points originally pre-

[4] J. A. Gann, *Kolloid-Bichefte,* **8,** 64 (1916).

[5] H. R. Kruyt and J. Postma, *Rec. trav. chim.,* **44,** 765 (1925).

vented from coming together by the electrokinetic potential existing in these regions of the surface.

The conversion of a heterostabilized sol to a gel by destruction of the electrokinetic stabilizing factor [6] is illustrated diagrammatically in Fig. 66. An asymmetric particle of the sol is represented at *a*. The regions along the particle which are solvated are represented by the thin lines enclosing these regions of the particle. The regions which are unsolvated and dependent on a charge for stability are represented by a thickening of the heavy line of the particle. The regions of solvation and

TABLE 30. THE EFFECTIVENESS WITH WHICH VARIOUS ELECTROLYTES BRING ABOUT GELATION OF A POSITIVELY CHARGED Al_2O_3 SOL (DATA OF GANN) AND A NEGATIVELY CHARGED SiO_2 SOL (DATA OF KRUYT)

Al_2O_3		SiO_2	
Salt	Conc. in millimoles per liter required to cause gelation	Salt	Conc. in millimoles per liter required to cause gelation
NaCl	77	NaCl	100
KCl	80	$\frac{1}{2}Na_2SO_4$	100
K_2SO_4	0.28	$BaCl_2$	15
$K_3Fe(CN_6)$	0.10		

non-solvation may occur regularly or haphazardly along the surface of the particle. Addition of electrolyte destroys the potential and the stability of these regions as represented at *b*. This is followed by an agglomeration of particles shown at *c* in which the particles of the sol are anchored (flocculated) together in the regions of no solvation but remain apart in the solvated regions. Thus a framework or structure results in which there remains nevertheless a strong affinity for the solvent. The relative extent of these two areas on the sol particles, the relative intensity of the stabilizing factors involved, and the degree of asymmetry of the particles would be the factors determining the actual concentration of disperse phase required for a gel to be formed from a sol by this process or, conversely, the amount of liquid which could be incorporated within a gel so formed. The sol → gel process is not reversible upon removal of the flocculating electrolyte for the same reasons that a true flocculum is not ordinarily reversed to a sol upon removal of the pre-

[6] N. G. Bungenberg de Jong, *Z. physik. Chem.*, **130**, 205 (1927).

cipitating electrolyte. The work which would have to be done on the system in order to bring about redispersion of the gel is greater than that available from its kinetic energy at the temperature of the system.

An entirely analogous picture can be drawn for the temperature-gelating gels, with the difference that in this case the non-lyophilic regions of the particles are not stabilized by a charge but must rather be considered either (1) as regions of lyophobic groups which show a tendency toward mutual solubility in each other, or (2) as regions capable of

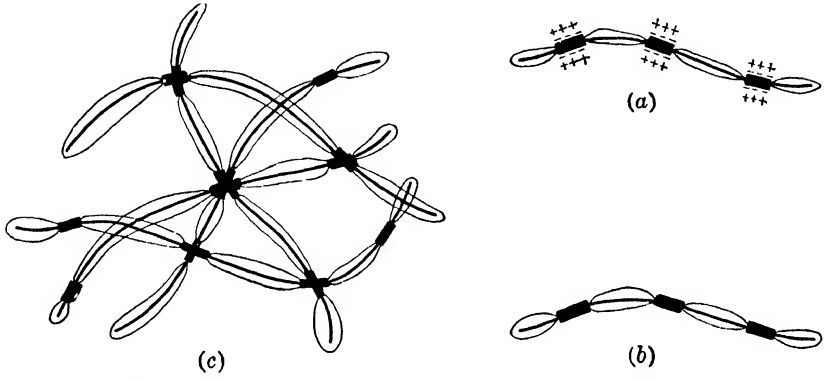


FIG. 66. Diagrammatic representation of the mechanism of gel formation from a hetero-stabilized sol by the action of electrolytes. (*a*) Asymmetric sol particle stabilized in some regions by solvation and in other regions by an electrokinetic potential. (*b*) Addition of electrolyte renders unstable the regions originally stabilized by the electrokinetic potential, so that (*c*) flocculation of sol particles at points of unstabilized surface regions results in gel formation in which the dispersions medium remains strongly attracted to the particle surface and is incorporated within the structural volume of the gel.

weak van der Waals or hydrogen-bridge attractions for each other. The aggregate effect of a large number of such weakly anchoring forces can result in a stiff gel, but individually each such region of attraction must be weak enough so that the kinetic energy of the system above a certain temperature can supply the energy of activation required to break the bonds (change the gel to a sol). Thus, below a temperature at which the bond is stable a gel is formed, the process being reversed at some higher temperature.

That the individual bonds of attraction functioning in a temperature-gelating gel are weak is indicated by experiments of Freundlich and Abramson,[7] who found the electrophoretic mobility of microscopically visible particles, such as quartz or blood corpuscles, to be the same in a

[7] H. Freundlich and H. A. Abramson, *Z. physik. Chem.*, **128**, 25 (1927); **133**, 51 (1928).

gelatin gel as in the sol before gelation. Thus the gel is thixotropic in small dimensions where it does not appear to be so in bulk dimensions. This observation may serve as a clue in explaining another broad difference between temperature-gelating and electrolyte-gelating gels, *i.e.*, the former are swelling gels and the latter are not. In the swelling gels as the liquid phase is removed by evaporation, for example, the body of the gel changes in volume in such a way that the gel components move into and fill up all cavities resulting from liquid removal, and this process continues until all fluid is removed and only the horny, compact structural phase remains. A non-swelling gel, on the other hand, may follow a similar course of volume change as the first part of the liquid is removed, but it soon reaches a volume where further removal of liquid results in a replacement by air of the volume lost and the volume of the gel structure no longer changes with liquid removal. With the swelling gels the reverse process, *i.e.*, that of supplying liquid again to the dried gel, results in an increase in volume of the gel where the final volume may approximate that of the original gel at the time of formation. The non-swelling gel, in contrast, having once been dried to its minimum volume, will take up only the amount of liquid required to fill its cavities and will not swell much beyond this minimum volume.

It would appear that the individual anchoring forces functioning in the non-swelling gels are so strong as to resist tensions caused by lowering the liquid content beyond a point where the distortion of the relatively inflexible particles reaches a limit, and that adsorption or osmotic forces are not capable of overcoming the anchoring forces to any marked extent as the liquid is again made available to the system. In the swelling gels, on the other hand, a sufficient number of individual linkages may be broken, under the tension developed upon liquid removal or under the forces of adsorption and osmotic uptake of liquid, to allow for a more nearly reversible rearrangement within the structural phase. The structural elements of a swelling gel must also be considered as probably of a more flexible nature than non-swelling gels.

Hysteresis and Syneresis in Gels. The phenomenon of hysteresis in gels would place a limitation upon the extent to which the above explanation of the difference between swelling and non-swelling gels is valid. *Hysteresis may be defined as the effect of previous treatment, thermal, mechanical, etc., upon the present properties of a system.* Such effects are very prominent in gel systems. An example is supplied in van Bemmelen's [8] classical study on the water content-relative vapor pressure relationships in hydrous silica gel. He showed that the amount of water retained by a silica gel is dependent on the past history of the gel. He

[8] J. M. van Bemmelen, *Die Adsorption*, T. Steinkopf, Dresden, 1910.

found that the dehydration curve of silica gel is a continuous process but was not an equilibrium. A silica gel which has been dehydrated to a certain point will again take up water if placed in a suitable environment, but the rehydration curve will not follow the dehydration path. Figure 67 shows the familiar dehydration-hydration curves of silicic acid gels as found by van Bemmelen. The arrows pointing downward indicate the dehydrating processes, the arrows pointing upward indicate the hydrating processes.

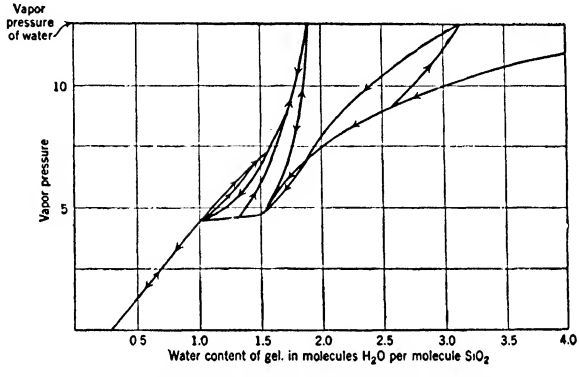


FIG. 67. Hydration and dehydration curves of silica gel. (Data of van Bemmelen.)

The older a given gel may be (after gel formation) and the greater the extent to which it may have been desolvated, the lower will be its tendency to take up the solvent when this becomes again available to the gel. The tendency of a gel structure to pull more closely together with aging is demonstrated by the phenomenon of *syneresis*, the squeezing out of the liquid phase from the gel structure, which is characteristic of most gels. This process is variously explained [9] as due (1) to the slow increase in the number or strength, or both, of the bonds existing between the structural elements of the gel (progressive flocculation), or (2) to a slow decrease in the attraction of the structural elements for the solvent (progressive desolvation). These characteristics of a gel denoted as hysteresis and syneresis, progressive changes of many types with time, emphasize the fact that only in the exceptional case can we consider a gel system as having attained a complete and true equilibrium state.

Solvation or Imbibition in Gels. The property of incorporating relatively large amounts of liquid within their structure, their tendency to retain a portion of this liquid against strong desolvating forces, and their ability, when desolvated, to take up the liquid again with strong attracting forces are the most characteristic and important properties

[9] W. Heller, *Compt. rend.*, **204**, 43 (1937).

of gels. The process of taking up water by strongly dried hydrogels is accompanied by the evolution of heat and will be accomplished against very high pressures. A method used by the Egyptians to quarry stone consisted in drilling a series of holes on the face of the stone along the line of desired fracture, pounding dried wooden pegs into the holes, and then keeping the pegs moistened. Water is taken up with such force as to cause the stone to be broken apart by the pressure generated as the wood (a gel) swells.

An aluminum oxide gel which has been heated to 700°–750° is one of the most efficient drying agents known and is used in organic combustions in place of calcium chloride or P_2O_5. Johnson [10] states that "1 gram of Al_2O_3 can practically completely absorb all moisture from approximately 10 liters of air saturated with water vapor at 18°." Aluminum oxide which is perfectly anhydrous will not act as an efficient adsorbent for water, for the adsorbing properties depend on an alumina gel from which not all the original water has been removed.

The process of liquid uptake by gels is variously denoted by the terms solvation and imbibition or, in the case of biocolloids where water is the liquid phase taken up, by the term hydration.

Methods for Measuring Imbibition. Three general methods have been used for the measurement of imbibition changes by hydrophilic colloids: (1) the volume swelling of granules of gel when placed in water, (2) the increase in weight of gel when placed in water or in atmospheres of varying humidity, and (3) in cases where the gel particles are particulate within a large volume of the liquid, by viscosity measurements (see p. 40).

The method for increase in volume was used by Martin Fischer in his studies of the hydration of fibrin and other animal proteins. A given amount of the protein is placed in test tubes of a uniform bore, and the various solutions whose effect on imbibition it is desired to study are added. At varying intervals of time the height of the swollen column of protein is measured, and the amount of imbibition which has taken place is calculated from the height of the swollen column.

Hofmeister was probably the first to use the increase-in-weight method. A sheet of gelatin or some similar lyophilic colloid is placed in the liquid to be studied and from time to time is removed and weighed. The increase in weight plotted against time will give an imbibition curve. This method has been used in studying the quality of wheat-flour proteins.[11, 12]

[10] F. M. G. Johnson, *J. Am. Chem. Soc.*, **34**, 911 (1912).

[11] F. W. Upson and J. W. Calvin, *J. Am. Chem. Soc.*, **37**, 1295 (1915).

[12] R. A. Gortner and E. H. Doherty, *J. Agr. Research*, **13**, 389 (1918).

Imbibition Pressure. Many organic colloids show a great affinity for certain liquids and take up such liquids against relatively enormous pressures. The pressure against which such a colloid will imbibe a liquid or, conversely, the pressure which is required to force the dispersions medium out of a gel, is known as the imbibition pressure. Imbibition pressures should not be confused with osmotic pressure. In many instances the magnitude of imbibition pressure greatly exceeds that of osmotic pressure. For example, seeds of *Xanthium glabratum* containing 8 to 9 per cent of water will, through imbibition pressure, withdraw water from a saturated solution of lithium chloride which has an osmotic pressure of approximately 965 atmospheres,[13] although the salt content of the seed is sufficient to account for only a few atmospheres of osmotic pressure. Similarly, if a sheet of dried gelatin is placed in a saturated solution of sodium chloride, water will be withdrawn by imbibition forces against the osmotic pressure of the sodium chloride solution, and sodium chloride will crystallize out in the solution. The sap of plants in salt marshes or alkali flats may reach an osmotic pressure as high as 172 atmospheres [14] (14.4° depression of freezing point), but such plants are not characteristic of the more extreme xerophytes, such as the cacti, where the osmotic pressure plays a minor role and imbibition pressure becomes all-important.

Newton [15] and his co-workers, in July, 1925, removed some stems from the Alberta *Opuntia*, sealed the cut surfaces with grafting wax, and then placed the segments in a desiccator over concentrated sulfuric acid. At the end of six months in an atmosphere of almost zero humidity, the stems had lost less than 10 per cent of the water which they originally contained. At the end of 94 days of desiccation, one of the stems was removed from over the sulfuric acid and placed in another desiccator over a free water surface. Figure 68 shows this stem 54 days after it had been placed in the atmosphere of higher humidity. Two new shoots were already well developed. A change in the relative humidity of the air was sufficient to produce a growth response. The ability of the *Opuntia* to resume growth almost as soon as favorable conditions appeared is abundant evidence that the apparently drastic desiccation over sulfuric acid was of no real significance so far as the vital activities of the cactus were concerned but that the imbibitional forces with which the colloids of the stem retained and imbibed water constitute a vital factor in the ability of the plant to grow under extremely adverse conditions.

[13] C. A. Shull, *Botan. Gaz.*, **56**, 169 (1913); *Ecology*, **5**, 230 (1924).

[14] J. A. Harris, R. A. Gortner, W. F. Hoffman, and A. T. Valentine, *Proc. Soc. Exptl. Biol. Med.*, **18**, 106 (1921).

[15] R. Newton and W. McK. Martin, *Canadian J. Research*, **3**, 336 (1930).

Similar examples of the force with which water is taken up or retained by organic colloids are numerous. Reinke,[16] in 1879, measured the swelling pressure of dried disks of the sea algae, *Laminaria*, against water. He placed disks of the dried *Laminaria* in a hollow metal cylinder and above the disks fitted a metallic piston carrying a platform upon which weights could be placed. The disks of dried algae were then allowed to come in contact with water, and the amount of swelling against known weights was measured by the movements of the piston. When no further swelling took place, a part of the weights could be removed, and swelling again set in, reaching a new equilibrium determined by the weight of the piston.

FIG. 68. *Opuntia* sp. removed from desiccator after 94 days over concentrated H_2SO_4 and placed in another desiccator over pure water. Growth was observed within two weeks of transference. Photograph taken 54 days after transference. (Photograph, courtesy of Dr. Robert Newton.)

That imbibition pressures are not limited to hydrophilic colloids is evidenced by the work of Posnjak,[17] who studied the swelling pressures of rubber in organic liquids and of gelatin in water. He found the same general phenomena in both types of systems.

MacDougal [18] refined the methods of measuring the swelling of biocolloids. The general arrangement of the apparatus is shown in Fig. 69. By means of this instrument MacDougal and his co-workers studied various problems involving the swelling of plant colloids, and in his monograph he attempted to correlate the observations with the growth processes.

Various attempts have been made to express in mathematical terms the forces of imbibition. Perhaps the most successful of these have been

[16] J. Reinke, *Hanstein's botanische Abhandl.*, **4**, 1 (1879).

[17] E. Posnjak, *Kolloidchem. Beihefte*, **3**, 417 (1912).

[18] D. T. MacDougal, *Hydration and Growth*, Carnegie Institution of Washington Publication 297, Washington, 1920.

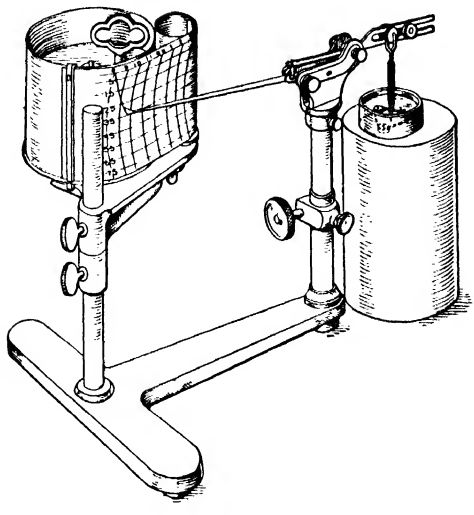


FIG. 69. MacDougal's auxograph arranged for recording changes in thickness, due to swelling, of trio of sections of plant tissues and of biocolloids. The vertical arm, which is set in position on the horizontal arm to give an amplification of 20, rests on a triangle of glass laid on top of the sections. The dish containing the sections rests on an iron cylinder to secure stability, and a weight is placed on the T base of the instrument. The record sheet is ruled to millimeters (not shown) with heavier horizontal lines 1 cm. apart. The curved lines represent hour intervals. The space is ruled to 15-minute intervals (not shown). Height of clock and lever supports adjustable.

the equations proposed by Wo. Ostwald [19] and Dorothy Jordan Lloyd.[20] Ostwald's equation is

$$P = \left(\frac{RT}{M}\right) C + KC^n \qquad (127)$$

where P = the swelling pressure
C = the concentration of the sol
M = particle weight
K and n = constants.

The first part of equation (127) is essentially the true van't Hoff osmotic pressure which increases in direct proportion to the concentration and the absolute temperature. The second portion of the equation is an adsorption isotherm, and it indicates that the swelling pressure increases disproportionately with an increase in concentration. The con-

[19] Wo. Ostwald, *Kolloid-Z.*, **24**, 7 (1919); **49**, 60 (1929); *Z. physik. Chem., Abt. A*, **159**, 375 (1932).

[20] Dorothy Jordan Lloyd, *Biochem. J.*, **24**, 1460 (1930); **25**, 1580 (1931).

stants K and n vary greatly for different colloid systems. A graph of this equation as compared with ordinary osmotic pressure effects is shown in Fig. 70.

Miss Lloyd's equation is

$$S = K_1\left(\frac{1}{\log C}\right) + K_2 \qquad (128)$$

where S = the swelling
C = the original concentration of the gel
K_1 and K_2 = constants, K_2 equaling some function of K_1.

She found this equation to express rather exactly the swelling of gelatin gels where the original concentration of the gelatin did not exceed 20 per cent. For an individual gelatin gel in salt solutions she found that swelling was independent of the volume of the salt solution but was directly proportional to the logarithm of the concentration of the salt.

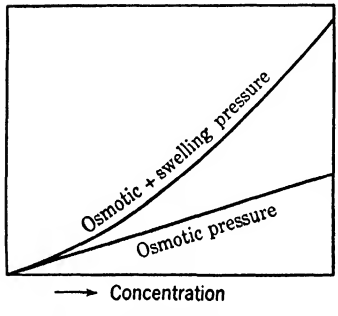


FIG. 70. A diagrammatic representation of the effect of concentration of lyophilic sols on osmotic pressure and on swelling pressure. (After Ostwald.)

$$S = a \log C_1 \qquad (129)$$

where C_1 = the concentration of the salt
a = a constant depending on the nature of the salt.

As long as the swelling process does not proceed so far as to induce solution, a is independent of temperature. She lists three mechanisms as influencing swelling: (1) true imbibition which she defines as the taking up of the dispersions medium until the gel reaches a preformed structure, (2) osmotic swelling due to unequal solvent pressure in the solution and in the gel, and (3) swelling of hydration characteristic of salt solutions and of acid or alkaline solutions some distance removed from the isoelectric point of the colloid.

A somewhat converse and perhaps more satisfactory method of representing the imbibition properties of a gel is that in which the amount of solvent (*e.g.*, water) which is taken up by unit weight of dry gel is followed in relationship to the activity coefficient of the solvent in temporary equilibrium with the gel. For example, if samples of a dry isoelectric casein gel [21] are allowed to equilibrate against atmospheres of

[21] D. R. Briggs, *J. Phys. Chem.*, **35**, 2914 (1931).

varying relative vapor pressure of water at a given temperature, a curve is obtained such as that shown in Fig. 71, when the water taken up by unit weight of gel is plotted against the relative vapor pressure of the water. Such S-shaped curves are typical for most biocolloids. It is readily apparent that such a curve is identical in form to the adsorption isotherm described by the Brunauer, Emmett, and Teller [22] equation for polylayer adsorption (see p. 181).

It will be remembered that in the B.E.T. isotherm the region of the curve *OA* (Fig. 71) describes the region of adsorbate binding correspond-

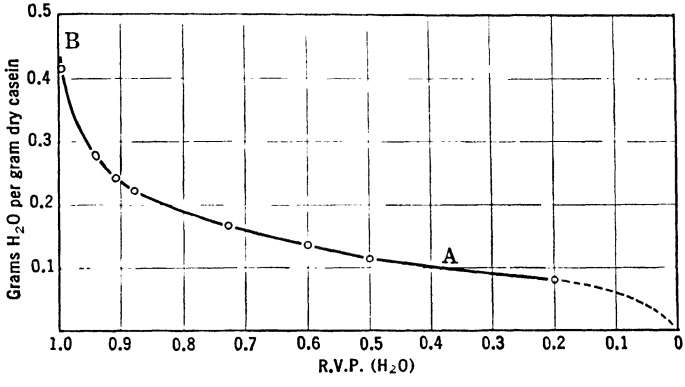


FIG. 71. Water activity (R.V.P.)—water content curve for isoelectric electrodialyzed casein.

ing to the formation of a monolayer of adsorbate molecules on the available surface of the solid. In this region, 0*A*, most of the heat of adsorption is evolved, and the adsorbate is bound by the solid against very strong forces, *i.e.*, the bound adsorbate is brought to a low degree of activity (low escaping tendency). In this region of the imbibition curve of a gel, the volume of the solid plus the volume of adsorbed solvent is less than the sum of their volumes when uncombined, indicating a strong orientation and accompanying volume decrease in the solvent as a result of the adsorption forces acting upon its molecules. In the instance of a gel imbibing water it is in this region of imbibition that the heat of imbibition is positive and high, the imbibition pressure is high, the strength of binding of the solvent is high. Only in this region of imbibition will swelling forces of great magnitude be encountered.

The region *AB* is depicted in the B.E.T. adsorption equation as a region in which more adsorbate is bound by the adsorbent in the form of additional layers adsorbed on the surface of the strongly bound primary

[22] S. Brunauer, P. H. Emmett, and E. Teller, *J. Am. Chem. Soc.*, **60**, 309 (1938).

layer of adsorbate molecules. Relatively low heats of adsorption occur in this region of the isotherm. Whereas relatively much more adsorbate may be taken up as the equilibrium relative vapor pressure of the adsorbate approaches a value of 1, the strength of binding is of a much lower order of magnitude than is characteristic of that in region $0A$. In gels, relatively small applied pressures can cause the removal of this fraction of bound solvent. It may well be questioned that the solvent imbibed in region AB is in truth adsorbed solvent. A large fraction of it may be attracted into the region of the gel by osmotic forces which the components of the gel may exert upon the solvent. The part of the curve AB does approximate, in extent and contour, the relative vapor pressure-concentration relationships of dilute solutions. The region of the curve $0A$ is sometimes [23] spoken of as representing the solution of solvent (liquid) in the solid phase of the gel, and region AB represents a solution of the primarily lyated solid in the solvent (liquid).

On the basis of such an interpretation of the factors which function in the imbibition process we might expect that *osmotic factors* in gels, such as (1) gegen ions, (2) relatively small but non-diffusible molecules, and (3) the absorption of kinetic energy by vibrational or rotational degrees of freedom of the long chains existing between points of adhesion in the gel structural elements themselves,[24,25] would modify the imbibition curve in the region AB to marked degrees while exerting little or no influence upon the region of true adsorption $0A$. The amount of solvent fixed in region $0A$, on the other hand, would be determined primarily by the degree of surface available in the gel upon which adsorption could occur and by the specific attraction which these surfaces may exhibit toward the solvent molecules. Effects of hydrogen-ion concentration changes and the lyotropic effects of salts upon gel (such as protein) imbibition are indeed restricted primarily to effects upon the extent to which water is bound in region AB.[21] Changes in the chemical nature of the gel and in its degree of dispersion (specific surface) will change the course of the curve in region $0A$.

The Aging of Gels. The *time* factor affects all colloid systems but is particularly noticeable in gels. The micelles may aggregate into larger units, or crystal growth may be a factor, and the larger particles may grow at the expense of the smaller ones, owing to an unequal distribution of surface energy forces. With the biocolloids there is a decrease in imbibition capacity with age. Probably the changes involved in senescence are in a large measure changes characteristic of aged colloid systems.

[23] P. H. Hermans, *Contributions to Physics of Cellulose Fibers*, pp. 11, 188, Elsevier, New York, 1946.

[24] M. L. Huggins, *J. Phys. Chem.*, **46**, 151 (1942).

[25] P. J. Flory, *J. Chem. Phys.*, **10**, 51 (1942).

The older leaves on a tree are characteristically more lignified and contain a lower content of colloids in the sap which can be expressed than the leaves which are just unfolding. Old tissues, both of plants and of animals, are in general less highly hydrated than the younger tissues. The walls of the blood vessels of older animals contain a higher proportion of dry matter, a higher proportion of inorganic constituents, notably calcium, and a much lower proportion of water than walls of the blood vessels of young animals.

The body tissues can be looked upon as examples of colloid gels, the behavior of which may be expected to be intimately related to the gel structure and the water content. In the hardening of the arteries (arteriosclerosis) in old age, we have a striking example of the loss of flexibility of the arterial wall which is associated with a higher content of dry matter and a reduced imbibition capacity. Thoenes [26] has likewise shown that the muscle tissues of dogs and guinea pigs have a progressively lowered imbibitional capacity as the animals become older. If the problems of rejuvenescence are ever solved, they will be solved very largely through colloid chemical studies designed to bring about an increased, or to maintain a high, imbibition capacity of the tissue colloids.

The Role of Water in the Living Organism. Inasmuch as living organisms are composed very largely of water, a study of the gels which comprise the living organism is extremely important.

Table 31 gives the approximate elementary composition of the human body, as well as the relative proportion of water, protein, fat, salts, and

TABLE 31. COMPOSITION OF THE HUMAN BODY

Elementary Composition, per cent		*Group Composition, per cent*	
Oxygen	66.0	Water	65.0
Carbon	17.5	Protein	15.0
Hydrogen	10.2	Fat	14.0
Nitrogen	2.4	Salts	5.0
Calcium	1.6	Other organic compounds	1.0
Phosphorus	0.9		
Potassium	0.4		
Sodium	0.3		
Chlorine	0.3		
Sulfur	0.2		
Magnesium	0.05		
Iron	0.004		
Iodine, Fluorine, Silicon, Manganese, Arsenic, etc.	Present		

[26] F. Thoenes, *Biochem. Z.*, **157**, 174 (1925).

other organic compounds. It will be noted that the greater percentage of the human body is composed of water. This is probably true of every living organism. In some organisms, as the jellyfish, only an insignificant fraction is composed of organic material, as little as 1 per cent of the jellyfish being organic matter. Figure 72 illustrates what a small proportion of dry matter enters into the composition of a jellyfish. In this figure the "bell" of a jellyfish has been dried down upon the page of a

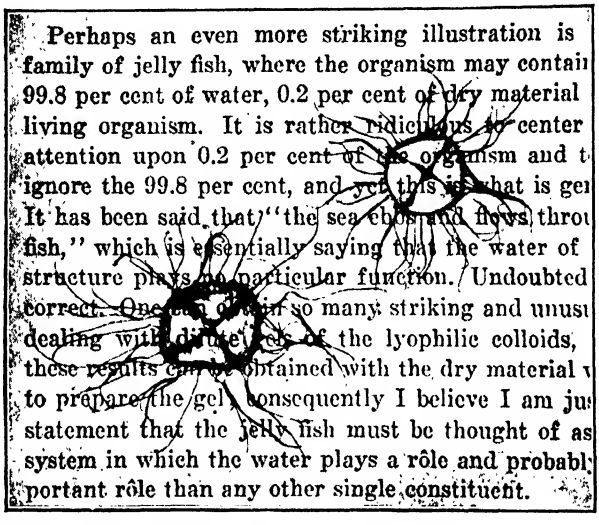

Perhaps an even more striking illustration is
family of jelly fish, where the organism may contai
99.8 per cent of water, 0.2 per cent of dry material
living organism. It is rather ridiculous to center
attention upon 0.2 per cent of the organism and t
ignore the 99.8 per cent, and yet this is what is ge
It has been said that "the sea ebbs and flows throu
fish," which is essentially saying that the water of
structure plays no particular function. Undoubted
correct. One can obtain so many striking and unusu
dealing with dilute gels of the lyophilic colloids,
these results cannot be obtained with the dry material
to prepare the gel, consequently I believe I am ju
statement that the jelly fish must be thought of as
system in which the water plays a rôle and probabl
portant rôle than any other single constituent.

FIG. 72. Showing adult medusae (*Gonionemus* sp.) dried down upon a magazine page. It will be noted that the amount of dry material in the umbrella is so slight that the print can easily be read through the dried organism. The dark cross bars underneath the umbrella are the radial canals to which are attached the reproductive organs, which in these specimens are filled with ripe ova.

magazine, and the print can be rather readily seen through the dry material of what was formerly a living organism. It has been said that "the sea ebbs and flows through a jellyfish," which is essentially saying that the water of a jellyfish structure plays no particular function. Undoubtedly this is incorrect. The many striking and unusual results which one obtains in dealing with dilute gels of the lyophilic colloids cannot be duplicated when one works with the dry materials, and the jellyfish must be thought of as a biological system in which water plays a role and probably a more important role than any other single constituent.

In most organisms a very considerable part of the growth process is concerned with the building of water into the living structure. For example, a frog's egg, weighing on the dry basis only a few milligrams, can,

after fertilization, be placed in a dish of filtered, sterile water and allowed to undergo the process of development. Such an egg will undergo cell division, giving rise at the end of several weeks to a living tadpole which may be as much as 2 cm. in length and weigh several grams. Such a tadpole has never partaken of food other than utilizing the nutrients already present in the original egg and will be found on analysis to contain less dry matter than the original egg, owing to the fact that a certain amount of the organic materials present has been utilized as a source of energy and eliminated as carbon dioxide and water. The increase in weight during these several weeks has all been due to the intake of liquid water which has become "living water," so the tadpole is actually more than 95 per cent water. It would be ridiculous to speak of this organism as being composed of only 5 per cent of vital materials. The water is as much a part of the tadpole as the fats, proteins, etc., which serve to form the gel structure, and the biochemical and biophysical reactions which take place within the cells and tissues of the tadpole are determined probably more by the water which is present than by any or all of the other constituents.

Mammalian embryos during development show similar water relationships. Table 32 [27] shows the hydration capacity of pig embryos at

TABLE 32. HYDRATION CAPACITY OF THE EMBRYONIC TISSUE OF THE PIG AT VARIOUS STAGES DURING EMBRYONIC GROWTH

Length of Embryo, mm.	*Average Weight, grams*	*Water Content, per cent*	*Hydration of Tissues, per cent*
2–4	. . . . .	97.40	3,750
6–7	0.313	94.07	1,640
15	0.93	91.38	1,062
30	2.21	91.14	1,030
100	72.20	91.18	1,033
160	238.57	91.71	1,106
240	725.0	88.70	784

various stages during embryonic growth. In the 15-day-old embryos (2–4 mm.) there is a 3,750 per cent hydration of the tissue colloids. At the 6–7 mm. stage this has dropped to 1,640 per cent. These embryos, each containing only 0.018 gram of dry matter, are nevertheless fairly well-differentiated biological organisms and actively respond by move-

[27] V. A. Wilkerson and R. A. Gortner, *Am. J. Physiol.*, **102,** 153 (1932); *cf.* also M. Rubner, *Biochem. Z.*, **148,** 187 (1924), who has critically discussed various factors regarding the interrelationships of water and vital elements in physiological processes and growth.

ment to tactile stimuli. Shortly before the 15-mm. stage the hydration capacity falls to a new level of approximately 1,000 per cent and remains at this new level until shortly after reaching the 160-mm. stage when it again falls to a lower level which persists until near birth. After birth a still lower level is attained.

The constituents that go to make up a living organism may be classified into five great groups, (1) proteins, (2) carbohydrates, (3) fats and lipids, (4) the inorganic salts, and (5) water. There are, to be sure, a number of compounds which do not fall definitely within one or another of these groups, but the actual amounts of such compounds are extremely small.

Of these various groups, the proteins belong to the class of lyophilic colloids. They possess the power of becoming strongly hydrated, and the large amount of water present in living organisms is probably mainly held through this affinity of the proteins for water.

As far as cell protoplasm is concerned, carbohydrates can be largely looked upon as a source of energy, the carbohydrate content of true protoplasm being relatively low. In the plant kingdom, however, the structural elements and intracellular constituents are largely of carbohydrate nature, polysaccharides in most instances. Many of these polysaccharides are highly hydrophilic and combine with relatively large amounts of water.

The fats and lipids of living organisms exist in the form of emulsions. In the study of such emulsions, attention has been chiefly directed toward the fats, relatively little attention having been devoted to the aqueous phase. Adipose tissue usually contains as much, if not more, water than it does fat, and in some instances it is possible to break down adipose tissue by limiting the intake of water or by the use of large doses of drastic diuretics. Certain of the phospholipids, such as lecithin, are in themselves extremely hydrophilic and exist in the colloidal state and react as typical lyophilic colloids.

The salts and other true solutes present in the biological organism may be, to a very considerable extent, molecularly dispersed in the water and, on the other hand, may exist in a more or less fixed condition by being adsorbed upon the surface of the biocolloids. If they are adsorbed, they must be regarded as part of the colloid system, the reaction of the surface upon which they are adsorbed being modified accordingly.

State of Water in Biocolloids. Many phenomena observed in living systems indicate that not all the water present can function in the capacities which would be expected of ordinary water. For example, a part of the water molecules present appears to be unable to act as solvent or medium in which reactions can take place. Seeds containing perhaps

50 per cent of their weight as removable water will remain dormant, but if the water content is increased by half again as much the seed will start to germinate and respiration will increase many fold. Certain fungus spores will remain dormant when placed on a substrate containing water sufficient to give a relative vapor pressure of 0.75 but will start to grow if the relative vapor pressure is raised to 0.80.[28,29] If added water contains electrolyte or sugar in sufficient concentration to retain the relative vapor pressure at 0.75, this addition of water in any amount still will not serve to allow the mold to start growing. Bacteria show similar relationships, in general requiring that the activity coefficient of the water be somewhat higher (0.90–0.95) than that required by molds before growth can take place.

Such observations have led to the idea that the water in an organism (or in any colloid system) may be partially or entirely "bound" water, with the obvious connotation that it is not "free" to act as water is required to act in order to promote those processes which are dependent on the presence of "free" water. In the above cases, perhaps water has to be "free" enough so it can act as a medium or solvent in which certain enzymatic processes are able to proceed. In so far as the water requirements for these specific processes in fungi are concerned, any water present in the substrate but possessing an activity less than 75 per cent of that of pure water would be "bound" (unavailable) water.

The idea of "bound" and "free" water has been applied in attempts to supply a basis of explanation for such physiological processes as the swelling of animal tissues as in edema, the drought- and frost-hardiness in plants, and for the imbibition process in colloidal systems generally. Many methods have been devised and used for the estimation of "bound" water in colloid systems.[30] The water present which is not "free" to act as a solvent for an added crystallite can be calculated from the observed increase in freezing point depression caused by adding a given molar quantity of sugar or alcohol to the colloid system in comparison to the increase in freezing point depression which would result if all the water present were free to act as a solvent medium for the

[28] L. D. Galloway, *J. Textile Inst.*, **26,** T 123 (1935).

[29] M. Milner and W. F. Geddes, *Cereal Chem.*, **23,** 225 (1946).

[30] For general references on the concept of and methods used in determining "bound" water, the student is referred to: R. A. Gortner, *Trans. Faraday Soc.*, **26,** 678 (1930); I. D. Jones and R. A. Gortner, *J. Phys. Chem.*, **36,** 387 (1932); R. A. Gortner, *Ann. Rev. Biochem.*, **1,** 21 (1932); D. Jordan Lloyd, *Biol. Rev.*, **7,** 254 (1932); D. Jordan Lloyd, *Biol. Rev.*, **8,** 463 (1933); R. A. Gortner, *Ann. Rev. Biochem.*, **3,** 1 (1934); T. C. Barnes and T. L. Jahn, *Quart. Rev. Biol.*, **9,** 292 (1934); R. A. Gortner, Chap. VIII, in *Selected Topics in Colloid Chemistry,* Cornell University Press, 1937.

crystallite (on the basis of the dilute solution laws).[31] Similarly, water which cannot be frozen in a gel at some chosen temperature below 0°C.[32] or which cannot be removed by subjecting the gel to some chosen applied squeezing pressure [33] can be estimated and designated as water "bound" by the gel. These various methods could lead to different values of "bound" water when applied to identical systems. Similar methods employing different desiccation forces but using identical samples of a colloid system (*e.g.*, measuring the amount of ice formed when subjecting the system to −5°C. in one case and −20°C. in another, or measuring the liquid removed from a gel under a pressure of 100 lb. per sq. in. in one case and under 8,000 lb. per sq. in. in another) would lead to different values of "bound" water determined.

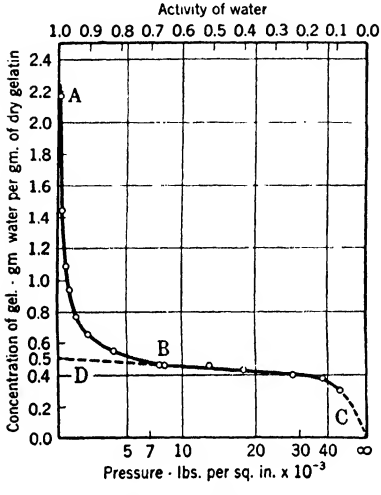


FIG. 73. Water content and "activity" of water in gelatin gels as a function of pressure. (Data of Lloyd and Moran.)

These various methods for obtaining values of the bound water in a colloid system are empirical in nature, and many of them are of such complexity as to be subject to high experimental errors. However, each can be interpreted [34] as yielding, for the colloid system being investigated, a point on the water content-water activity curve for that system such as is illustrated in Fig. 73. To illustrate, the activity a of water in equilibrium with pure ice at various temperatures T below 0°C. can be calculated from the equation [35]

$$\log a = -0.004211T - 0.0000022T^2 \qquad (130)$$

Water in equilibrium with pure ice at a temperature of −20°C. will exhibit an activity or relative vapor pressure of 0.822. Likewise the activity of water which will remain in a gel, for instance, against an ap-

[31] R. Newton and R. A. Gortner, *Botan. Gaz.*, **74,** 442 (1922).
[32] W. Robinson, *J. Biol. Chem.*, **92,** 699 (1931).
[33] D. J. Lloyd and T. Moran, *Proc. Roy. Soc.* (*London*), **A147,** 382 (1934).
[34] D. R. Briggs, *J. Phys. Chem.*, **36,** 367 (1932).
[35] G. N. Lewis and M. Randall, *Thermodynamics*, McGraw-Hill Book Co., New York, 1923.

plied squeezing pressure (P) can be calculated from the equation [35]

$$\log a = -\frac{PV}{2.303RT} \tag{131}$$

where P = pressure in atmospheres
V = molar volume of water = 18 cc.
R = gas constant (cc. atm.)
T = absolute temperature.

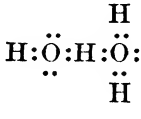
An activity of 0.822 is thus equivalent (at 25°C.) to a pressure of 3,900 lb. per sq. in. (26.4 atmospheres). If a given gel were allowed to equilibrate against water vapor of a relative vapor pressure of 0.822 it would reach a water content which should be identical with that which it would retain unfrozen if subjected to a temperature of −20°C., or if it were placed under a squeezing pressure of 3,900 lb. per sq. in. Other freezing temperatures and other pressures would correspond to other points on the relative vapor pressure-water binding curve. Figure 73 shows the variation in the water content of gelatin gels, per gram of gelatin, after equilibration under various pressures, as obtained by Lloyd and Moran.[33] It is apparent that there is a strongly bound fraction (about 0.5 gram per gram of gelatin) and that the remainder is relatively lightly bound and removable by relatively low applied pressures. The former corresponds to the strongly adsorbed water (monolayer adsorption), and the latter corresponds to water adsorbed with decidedly lower heats of adsorption and to water retained in the gel under the relatively weak forces of osmosis.

From the above considerations it is to be concluded that, certainly, water is bound in gels, biocolloids, and colloid systems generally, but not in the sense that it constitutes a definite and invariant fraction of the total water present in a given case. In order properly to define bound water in such systems it is necessary to do so in terms of the activity above which the water is to be considered "free" and below which it is to be considered "bound." Water which has a sufficiently high activity (degree of "freeness") to serve one physiological function (*e.g.*, allow a fungus spore to start growing) may not be sufficiently active for another (*e.g.*, allow bacteria to grow). In order to know what the water-binding properties of a gel may be it will be necessary to obtain the entire water activity-water content curve for that gel. We have already pointed out that this relationship may be expected to change slowly with such processes as aging of a gel and other hysteresis changes.

Theoretical Considerations of the State of Adsorbed Water. We have discussed the hydration of hydrophilic colloids and some of the ex-

perimental evidence that the water molecules in the inner hydration shell differ from the state which characterizes water molecules in ordinary liquid water. Heats of hydration, dielectric-constant measurements, and the contraction of the system, colloid-water, including the pressure studies of Lloyd and Moran, all indicate that the H_2O molecules in adsorbed water are more closely packed and are probably specifically oriented in relationship to each other, possibly in a more or less true crystal lattice which is more densely packed than the crystal lattice of ice. The water molecule is unsymmetrical and possesses a high dipole moment. Smyth [36] notes that "In the water molecule, the positive ends of two large doublets lie near the surface causing a very strong field of force around the molecule, so that the molecules affect one another greatly, strong association occurs, and the liquid is highly abnormal." When water comes in contact with a polar surface, the electrical forces of that polar surface will cause the water dipole to be "associated" with it under the same forces that cause the association of water in bulk. If the surface attractions are more intense than the attractions between the water molecules themselves, the intensity of the water binding at surfaces will be greatly increased and the water molecules will be held more rigidly and more closely packed than they are in liquid water.

The phenomenon of the association of liquids has been interpreted by Latimer and Rodebush [37] and by Huggins [38] on the basis of a "hydrogen bond." A hydrogen bond may be visualized as a hydrogen atom which is strongly associated with two electronegative atoms and thus acts as a "bridge" to hold these electronegative atoms relatively close together. Latimer and Rodebush state, "In terms of the Lewis theory, a free pair of electrons on one water molecule might be able to exert sufficient force on a hydrogen held by a pair of electrons on another water molecule to bind the two molecules together. Structurally this may be represented as



Such combinations need not be limited to the formation of double or triple molecules. Indeed, the liquid may be made up of large aggregates of molecules, continually breaking up and reforming under the influence of thermal agitation.

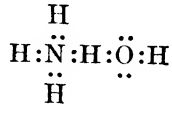
"Such an explanation amounts to saying that the hydrogen nucleus

[36] C. P. Smyth, p. 180, Am. Chem. Soc. Monograph 55, Chemical Catalog Co., New York, 1931.

[37] W. M. Latimer and W. H. Rodebush, *J. Am. Chem. Soc.*, **42**, 1419 (1920).

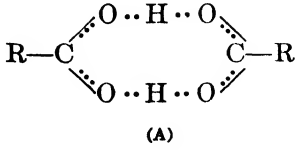
[38] M. L. Huggins, *J. Phys. Chem.*, **40**, 723 (1936); *J. Org. Chem.*, **1**, 407 (1936)

held between 2 octets constitutes a weak 'bond.' Ammonium hydroxide,

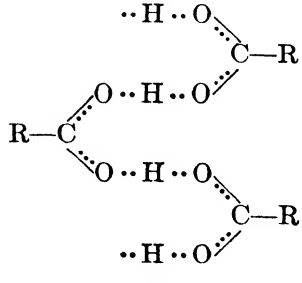


is an example in which the union is fairly strong. . . . There seems to be no reason for believing that gradations may not exist all the way from the case of ammonium chloride, where the hydrogen is definitely transferred from the chlorine to the ammonia, to the case in the association of water where the hydrogen is still held quite firmly to the original water molecule."

Huggins points out that the hydrogen bond is rather readily formed between atoms of fluorine, oxygen, nitrogen, phosphorus, and sulfur, and that the strength of the bond is in the order noted. The phenomena that result are essentially the same as would result if hydrogen were considered as not only univalent but also divalent. Thus, in hydrofluoric acid we must consider the undissociated molecule as having the composition H_2F_2 which gives rise to a hydrogen ion (H^+) and a hydrofluoride anion HF_2^-  . In this anion the hydrogen appears to be held extremely rigidly and essentially equidistant from the two atoms of fluorine. Huggins has calculated the dissociation energy of the FHF bridge as 6.2 kg.-cal. per mole with the energy of the OHO bridge somewhat less but still very appreciable. The hydrogen bridge explains the x-ray structure of the ice crystal where each oxygen atom is surrounded by four equidistant hydrogen atoms tetrahedrally arranged, and these in turn are "bridged" to other oxygen atoms, thus building up a symmetrical structure. The hydrogen bridge likewise explains the association and the different crystal forms of the aliphatic acids. It is well known that the fatty acids even in the vapor state appear to be bimolecular. They probably exist in the following form (*A*), although a second form (*B*) in which the hydrogen bridge permits the formation of



(A)



(B)

strings of molecules is likewise probable. Oxalic acid crystals of each form are known.

When the hydrogen is essentially equidistant between the two electronegative atoms, the bridge is said to be symmetrical. If the hydrogen atom oscillates about a position closer to one of the electronegative atoms than to the other, the bridge is unsymmetrical and, of course, is weaker between the hydrogen atom and the electronegative atom which is at the greater distance. The strength of the bridge increases as the strength of the attraction between the hydrogen atom and the electronegative atoms is increased. We have seen that the forces of adsorption may be of relatively enormous magnitudes, *e.g.*, holding water on SiO_2 or Al_2O_3 at temperatures exceeding 500°C. This indicates that the force bonding the water on such surfaces is very intense, and a possible explanation for these forces may be found in the strength of the hydrogen bridge.

A partial list of the energy values of the bonds and the corresponding interatomic distances across which the bond is effective for hydrogen bonds between various electronegative atoms is given in Table 33. Bond energies and bond distances for a few covalent bonds are included in this table for comparison.[39] From the values of bond energy shown in this table it is evident that the hydrogen bond constitutes a weak chemical linkage. Should ten or so such bonds occur between two molecules of such long chain compounds as protein, starch, or cellulose, they would be held together with the strength equal to a chemical linkage, and it is understandable that such complex molecules would act as though they were physically a single molecule. It is understandable, too, how water molecules may be strongly bonded to such substances which contain numerous nitrogen, oxygen, and sulfur atoms capable of hydrogen bonding with the water molecule.

Lloyd and Moran [33] give a somewhat similar picture for the forces holding bound water in their gelatin gels. They point out that "The hydration of proteins can occur by the formation of a co-ordinate link between a water molecule and certain groups in the protein structure, for instance, the positively charged basic groups of the proteins readily accept a pair of electrons from the oxygen of a water molecule on to one (or more) of the hydrogen atoms.

"Conversely the negatively charged acidic groups of the proteins readily donate a pair of electrons from the singly bound oxygen atom to the hydrogen of a water molecule.

"Neutral groups, such as OH, NH_2, NH, COOH, can similarly form co-ordinate links with water molecules either by the donation or the ac-

[39] L. Pauling, *Nature of the Chemical Bond*, Cornell University Press, Ithaca, 1940.

ceptance of a pair of electrons. . . . It should be noticed, however, that there is no way of constructing a co-ordinate link between a protein molecule and a water molecule except by donating a pair of electrons on to a chemically combined hydrogen atom. . . .

TABLE 33. ENERGY VALUES AND BOND DISTANCES OF VARIOUS HYDROGEN BONDS AND COVALENT BONDS

Hydrogen Bonds

Bond	Substance	Bond Energy, cal./mole	Bond Distance, Å.
F—H—F	H_6F_6	6.7	2.26
O—H···O	H_2O, H_2O_2	4.5	2.76
O—H···O	CH_3OH	6.2	
O—H···O	$(CH_3COOH)_2$	8.2	
C—H···N	$(HCN)_2$	3.28	
N—H···N	NH_3	1.3	
N—H···O	Urea		2.98
N—H···O	Diketopiperazine		2.85
N—H···F	NH_4F	5	2.63

Covalent Bonds

H—H	H_2	103.4	0.74
N—H	NH_3	83.7	1.01
O—H	H_2O	110.2	0.97
S—H	H_2S	87.5	1.35
C—H	CH_4	87.3	1.09
C—C	C_2H_6	58.6	1.55
C—O	$(CH_3)_2O$	70.0	1.43
C—N	$(CH_3)_3N$	48.6	1.47

"The experimental data . . . show that 1 gram of dry gelatin carries 0.5 gram of closely bound water. One molecule of gelatin (mol. wt. 34,500) is therefore closely bound to 960 water molecules, a figure very close to the calculated number of possible co-ordination centres."

We may conclude therefore that the forces which bind water on the surface of the lyophilic colloids are of the same nature as the forces which cause the association of water in bulk and which immobilize water molecules in the ice crystal lattice. However, there is evidence that these forces on a surface or at an interface may be of greater magnitude than

the forces of association of water molecule for water molecule or the forces which tend to arrange water molecules in the ordinary ice crystal lattice.

Imbibition of Lyophilic Colloids in Normal Physiological Processes. Imbibition and syneresis play very important roles in the digestive and assimilative processes. In fact, osmotic phenomena within the living organism may be regarded, at least in part, as dependent on imbibition forces.

The osmotic membranes or semi-permeable membranes of living organisms are invariably lyophilic gels. Much has been written about the chemical nature of such membranes. Overton postulated that the plasma membrane is of a lipid nature, because anesthetics and fat-soluble materials pass readily into a cell, whereas salt solutions and substances which are typically water-soluble, as contrasted with lipid-soluble, pass with difficulty.

The plasma membrane is not necessarily a structure formed by vital forces. A purely mechanistic explanation can be given for its formation. It would seem as though the formation of the plasma membrane was the inevitable consequence of the chemical composition of protoplasm. As already noted, protoplasm is composed, so far as the organic portion is concerned, largely of proteins together with a certain proportion of fats and lipids. If a protein-fat-lipid mixture is allowed to come in contact with water or a salt solution at an interface, *interfacial tension relationships will, according to Gibbs' equation (122), determine what compounds are present in the interface.* Substances which decrease interfacial tension will be concentrated in the interface. Since lipids, such as lecithin, and the other phospholipids, as well as fats and salts of the fatty acids, markedly lower interfacial tension, these compounds will be concentrated in an interface, and consequently Overton would be correct in postulating a lipid layer. On the other hand, proteins are likewise efficient depressors of interfacial tension and on theoretical grounds it is impossible to picture an interface between protoplasm and either water or a salt solution where an increased concentration of proteins, as well as fats and lipids, is absent. Accordingly Overton must be only partly correct in his lipid theory, and *the plasma membrane, from a purely mechanistic standpoint, must contain any and all of the constituents of protoplasm which effect a lowering of the surface energy at an interface.* Such a mechanistic picture of the plasma membrane would consist of a more or less completely denatured (surface energy coagulated) protein gel, probably in the form of a fibrillar structure with fats, soaps, and lipids immeshed in the protein network. The transfer of lipid-soluble materials would be through the fat-soap-lipid portion of the structure, whereas the passage of water

and such water-soluble materials as actually do pass in and out of the cells would be through the hydrated filaments of the protein network.

If a semi-permeable membrane has the structure noted above, the transference of water across the membrane, such as occurs in osmotic phenomena, can be considered to occur, at least in part, through the intermediate state of water of imbibition in the membrane. Let us assume, for example, pure water on one side of such a membrane and a salt solution on the opposite side. Proteins have a lowered imbibition capacity when in contact with a salt solution. Accordingly we would expect the protein fibrils in contact with the salt solution to have a decreased water content, whereas the fibrils on the opposite side of the membrane in contact with pure water or with a solution of lower concentration would be expected to have a higher water content. This would provide a hydration gradient across the fiber, and the normal process of diffusion in re-establishing an equilibrium within the protein fiber would cause a transfer of water from the point of lower salt concentration to the point of higher salt concentration. This suggestion was first made by Graham,[40] who wrote, "It now appears to me that the water movement in osmose is an affair of hydration and of dehydration in the substance of the membrane or other colloid septum. . . . The equilibrium of hydration is different on the two sides of the membrane of an osmometer. The outer surface of the membrane being in contact with pure water tends to hydrate itself in a higher degree than the inner surface does, the latter surface being supposed to be in contact with a saline solution. When the full hydration of the outer surface extends through the thickness of the membrane and reaches the inner surface, it there receives a check. The degree of hydration is lowered, and water must be given up by the inner layer of the membrane, and it forms the osmose."

Such a viewpoint attributes more than a passive role to the membrane, and, although the ultimate equilibrium is determined by the relative concentrations on the two sides of the membrane, the membrane itself, according to the above hypothesis, plays a definite role in osmotic phenomena.

The transfer of fluids from the digestive tract to the body tissues and the transfer of liquid from the body tissues to the digestive tract involve both imbibition and syneresis. A normal individual secretes from 700 to 1,000 ml. of saliva per day, from 600 to 900 ml. of bile, from 600 to 800 ml. of pancreatic juice, from 1,000 to 2,000 ml. of gastric juice, whereas the water intake is usually only from 1,000 to 1,500 ml. Accordingly there passes into the digestive tract from 3.9 to 6.2 liters per

[40] Thomas Graham, *Phil. Trans.* (1861), p. 183.

day or from 4 to 6 times the amount of liquid that is taken in the form of liquid food, indicating that the water absorbed in the digestive tract is used several times over in the form of liquids which are secreted into the digestive tract. The passage of this relatively enormous volume of fluids from the digestive tract into the blood stream cannot be adequately accounted for by pure osmosis, since in many instances the contents of the digestive tract have a higher osmotic pressure than the blood serum.

Imbibition, however, may account for this transfer of liquid. The proteins of the intestinal wall take up the water from the intestinal tract and transfer it to the proteins of the blood stream, which in turn carry it to other portions of the body where their imbibitional capacity is lowered and where they yield the water to other tissues or glands. The crystalloids which pass from the digestive tract to the blood stream and from the blood stream to the various cells and tissues of the body can be regarded as diffusing through a swollen gel rather than moving by osmotic processes across a membrane. Chemicals which favor swelling of proteins favor absorption from the intestinal tract, and substances which hinder the swelling of proteins hinder or prevent absorption from the intestinal tract. Magnesium sulfate more or less inhibits protein imbibition, and the effect of magnesium sulfate as a saline cathartic is to a large extent due to the prevention of the absorption of liquid from the intestinal tract, retaining the liquid within the tract. Similarly, agar, which furnishes bulk in cases of chronic constipation, is efficient because of its very pronounced hydrophilic nature, which allows it to hold the water by imbibition forces against the pull of the intestinal wall and of the blood stream.

Imbibition of Lyophilic Colloids as Related to Medical Problems. It would be beyond the scope of this discussion to more than mention certain of the pathological problems in which lyophilic colloid-water relationships appear to be involved. Martin Fischer [41] has ably defended the viewpoint that edema and nephritis are diseases characterized by abnormal imbibition of the body colloids. He notes that in the normal organism the degree of imbibition of the tissues is very finely regulated. For example, the brain may swell as much as 1,000 per cent if removed from the body and placed in various solutions. If, however, the brain were to swell 3 per cent in the living organism, the volume of the brain would become greater than the volume of the skull cavity, and the pressure of the brain on the skull cavity would produce intense pain followed by death.

The role of the salt content in the blood and tissues is at least in part

[41] M. H. Fischer, *Oedema and Nephritis*, John Wiley & Sons, New York, 1921 (out of print).

the repression of the imbibition which would take place in the absence of such electrolytes. Bottazzi [42] suggests that the function of sodium chloride in the blood is to decrease the viscosity of the blood. If the inorganic salt content of the blood were decreased, as it sometimes is by excessive perspiration, the hydration of the blood proteins would be increased, and accordingly the viscosity of the blood likewise would be increased.

This problem of excessive perspiration with the loss of relatively large quantities of sodium chloride from the body may have a rather important bearing on industry. Moss [43] notes that, under muscular exertion and relatively high temperature, the sodium chloride content of perspiration ranges from 0.118 to 0.325 per cent, averaging 0.224, and that under such conditions more sodium chloride may be lost through perspiration than in the urine. He suggests that workers in hot mines need more highly salted foods.

In certain of the very deep mines where the temperature regularly exceeds 100°F. a very considerable proportion of the miners developed intense headaches and required hospital treatment at frequent intervals. In these mines the miners worked practically nude, drank large quantities of water, and, owing to the high temperatures and humidities, were continuously bathed in perspiration. It occurred to the physician in charge that possibly the headaches might be attributed to an excessive loss of sodium chloride through the perspiration, causing an excessive imbibition of the body tissues, including the nervous tissue. Accordingly the recommendation was made that, instead of the usual supply of water, the miners be furnished drinking water containing salts in approximately the proportion of a physiological salt solution. The effect of such substitution was striking in the extreme. The epidemic of headaches, which had persisted for years, disappeared as if by magic, indicating that the excessive loss of salts in the perspiration had been the determining factor.

Thomas and Andrews [44] point out that the serum proteins from edematous individuals have a greater affinity for water than those from normal individuals. Normal sera never swell more than 9 per cent, whereas sera from sufferers from uremia and edema swell enormously, often more than 50 per cent. They suggest that these observations be made a method of diagnosis.

A similar method of diagnosis is used in certain hospitals, the technic being to inject small amounts of water intradermally, noting the length

[42] F. Bottazzi, *Arch. fisiol.*, **7**, 579 (1909).
[43] K. N. Moss, *Proc. Roy. Soc.* (*London*), **B95**, 181 (1923).
[44] W. A. Thomas and E. Andrews, *Proc. Soc. Exptl. Biol. Med.*, **25**, 773 (1928).

of time required for the absorption of the water by the tissues. An extremely rapid absorption is characteristic of edematous individuals or of individuals who will shortly develop edemas, and such technic has permitted the detection of the onset of edema considerably in advance of the time when it could be observed by the usual clinical methods.[45,46]

McQuarrie [47] made the important discovery that at least certain types of epilepsy in children are either caused by, or are associated with, a disturbed water balance, and he has had remarkable success in the prevention of "grand mal" seizures by withholding water from the patient so that the patient becomes somewhat dehydrated. Certain patients who came to the hospital having forty or more grand mal seizures per day were freed of seizures after having been dehydrated to a point where they had lost 1–2 kg. of body water. If water was then returned to the diet, or if the antidiuretic hormone of the pituitary was given to such patients, they promptly went into a positive water balance with a recurrence of the seizures. However, if sufficient urea was dissolved in the ingested water, positive water balance would not take place, and seizures did not develop. With the dehydrated epileptic children typical grand mal seizures recurred when the antidiuretic hormone was given and water was taken at the rate of 2 to 5 ml. per kg. of body weight per hour. Under similar technic normal children did not develop epileptic seizures. McQuarrie does not believe that the epileptic seizures are the resultant of a brain edema but finds that the seizure is correlated with a marked negative potassium balance and a striking increase in the K:Na ratio in the urine, which indicates that there is an apparent "leakage" of potassium from the cells of the nervous tissue. He suggests that perhaps the epileptic state is characterized by an inherent defect in the mechanism regulating the permeability of the brain cell membranes and that, at the time when potassium leaks out, abnormal amounts of water and perhaps sodium ions pass into the cells of the nervous tissue. That there may be a shift in the water relationships of the brain is supported by the observations of Barbour [48] that during anesthesia the cerebrum loses water and the medulla gains water and that the shift appears to be directly associated with the phenomena of narcosis.

[45] W. B. McClure and C. A. Aldrich, *J. Am. Med. Assoc.*, **81**, 293 (1923).

[46] C. A. Aldrich and W. B. McClure, *J. Am. Med. Assoc.*, **82**, 1425 (1924). *Cf.* also *J. Am. Med. Assoc.*, **83**, 1566 (1924); **84**, 1258 (1925); *Arch. Internal Med.*, **37**, 281 (1926); **41**, 102 (1928); *Surg., Gynecol. Obstet.*, July, 1926, p. 40.

[47] I. McQuarrie, *J. Pediatrics*, **3**, 539 (1933); *Am. J. Diseases Children*, **38**, 451 (1929); (with D. B. Peeler) *J. Clin. Investigation*, **10**, 915 (1931); (with R. C. Manchester and C. Husted) *Am. J. Diseases Children*, **43**, 1519 (1932); and *Ann. Internal Med.*, **6**, 497 (1932).

[48] H. G. Barbour, *Science*, **73**, 346 (1931).

The problem of the nature of complement has interested immunologists for many years. Complement is some property or some component of normal serum which is destroyed by heating to 55°C. That this change may be in part colloid-chemical is suggested by the findings of du Nouy,[49] who studied the effect of heat on the viscosity of normal rabbit serum. The viscosity fell during the temperature rise from 24.2° to 55°C. at the same rate as the fall in viscosity of water over a similar temperature range. The relative viscosity of the serum then began to increase sharply. When the hydration capacity of the serum proteins was calculated by equation (8) each gram of protein was found to occupy 2.09–2.19 ml. over the range of 24.2° to 55°C. At 57° the volume had increased to 2.58 ml., at 59° to 2.90 ml., and at 62° to 3.37. When held for 4 hours at 60°, the volume was 3.88 ml. or a hydration capacity 169 per cent of the normal. Du Nouy notes that this is an irreversible change, the protein micelles becoming hydrated at the expense of the water in the serum.

The field of pathology offers a great opportunity to the colloid chemist who is interested in the water relationships of the lyophilic colloids.

The Liesegang Phenomena. Liesegang [50] first described the peculiar reactions which take place when certain chemical processes are carried out in colloid gels. If potassium chromate is dissolved in a gelatin gel and a solution of silver nitrate is allowed to diffuse into the gel, the silver chromate which is formed is not dispersed uniformly throughout the gel but rather separates in a series of concentric rings, separated by more or less clear portions of the gel. Similarly, if potassium chromate is dissolved in sodium silicate and if the liquid is then acidified to form a silicic acid gel, when copper sulfate diffuses into such a gel, banded precipitates of copper chromate will be produced. Such banded precipitates are known as Liesegang rings.

In other instances, for example, in the reaction between potassium iodide and lead acetate, bands are not formed, but the lead iodide slowly separates in the form of very large crystals. Figure 74 illustrates the formation of copper chromate rings in a silicic acid gel, and the large crystals of lead iodide, likewise in silicic acid gel.

Much has been written about this phenomenon, and numerous theories have been proposed. It is extremely difficult to evaluate the theories, and we still lack exact data by which one or the other of the theories can be proved.

It has been suggested that the silver ions, diffusing into a chromate-gelatin gel, form silver chromate which remains in a supersaturated

[49] P. L. du Nouy, *Ann. Inst. Pasteur*, **42**, 742 (1928).

[50] R. E. Liesegang, *Naturw. Wochschr.*, **11**, 353 (1896).

state, owing to the protective action of gelatin. As the concentration of the silver increases, the silver chromate micelles are coagulated, and they crystallize out in the gel, forming a more or less impermeable layer.

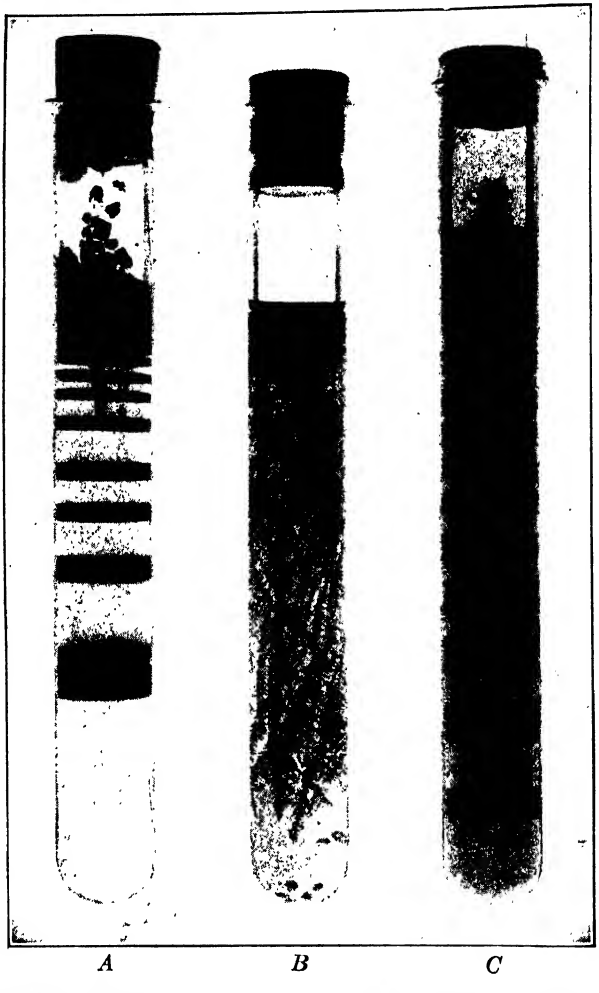


FIG. 74. Liesegang phenomena. *A* and *C*, copper chromate in silica gel; *B*, lead iodide crystals in silica gel.

This silver chromate band will adsorb silver ions and retard the diffusion of such ions through the gel but will not adsorb nitrate ions which pass through, possibly in the form of nitric acid, and such ions will advance ahead of the silver ions. As the silver chromate gel ages, it will break down, become coarser in structure, allow the silver ions to pass through,

and eventually a new point of supersaturation with subsequent precipitation which causes the formation of a new band will be reached.

Such a theory, however, does not explain the specificity of the gel structure. Banded precipitates will be formed in a number of instances in silicic acid gels which are not formed in gelatin gels. Similarly, banded precipitates are formed in other instances in gelatin gels, whereas no Liesegang rings are formed by the same reaction in silica gel. The nature of the gel in which the precipitation takes place plays an important role, and no theory as yet proposed accounts for the specificity of the gel.

It appears possible that the reaction may be due to differences in concentration of the reacting ions which bring about progressive peptization and coagulation. Thus, in Fig. 4 we have shown that silver bromide is held in colloidal dispersion by an excess of either silver or bromide ions, and that complete precipitation occurs only when the concentrations of silver and bromide ions are approximately equivalent. One could postulate, therefore, that in the diffusion of silver nitrate into a chromate gel there would be for a time an excess of chromate ions, which would result in the peptization of the precipitate until eventually a point would be reached at which the chromate and silver ions are equivalent. At this point one would expect a precipitate to be formed to make a band of silver chromate. A repetition of this process would account for succeeding bands. Here again, however, the specificity of the gel is not explained.

Liesegang has suggested that the bands characteristic of agates are due to the Liesegang phenomena. This appears to be so. Figure 75 is the photograph of an artificial agate produced under laboratory conditions by allowing copper sulfate to diffuse into a silicic acid gel containing potassium ferrocyanide. The silicic acid gel was allowed to set in a collodion bag. The collodion bag was immersed in a dilute solution of copper sulfate and removed after several weeks, and the ball of gel was cut in two. It will be noted that the laboratory product resembles the natural agate exactly. Bhatnagar and Mathur [51] have prepared such artificial agates and slowly dried them under pressure, in this way obtaining preparations which have a hardness of approximately 5.0 on the mineralogical scale and which can be cut and polished so as to make them practically indistinguishable from natural agates.

It frequently happens that precipitates which are not in exact rings form in gels. Under such conditions it almost invariably happens that the figures which do form are very symmetrical, *i.e.*, a line drawn down the center of the figure will yield two halves which are mirror images of each other. This problem of symmetry of structures formed in gels in-

[51] S. S. Bhatnagar and K. K. Mathur, *Kolloid-Z.*, **30**, 368 (1922).

dicates relationships which may be of biological importance. However, here again no adequate explanation has been offered to account for the production of symmetrical figures.

Whether or not the Liesegang phenomena play a role in living processes is still an open question. Liesegang has suggested that the coloration patterns on butterfly wings, the stripes which occur on elytra of

FIG. 75. A photograph of an artificial agate formed by allowing copper sulfate to diffuse into a silica gel containing potassium ferrocyanide. An example of a Liesegang ring phenomenon.

beetles, etc., may be due to these phenomena. Such explanations appear probable but are still unproved. Many interesting and beautiful museum specimens may be prepared by causing precipitates to form in the various gels. For those who are interested in following the literature in this field certain references are noted.[52-61]

[52] E. Hatschek, *Second Report on Colloid Chemistry*, p. 21, *Brit. Assoc. Advancement Sci.*, London (1918).

[53] H. N. Holmes, *J. Am. Chem. Soc.*, **40**, 1187 (1918).

[54] E. C. H. Davies, *J. Am. Chem. Soc.*, **45**, 2261 (1923).

[55] S. C. Bradford, *Biochem. J.*, **10**, 169 (1916); **11**, 14 (1917).

[56] Wo. Ostwald, *Kolloid-Z.*, **36**, 380 (1925).

[57] P. B. Ganguly, *J. Phys. Chem.*, **31**, 481 (1927).

[58] C. A. Schleussner, *Kolloid-Z.*, **34**, 338 (1924).

[59] H. Handovsky and E. du Bois-Reymond, *Kolloid-Z.*, **33**, 347 (1923).

[60] T. R. Bolam and M. R. MacKenzie, *Trans. Faraday Soc.*, **22**, 151 (1926).

[61] E. S. Hedges, *Liesegang Rings and Other Periodic Structures*, Chapman and Hall, Ltd., London, 1932; also *J. Chem. Soc.*, 2781 (1929).

So far as the author is aware only one study has dealt directly with an attempt to apply the Liesegang phenomenon to biochemical or medical problems. This is a study by Bucher [62] on the Liesegang phenomenon in blood plasma gels. Citrated blood plasma from a great variety of sources was mixed under rigidly controlled conditions with potassium bichromate solution, and the mixture was allowed to set to a gel in special test tubes. The diffusion of silver nitrate into such gels then resulted in the formation of Liesegang rings. The time required for each ring to form and the distance separating the rings were recorded. A marked individuality was found for these data, blood serum from one individual being sharply differentiated from that of another individual. Characteristic species differences were also observed. Within a species there were apparently characteristic differences between young and old individuals and between normal and pathological individuals of the same age. Serums from individuals with carcinomas apparently gave reactions different from normal serums. Perhaps the most striking differences observed were those existing between normal cattle blood sera and the sera of cattle fetal blood. In the former typical sharply defined Liesegang rings formed; in the fetal blood gels, however, no rings formed but instead characteristic "flame-like" precipitations were produced. Repeated hemorrhages likewise produce characteristic changes in the pattern so that eventually no rings formed in the sera from the later bleedings. No explanation is available for the peculiar behavior observed, but unquestionably these observations will stimulate further investigation of the biochemical factors involved.

[62] R. Bucher, *Die Diffusionsanalyse am Blutplasmagel*, Benno Schwabe and Co., Basel, 1937. This paper appeared as Suppl. II to *Helv. Medica Acta*, **4**, Heft 5 (1937).

CHAPTER 10

The Gibbs-Donnan Equilibrium

Two gases separated by a membrane permeable to both will diffuse through the membrane in both directions, so that at equilibrium the mixture will have the same composition on both sides of the membrane. Similarly, if a membrane separates two solutions of different concentration and if the membrane is permeable to both the solute and the solvent, equilibrium will be attained only when the concentration of the solute on both sides of the membrane is identical. Thus, if a solution of potassium chloride is placed on one side of a permeable collodion membrane and a solution of equivalent concentration of sodium sulfate is placed on the opposite side of the membrane, at equilibrium it will be found that there is an equal concentration of potassium chloride and sodium sulfate on both sides of the membrane, if it is assumed that the membrane is chemically inactive, merely acting as a septum to keep the initial solutions from mixing mechanically. Under such conditions, the osmotic pressure of the solutions in both compartments will be identical.

That very different conditions prevail, however, when the membrane is impermeable to one of the ions was first predicted by Gibbs [1] and has been described and experimentally verified by Donnan.[2]

If we picture a membrane separating two solutions, one of which contains a salt such as NaCl and the other an electrolyte, NaR (where R represents a monovalent anion so large that it is unable to pass through the pores of the membrane), and if we assume complete ionization of both electrolytes, we can represent the hypothetical initial state by

Na^+	Na^+
R^-	Cl^-
(1)	(2)

which at equilibrium and constant volume becomes

Na^+	Na^+
R^-	Cl^-
Cl^-	
(1)	(2)

[1] J. W. Gibbs, *Trans. Connecticut Acad. Arts Sci.*, **3**, 228 (1876).

[2] F. G. Donnan, *Z. Electrochem.*, **17**, 572 (1911); *Chem. Revs.*, **1**, 73 (1924).

owing to the diffusion of NaCl from compartment (2) through the membrane to compartment (1), the NaR being restrained from diffusing by impermeability of the membrane to the anion.

At equilibrium the necessary work for the isothermal reversible transference of a molecule of Na^+ from (2) to (1) is just as great as the work which is gained for the corresponding isothermal reversible transference of a molecule of Cl^-. Thus, if we consider the infinitely small isothermal and reversible change of the system at equilibrium in which dn molecules of Na^+ and dn molecules of Cl^- are transferred from (2) to (1), the work gained through this transference (the increase in free energy) is *nil*, and we can therefore write

$$dn \cdot RT \log_e \frac{(Na^+)_2}{(Na^+)_1} = -dn \cdot RT \log_e \frac{(Cl^-)_2}{(Cl^-)_1} \tag{132}$$

and

$$\log_e \frac{(Na^+)_2}{(Na^+)_1} = -\log_e \frac{(Cl^-)_2}{(Cl^-)_1} \tag{133}$$

or

$$\log_e (Na^+)_2 - \log_e (Na^+)_1 = -\log_e (Cl^-)_2 + \log_e (Cl^-)_1 \tag{134}$$

or

$$\log_e (Na^+)_2 + \log_e (Cl^-)_2 = \log_e (Cl^-)_1 + \log_e (Na^+)_1 \tag{135}$$

or

$$(Na^+)_2 \cdot (Cl^-)_2 = (Cl^-)_1 \cdot (Na^+)_1 \tag{136}$$

where the quantity within the parentheses indicates the molar concentration of the respective ions.

Equation (136) states that, *at equilibrium, the product of the sodium- and chloride-ion concentrations in compartment (2) is equal to the product of the sodium- and chloride-ion concentrations in compartment (1)*. Since compartment (2) contains only sodium chloride at equilibrium, the chloride-ion concentration in compartment (2) must be equal to the sodium-ion concentration. Accordingly,

$$[(Cl^-)_2]^2 = (Na^+)_1 \cdot (Cl^-)_1 \tag{137}$$

A kinetic picture for this equilibrium can be drawn as follows. Since the total number of positive and negative charges must be equal at all times on a given side of the membrane, *i.e.*, $(+)_1 = (-)_1$ and $(+)_2 = (-)_2$, it is evident that Na^+ ion, for example, cannot move in either direction across the membrane at any time without being accompanied by an ion of equal charge but of opposite sign of charge, *e.g.*, a Cl^- ion. The probability of a Na^+ ion and a Cl^- ion striking the pore area on one side of the membrane at the same instant, and therefore being able to pass through, will be equal to the product of their respective concentra-

tions on that side of the membrane. Equilibrium, *i.e.*, the condition where equal amounts of diffusible electrolyte will move in either direction across the membrane, will be attained when the probabilities of their striking the pore area of the membrane from each side are the same. Equilibrium will exist, obviously, when

$$(Na^+)_1 \cdot (Cl^-)_1 = (Na^+)_2 \cdot (Cl^-)_2$$

Since at equilibrium NaR, as well as NaCl, is present in compartment (1), the concentration of $(Na^+)_1$ will be greater than the concentration of $(Cl^-)_1$. This means that there will be an unequal distribution of ions on the two sides of the membrane at equilibrium, and accordingly *at equilibrium* the liquid on one side of the membrane may have an appreciably higher osmotic pressure than the liquid on the opposite side of the membrane.[3]

Assuming that x molecules of sodium chloride diffused from compartment (2) to compartment (1) we would have, assuming constant volume, the following concentrations of ions on the two sides of the membrane:

Original State				State of Equilibrium			
Na^+	R^-	Na^+	Cl^-	Na^+	R^-	Cl^-	Na^+ Cl^-
C_1	C_1	C_2	C_2	$C_1 + x$	C_1	x	$C_2 - x$ $C_2 - x$
(1)		(2)		(1)			(2)

where C_1 and C_2 are the original molar ion concentrations in compartments (1) and (2).

Equation (136) affords the following algebraic solution for the quantity of NaCl which is transferred from compartment (2) to compartment (1):

$$(C_1 + x) \cdot x = (C_2 - x)^2 \tag{138}$$

or

$$x = \frac{(C_2)^2}{C_1 + 2C_2} \tag{139}$$

and

$$\frac{x}{C_2} = \frac{C_2}{C_1 + 2C_2} \tag{140}$$

or

$$\frac{C_2 - x}{x} = \frac{C_1 + C_2}{C_2} \tag{141}$$

[3] The student should remember that all the thermodynamic considerations discussed in this chapter and the generalizations apply to systems *where the volume of the liquid on the two sides of the membrane is kept constant and where only the transference of ions takes place.* Obviously, if the liquid molecules are allowed to move so that the volume of the liquid on the two sides of the membrane can change, it will be impossible to build up an *equilibrium* osmotic pressure which will be greater on one side of the membrane than on the other.

The percentage of NaCl which will diffuse from (2) to (1) is accordingly $(x/C_2)100$, and the ratio of the division of sodium chloride between compartment (2) and compartment (1) at equilibrium is given by $(C_2 - x)/x$.

Tables 34 and 35 show the distribution of NaCl at equilibrium under the conditions that we have been discussing, for various initial concentrations of NaR and NaCl in compartment (1) and compartment (2). These tables show that the action of non-dialyzable but electrolytically dissociated NaR is very important. Although the membrane is com-

TABLE 34. DISTRIBUTION OF NaCl ON THE TWO SIDES OF A MEMBRANE PERMEABLE TO NaCl BUT IMPERMEABLE TO A COLLOIDAL ION (R^-) IN THE SOLUTION ON ONE SIDE OF THE MEMBRANE

(Calculations of Donnan)

Original State				State of Equilibrium				
Na^+ C_1	R^- C_1	Na^+ C_2	Cl^- C_2	Na^+ $C_1 + x$	R^- C_1	Cl^- x	Na^+ $C_2 - x$	Cl^- $C_2 - x$
1,000	1,000	1,000	1,000	1,333	1,000	333	666	666
1,000	1,000	100	100	1,008	1,000	8.3	92	92
100	100	1,000	1,000	576	100	476	524	524

TABLE 35. DISTRIBUTION RATIO OF NaCl ON THE TWO SIDES OF A MEMBRANE PERMEABLE TO NaCl BUT IMPERMEABLE TO A COLLOIDAL ION (R^-) IN THE SOLUTION ON ONE SIDE OF THE MEMBRANE

(Calculations of Donnan)

Original Concentration of NaR in (1) C_1	Original Concentration of NaCl in (2) C_2	Original Ratio of NaR to NaCl $\frac{C_1}{C_2}$	Per Cent NaCl Going from (2) to (1) $\frac{100x}{C}$	Ratio of NaCl between (2) and (1) $\frac{C_2 - x}{x}$
0.01	1.00	0.01	49.7	1.01
0.10	1.00	0.10	47.6	1.10
1.00	1.00	1.00	33.0	2.00
1.00	0.10	10.00	8.3	11.00
1.00	0.01	100.00	1.0	99.00(?)

pletely permeable to NaCl, a high enough concentration of NaR will practically inhibit the diffusion of NaCl through the membrane. Conversely, if NaCl were added to the compartment containing the non-diffusible ion, the presence of the non-diffusible ion would cause the *excretion* of NaCl from this compartment, *such excretion taking place even against an opposing osmotic pressure.* The membrane which is freely permeable to NaCl thus shows an apparent differential permeability for the completely dialyzable NaCl, this differential permeability being due not to the membrane but to the ionic micelles in the sol on one side.

As Donnan points out, this phenomenon must play a role in physiology. Non-dialyzable anions and cations are present in the cell contents and in the intracellular fluids. The membranes may themselves be completely permeable to molecularly dissolved solutes, but, owing to the influence of the ionic micelles, an unequal distribution of ions occurs on the two sides of the membrane. The magnitude of this unequal distribution of total ions at equilibrium and its dependence on the ratios of original concentrations of diffusible and non-diffusible ions on the two sides of the membrane is illustrated in Table 36.

TABLE 36. CHANGE IN THE EQUILIBRIUM CONCENTRATION OF IONS ON THE TWO SIDES OF A MEMBRANE WHEN THE MEMBRANE IS IMPERMEABLE TO ONE OF THE IONS ORIGINALLY PRESENT IN COMPARTMENT C_1

Ratio of Original Concentration, $C_1:C_2$	*Ratio of Equilibrium Concentration,* $C_1:C_2$
1:1	2:1
10:1	11:1
1:10	1.1:1

On the basis of similar considerations Donnan points out that at equilibrium, where several species of diffusible ions are present, their distribution ratios will be, for example,

$$\frac{(Na^+)_1}{(Na^+)_2} = \frac{(H^+)_1}{(H^+)_2} = \frac{(Cl^-)_2}{(Cl^-)_1} = \frac{(OH^-)_2}{(OH^-)_1}$$

$$= \frac{\sqrt{(Ca^{++})_1}}{\sqrt{(Ca^{++})_2}} = \frac{\sqrt[3]{(Al^{+++})_1}}{\sqrt[3]{(Al^{+++})_2}} = \frac{\sqrt{(SO_4^{=})_2}}{\sqrt{(SO_4^{=})_1}}, \text{ etc.} \quad (142)$$

where the quantities inside the parentheses refer to the molar activities of the ions denoted. Thus, from a determination of the activity ratio of one ionic species, *e.g.*, the hydrogen-ion activities, at equilibrium, the ratios of all other diffusible species of ions may be obtained.

Donnan continues his discussion with the question of the hydrolytic decomposition of salts through the action of a membrane. If NaR is placed on one side of a membrane and pure water is placed on the opposite side, owing to the presence of hydrogen and hydroxyl ions in the water, and since, at equilibrium,

$$\frac{(Na^+)_1}{(Na^+)_2} = \frac{(OH^-)_2}{(OH^-)_1}$$

we should obtain the following original and equilibrium states:

Original State			State of Equilibrium				
Na^+	R^-	Pure	Na^+	H^+	R^-	Na^+	OH^-
C_1	C_1	water	$C_1 - x$	x	C_1	x	x
(1)		(2)	(1)			(2)	

The value of x can be determined from the dissociation constant of water.

$$x \cdot (OH^-)_1 = K_w \tag{143}$$

from which Donnan obtains the value of (x) as

$$x = \sqrt[3]{K_w C_1} \tag{144}$$

Table 37 shows Donnan's calculations for the value of x at various concentrations of NaR and the percentage of Na^+ which is transferred across the membrane to form NaOH.

TABLE 37. HYDROXYL-ION CONCENTRATION (x) OUTSIDE A MEMBRANE AT EQUILIBRIUM AND THE PERCENTAGE OF Na^+ ($100x/C_1$) TRANSFERRED ACROSS THE MEMBRANE FOR VARIOUS ORIGINAL CONCENTRATIONS OF NaR ON ONE SIDE OF THE MEMBRANE AND OF PURE WATER ON THE OTHER SIDE OF THE MEMBRANE

C_1	x	$\frac{100x}{C_1}$
		Per Cent
0.01	5.10^{-6}	0.05
0.10	1.10^{-5}	0.01
1.00	2.10^{-5}	0.002

In this case, one would have alkali excreted through the membrane, the liquid on the inside of the membrane becoming more acid. On the

other hand, assuming a positively charged ionic micelle, we might very well have the following initial state and equilibrium state:

Original State		State of Equilibrium	
R^+	H^+	R^+	H^+
Cl^-	OH^-	Cl^-	Cl^-
		OH^-	
(1)	(2)	(1)	(2)

In this case we should have *acid excreted* across the membrane.

Donnan points out that *the proper ampholyte can easily give rise by this means alone to a concentration of hydrogen ions in the external liquid as great as that found in the gastric juice.*

As a consequence of the unequal concentrations (activities) of a given ionic species on the two sides of the membrane at equilibrium, there must exist an electrical potential across the membrane. Such a potential can, indeed, be measured when liquid junction electrodes (such as calomel half cells) are brought into contact with the two solutions, or it can be calculated from the measured activities of any diffusible ionic species in the two phases of the system, from the relationship

$$E = \frac{RT}{nF} \log \frac{(H^+)_1}{(H^+)_2} = \frac{RT}{nF} \log \frac{(Cl^-)_2}{(Cl^-)_1}, \text{ etc.} \qquad (145)$$

Such membrane potentials are equilibrium potentials and incapable, therefore, of doing work. The sign of the membrane potential will be such that the solution containing the non-diffusible ion will acquire a sign of potential the same as that carried by this ion.

Since the Donnan membrane equilibrium will not be attained unless an externally applied pressure is maintained on the solution containing the non-diffusible ionic component, there will exist a difference in the osmotic pressures of the two solutions bathing the membrane at equilibrium which will be equal to the applied pressure. The osmotic pressure of the solution containing the non-diffusible component will be higher than that of the equilibrium solution on the other side of the membrane, but this difference will not be due to the osmotic increment of the non-diffusible ion alone. The unequal concentrations of the diffusible ionic species on the two sides of the membrane will be important determinants in the osmotic pressure relationships.

For purposes of simplification, in the foregoing discussion, the non-diffusible ion has been considered as carrying a single charge. Osmotically such an ionic micelle should be equivalent to any other monovalent

ion in the system. If we assume that the individual osmotic pressures of the equilibrium solutions are proportional to their concentrations of total ions, then, in the three examples given in Table 36, the ratios of the osmotic pressures of the equilibrium solutions, for example, should be 2:1, 11:1 and 1.1:1. The percentage of the total osmotic pressure of the solution containing the non-diffusible ion which is due to it will be 37.5 per cent, 49.6 per cent, and 11.3 per cent, respectively, and the increment with respect to the *difference* in osmotic pressure between the equilibrium solutions which is due to the non-diffusible ion will be 75 per cent, 54.5 per cent, and 96.1 per cent, respectively. It is obvious that in order to obtain the molecular weight of the non-diffusible ion by osmotic pressure measurements, such measurements would have to be made under conditions where the Donnan effect is eliminated or where the increment due to diffusible ions can be accurately accounted for.

That the osmotic increment due to the unequal distribution of diffusible ions may be difficult to calculate from their observed concentration differences is indicated by experiments of Hammarsten [4] with nucleic acid and Briggs [5] with gum arabic. In general, it was observed that the osmotic pressure differences for the solutions on the two sides of the membrane, as calculated from the electrometrically measured activities of the diffusible ions alone, were always greater than the experimentally observed osmotic pressures, although the equilibrium distribution of these ions, as measured electrometrically, was in accord with the Donnan theory. An obvious conclusion is that the osmotic activity of the diffusible ions in the colloid-containing solution was lower than their electrometric activity. By some mechanism, as yet not clarified, part of their osmotic activity was "neutralized" by their association with the colloid constituent.

On the other hand, Adair and co-workers [6,7] found in their studies on the osmotic pressures of hemoglobin solutions that, within certain ranges of *p*H and colloid concentration,

$$\pi_o = \pi_p + \pi_i \tag{146}$$

where π_o was the observed osmotic pressure, π_p the fraction of the observed osmotic pressure which was due to the protein component, and π_i was that due to the unequal distribution of diffusible ions as determined from electrometric activity measurements. An excellent discussion of the methods to be employed in calculating π_p from measured

[4] E. Hammarsten, *Biochem. Z.*, **144,** 383 (1924).
[5] D. R. Briggs, *J. Phys. Chem.*, **38,** 1145 (1934).
[6] G. S. Adair, *Proc. Roy. Soc.* (*London*), **A120,** 573 (1928); **A126,** 16 (1929).
[7] G. S. Adair and M. E. Robinson, *Biochem. J.*, **24,** 1864 (1930).

values of π_o and calculated values of π_i in such systems is given by Wagner.[8]

In experiments where it is desired to attain an estimate of the particle (molecular) weight of the colloid components of a colloid electrolyte (such as a protein) it is usually better to eliminate the Donnan effect rather than try to take it into account. In general the experiments should be carried out (1) under conditions where the degree of ionization of the colloid ion is near a minimum (as at the isoelectric point of a protein or at a very low *p*H in the case of an anionic colloid electrolyte such as gum arabic), and (2) in the presence of a high enough salt or buffer (diffusible ions) concentration to suppress to a low value the contribution of these ions to π_o (usually an ionic strength of 0.3–0.5 is sufficient).

The swelling of protein gels has been assumed by many investigators to be due to a Donnan equilibrium. Wilson [9] and Loeb [10] and his school have been especially prominent proponents of the Donnan equilibrium. On the other hand, Miss Lloyd [11] finds no evidence that a Donnan equilibrium influences the swelling of silk gut in alkaline solutions. Instead, the postulate is made that coordinate linkages are opened between the carbonyl group of one peptide chain and the imino group of another and that there is a "binding" of water on the groups so released.

From an inspection of the preceding tables, which show the unequal distribution of diffusible ions on the two sides of the membrane, we may conclude that the Donnan effect probably accounts in a considerable measure for the concentration of electrolytes within plant or animal cells and tissues. Contrary to what one would expect, *the Donnan equilibrium provides a mechanism whereby such cells and tissues may possess a considerably higher osmotic pressure than the liquid which bathes them and at the same time be in equilibrium with the external liquid.*

However, one must not consider that the Donnan equilibrium offers a complete explanation for the adsorption of ions from the soil solution by the root hairs of plants. Even though probably it plays a role, other factors, as yet unknown, are involved. *The cell membrane, at least in certain instances, is more than a mechanical sieve restraining the diffusion*

[8] R. H. Wagner, p. 253, *Physical Methods of Organic Chemistry*, Vol. I, Interscience Publishers, New York, 1945.

[9] J. A. Wilson, *The Chemistry of Leather Manufacture*, 2nd ed., Chemical Catalog Co., New York, 1928–1929.

[10] Jacques Loeb, *Proteins and the Theory of Colloidal Behavior*, McGraw-Hill Book Co., New York, 1922. (*Cf.* also numerous articles in the *Journal of General Physiology*, Vols. 1–6, inclusive.)

[11] D. J. Lloyd and R. H. Marriott, *Trans. Faraday Soc.*, **29**, 1228 (1933).

of ionic micelles. Harris and his co-workers [12-16] have shown that the leaf tissue fluids of Egyptian and Upland cotton in the Gila River valley of Arizona differ very markedly in the content of chloride and sulfate ions, the Egyptian type having a chloride content from 28 to 189 per cent higher than the Upland cottons, whereas the Upland cottons have a sulfate content 18 to 28 per cent higher than the Egyptian cottons.

The surprising feature of these investigations is that *the ability of the plant to absorb selectively sulfates and chlorides from the saline soil solution is heritable.* The F_1 crosses between Egyptian and Upland types yielded progeny which were intermediate between the parents in this respect, whereas the ability to absorb selectively either chlorides or sulfates segregated in the F_2 generation.

In these experiments there was no possibility that the results were due to soil heterogeneity. The plants under experiment were alternated in the row and almost invariably showed the higher sulfate or higher chloride content characteristic of the variety. Similarly, the F_1 generation and the F_2 generation were spaced methodically over the fields and still showed in the F_1 intermediate characteristics and in the F_2 the characteristic segregation. The physical mechanism responsible for the selective accumulation of sulfates in the tissue fluids of one form and of a preponderance of chlorides in the tissue fluids of the other form still remains to be worked out.

We have covered in mere outline the basic considerations underlying the Donnan equilibrium. For a further elaboration of the theory and a survey of the applications of the Donnan equilibrium to chemical, physiological, and technological processes, the student is referred to the most excellent review by Bolam.[17]

[12] J. A. Harris, J. V. Lawrence, and Z. W. Lawrence, *J. Agr. Research*, **28**, 695 (1924).

[13] J. A. Harris, C. T. Hoffman, and W. F. Hoffman, *J. Agr. Research*, **31**, 653 (1925).

[14] J. A. Harris, W. F. Hoffman, W. B. Sinclair, A. H. Johnson, and R. D. Evans, *J. Agr. Research*, **31**, 1027 (1925).

[15] J. A. Harris, W. F. Hoffman, and J. V. Lawrence, *Proc. Soc. Exptl. Biol. Med.* **22**, 350 (1925).

[16] J. A. Harris, *Proc. Soc. Exptl. Biol. Med.*, **22**, 415 (1925).

[17] T. R. Bolam, *Kolloid-Beihefte*, **39**, 140 (151 refs.) (1934).

II

PROTEINS

Da die Proteïnstoffe bei allen chemischen Prozessen im lebenden Organismus auf die eine oder andere Weise beteiligt sind, so darf man von der Aufklärung ihrer Struktur und ihrer Metamorphosen die wichtigsten Aufschlüsse für die biologische Chemie erwarten.

EMIL FISCHER (1906)

CHAPTER 11

The Amino Acids, the Primary Decomposition Products of Proteins

Early Investigations. With the remarkable development of organic chemistry which began early in the nineteenth century it is not surprising that the attention of the chemist was directed toward a study of the proteins, inasmuch as proteins are a major constituent of every living cell.

The methods which the organic chemist developed for the elementary analyses of C, H, N, O, and S were early applied, but it was impossible, owing to the complex nature of the proteins, to differentiate the proteins on the basis of their elementary analysis. Mulder [1] concluded that the radical $C_{16}H_{24}N_4O_5$ was common to all proteins; Liebig [2] postulated that the various proteins differed only in physical state and not in composition, and concluded that all plant and animal proteins were essentially similar.

However, the early chemists reverted to other procedures of the organic laboratory. Mann [3] discussed some of these earlier investigations at considerable length. When several proteins were oxidized in acid media with potassium permanganate, products such as formic, acetic, propionic, valeric, caproic, and benzoic acids were obtained. From an alkaline oxidation the same products were isolated, together with guanidine, butyric, and succinic acids. By using hydrogen peroxide with acid ferrous sulfate, workers obtained acetone and isovaleric acid. Bromination in an autoclave led to the isolation of bromoform, bromacetic acid, oxalic acid, leucine, aspartic acid, and bromanil. Suffice it to say that these and many other studies, similar to those which have been noted, showed the great complexity of the protein molecule, but did not contribute much in the way of establishing the essential chemical nature of the proteins or the structure of the protein molecule.

Now that we have an insight into the chemical nature of the proteins, and accordingly have some idea as to what products might be formed by

[1] G. J. Mulder, *J. prakt. Chem.*, **16**, 129 (1839).

[2] J. Liebig, *Ann.*, **39**, 129 (1841).

[3] G. Mann, *Chemistry of the Proteids*, Macmillan and Company, London, 1906.

the actions of the various reagents, it would be highly desirable to repeat many of these older studies in the light of modern information and modern technic. To some extent this has been done. Johnson [4] successfully nitrated various proteins; Johnson [5] and Pictet [6] subjected proteins to destructive distillation, and secured valuable data. Similarly, Vandevelde [7] studied the halogen derivatives of the proteins.

Amino Acids as Decomposition Products of Proteins. We now know the protein molecule to be made up primarily of amino acids linked together through the amino and carboxyl groups or through other reactive groups present in the amino acids concerned. The real clue to the structure of the proteins came from investigations in which hydrolytic methods were employed, the protein being hydrolyzed with acids, enzymes, or alkalies.

Apparently Braconnot [8] was the first investigator to use acid hydrolysis. He boiled gelatin and meat with dilute sulfuric acid and isolated what was later identified as glycine, although he did not at that time recognize that it contained nitrogen. However, this was the first instance in which an amino acid was shown to be a primary decomposition product of proteins. Earlier (1810) Wollaston had prepared cystine from urinary calculi, and Proust (1818) had found leucine in crystalline form in the holes of cheese, but the latter observation did not prove that the leucine was a unit of the protein molecule.

Braconnot's method of acid hydrolysis was not generally recognized as one that afforded a means by which primary protein decomposition products could be obtained, and no further amino acids were discovered until Liebig, in 1846, isolated a crystalline compound from casein which had been hydrolyzed by strong alkali. This compound was later identified as tyrosine. The next amino acid to be discovered was serine, isolated by Cramer by hydrolyzing silk with sulfuric acid.

Kühne, in 1867, introduced a new method for the study of proteins, *i.e., digestion with tryptic enzymes,* and identified tyrosine and leucine as primary decomposition products. Here again, however, the fact that tryptic digestion afforded a new technic for the study of proteins was not generally recognized.

[4] T. B. Johnson, *et al., J. Am. Chem. Soc.*, **37**, 1863, 2164, 2170, 2598 (1915); **38**, 1392 (1916).

[5] T. B. Johnson and P. G. Daschavsky, *J. Am. Chem. Soc.*, **41**, 1147 (1919).

[6] A. Pictet and M. Cramer, *Helv. Chim. Acta.*, **2**, 188 (1919).

[7] A. J. J. Vandevelde, *Rec. trav. chim.*, **43**, 158, 326, 702 (1924); **44**, 224, 900 (1925); **45**, 825 (1926); **46**, 133, 590 (1927).

[8] H. Braconnot, *Ann. chim. phys.*, [2] **13**, 113 (1820).

At about this time, Ritthausen [9] began his extensive series of investigations of the vegetable proteins, and in 1868 added aspartic acid and glutamic acid to the list of known amino acids. These acids were isolated from vegetable proteins by acid hydrolysis. Schützenberger and Bourgeois isolated alanine from silk by hydrolysis with barium hydroxide, and its presence was later confirmed by Weyl, who used acid hydrolysis. Schulze and Barbieri isolated phenylalanine from the sap of germinated seeds of *Lupinus luteus*, and, in 1889, Drechsel identified lysine as a constituent of proteins.

Thus, in the period from 1800 to 1890, eleven amino acids were isolated and identified. During the decade 1890–1900, three additional amino acids were isolated. In the decade 1900–1910, the discovery of eight amino acids was announced, of which tryptophan, proline, hydroxyproline, isoleucine, valine, and cystine, together with cysteine, are of rather common occurrence. β-Alanine has never been found in unaltered proteins but is present in carnosine and pantothenic acid.

In the decade 1910–1920, six new amino acids were added to the list. Since 1920, additional amino acids have been isolated from biological sources, and it is not unlikely that still other amino acids remain to be discovered in biological materials.

Table 38 lists the various amino acids, the year of their discovery, their common names, and their scientific names, with the names of the persons who announced their discovery. Not all these amino acids have been isolated from proteins, but all have been isolated from some biological material. Some are found in aqueous extracts from plant or animal tissues.

In 1941 Vickery [10] reclassified the amino acids on the basis of the criteria established ten years earlier by himself and Schmidt.[11] They proposed that an amino acid should be considered a component of proteins when it has been isolated by some worker other than the discoverer, when its constitution has been established by synthetic means, and when it has been derived from a pure protein. It is recognized that glycine, leucine, tyrosine, serine, glutamic and aspartic acids, phenylalanine, alanine, lysine, arginine, histidine, valine, proline, tryptophan, hydroxyproline, isoleucine, methionine, and threonine occur in many proteins. Cystine and cysteine are not included in the above list, but recent work on the denaturation of proteins indicates that one or both occur in many proteins. Certain amino acids, like thyroxine, iodogorgoic acid, and nor-

[9] H. Ritthausen, *Die Eiweisskörper der Getreidearten, Hülsenfrüchte und Ölsamen*, Max Cohen und Sohn, Bonn (1872).

[10] H. B. Vickery, *Ann. N. Y. Acad. Sci.*, **41**, 87 (1941).

[11] H. B. Vickery and C. L. A. Schmidt, *Chem. Revs.*, **9**, 169 (1931) (415 refs.).

TABLE 38. THE NATURALLY OCCURRING AMINO ACIDS

(The existence of a few of these amino acids is doubtful, as noted by "?")

Year	Common Name and Scientific Name (When Known)	Discoverer
1810	Cystine, β-β′dithio-di(α-aminopropionic acid)	Wollaston
1818	Leucine, β-isopropyl-α-aminopropionic acid	Proust
1820	Glycine or glycocoll, α-aminoacetic acid	Braconnot
1846	Tyrosine, β-(*p*-hydroxyphenyl)-α-amino-propionic acid	Liebig
1865	Serine, β-hydroxy-α-aminopropionic acid	Cramer
1868	Aspartic acid, α-aminosuccinic acid	Ritthausen
1868	Glutamic acid, α-aminoglutaric acid	Ritthausen
1875	Alanine, α-aminopropionic acid	Schützenberger and Bourgeois
1877	Ornithine, α-δ-diaminovaleric acid	Jaffe
1883	Phenylalanine, β-phenyl-α-aminopropionic acid	Schulze
1889	Lysine, α-ε-diaminocaproic acid	Drechsel
1895	Arginine, α-amino-δ-guanidinovaleric acid	Hedin
1896	Histidine, β-imidazole-α-aminopropionic acid	Kossel
1896	Iodo-gorgoic acid, 3,5-diiodotyrosine	Drechsel
1901	Tryptophan, β-indole-α-aminopropionic acid	Hopkins and Cole
1901	Proline, pyrrolidine-α-carboxylic acid	Fischer
1901	Cysteine, α-amino-β-thiolactic acid	Embden
1902	Hydroxy-proline, γ-hydroxypyrrolidine-α-carboxylic acid	Fischer
1905	Isoleucine, β-methyl-α-aminovaleric acid	Winterstein
1906	Valine, α-aminoisovaleric acid	Fischer
1907	Hydroxytryptophan [position of (—OH) uncertain (?)]	Abderhalden and Kempe
1908	β-Alanine, β-aminopropionic acid	Engeland
1913	Nor-leucine, α-aminocaproic acid	Abderhalden and Weyl
1913	α-Aminobutyric acid	Foreman
1913	Dopa, 3,4-dihydroxyphenylalanine	Guggenheim
1913	3,5-Dibromotyrosine	Mörner
1914	Citrulline, α-amino-δ-carbamidovaleric acid	Koga and Odake (Wada, 1930)
1919	Thyroxine, β-(3,5,3′,5′-tetra-iodo-4′-hydroxydi-phenyl-ether)-α-aminopropionic acid	Kendall
1922	Methionine, γ-methylthiol-α-aminobutyric acid	Mueller
1925	Hydroxylysine	Schryver
* 1925	$C_4H_{11}O_3N$ (?)	Gortner and Hoffman
* 1926	$C_4H_9O_3N$, hydroxyaminobutyric acid	Schryver and Buston
* 1935	Threonine, β-hydroxy-α-aminobutyric acid	Rose
1926	Hydroxyvaline	Schryver and Buston
1929	Canavanine, α-amino-γ-hydroxy-guanidino-butyric acid	Kitagawa and Tomiyama
1930	Norvaline, α-aminovaleric acid	Abderhalden and Bahn
1932	Canaline, α-amino-γ-(*o*)hydroxylaminobutyric acid	Kitagawa and Yamada
1933	Djenkolic acid, L-cysteine-thiolformacetal	van Veen and Hijman

* The compounds within the bracket are probably identical.

leucine, have been found in only a few places. Others like β-alanine, canavanine, and djenkolic acid, have never been found in proteins. However, the amino acids which have been actually isolated seldom total more than 70 per cent of the original protein.

Certain of the more recently discovered amino acids deserve special mention because they possess unusual organic structures. In 1929, Kitagawa and Tomiyama [12] isolated canavanine from the jack bean and showed that it possessed the structure NH_2—CH(=NH)—NH—O—CH_2—CH_2—CH(NH_2)—COOH. It is thus a guanidine derivative of an α-amino-γ-hydroxy acid in which the linkage between the guanidine group and the acid is through an oxygen; or it may be considered a derivative of hydroxylamine, for, when it is acted upon by alkalies, urea is split off and canaline is formed; canaline possesses the unusual structure of an amino group linked in the γ position through an oxygen to an amino acid.

Djenkolic acid was found by van Veen and Hijman [13] in the djenkol nut. This compound has the structure which would result from the condensation of two molecules of cysteine with one molecule of formaldehyde and the splitting off of water from the —SH groups and the oxygen of the formaldehyde. The structure has been confirmed by synthesis, and on hydrolysis djenkolic acid breaks down into cysteine and formaldehyde, as shown by Lillevik and Sandstrom.[14]

Two unusual compounds have been found in animal tissues. Asterubine was isolated from starfish by Ackermann [15] and shown to have the formula $(CH_3)_2$=N—C(=NH)—NH—$(CH_2)_2SO_3H$. This is the first instance in which a sulfonic acid has been found in natural products. Octopine has been isolated by Moore and Wilson from octopus muscle and appears to be composed of alanine and arginine which share one —NH— group in place of their respective amino groups. Penicillamine, β-β-dimethyl cysteine, has been found to be one of the components of the several kinds of penicillin.

Development of the Present View of the Nature of Proteins. As already indicated, amino acids are the primary decomposition products of proteins. This fact was established by the work of Emil Fischer who contributed so extensively to this field during the period from 1889 to 1918. He was preeminent not alone in the field of proteins but also in the study of carbohydrates, the purine and pyrimidine derivatives,

[12] M. Kitagawa and T. Tomiyama, *J. Biochem. (Tokyo)*, **11**, 265 (1929).

[13] A. G. van Veen and A. J. Hijman, *Geneesk. Tijdschr. Nederland.-Indië*, **73**, 991 (1933); *Rec. trav. chim.*, **54**, 493 (1935).

[14] H. A. Lillevik and W. M. Sandstrom, *J. Am. Chem. Soc.*, **63**, 1028 (1941).

[15] D. Ackermann, *Z. physiol. Chem.*, **232**, 206 (1935).

TABLE 38. THE NATURALLY OCCURRING AMINO ACIDS

(The existence of a few of these amino acids is doubtful, as noted by "?")

Year	Common Name and Scientific Name (When Known)	Discoverer
1810	Cystine, β-β′dithio-di(α-aminopropionic acid)	Wollaston
1818	Leucine, β-isopropyl-α-aminopropionic acid	Proust
1820	Glycine or glycocoll, α-aminoacetic acid	Braconnot
1846	Tyrosine, β-(*p*-hydroxyphenyl)-α-amino-propionic acid	Liebig
1865	Serine, β-hydroxy-α-aminopropionic acid	Cramer
1868	Aspartic acid, α-aminosuccinic acid	Ritthausen
1868	Glutamic acid, α-aminoglutaric acid	Ritthausen
1875	Alanine, α-aminopropionic acid	Schützenberger and Bourgeois
1877	Ornithine, α-δ-diaminovaleric acid	Jaffe
1883	Phenylalanine, β-phenyl-α-aminopropionic acid	Schulze
1889	Lysine, α-ε-diaminocaproic acid	Drechsel
1895	Arginine, α-amino-δ-guanidinovaleric acid	Hedin
1896	Histidine, β-imidazole-α-aminopropionic acid	Kossel
1896	Iodo-gorgoic acid, 3,5-diiodotyrosine	Drechsel
1901	Tryptophan, β-indole-α-aminopropionic acid	Hopkins and Cole
1901	Proline, pyrrolidine-α-carboxylic acid	Fischer
1901	Cysteine, α-amino-β-thiolactic acid	Embden
1902	Hydroxy-proline, γ-hydroxypyrrolidine-α-carboxylic acid	Fischer
1905	Isoleucine, β-methyl-α-aminovaleric acid	Winterstein
1906	Valine, α-aminoisovaleric acid	Fischer
1907	Hydroxytryptophan [position of (—OH) uncertain (?)]	Abderhalden and Kempe
1908	β-Alanine, β-aminopropionic acid	Engeland
1913	Nor-leucine, α-aminocaproic acid	Abderhalden and Weyl
1913	α-Aminobutyric acid	Foreman
1913	Dopa, 3,4-dihydroxyphenylalanine	Guggenheim
1913	3,5-Dibromotyrosine	Mörner
1914	Citrulline, α-amino-δ-carbamidovaleric acid	Koga and Odake (Wada, 1930)
1919	Thyroxine, β-(3,5,3′,5′-tetra-iodo-4′-hydroxydi-phenyl-ether)-α-aminopropionic acid	Kendall
1922	Methionine, γ-methylthiol-α-aminobutyric acid	Mueller
1925	Hydroxylysine	Schryver
1925	$C_4H_{11}O_3N$ (?)	Gortner and Hoffman
*1926	$C_4H_9O_3N$, hydroxyaminobutyric acid	Schryver and Buston
1935	Threonine, β-hydroxy-α-aminobutyric acid	Rose
1926	Hydroxyvaline	Schryver and Buston
1929	Canavanine, α-amino-γ-hydroxy-guanidino-butyric acid	Kitagawa and Tomiyama
1930	Norvaline, α-aminovaleric acid	Abderhalden and Bahn
1932	Canaline, α-amino-γ-(*o*)hydroxylaminobutyric acid	Kitagawa and Yamada
1933	Djenkolic acid, L-cysteine-thiolformacetal	van Veen and Hijman

* The compounds within the bracket are probably identical.

leucine, have been found in only a few places. Others like β-alanine, canavanine, and djenkolic acid, have never been found in proteins. However, the amino acids which have been actually isolated seldom total more than 70 per cent of the original protein.

Certain of the more recently discovered amino acids deserve special mention because they possess unusual organic structures. In 1929, Kitagawa and Tomiyama [12] isolated canavanine from the jack bean and showed that it possessed the structure NH_2—CH(=NH)—NH—O—CH_2—CH_2—CH(NH_2)—COOH. It is thus a guanidine derivative of an α-amino-γ-hydroxy acid in which the linkage between the guanidine group and the acid is through an oxygen; or it may be considered a derivative of hydroxylamine, for, when it is acted upon by alkalies, urea is split off and canaline is formed; canaline possesses the unusual structure of an amino group linked in the γ position through an oxygen to an amino acid.

Djenkolic acid was found by van Veen and Hijman [13] in the djenkol nut. This compound has the structure which would result from the condensation of two molecules of cysteine with one molecule of formaldehyde and the splitting off of water from the —SH groups and the oxygen of the formaldehyde. The structure has been confirmed by synthesis, and on hydrolysis djenkolic acid breaks down into cysteine and formaldehyde, as shown by Lillevik and Sandstrom.[14]

Two unusual compounds have been found in animal tissues. Asterubine was isolated from starfish by Ackermann [15] and shown to have the formula $(CH_3)_2$=N—C(=NH)—NH—$(CH_2)_2SO_3H$. This is the first instance in which a sulfonic acid has been found in natural products. Octopine has been isolated by Moore and Wilson from octopus muscle and appears to be composed of alanine and arginine which share one —NH— group in place of their respective amino groups. Penicillamine, β-β-dimethyl cysteine, has been found to be one of the components of the several kinds of penicillin.

Development of the Present View of the Nature of Proteins. As already indicated, amino acids are the primary decomposition products of proteins. This fact was established by the work of Emil Fischer who contributed so extensively to this field during the period from 1889 to 1918. He was preeminent not alone in the field of proteins but also in the study of carbohydrates, the purine and pyrimidine derivatives,

[12] M. Kitagawa and T. Tomiyama, *J. Biochem. (Tokyo)*, **11**, 265 (1929).

[13] A. G. van Veen and A. J. Hijman, *Geneesk. Tijdschr. Nederland.-Indië*, **73**, 991 (1933); *Rec. trav. chim.*, **54**, 493 (1935).

[14] H. A. Lillevik and W. M. Sandstrom, *J. Am. Chem. Soc.*, **63**, 1028 (1941).

[15] D. Ackermann, *Z. physiol. Chem.*, **232**, 206 (1935).

and the tannins. Probably no one person has influenced the fields of biochemistry and organic chemistry more than Emil Fischer.

His work began on the proteins at a time when nine monoamino acids and four basic amino acids were known. Eight of the monoamino acids had previously been synthesized and their structures thus definitely proved. Serine and the basic amino acids had not been synthesized. It is an axiom that, until an organic compound has been synthesized by known chemical reactions, its structure is still uncertain. Fischer,[16] in 1906, presented an outline of the plan which he developed for the study of the proteins. The first item of the outline was these studies of syntheses.

As a second field, he studied amino acid derivatives, including the preparation and study of the esters and acetyl derivatives, the phenylisocyanates, etc., with a view to finding chemical derivatives which would permit the separation of mixtures of amino acids. He then proceeded to the experimental separation of mixtures of amino acids, such as result from the hydrolysis of unaltered proteins.

Simultaneously with studies on the separation of the amino acids, he attempted to recombine two or more amino acids into compounds which he designated by the name of "polypeptides" and studied the chemical and physical properties of such derivatives, particularly in their relation to protein structure. These polypeptides were subjected to hydrolysis, were tested as to their behavior toward enzyme action, and in various ways contributed much to our knowledge of the chemistry of the proteins.

He further investigated the proteins themselves. He was interested in the problem of protein structure, in a critical study of the various methods for protein hydrolysis, he conducted extensive series of analyses of various proteins, and he studied in a general way the protein molecule as an entity. Remarkable as it may seem, he completed very satisfactorily and in a most extraordinary manner the entire program which he planned, and today we are indebted to his technics for much of the available information on the organic chemistry of the protein molecule.

Four other workers may be credited with laying the foundation of the modern viewpoint. Kossel, during the period from 1895 to 1915, made major contributions to protein chemistry in his study of the basic amino acids, including the Kossel method for their quantitative separation and identification (see p. 338). He contributed much to the general field of protein chemistry but was particularly interested in the histones and the protamines, proteins characterized by a high content of basic amino acids.

[16] E. Fischer, *Ber.*, **39**, 530 (1906).

Abderhalden received his early training in the field of proteins under the guidance of Emil Fischer and is probably the most prolific research worker in the field. Since 1904 he has published literally hundreds of papers dealing with one or another phase of protein chemistry, in addition to many papers in other fields. No consideration of workers in the field of proteins would be complete without the acknowledgement of the advances made by Abderhalden and his students.

As already noted, Ritthausen early began an investigation of the vegetable proteins. Owing to the lack of organized technic and any definite knowledge of protein structure, Ritthausen's contributions, though important, were inadequate.

Thomas B. Osborne began work with vegetable proteins about 1891 and from that time until his death in 1929 was recognized as the outstanding authority in this field of protein research. It is to Osborne that we owe many improvements in the methods of protein analysis and much of our information in regard to the methods of isolation and purification of the vegetable proteins.

During the nineteenth century, organic chemistry dominated chemical science. Toward the end of the nineteenth century, however, physical chemistry began more and more to assume a supremacy. It is not surprising, therefore, that many of the more important recent contributions in the field of protein study should be characterized by the application of physicochemical methods. In 1909, S. P. L. Sörensen demonstrated the importance of hydrogen-ion concentration to biological and biochemical reactions, and, in 1917, he published a series of papers [17] in which the most exact physicochemical technics, including the influence of hydrogen-ion concentration, were applied to the problems of protein behavior. In the intervening years since that time there have come from his laboratory numerous contributions dealing with the physicochemical properties of protein systems, and because many workers in other laboratories are now utilizing physicochemical technics, it seems fitting to include Sörensen in the list of workers in the field of proteins who have given us new tools and new concepts.

Amino Acids as Zwitterions. Apparently the first suggestion that amphoteric compounds (compounds which contain both acidic and basic groups) exist in a special state was made by Adams [18] in 1916. He pointed out that aminoacetic acid exists almost exclusively as the inner salt, $^{+}NH_3—CH_2—CO_2^{-}$, with probably less than one part per million of the true amino acid, $NH_2—CH_2—COOH$. Bjerrum,[19] in a more com-

[17] S. P. L. Sörensen, *Compt. rend. lab. Carlsberg*, **12** (1917).

[18] E. C. Adams, *J. Am. Chem. Soc.*, **38**, 1503 (1916).

[19] N. Bjerrum, *Z. physik. Chem.*, **104**, 147 (1923).

plete discussion of the constitution of ampholytes, confirmed and extended Adams' theory and pointed out that all the amino acids exist in salt-like dipolar ions, $^{+}NH_3RCO_2^{-}$, and that therefore they are not true amino acids but rather ionized internal ammonium salts. Bjerrum coined the name *zwitterion* for this particular type of substance.

In 1930, Harris [20] extended the zwitterion theory, and later investigations of C. L. A. Schmidt, E. J. Cohn, and others have firmly established it. The zwitterion theory has materially changed our conceptions of the relationships regarding the dissociation constants of the amino acids. We have already seen that the dissociation constants of acids and bases are the product of the concentration of the anion and the cation divided by the concentration of the undissociated acid. These constants have been expressed by K_a and K_b and are usually referred to as apparent dissociation constants.

Bjerrum pointed out that in the amino acids these constants were not true dissociation constants but rather hydrolysis constants arising from the hydrolysis of the internal salt. He suggested that the real dissociation constants of the amino acids could be expressed by K_A and K_B where

$$K_A = \frac{K_w}{K_b} \tag{147}$$

and

$$K_B = \frac{K_w}{K_a} \tag{148}$$

The inner-salt nature of amino acids is abundantly proved by many of the properties of the solutions, which are essentially neutral in reaction. The amino acids, in general, are insoluble in organic solvents, are soluble in water, are often more soluble in salt solutions than in pure water, and they melt with decomposition in the range from 217° to 344°C. These properties are characteristic of true salts rather than of an un-ionized amino acid.

Several additional lines of evidence point to the zwitterion state. Formaldehyde produces a shift in the titration curve of an amino acid in the alkaline range, but not in the acid region. It is known that formaldehyde reacts with amines but does not react with the carboxyl group. It can be concluded that the amino group is involved in the titration with a base, and hence the amino acid must exist as a zwitterion. This principle is the basis of the formol titration (see p. 322).

[20] L. J. Harris, *Biochem. J.*, **24**, 1080 (1930); (with T. W. Birch) *Biochem. J.*, **24**, 564, 1086 (1930).

Bull [21] points out that the pK_a (the pH at which half of the acid present exists as its salt) calculated for glycine from its apparent dissociation constant would be 9.60, which is improbable when compared with the value of 4.73 for acetic acid. On the basis of the zwitterion structure the true pK_a is 2.31, which is in the expected range.

If the amino acids exist as dipolar ions one would expect them to show high dipole moments (the product of the charge times the distance between the charges). Because the amino acids are insoluble in non-polar solvents, direct determinations of their dipole moments cannot be made. However, solutions of amino acids in water possess dielectric constants higher than that of water. It is found [22] that the dielectric constant D in all solutions is a linear function of the concentration C of the amino acid expressed as moles per liter.

$$D = D_0 + \delta C \tag{149}$$

where D_0 is the dielectric constant for water. The increment δ is a constant, characteristic of the particular amino acid. This increment ranges from 22 to 28 for the α-amino acids and increases when the amino group is further removed from the carboxyl group as in the case in β-alanine (34.6) and in glycyl-glycine (70.6). This phenomenon could be displayed only by substances of a dipolar nature.

Studies with the Raman spectra also lend support to the existence of zwitterions. When a solution is strongly illuminated most of the light is transmitted unchanged. A small amount is scattered in other directions by the molecules of the solution. The scattered light is examined in a spectrograph to determine the frequency shifts which lie in the range 200 to 4,000 cm.$^{-1}$. It is observed that the spectrum for acetic acid [23] gives a band near 1,700 which is absent in sodium acetate. Glycine hydrochloride gives a band similar to that of acetic acid, whereas glycine, like sodium acetate, shows no band in the 1,700 cm.$^{-1}$ region (*cf.* Fig. 76). It is to be remembered that the particular frequencies observed are characteristic of groups bound by covalent linkages; thus the carboxyl group is essentially intact (un-ionized) in acetic acid and in the hydrochloride salt of glycine, whereas the un-ionized carboxyl group is not present in sodium acetate and in glycine.

The zwitterion theory changes decidedly the viewpoint on the reactions of amino acids with acids and bases. Thus, in the older view-

[21] H. B. Bull, *Physical Biochemistry*, John Wiley & Sons, New York, 1943.

[22] E. J. Cohn and J. J. Edsall, *Proteins, Amino Acids and Peptides*, Reinhold Pub. Corp., New York, 1943.

[23] J. T. Edsall, *J. Chem. Phys.*, **4**, 1 (1936); *Cold Spring Harbor Symposia Quant. Biol.*, **6**, 40 (1938).

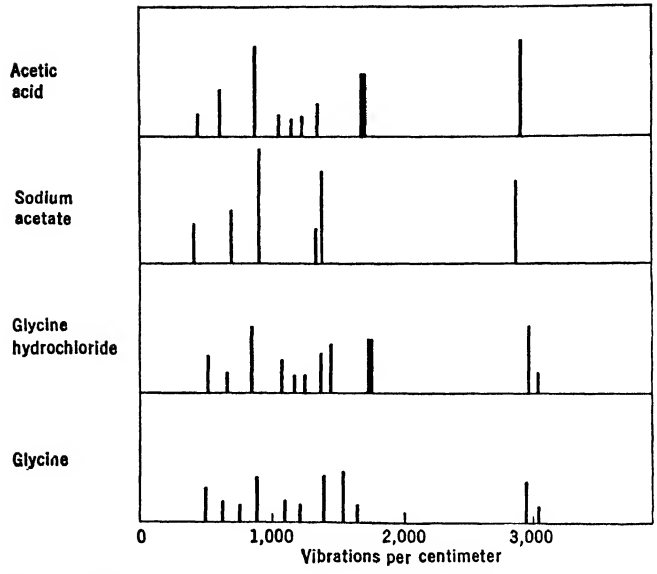


FIG. 76. Raman spectra of acetic acid and its sodium salt, and glycine and its hydrochloride.

point the reaction of glycine with hydrochloric acid would have been written

$$NH_2CH_2COOH + HCl \rightleftharpoons {}^{+}NH_3CH_2COOH + Cl^{-}$$

whereas under the zwitterion theory the reaction would be expressed

$${}^{+}NH_3CH_2COO^{-} + HCl \rightleftharpoons {}^{+}NH_3CH_2COOH + Cl^{-}$$

and, in bases, the reaction

$$NH_2CH_2COOH + NaOH \rightleftharpoons NH_2CH_2COO^{-} + Na^{+}$$

would be altered to

$${}^{+}NH_3CH_2COO^{-} + NaOH \rightleftharpoons NH_2CH_2COO^{-} + Na^{+}$$

It will be noted that the net result so far as the nature of the final product is concerned is the same, but that there are fundamental differences in the mechanism of the reactions. Under the older theories the acids reacted with the amino group to form an ammonium salt. Under the zwitterion theory the acid reacts to set free, from a salt, the carboxyl group of a weak acid. Similarly under the old viewpoint the strong base reacted with the acid group to form a salt, whereas under the zwitterion theory the strong base liberates the basic group of the weaker base in the salt in the same way that sodium hydroxide liberates ammonia from

ammonium acetate. *Fundamentally, therefore, acids react with the acidic groups of amino acids and proteins, and bases react with the basic groups of amino acids and proteins.*

Kirk and Schmidt [24] have summarized the literature on the apparent dissociation constants of the amino acids and have discussed the theories involved. In Table 39 are given the data for the amino acids which they regard as of greatest validity.

TABLE 39. APPARENT DISSOCIATION CONSTANTS OF THE AMINO ACIDS

Compound	Temperature, °C.	K_a	K_b
Alanine	25	1.9×10^{-10}	2.3×10^{-12}
β-Alanine	25	6.5×10^{-11}	4.0×10^{-11}
Arginine	25	3.3×10^{-13}	(1) 1.1×10^{-5}
			(2) 1.1×10^{-12}
Aspartic acid	25	(1) 2.0×10^{-4}	1.0×10^{-12}
		(2) 2.0×10^{-10}	
Cysteine	30	(1) *ca.* 7×10^{-9}	*ca.* 8×10^{-13}
		(2) *ca.* 5×10^{-11}	
Cystine	30	(1) 3.3×10^{-8}	(1) 5×10^{-13}
		(2) 9.6×10^{-10}	(2) $< 1.5 \times 10^{-13}$
Glutamic acid	25	(1) 6.0×10^{-5}	1.4×10^{-12}
		(2) 2.5×10^{-10}	
Glycine	25	1.8×10^{-10}	2.6×10^{-12}
Histidine	25	3.9×10^{-10}	(1) 1.2×10^{-8}
			(2) 2.9×10^{-13}
Hydroxyproline	25	1.9×10^{-10}	8.3×10^{-13}
Isoleucine	25	2.1×10^{-10}	2.3×10^{-12}
Leucine	25	2.5×10^{-10}	2.3×10^{-12}
Lysine	25	3.0×10^{-11}	(1) 2.0×10^{-5}
	25		(2) 1.0×10^{-12}
Norleucine	25	1.7×10^{-10}	2.5×10^{-12}
Norvaline	25	1.9×10^{-10}	2.3×10^{-12}
Ornithine	25	1.7×10^{-11}	(1) 4.5×10^{-6}
			(2) 8.7×10^{-13}
Phenylalanine	25	7.5×10^{-10}	4.0×10^{-13}
Proline	25	2.5×10^{-11}	1.0×10^{-12}
Serine	25	7.1×10^{-10}	1.6×10^{-12}
Tryptophan	25	4.1×10^{-10}	2.2×10^{-12}
Tyrosine	25	(1) 7.0×10^{-10}	1.7×10^{-12}
		(2) 7.0×10^{-11}	
Valine	25	2.3×10^{-10}	2.0×10^{-12}

[24] P. L. Kirk and C. L. A. Schmidt, *Univ. Calif. Pubs. Physiol.*, **7**, 57 (1929).

The Isoelectric Point of Amino Acids and Proteins. We have seen that in the presence of strong acids the amino acid residue exists largely as a cation and in the presence of strong bases it exists as an anion. In pure water, cations, anions, and the zwitterion are all present, with the latter in great excess. Accordingly there must be some hydrogen-ion concentration at which the zwitterion has a maximum concentration and at which the sum of the anions plus the cations is minimal. This hydrogen-ion concentration has been defined as the isoelectric point, and in terms of electrical transport it may be defined as *that hydrogen-ion concentration at which there will be a tendency for as many cations to migrate toward the cathode as there are anions migrating toward the anode.* This point will be reached when

$$-\frac{K_a}{[H^+]^2} + \frac{K_b}{K_w} = 0 \tag{150}$$

or

$$[H^+]_{I.P.} = \sqrt{\frac{K_a}{K_b} K_w} \tag{151}$$

where $[H^+]_{I.P.}$ = the hydrogen-ion concentration at which the ampholyte is isoelectric.

This definition of the isoelectric point was originally proposed by Michaelis [25] and has been accepted and somewhat expanded by Eckweiler, Noyes, and Falk [26] and by Levene and Simms.[27] Levene and Simms point out that, although equation (151) was originally developed for simple monobasic and monoacidic ampholytes, it has a wider application and can be applied to complex ampholytes, such as proteins, by taking into consideration the *strongest acid group* and the *strongest basic group* of the complex ampholyte. For a complex ampholyte the isoelectric point would be defined as

$$[H^+]_{I.P.} = \sqrt{\left(\frac{K_{a1} + K_{a2} + K_{a3} \cdots + K_{an}}{K_{b1} + K_{b2} + K_{b3} \cdots + K_{bn}}\right) K_w} \tag{152}$$

or

$$[H^+]_{I.P.} = \sqrt{\frac{\Sigma K_a}{\Sigma K_b} K_w} \tag{153}$$

It is obvious that, if the second, third, etc., dissociation constants contribute only small amounts to the numerator and the denominator

[25] L. Michaelis, *Die Wasserstoffionenkonzentration*, Julius Springer, Berlin, 1914
[26] H. Eckweiler, Helen M. Noyes, and K. G. Falk, *J. Gen. Physiol.*, **3**, 291 (1921)
[27] P. A. Levene and H. S. Simms, *J. Biol. Chem.*, **55**, 801 (1923).

of equation (152), equation (151), where only the primary dissociations are considered, is approximated.

In this consideration of the isoelectric point we have used the "apparent dissociation constants" and considered the ampholyte to be isoelectric at that hydrogen-ion concentration where it is ionized equally as an acid and as a base and where the isoelectric point may be considered as the point of *minimum* dissociation. As a matter of fact, the zwitterion theory leads to an exactly opposite viewpoint and states that *the isoelectric point is at that hydrogen-ion concentration where the ampholyte exists to the maximum degree as a zwitterion, and where the positive charges on the zwitterion exactly balance the negative charges on the zwitterion.* It may seem somewhat of a paradox that equations (151) and (153) are valid under either consideration and that the isoelectric point calculated by these equations is actually the true isoelectric point of the ampholyte. It should be remembered, however, that the apparent dissociation constants, K_a and K_b, are in reality hydrolysis constants; this fact explains why these equations are valid when the ampholyte is considered a zwitterion.

The Reactive Groups in Proteins. In all his studies on the proteins, Fischer emphasized the reactions which take place between a primary amino group and the carboxyl group of an organic acid. In his synthesis of the polypeptides, he formed the peptide linkage

$$RNH_2 + R'COOH \rightarrow RNH{-}COR' + H_2O$$

and it was definitely proved to occur in the unaltered proteins. Probably because Fischer emphasized this particular linkage, attention has been focused by research workers almost entirely upon the amino and the carboxyl groups of the proteins, in an attempt to explain protein structure and protein reactions. There is a certain amount of evidence, however, that even though the amino groups and carboxyl groups are of primary importance, groups and linkages other than the peptide linkage may and probably do occur in the unaltered proteins. A consideration of the reactive groups present in the various amino acids should indicate to any organic chemist the possibility of other reactions and other linkages. The following groups are present in at least one of the amino acids noted in Table 38.

1. Primary amino group (—NH_2).
2. Carboxyl group (—COOH), especially in the dicarboxylic acids.
3. Aliphatic alcohol group (—OH).
4. Phenolic group (—OH), (aromatic alcohols).

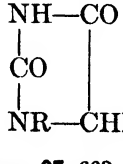
5. Alcohol group intermediate between the aliphatic (—OH) and the aromatic (—OH), as in hydroxyproline.
6. Imino group (=NH).
7. Acid amide group (—CO—NH_2).
8. Sulfhydryl group (—SH) in cysteine, or the disulfide group (—S—S—) of cystine.
9. α-Hydrogen of tryptophan.
10. The guanidine nucleus.

Several papers in the literature indicate the presence of methoxy (—OCH_3) and N-methyl groups in proteins. However, the discovery of methionine which contains an —S—CH_3 group has apparently accounted [28] for all the methyl groups reported by the earlier workers.

That the imino group of proline does enter into combination is shown by the presence of polypeptides such as glycyl-L-proline where the carboxyl group of glycine is united to the imino group of proline. Bergmann [29] states that more than 25 per cent of the "peptide" linkages in gelatin and collagen are of this type and that this linkage is split by a specific proteolytic enzyme [30] which occurs in erepsin but not in pancreatic juice.

Theoretically the ether linkage is a possibility. The presence of such a linkage, except in the case of thyroxine, still remains to be proved, although the —NH—O—CH_2— linkage in canavanine may be looked upon as a pseudo ether. The —OH groups of the hydroxyamino acids do play an important role in protein structure, since, when they are esterified with phosphoric acid in the diester type, the phosphoric acid forms a bridge between two polypeptide chains. Linkages of this sort are all in the phosphoproteins, *e.g.*, casein where hydroxyamino acids, such as serine, are involved.

Dunn [31] points out that there is a liberation of carbon dioxide during the acid hydrolysis of a protein, amounting in casein to 0.71 per cent of the weight of the casein. He suggests that the sources of this carbon dioxide may be uramino acids, hydantoins, or cyclic diacipiperazines, possibly in such structures as



[28] H. D. Baernstein, *J. Biol. Chem.*, **97**, 663 (1932).
[29] M. Bergmann, *J. Biol. Chem.*, **110**, 471 (1935).
[30] M. Bergmann and J. S. Fruton, *Science*, **83**, 306 (1936).
[31] M. S. Dunn, *J. Am. Chem. Soc.*, **47**, 2564 (1925).

Ssadikow [32] had earlier observed the formation of carbon dioxide when proteins were hydrolyzed in an autoclave but assumed that it was due to oxidation processes inasmuch as no oxygen remained in the residual gases in the autoclave. Incidentally his studies yielded another observation which still remains to be explained. In a second paper [33] he observed that, when proteins were hydrolyzed in an autoclave in a nitrogen atmosphere, the nitrogen was in some manner "fixed" and disappeared from the gas phase.

Synthesis of Amino Acids. Six general types of reactions have been utilized for the synthesis of α-amino acids.

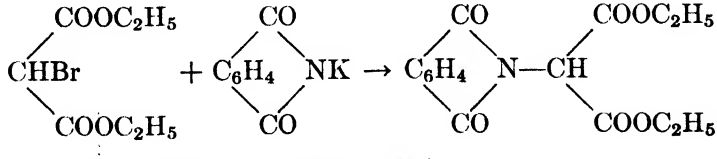
1. The synthesis of an α-amino acid from an *α-halogen acid* and ammonia:

$$RCH(Br)COOH + 2NH_3 \rightarrow RCH(NH_2)COOH + NH_4Br$$

There are several variations of this method. The α-bromo acids are most often used since they are readily prepared and are more reactive than are the α-chloro acids; also the ammonium bromide is the more easily removed by aqueous alcohol. However, monochloroacetic acid is generally used to prepare glycine. Concentrated ammonia is employed in great excess in order to minimize the formation of secondary and tertiary amine derivatives.

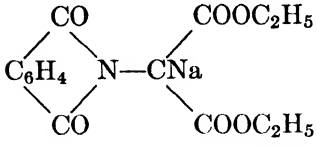
This reaction, as a rule, works smoothly; the chief difficulty lies in securing the proper parent acid. At times this can be achieved by using the appropriate alkyl derivative of malonic ester (see p. 292). Marvel [34] has applied this modification to recent syntheses of valine, leucine, isoleucine, and phenylalanine.

In order to insure the formation of the primary amino group, the Gabriel synthesis may be employed. The sodium or potassium salt of phthalimide reacts readily with a halogenated acid to yield the corresponding phthalimido derivative, from which the phthalic acid is removed by hydrolysis. Sörensen [35] combined the phthalimide method of amination with the malonic ester synthesis. Bromomalonic ester was combined with phthalimide potassium to form phthalimidomalonic ester:

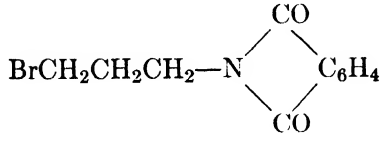


[32] W. S. Ssadikow, *Biochem. Z.*, **143**, 492 (1923).
[33] W. S. Ssadikow, *Biochem. Z.*, **143**, 496 (1923).
[34] C. S. Marvel, *Org. Syntheses*, **21**, 60, 74, 99, 106 (1941).
[35] S. P. L. Sörensen, *Compt. rend. trav. lab. Carlsberg*, **6**, 1 (1903).

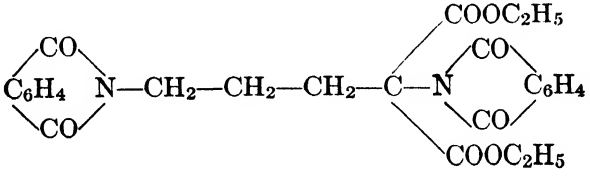
This compound was converted into phthalimidosodium malonic ester,



which was combined with γ-bromopropylphthalimide,



yielding γ-phthalimidopropylphthalimidomalonic ester,



The ester groups were split off by saponification, the free acid lost carbon dioxide on heating, and the phthalic acid residues were removed by acid hydrolysis, forming the desired *ornithine* in good yield.

2. The synthesis from an *aldehyde* having one less carbon than the desired amino acid, by the addition of hydrocyanic acid and ammonia, with the subsequent saponification of the nitrile:

$$RCHO + HCN \rightarrow RCH(OH)CN$$

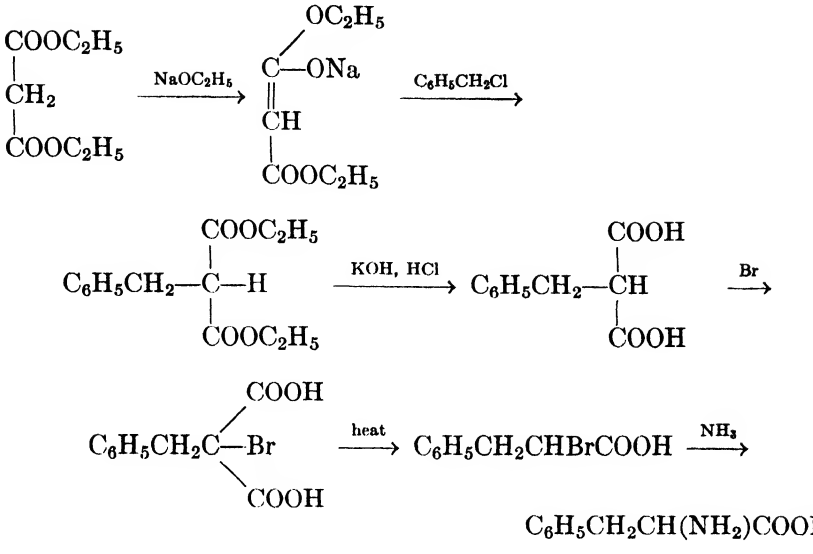
$$RCH(OH)CN + NH_3 \rightarrow RCH(NH_2)CN + H_2O$$

$$RCH(NH_2)CN + 2H_2O \rightarrow RCH(NH_2)COOH + NH_3$$

The reaction is carried out in various ways, particularly in the order of adding the cyanide and ammonia. Most aldehydes are not readily available, except formaldehyde for the synthesis of glycine.[36]

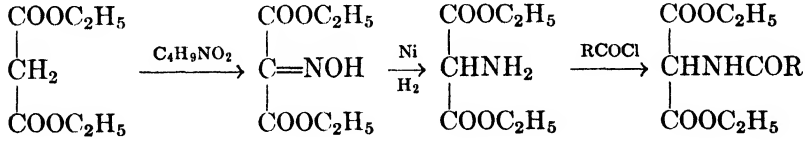
3. The syntheses from *malonic ester* depend upon the fact that the hydrogens of the —CH_2— group are reactive and may be replaced in turn by alkyl or aryl radicals and by the amino group after first introducing sodium and a halogen, respectively.

[36] W. K. Anslow and H. King, *Org. Syntheses*, **1**, 292 (1932).



The above reaction is that employed by Marvel.[34] The sequence of steps may be varied in the several preparations as illustrated by Sörensen's synthesis of ornithine, cited earlier, where phthalimide was employed to introduce the amino group before the alkyl group was introduced.

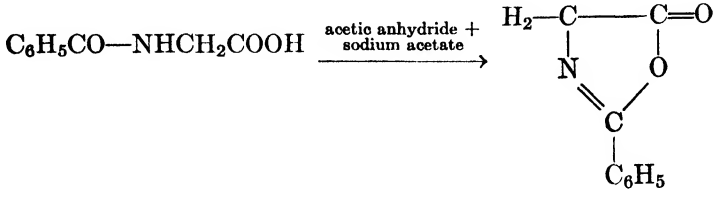
Because the alkylation of malonic ester with low molecular weight halides often gives disubstituted products, recent workers have used an acylamino malonic ester. The first of these was the phthalimido derivative used by Sörensen. Redemann and Dunn [37] and Painter [38] employed the benzoylamino malonic ester; whereas Dakin,[39] as well as Snyder and co-workers,[40] used the corresponding acetyl derivative. The acylamino malonic ester is prepared thus:



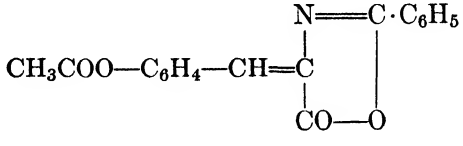
The introduction of the desired alkyl or aryl radical is then effected as in any malonic ester synthesis.

[37] C. E. Redemann and M. S. Dunn, *J. Biol. Chem.*, **130**, 341 (1939).
[38] E. P. Painter, *J. Am. Chem. Soc.*, **62**, 232 (1940).
[39] H. D. Dakin, *J. Biol. Chem.*, **154**, 549 (1944).
[40] H. R. Snyder, J. F. Shekleton, and C. D. Lewis, *J. Am. Chem. Soc.*, **67**, 310 (1945).

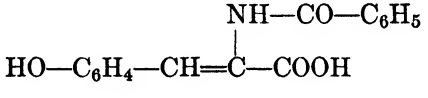
4. The *hippuric acid* or azlactone synthesis, first introduced by Erlenmeyer in 1882, has been improved by Lamb and Robson.[41] When hippuric acid is heated with sodium acetate and acetic anhydride, an azlactone is formed in which the $-CH_2-$ group is very reactive.



This azlactone condenses readily with an aromatic aldehyde to form a substituted azlactone; such glycine derivatives as the hydantoin and the diketopiperazine react similarly to the azlactone,



which on hydrolysis with sodium hydroxide yields a substituted benzoylaminoacrylic acid:



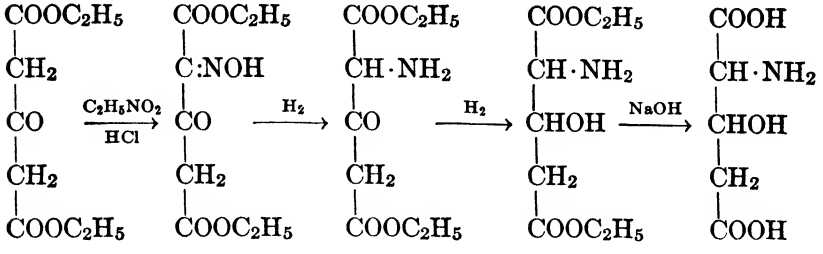
and this on reduction yields benzoyl-DL-tyrosine from which the benzoyl group can be removed by acid hydrolysis. Yields by the hippuric acid method are good, and it is a suitable method for the synthesis of phenylalanine, tyrosine, tryptophan, serine, etc.

5. The *oximino* synthesis. Isoleucine, leucine, aspartic, glutamic, and hydroxyglutamic acids have been synthesized by the reduction of the appropriate oximino compound. In the synthesis of hydroxyglutamic acid, Harington and Randall [42] treated the ester of acetone dicarboxylic acid (β-ketoglutaric acid) with ethyl nitrite to form the corresponding α-isonitroso derivative. This was then catalytically reduced in two stages, a palladium-charcoal catalyst being used and in the second

[41] J. Lamb and W. Robson, *Biochem. J.*, **25**, 1231 (1931).

[42] C. R. Harington and S. S. Randall, *Biochem. J.*, **25**, 1917 (1931).

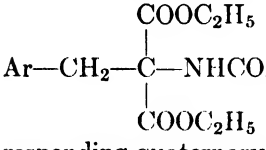
stage a platinum catalyst being added. On saponification of the ester, *i*-hydroxyglutamic acid resulted.



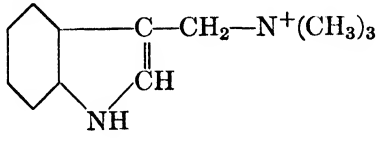
β-Hydroxyglutamic acid has not been found among the products of a protein hydrolysate, but this synthesis is a recent illustration of the principle. If available, α-keto acids may be converted to their oximes or phenylhydrazones which are then reduced. The latter reaction is somewhat difficult to carry out in good yields.

6. Other syntheses. In addition to those noted, a number of special syntheses have been used for the preparation of some particular amino acid. Those amino acids which contain the guanidine residue are usually prepared by the action of cyanamide upon the corresponding amine. Thus, cyanamide reacting on ornithine would produce arginine.

Tryptophan has recently been synthesized by an ingenious method. Snyder and Smith [43] found that quaternary ammonium compounds like $Ar—CH_2N^+(CH_3)_3$ will alkylate acylamino malonic ester to yield



Indole could readily be converted to the corresponding quaternary compound, gramine methiodide,

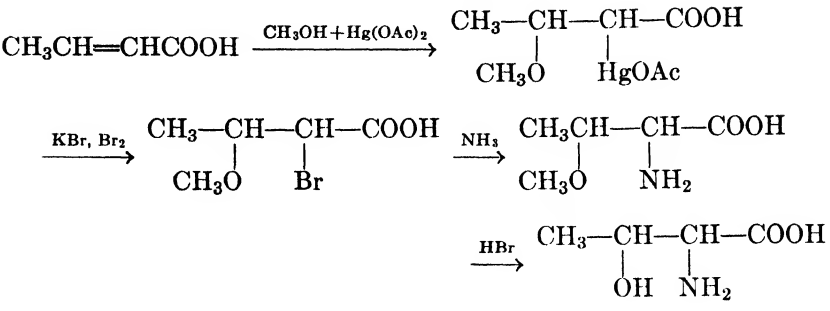


by the Mannich reaction of dimethylamine plus formaldehyde upon indole. From this point the gramine methiodide is processed through the malonic ester synthesis. A similar procedure is reported by Albertson [44] and others.

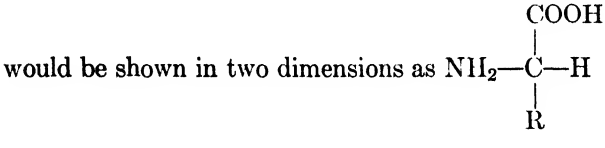
[43] H. R. Snyder and C. W. Smith, *J. Am. Chem. Soc.*, **66**, 350 (1944).

[44] N. F. Albertson, S. Archer and C. M. Suter, *J. Am. Chem. Soc.*, **67**, 36 (1945)

Threonine has been prepared by Carter and West [45] thus:



Separation of the Racemic Mixture into Its Active Components. With the exception of glycine (aminoacetic acid), all the naturally occurring amino acids contain an asymmetric α-carbon atom. As is usual in biological compounds, the naturally occurring amino acids are optically active; most are levorotatory. However, irrespective of their optical behavior all, except glycine, have the same configuration with regard to the asymmetric α-carbon atom. This configuration is similar to that of L-lactic acid, and hence the amino acids in nature may be said to possess the L-configuration; according to the convention generally employed this



would be shown in two dimensions as

Dunn [46] has reviewed the literature in this field. First the relationship of alanine to L-lactic acid was established. All other amino acids were shown by a series of physical and chemical reactions to be related to L-alanine. Care was taken to avoid reactions which would involve the valence bonds on the asymmetric carbon atom and those in which a Walden inversion might occur.

The fact that all amino acids have the same spatial configuration around the α-carbon atom becomes an important consideration in problems involving the geometry of the protein molecule. It also speaks strongly for a single primary mechanism involved in the synthesis of amino acids in nature.

At this time it should be noted that the letters D and L have only recently been approved by the Committee on Nomenclature, Spelling

[45] H. E. Carter and H. D. West, *Org. Syntheses*, **20**, 101 (1940).
[46] M. S. Dunn, *Ann. Rev. Biochem.*, **10**, 91 (1941).

and Pronunciation of the American Chemical Society [47] for designating the configuration of the asymmetric α-carbon atom of amino acids. Heretofore the letters *d* and *l* have been universally used to indicate spatial relationships, but since they sometimes have been used also to indicate the direction of rotation of plane polarized light (as abbreviations of *dextro* and *levo*) the new nomenclature is superior. When the configuration is designated by D or L, an additional symbol to indicate the direction of rotation is not necessary; if the configuration of the α-carbon is not certainly established, the direction of rotation in a specified solvent (preferably water) should be designated by the prefixes *dextro* or *levo* or by a plus or minus sign in parentheses.

Almost invariably, compounds which are synthesized in the organic laboratory, and which contain an asymmetric carbon atom, are what is known as a *racemic mixture*, where equal quantities of the D- and L-forms are present. Such a mixture is optically inactive, the levorotation of the one form being neutralized by the dextrorotation of the other form.

As we shall see later, there may be pronounced differences in solubility and in physiological action between the D- and L-forms of a chemical compound. Accordingly the task of a biological chemist is not completed when one of the syntheses noted above has yielded the pure amino acid. He must still separate the racemic mixture into its optically active components.

Three general methods are available for such separation.

Mechanical Separation. Since the crystals of the D- and L-forms are mirror images of each other, it is possible to pick out the different crystals from a crystal mixture by hand, using forceps under a lens. Pasteur used this method to separate *dextro*- and *levo*-tartaric acids. Unfortunately, amino acids rarely or never crystallize in large enough crystals to allow one to use this method of separation.

The Biological Method. Biological organisms show a surprisingly high degree of specificity toward organic molecules. The yeasts, molds, and bacteria, as well as the higher animals, are with a few exceptions capable of utilizing only one form of an optically active amino acid. In general, the optically active form which occurs in proteins is the one which is attacked by the biological organism.

The mixture of amino acids is inoculated with a pure culture of a bacterium, a fungus, or a yeast, in a culture medium, and the organism is allowed to grow and develop until one optically active isomer has been completely destroyed. The solution is then worked up for the isolation of the optically active isomer which was not attacked by the organism

[47] See *Chem. Eng. News*, **25**, 1364 (1947).

Unfortunately, although this method yields one of the optically active isomers, it is usually the isomer which does not occur in nature. Accordingly, the biological method is rarely of great value in the isolation of the naturally occurring compound.

The Chemical Method. The chemical method for separating racemic mixtures of amino acids depends on the formation of a compound of the amino acid with some optically active substance. Alkaloids are very generally used for this purpose, inasmuch as they can be obtained in quantity and in a high degree of purity at a comparatively low cost. Quinine, strychnine, brucine, and cinchonine are the alkaloids usually employed, inasmuch as they can be readily crystallized and purified. The alkaloids possess pronounced basic properties, but the amino acids are not sufficiently acidic to combine directly with the alkaloids. It is accordingly necessary to intensify the acidic properties of the amino acid molecule. This is usually done by forming the benzoyl derivative of the amino acid.

$$RCH(NH_2)COOH + C_6H_5COCl \rightarrow RCH(NHCOC_6H_5)COOH + HCl$$

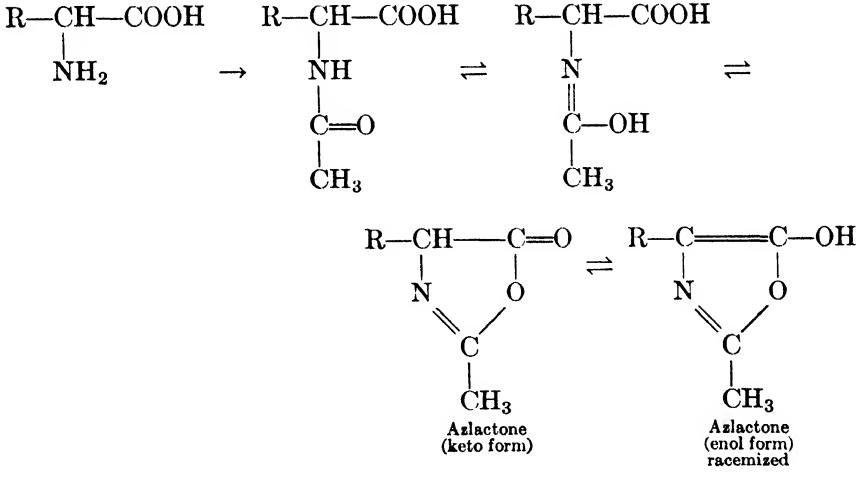
In this way the basicity of the amino group is masked, the acidity of the benzoyl derivative being intensified so that it forms a salt with the basic alkaloid.

The alkaloid selected and the benzoyl derivative are then mixed in the stoichiometrical proportions necessary for the formation of the alkaloid salt. Assuming that strychnine is the alkaloid chosen and that racemic phenylalanine is the amino acid which it is desired to separate into the optically active isomers, we would have a mixture of *levo*-strychnine-benzoyl-L-phenylalanine and *levo*-strychnine-benzoyl-D-phenylalanine. Such compounds will be found to differ in solubility. In some instances this difference is only slight; in others it may be very appreciable. The mixture is accordingly separated by a series of fractional crystallizations, retaining at one end of a series the most insoluble fraction and at the other end of the series the most soluble fraction. It will usually be found that one end of the series yields a constant dextrorotation, the other end of the series a constant levorotation, indicating that a complete separation has been obtained. When the separation is complete, the alkaloidal base is removed by appropriate means, the benzoyl derivative is hydrolyzed off by acid, and the optically active amino acid separated and recrystallized until it shows constant physical properties. If the crystals and the physical and chemical properties of an amino acid so prepared are identical with the amino acid prepared from proteins, then and only then can one state with certainty that the naturally occurring amino acid has been synthesized.

A few of the naturally occurring amino acids, *e.g.*, cystine and the hydroxyamino acids, contain two asymmetric carbon atoms. Thus, in their synthesis four stereoisomers would be formed. Cystine itself, since the molecule is symmetrical, possesses the unique possibility of being able to exist in the *meso* form, and L-cystine, which had been racemized by boiling with hydrochloric acid, has been resolved [48] into D-cystine, L-cystine, and *meso*-cystine, in the latter compound one end of the molecule being derived from D-cysteine and the other from L-cysteine. *Meso*-cystine is strictly analogous in its optical behavior to *meso*-tartaric acid.

Racemization of Optically Active Amino Acids. As we shall see later when we discuss the hydrolysis of proteins, hydrolysis with alkalies usually results in the racemization of part or all of the amino acids, and it is for this reason that alkaline hydrolysis is so seldom used. However, amino acids may be racemized by forming the acetyl derivative, the salts of which readily racemize.[49]

Bergmann and Zervas [50] have studied the mechanism of this reaction which occurs by heating with sodium acetate and acetic anhydride. The reactions involved are apparently



The essential characteristic of this reaction is that there must remain one hydrogen on the amino group after the acetyl derivative has been

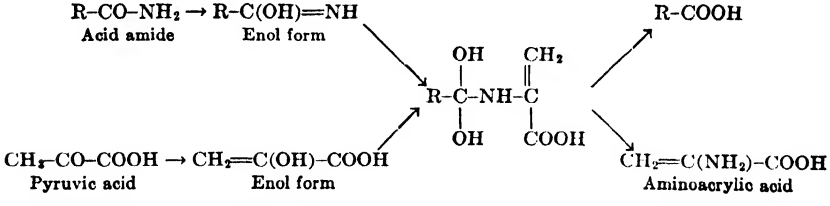
[48] L. Hollander and V. du Vigneaud, *J. Biol. Chem.*, **94**, 243 (1931); and H. S. Loring and V. du Vigneaud, *Proc. Soc. Exptl. Biol. Med.*, **29**, 41 (1931).

[49] V. du Vigneaud and R. R. Sealock, *J. Biol. Chem.*, **96**, 511 (1932); V. du Vigneaud and C. E. Meyer, *J. Biol. Chem.*, **98**, 295 (1932); *J. Biol. Chem.*, **99**, 143 (1932).

[50] M. Bergmann and L. Zervas, *Biochem. Z.*, **203**, 280 (1928).

formed. Proline does not give the reaction. Neither does acetyl-N-methylphenylalanine, inasmuch as the enolization of the acetyl derivative cannot take place. This reaction appears to afford a very convenient method for the preparation of that form of the amino acid which does not occur in nature.

Synthesis of Amino Acids in Nature. The mechanism of the synthesis of amino acids in nature still remains to be proved. Animals are, to a very large extent, dependent on plant sources for the amino acids which they build into their body tissues. Although there is definite evidence that the animal organism can synthesize glycine and there is indirect evidence that the animal organism may be able to synthesize certain of the other amino acids, there are at least nine or ten amino acids which apparently the animal organism cannot synthesize and which it must obtain from its food sources, which in the last analysis are plant proteins. Numerous theories have been proposed for the mechanism involved in the synthesis of amino acids in the plant cell. All these have been critically reviewed by Björksten,[51] who concludes that the most probable mode of synthesis involves the interaction of the enol form of pyruvic acid with an acid amide. Theoretically ammonia could be substituted for the acid amide, but in a second paper [52] he rules out ammonia as a probable intermediate. Björksten suggests that the enol form of pyruvic acid condenses with the enol form of an acid amide and that the resulting product then breaks down to form an organic acid and aminoacrylic acid. The aminoacrylic acid is unstable and readily condenses with other compounds; this theory would account for the various amino acids which occur in nature.



The primary question involves the transformation of nitrogen from the elemental form or from nitrates, nitrites, and ammonia to the amides mentioned above, probably glutamine and asparagine. The voluminous literature in the field has been reviewed by Chibnall [53] and Wilson.[54]

[51] J. Björksten, *Biochem. Z.*, **225**, 1 (1930).

[52] J. Björksten and I. Himberg, *Biochem. Z.*, **225**, 441 (1930).

[53] A. C. Chibnall, *Protein Metabolism in Plants*, Yale University Press, New Haven, 1939.

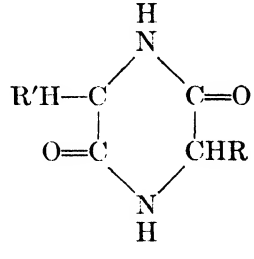
[54] P. W. Wilson, *The Biochemistry of Symbiotic Nitrogen Fixation*, University of Wisconsin Press, Madison, 1940.

CHAPTER 12

Polypeptides

The name "polypeptide" was given by Fischer to a compound of two or more amino acids joined together by the peptide linkage. This is really a substituted amide linkage resulting from the elimination of water between the amino group of one amino acid and the carboxyl group of another. If only two amino acids are involved, one peptide linkage results, and the resulting compound is called a dipeptide. Two general methods lead to the preparation of polypeptides: (1) the synthesis of peptides of known structure by the condensation of amino acids or their derivatives; and (2) the partial hydrolysis of proteins by enzymatic action, or by controlled acid or basic hydrolysis.

At this point a second type of condensation product should be noted. If the two free end groups of the dipeptide were to condense in a second peptide linkage, a ring compound would result. In this case we should obtain



a mixed diketopiperazine: 3-R-6-R'-2,5-diketopiperazine. Such a compound is often called an amino acid anhydride; the first of these to be isolated was that of leucine which was named leucinimide almost 100 years ago.

Synthesis of Polypeptides. Before 1900, several workers attempted to form condensation products from a single amino acid or from mixtures of acids by heating with such agents as phosphorus pentoxide, formaldehyde, glycerol, or potassium pyrosulfate. All the products obtained were amorphous materials having indifferent chemical characteristics. It was impossible to classify them or to determine their true relationship to the proteins.

Fischer early pointed out that, in order to arrive at a definite conclusion, some method must be devised whereby one could build up or tear down a chain of amino acids at will, and that the reactions must be so controlled that the various steps of the process could be traced by the usual procedures of the organic laboratory.

Previous to Fischer, Curtius had busied himself with linking amino acids together, but in practically every instance he had used not the free amino acid but the benzoyl derivative. In his studies, he obtained definite crystalline organic compounds. These, because they contained the benzoyl radical, were very different from the true polypeptides which were later to be synthesized.

As early as 1882, Curtius treated glycine silver with benzoyl chloride, obtaining, beside hippuric acid (benzoylaminoacetic acid), a substance which he called "hippurylaminoacetic acid" and which Fischer later showed to be the benzoyl derivative of the dipeptide glycyl-glycine. By the union of hippuric acid ester and glycine Curtius obtained a second compound, which he called "γ acid." This compound gave the biuret reactions, although the color was somewhat different from the buiret reaction as given by unaltered protein. Twenty-one years later, in the light afforded by Fischer's studies, Curtius repeated his experiments and found that the "γ acid" was in reality benzoyl-pentaglycyl-glycine. As early as 1884, Curtius had stated that besides hippuric acid a number of compounds could be obtained from glycine silver and benzoyl chloride, each compound differing from the other by one molecule of glycine with the elimination of one molecule of water. Accordingly Curtius must be credited with the first suggestion as to the peptide linkage and with the theoretical possibilities of linking amino acids together in chains to form polypeptides. Fischer pointed out that, though Curtius' statements were excellent theory, his experimental facts to prove the theory were lacking, except in the instances noted above.

Curtius observed another phenomenon, *i.e.*, that glycine ester in water solution could be transformed into glycine anhydride, the simplest member of the diketopiperazines.

$$2NH_2CH_2COOC_2H_5 \rightarrow \underset{\text{(Diketopiperazine)}}{\begin{matrix} NH\text{—}CO\text{—}CH_2 \\ | \qquad\qquad | \\ CH_2\text{—}CO\text{—}NH \end{matrix}} + 2C_2H_5OH$$

In 1901, Fischer and Fourneau [1] hydrolyzed glycine anhydride with acids and obtained the first peptide, *glycyl-glycine*. In a study of the various derivatives of glycyl-glycine, Fischer prepared the ethyl ester,

[1] E. Fischer and E. Fourneau, *Ber.*, **34**, 2868 (1901).

and the carbethoxy derivative of glycyl-glycine ester by the action of ethylchlorocarbonate:

$$C_2H_5OCOCl + NH_2CH_2CO—NHCH_2COOC_2H_5 \rightarrow$$

$$C_2H_5OCO—NHCH_2CO—NHCH_2COOC_2H_5$$

(Carbethoxy-glycyl-glycine ester)

A few months later he found that this carbethoxy ester would unite with the ester of another amino acid, *e.g.*, leucine ester, to give carbethoxy-glycyl-glycyl-leucine ester,

$$C_2H_5OCO—NHCH_2CO—NHCH_2CO—NHCH(C_4H_9)COOC_2H_5$$

the carbethoxy group of the glycyl-glycine ester intensifying the acidic properties of the compound, so that it unites directly with the leucine ester, with the elimination of a molecule of ethyl alcohol. At the same time Fischer announced a second peptide, leucyl-leucine, prepared by the partial hydrolysis of leucineimide.

Some months later, Curtius (1902) announced that by treating glycine with hippurylazide ($C_6H_5—CO—NH—CH_2—CO—N_3$) he could add one glycyl radical, and by repetition of the process he could lengthen the chain up to benzoyl-pentaglycyl-glycine. Thus, both Fischer and Curtius devised methods whereby chains of amino acids could be built up. In Fischer's method, however, the resulting compounds contained the carbethoxy group, and, in Curtius' method, the benzoyl radical. Accordingly, the resulting compounds did not show physical properties similar to those possessed by the partial hydrolysis products of proteins.

In 1903, Fischer for the first time was able to prepare the *acid chloride* of glycine derivatives by treating carbethoxyglycine with thionyl chloride. This acid chloride would condense with an amino acid ester to form a carbethoxydipeptide, which in turn could be converted into the acid chloride and further condensed with another molecule of an amino acid ester. He was able to saponify the ester group of the peptides so formed, and obtained carbamino-triglycyl-glycine, presumably having the formula

$$HOOCNHCH_2CO—(NHCH_2CO)_2—NHCH_2COOH$$

This compound was nearer to a true tripeptide than were the benzoyl derivatives prepared by Curtius, or the carbethoxy derivatives prepared by Fischer, but it still possessed a carboxyl group which could not be removed.

Accordingly, Fischer [2] again altered his technic. Using chloroacetyl chloride and glycyl-glycine ester, he obtained chloroacetyl-glycyl-glycine, and on treating this compound with ammonia the chlorine was replaced with $—NH_2$, yielding the tripeptide, diglycyl-glycine. He esterified this compound and prepared the benzoyl derivative which was identical with a benzoyl product that Curtius had synthesized several years before.

In his study of the synthesis of amino acids, Fischer prepared a number of α-bromo acids which could be easily converted into the corresponding acid chlorides by treatment with PCl_5. These acid chlorides he found would combine with amino acid esters to form derivatives containing bromine, and, when the bromine was removed by treatment with ammonia, polypeptides resulted. Thus, α-bromoisocaproyl chloride will combine with glycyl-glycine ester to form α-bromoisocaproyl-glycyl-glycine ester, which, on treatment with ammonia and saponification of the ester group, is converted into leucyl-glycyl-glycine. *Fischer further found that those derivatives of the polypeptides in which a halogen-containing acyl radical was attached to the amino group could be converted into the corresponding acid chloride by treatment with* PCl_5. Thus, α-bromoisocaproyl-glycyl-glycine could be converted by PCl_5 into α-bromoisocaproyl-glycyl-glycine chloride, and this in turn could be condensed with glycyl-glycine ester to form α-bromoisocaproyl-triglycyl-glycine ester; when this compound was treated with ammonia, the bromine was replaced with $—NH_2$, resulting in a pentapeptide, leucyl-triglycyl-glycine; or prior to the treatment with ammonia, the compound could be again converted into an acid chloride and further condensations carried out. *It is impossible to prepare the acid chloride of an unsubstituted amino acid or of an unsubstituted polypeptide.* The acid chlorides can be obtained, however, provided that the amino group is masked with such radicals as carbethoxy or benzoyl, or with radicals such as bromoisocaproyl and bromophenylacetyl. It is obvious, of course, that, if one is to prepare derivatives corresponding to the natural amino acids of proteins, the halogen must be in the α position.

Fischer prepared a number of α-halogen acids for use in his polypeptide syntheses:

Bromoacetyl chloride treated with ammonia yields glycine.
α-Bromopropionyl chloride yields alanine.
α-Bromobutyryl chloride yields α-aminobutyric acid.
α-Bromoisocaproyl chloride yields leucine.
α-Bromophenylacetyl chloride yields α-aminophenylacetic acid.

[2] E. Fischer and E. Otto, *Ber.*, **36**, 2106, 2982 (1903).

α-Bromophenylhydrocinnamyl chloride yields phenylalanine.

α-δ-Dibromovaleryl chloride, instead of yielding an α-δ-diamino acid, condenses to form proline.

$$BrCH_2CH_2CH_2CH(Br)COCl + NH_3 \rightarrow \begin{array}{l} CH_2{-}CH_2 \\ | \quad\quad | \\ CH_2 \quad CH{-}COOH \\ \quad\backslash / \\ \quad NH \end{array}$$

The use of α-halogen acids either to synthesize amino acids or as an aid in the synthesis of polypeptides was facilitated by the observation of Ward [3] that red phosphorus catalyzes the bromination of aliphatic acids in the α-position. He suggests that this may be due to the following reactions:

$$RCH_2COOH + \text{red P} + Br_2 \rightarrow RCH{=}C\begin{array}{l}/OH \\ \backslash OH\end{array} \rightarrow$$

$$RCHBr{-}C\begin{array}{l}/OH \\ \backslash Br\end{array}{-}OH \rightarrow RCH(Br)COOH + HBr$$

The three methods which Fischer devised for the synthesis of polypeptides are:

1. The splitting of amino acid anhydrides or diketopiperazines by mild acid hydrolysis.
2. Condensing chloroacetyl chloride with amino acid esters or esters of polypeptides and later treating the derivatives, so obtained, with ammonia in order to convert the chloroacetyl radical into a glycyl radical.
3. Forming acid chlorides of polypeptide derivatives which still contain in the molecule a halogenated acyl radical.

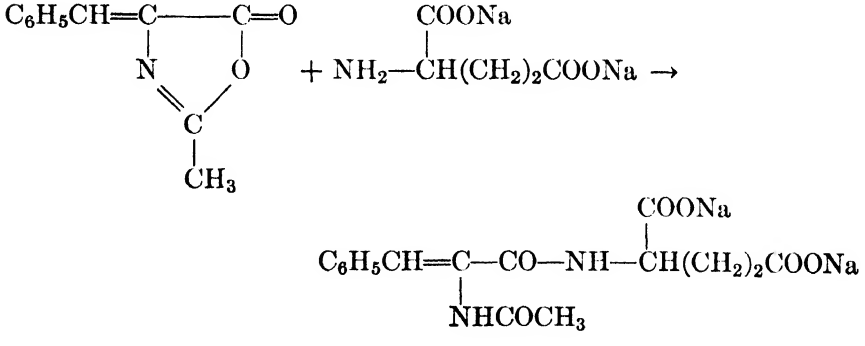
By use of the last process, Fischer,[4] in 1907, prepared an octadecapeptide (18 amino acids), leucyl-triglycyl-leucyl-triglycyl-leucyl-octaglycyl-glycine. He prepared this by combining bromoisocaproyl-diglycyl-glycine chloride with pentaglycyl-glycine, yielding, when treated with ammonia, leucyl-octaglycyl-glycine. This was further combined with another molecule of bromoisocaproyl-diglycyl-glycine chloride, yielding, when treated with ammonia, leucyl-triglycyl-leucyl-octaglycyl-glycine.

[3] C. F. Ward, *J. Chem. Soc.*, **121**, 1161 (1922).

[4] E. Fischer, *Ber.*, **40**, 1754 (1907).

A further treatment with an additional molecule of bromoisocaproyl-diglycyl-glycine chloride yielded, when the product was treated with ammonia, the octadecapeptide. This compound for many years had the highest molecular weight (1,213) of any compound prepared synthetically, the structure of which was accurately known. Later Abderhalden and Fodor [5] prepared a polypeptide containing 19 amino acids, L-leucyl-triglycyl-L-leucyl-triglycyl-L-leucyl-triglycyl-L-leucyl-pentaglycyl-glycine, with a molecular weight of 1,326. If either one of the above polypeptides had contained amino acids other than glycine, such as tyrosine or phenylalanine, the molecular weight would have been about 3,000. Fischer describes his compound as a colorless, amorphous powder, difficultly soluble in hot water to a solution which becomes turbid on standing; the solution foams, later is precipitated by saturation with ammonium sulfate; precipitated by tannins and by phosphotungstic acid, it gives the biuret test, but, as would be expected, no other characteristic color tests. The turbidity in aqueous solution and its ease of foaming indicate that it is approaching colloidal dimensions, if indeed the solution is not a typical colloidal sol.

Bergmann [6] discovered an interesting reaction of the azlactone condensation product of an aromatic aldehyde and an acetylated amino acid. This compound is stable in cold water but reacts readily with amines to open the lactone ring and yields an amide of the carboxyl group. If the sodium salt of an amino acid is employed, the resulting compound is an N-acetyl dipeptide:



Upon reduction and cautious hydrolysis with an acid, phenylalanyl-glutamic acid results.

Bergmann and Zervas [7] introduced a fifth method for the synthesis

[5] E. Abderhalden and A. Fodor, *Ber.*, **49**, 561 (1916).

[6] M. Bergmann, F. Stern, and C. Witte, *Ann.*, **449**, 277 (1926).

[7] M. Bergmann and L. Zervas, *Ber.*, **65B**, 1192 (1932).

of polypeptides which has proved to be very useful. The method consists essentially in combining the amino acid with carbobenzoxy chloride (C_6H_5—CH_2—O—CO—Cl) which readily unites with amino groups and protects them during the subsequent syntheses. The acid chloride of the carbobenzoxy derivative can then be formed by using PCl_5, and this acid chloride is condensed with the ester of another amino acid in the same way as Fischer condensed esters with his acid chlorides. The carbobenzoxy group can later be removed as toluene and carbon dioxide by catalytic hydrogenation at atmospheric pressure, leaving the free peptide. Bergmann has prepared a number of polypeptides by means of this reaction, and it is particularly adapted to the synthesis of polypeptides containing the dicarboxylic acids, such as D-glutamyl-D-glutamic acid and L-aspartyl-L-tyrosine. The method has also been used to synthesize the acid amides of both aspartic and glutamic acids and, in each instance, yielded either isoglutamine or isoasparagine, which differ from the natural products in that the acid amide is formed on the α-carboxyl group.

By the adaptation of methods detailed above, Barkdall and Ross [8] prepared di- and tri-tyrosyl-tyrosine. Corwin and Damerel [9] synthesized and studied several peptides containing the carbamic residue in the chain. Chloroacetyl-chloride was condensed with carbamyl-glycine and the product treated with liquid ammonia to yield glycyl-carbamyl-glycine.

An interesting biological synthesis of a polypeptide has been observed by Bovarnick.[10] From the action of *Bacillus subtilis* upon glutamic acid he isolated a polyglutamic acid. From its behavior during racemization he concluded, tentatively, that the δ-carboxyl group is involved in the linkage. If this is true the compound might be expected to show properties different from the usual polypeptides but more like those of nylon.

Bergmann and his associates have prepared a large number of peptides for use as enzyme substrates. Some, like benzoyl-arginine amide, are not peptides in the limited sense. All contain an amide linkage or a substituted amide linkage; the latter characterizes the peptide linkage. Several peptides were prepared as anilides and phenylhydrazine derivatives, and it is a striking fact that many were prepared by enzymatic synthesis, papain being used in most cases. A summary of these compounds is presented by Bergmann and Fruton.[11]

[8] A. E. Barkdall and W. F. Ross, *J. Am. Chem. Soc.*, **66**, 951 (1944).
[9] A. H. Corwin and C. I. Damerel, *J. Am. Chem. Soc.*, **65**, 1974 (1943).
[10] M. Bovarnick, *J. Biol. Chem.*, **145**, 415 (1942).
[11] M. Bergmann and J. S. Fruton, *Advances in Enzymol.*, **1**, 63 (1941).

du Vigneaud [12,13] and co-workers have prepared a series of peptides, one component of which is histidine.

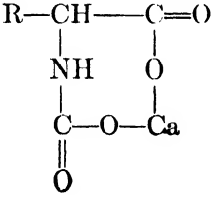
The Linkage in Polypeptides. The structure of the polypeptide molecule offers the possibility of the polypeptide group undergoing keto $\rightleftharpoons$ enol isomerism and of the terminal amino and carboxyl groups existing in the un-ionized state or ionized as zwitterions.

Certain polypeptides exist in more than one physical state. Thus, leucyl-diglycyl-glycine, when freshly prepared and in the amorphous state, is soluble in alcohol, but, if the alcoholic solution is warmed, the compound separates in the form of crystals which are then essentially insoluble in alcohol. The chemical composition is unchanged, but by this procedure evidently there has been a shift from one form to another. The probable explanation of this change in solubility is that the un-ionized molecule has shifted into the zwitterion form, although the keto $\rightleftharpoons$ enol shift may likewise be involved.

In order to show that various modifications of linkages exist, we have only to study the behavior of amino acids and polypeptides toward carbon dioxide in slightly alkaline aqueous solution. The sodium, calcium, and barium salts of the monoamino acids react with carbon dioxide to form carbamino acids and carbamino salts.

$$RCH(NH_2)COOH + CO_2 \rightarrow$$

$$RCH(NHCOOH)COOH \rightarrow$$



The monoamino acids have a ratio of nitrogen to added carbon dioxide of 1:1, tryptophan of 2:1, histidine of 3:1, and arginine of 4:1, indicating that only the free amino group reacts to form a carbamino acid and that the ring nitrogens or the nitrogen-containing groups of guanidine do not react to form carbamino acids. Siegfried and Liebermann [14] found that polypeptides also gave the carbamino reaction. Dipeptides, however, gave a ratio varying from 1.63:1 to 1.79:1, tripeptides a ratio of approximately 2.57:1, and tetrapeptides a ratio of 3.29:1

[12] V. du Vigneaud and M. Hunt, *J. Biol. Chem.*, **115**, 93 (1936); **125**, 269 (1938).

[13] M. Hunt and V. du Vigneaud, *J. Biol. Chem.*, **124**, 699 (1938); **127**, 43, 727 (1939).

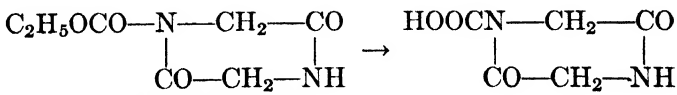
[14] M. Siegfried and H. Liebermann, *Z. physiol. Chem.*, **54**, 437 (1908).

instead of the expected 2:1, 3:1, and 4:1 ratios, showing that some group other than the single, free amino group was functioning to some extent.

It should be noted at this point that this carbamino reaction is of importance in laboratory technic. Proteins, peptones, and amino acids all undergo the addition of carbon dioxide in alkaline solution, with the formation of carbamino acid derivatives. Accordingly, if one hydrolyzes the protein with sulfuric acid, adds an excess of barium hydroxide to remove the sulfuric acid, and then bubbles carbon dioxide through the solution to precipitate barium carbonate, it invariably happens that very appreciable amounts of barium are retained in solution, owing to the formation of soluble barium salts of the carbamino acids. When such a solution is boiled, a part of the carbamino acid breaks down, precipitating barium carbonate. This reaction, however, does not immediately go to completion. Accordingly, it is very difficult to free a solution, containing proteins, peptones, or amino acids, from barium by converting the barium into barium carbonate. Barium will often remain in the final concentrates and will have to be removed by adding an equivalent of sulfuric acid.

Leuchs and Manasse [15] in a study of this reaction noted that carbethoxy-glycyl-glycine ester prepared from ethyl chlorocarbonate and glycine ester, when hydrolyzed, would not yield carbamino-glycyl-glycine ester as expected, but instead decomposed into glycine and carbon dioxide. They did, however, obtain from carbethoxy-glycyl-glycine chloride an anhydride which formed a barium salt which yielded an anhydride of glycine, having different properties from the diketopiperazine prepared by Fischer, and which was isomeric with Fischer's diketopiperazine. On the hydrolysis of this anhydride, instead of obtaining the stable glycyl-glycine carboxylic acid of Fischer, they obtained a very unstable compound. They represent the two isomeric acids and isomeric ring structures as follows, the *unstable* form of ring structure being derived from the acid ester:

$$C_2H_5OOC—NHCH_2CO—NHCH_2COOH \rightarrow$$

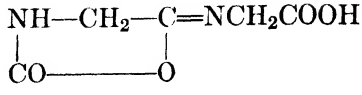


(N-Carbethoxydiketopiperazine)

[15] H. Leuchs and W. Manasse, *Ber.*, **40**, 3235 (1907); *cf.* also H. Leuchs, *Ber.*, **39**, 857 (1906); H. Leuchs and W. Geiger, *Ber.*, **41**, 1721 (1908); and H. Leuchs and F. B. La Forge, *Ber.*, **41**, 2586 (1908).

which at once loses carbon dioxide and goes to diketopiperazine. The *stable* form of ring structure is derived from the acid:

$$HOOC—NHCH_2C(OH)=NCH_2COOH \rightarrow$$



a stable compound which does not lose carbon dioxide.

It will be noted that the stable ring compound is a *lactone*. Therefore, the stable form of glycyl-glycine carbamino acid contains the grouping —N=COH— and is a *lactim* derivative, whereas the unstable form contains the grouping —NH—CO— and is a *lactam* derivative. The formulas of polypeptides are usually written in the lactam form. Inasmuch, however, as this form permits the addition of carbon dioxide and the formation of carbamino acids, whereas the lactim does not, it would appear that, in the light of the $N:CO_2$ ratios noted above, probably both the lactam and the lactim groupings occur in polypeptides. This is in agreement with the argument of Robertson [16] for a keto $\rightleftharpoons$ enol isomerism of the peptide linkage in proteins.

Polypeptides from Proteins. As already noted, a number of workers have partially hydrolyzed proteins and have isolated polypeptides from the decomposition products. In a number of instances the amino acids present have been identified and their quantities accurately determined. The following dipeptides have been isolated and identified by comparison with synthetic compounds. Glycyl-alanine, glycyl-tyrosine, and alanyl-leucine have been obtained from silk fibroin; glycyl-leucine and alanyl-leucine from elastin; and prolyl-phenylalanine from gliadin. In each component, except glycine (which is not optically active), the naturally occurring antipode was found to be a constituent of the dipeptide.

It is to be remembered that the groups known as the proteoses and peptones are split protein products. As far as the writer is aware their structures are not known. Similarly, such compounds as peptone roche [17] from silk, protalbinic and lysalbinic acids [18] from egg albumin, and the Vaughn [19] split proteins have not been fully characterized. Synge [20]

[16] T. B. Robertson, *The Physical Chemistry of the Proteins*, Longmans, Green and Co., New York, 1918.

[17] E. Abderhalden and E. Steinbeck, *Z. physiol. Chem.*, **63**, 312 (1910).

[18] C. Kennedy and R. A. Gortner, *J. Biol. Chem.*, **39**, 2734 (1917).

[19] B. White and O. T. Avery, *J. Med. Research*, **26**, 317 (1912).

[20] R. L. M. Synge, *Chem. Revs.*, **32**, 135 (1943).

has made a thorough review of the literature dealing with the many products isolated from proteins after partial hydrolysis.

Recently, Bergmann and his co-workers [21] have isolated several peptides, such as alanyl-glycine and glycyl-alanine, from silk fibroin by the selective use of aromatic sulfonic acids. Another approach is illustrated by a study of the split products from gelatin by Gordon, Martin, and Synge.[22] They were able to separate several acetyl peptides by employing chromatographic technics.

For peptides Fischer early employed napthalene sulfonic acid to characterize the acid carrying the free amino group of a peptide; this was possible because the sulfonamide obtained was fairly resistant to hydrolysis. More recently the method devised by Bergmann and Zervas [23] has been employed for a step-wise degradation of peptides. They prepared the azide of the N-benzoyl derivative which, with benzyl alcohol, yields the corresponding benzyl urethan. This is catalytically hydrogenated to the amine of the acid which originally carried the free carboxyl group. The latter acid is then split off as its aldehyde of one less carbon atom content.

Value of Polypeptide Studies. The study of polypeptides has thrown much light upon protein structure.

Inasmuch as the synthetic polypeptides have in some instances been shown to be identical with the polypeptides isolated by the partial hydrolysis of proteins, we know definitely that the —NHCO— or —N=COH— group is present in the proteins. This fact, however, does not prove that the peptide linkage is the only linkage in proteins.

The behavior of polypeptides toward proteolytic enzymes has shown in a striking manner the specificity of enzyme action. Trypsin hydrolyzes certain polypeptides; others it does not attack. In a racemic mixture of carbethoxyglycyl-D,L-leucine, it hydrolyzes the polypeptide of which L-leucine is a component but does not attack the polypeptide containing D-leucine. Alanyl-glycine is hydrolyzed by trypsin; glycyl-alanine is not.

The number of amino acids in the chain influences tryptic action. Triglycyl-glycine is not attacked by trypsin. Tetraglycyl-glycine is hydrolyzed. Triglycyl-glycine ester is likewise hydrolyzed.

Willstätter, Kuhn, and Waldschmidt-Leitz have been active in recent years in investigating the specific enzymes concerned in the hydrolysis of proteins, and erepsin has been shown to contain a mixture of enzymes. Among these are dipeptidases which attack only dipeptides, and in some

[21] W. H. Stein, S. Moore, and M. Bergmann, *J. Biol. Chem.*, **154**, 191 (1944).
[22] A. H. Gordon, A. J. P. Martin, and R. L. M. Synge, *Biochem. J.*, **37**, 92 (1943).
[23] M. Bergmann and L. Zervas, *J. Biol. Chem.*, **113**, 341 (1936).

instances specific dipeptides (note the dipeptidase of Bergmann, already discussed, which attacks dipeptides where the combination is with the imino group of proline). A further group of the polypeptidases likewise contain enzymes having marked specificity toward synthetic or natural polypeptides. Long polypeptide chains may be differentially hydrolyzed at various linkages, depending on the source of the enzyme employed. The net result of the enzyme studies appears to be a multiplicity of proteolytic enzymes.

In testing the action of proteolytic enzymes upon the various peptide substrates synthesized by Bergmann and his co-workers,[11] a definite specificity was found. Thus pepsin readily splits L-tyrosine or L-phenylalanine from a peptide, provided that amino acid is at the end of the chain, *i.e.*, that its carboxyl group is free and that its amino group is involved in a peptide linkage. Similar specificities have been found for other proteolytic enzymes, although several enzymes may attack the same linkage. For example, the amide grouping is hydrolyzed from benzoyl-L-arginineamide by trypsin, beef cathepsin, papain, and by ficin. The reader is referred to the summary cited above for additional specific data.

Natural proteins to some extent show similar differences toward enzymes. Fibrin, for example, is very easily digested by pepsin. Edestin is very slowly attacked, but is more easily attacked by trypsin than is fibrin.

Abderhalden studied the effect of enzymes other than those of the intestinal tract on many polypeptides. Nearly all the enzymes which he studied were "tryptic type" but were not obtained from the pancreatic juice, instead representing the cathepsin of various organs. He notes that, *as a rule, the cathepsin of animal organs has a greater hydrolyzing action and less of a selective action than pancreatic trypsin.* He fed polypeptides to dogs and rabbits or injected them into the blood stream and found that polypeptides which are not hydrolyzed by trypsin *in vitro* are burned in the animal body, the nitrogen being eliminated as urea. When, however, glycyl-glycine was injected subcutaneously into rabbits, it was eliminated in the urine as glycine, whereas, when glycine alone was injected, it was burned in the organism. Even the racemic compounds appear to be split and completely burned when fed, but how the hydrolysis and utilization come about is still uncertain.

Polypeptides have been of value in detecting the presence of proteolytic enzymes; those containing tyrosine, cystine, or tryptophan are particularly suited for this purpose. For example, glycyl-L-tyrosine is relatively soluble in water. The free tyrosine, however, is very insoluble. Similarly, the polypeptides containing cystine are usually relatively

soluble. Cystine is almost insoluble. Polypeptides containing tryptophan do not give the characteristic rose-red color test of free tryptophan when treated with dilute bromine water. When they are hydrolyzed, however, yielding free tryptophan, this very delicate color test can be used to indicate hydrolysis. Abderhalden has used rather extensively either the polypeptide, glycyl-L-tyrosine or "peptone roche" [24] for the identification of enzymatic action in tissues. If a section of an organ or tissue is covered with a solution of either glycyl-L-tyrosine or "peptone roche" and incubated at 37° for a time, tyrosine will crystallize out in stellate groups of needles upon those areas of the section where tryptic enzymes are present. Using this method, he found that proteolytic enzymes first made their appearance in chick embryos which were at the seven- or eight-day stage. Sections of 3.3-cm. pig embryos showed the presence of proteolytic enzymes in the liver and kidney areas. Sections of 3.2-cm. embryos, however, showed no evidences of the presence of proteolytic enzymes.

Polypeptides have been used to demonstrate the elaboration of proteolytic enzymes in the blood sera of animals injected with foreign protein. The normal blood sera of the horse or dog do not hydrolyze glycyl-L-tyrosine, although this polypeptide is hydrolyzed by the blood sera of the rabbit and the guinea pig. When, however, egg white or horse serum is injected into a dog and the dog is allowed to become an anaphylactic reactor (see p. 469), the blood serum acquires the property of hydrolyzing this polypeptide. Apparently the enzymes are elaborated in order to rid the blood stream of the foreign proteins.

In special cases the enzymatic decomposition of polypeptides may be followed by placing the optically active polypeptide solution, together with the enzyme, in the tube of a polarimeter and noting changes in optical rotation. With this method, care must be taken to have comparison samples in which the enzyme alone and the polypeptide plus boiled enzyme are observed.

Strepogenin. For a number of years it has been recognized that animals grow better when their dietary nitrogen is supplied as native protein than when it is obtained as completely hydrolyzed protein or a synthetic amino acid mixture. A few years ago Woolley [25-27] reported a substance in certain proteins (*e.g.*, purified casein and crystalline trypsinogen, but *not* in coagulated egg white) which serves as a growth stimulant for hemolytic streptococci and mice. The active substance, named

[24] E. Abderhalden and E. Steinbeck, *Z. physiol. Chem.*, **68**, 312 (1910).
[25] H. Sprince and D. W. Woolley, *J. Exp. Med.*, **80**, 213 (1944).
[26] D. W. Woolley, *J. Biol. Chem.*, **159**, 753 (1945).
[27] H. Sprince and D. W. Woolley, *J. Am. Chem. Soc.*, **67**, 1734 (1945).

strepogenin, is apparently a *peptide* (or peptide-like material) since its activity is lost on complete acid hydrolysis. Very active concentrates have been prepared from tryptic digests of casein.

More recently Woolley [28] has reported that synthetic seryl-glycyl-glutamic acid possesses streptogenin activity for *Lactobacillus casei*, but it is less effective than the concentrate from natural sources. Out of a large number of peptides tested [29] only the above and three other *glutamic acid-containing* tripeptides (glycyl-seryl-glutamic acid, alanyl-glycyl-glutamic acid, and glycyl-alanyl-glutamic acid) were found to contain strepogenin activity. The aspartic acid analogs of the active peptides are antagonistic to growth. Furthermore, the insertion of one additional amino acid into the peptides (*e.g.*, glycyl-alanyl-*leucyl*-glutamic acid) caused complete loss of activity; therefore more than just a glutamic acid peptide is required.

The existence of other nutritionally important "animal protein factors" of polypeptide nature is probable; this field, which is just opening, holds promise of being a fruitful one.

[28] D. W. Woolley, *J. Biol. Chem.*, **166**, 783 (1946).
[29] D. W. Woolley, *J. Biol. Chem.*, **172**, 71 (1948).

CHAPTER 13

Analysis of Proteins

Numerous methods have been proposed for the study of the amino acid content of proteins and the identification of the various linkages or of the individual amino acids. By certain of these methods we can prove either the presence or the absence of individual amino acids or the presence or the absence of certain groups of amino acids. Some of the methods give only qualitative information; others give quantitative.

It would be beyond the scope of the present work to consider any of these methods in sufficient detail to provide a complete manual for the laboratory worker. The discussion, therefore, will be limited to a consideration of the various technics which have been employed, with the literature citations so that the research worker may have the necessary information to obtain the exact technic and also some idea of the limitations of the proposed method.

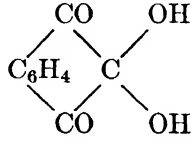
COLOR TESTS

Certain color reactions, characteristic of either amino acids or of linkages, have been proposed. These can be divided into two groups, those which are general for either linkages or for all amino acids, and those which are specific for some particular amino acid.

General. *The Biuret Reaction.* The biuret reaction is characteristic of the peptide linkage. When a solution of protein or of polypeptides containing this linkage is treated with a dilute solution of copper sulfate, after which an equal volume of a 20 per cent solution of sodium hydroxide is added, a characteristic blue-violet to violet-pink color reaction occurs. This test is given by all native proteins and most of their split products. The longer chains of proteins, in general, yield a blue-violet coloration, but, as the chain becomes shorter, the color grades more and more toward the pink. The reaction is given by any compound which contains two —CO—NH— groups separated by a carbon or a nitrogen atom. The former is the linkage in a protein or a peptide, whereas the nitrogen is present in biuret, NH_2—CO—NH—CO—NH_2, which is formed by heating urea and from which the test takes its name. Ammonium ions

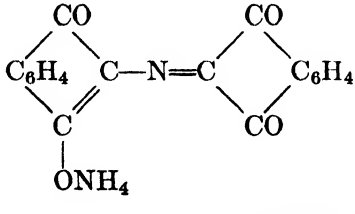
in excess interfere with the test. A single peptide linkage, like that present in glycyl-glycine, will not give the biuret reaction. Three or more amino acids must be present in order to produce at least two peptide linkages. The color is due to a complex in which the copper is coordinated with four nitrogen atoms.[1]

The Ninhydrin Reaction. When protein split products or α-amino acids are treated with "ninhydrin" (triketohydrindene hydrate),



in aqueous solution, color is developed, the color ranging from a clear, deep blue to a violet-pink or even red. This reaction was first noted by Ruhemann,[2] and was studied by Harding and MacLean [3] as a quantitative method for the determination of amino acids. At least one free $—NH_2$ group and a free —COOH group must be present in order that color may be developed. Ninhydrin is probably the most delicate reagent for detecting the presence of protein or of an α-amino acid. With some amino acids, 1 part in 100,000 parts of water can be detected.

The reaction is much more delicate if carried out in the presence of a small amount of pyridine. The reaction can be applied to the quantitative determination of an amino acid, provided that only a single amino acid is present in the solution. It is not, however, applicable to the quantitative estimation of amino acids when one is dealing with a mixture, since various amino acids give not only differences in the shade of color but differences in the depth of color per unit of the $—NH_2$ radical. Thus, using glycine as a standard, one can accurately determine quantitatively small amounts of glycine. The values for alanine or aspartic acid, however, determined by a glycine standard, will be erroneous. The color change which is developed is that of a substituted ammonium salt of diketohydrindylidenediketohydrindamine:



[1] M. B. Rising and P. S. Yang, *J. Biol. Chem.*, **99**, 755 (1933).
[2] S. Ruhemann, *J. Chem. Soc.*, **97**, 2025 (1910).
[3] V. J. Harding and R. M. MacLean, *J. Biol. Chem.*, **20**, 217 (1915).

The same reaction will be referred to later as a manometric method for the determination of amino nitrogen.

Folin's Reagent. Folin [4] proposed β-naphthoquinone sulfonic acid as a general reagent for the colorimetric estimation of amino acids in blood. In the presence of strong alkalies and of an amino acid, this reagent develops a striking, deep red color. Ammonia likewise yields a color, but ammonia is readily removed, so that its interference would not be serious. Folin notes that urea, uric acid, creatinine, creatine, and hippuric acid do not yield colors. A color is developed with most of the nitrogen bases, such as the alkaloids or aniline, but as a rule these compounds are not present in solutions where one wishes to determine the presence or absence of amino acids.

Specific Reactions. *The Millon Reaction.* A red color is developed in many instances when a protein or a protein hydrolysate is heated with a solution of mercurous and mercuric nitrate and nitrite. Chloride ions and an alkaline medium interfere with the reaction. Preservatives containing a phenolic group (such as phenol, thymol, and salicylic acid) may not be present, because the *red coloration is specific for the phenol group.* The reaction is, in general, considered as specific for tyrosine, although dopa and possibly the bromo- or iodotyrosines would give the reaction. The test is capable of great delicacy when properly carried out, *i.e.*, when a considerable excess of the reagent is avoided.

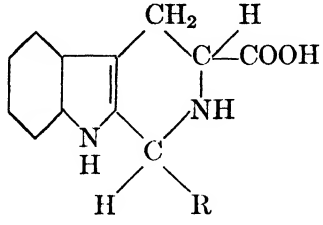
The Xanthoproteic Reaction. When protein is treated with nitric acid, a yellow coloration is produced which is intensified to orange on the addition of ammonium hydroxide. The yellow coloration so commonly produced on skin which has come in contact with nitric acid is an example of the xanthoproteic reaction and indicates the presence of *nitrated protein.* Nitration takes place on the benzene rings, and the reaction is, therefore, specific for *aromatic nuclei* which are easily nitrated (tryptophan or tyrosine). The benzene nucleus of free phenylalanine is not readily nitrated, but we do not know how this amino acid behaves when in a protein molecule.

The Tryptophan Reactions. Whenever a protein containing tryptophan is heated with strong sulfuric or hydrochloric acid in the presence of an aldehyde, a blue or violet color is produced. The color is due to a condensation of the α-hydrogen of the indole nucleus with an aldehyde. Harvey, Miller, and Robson [5] have shown that the first compound

[4] O. Folin, *J. Biol. Chem.*, **51**, 377 (1922).

[5] D. G. Harvey, E. J. Miller, and W. Robson, *J. Chem. Soc.* (*London*), **153** (1941).

formed is 2,3,4,5-tetrahydro-4-carboxyl-β-carboline:



If the heating is continued long enough, a black humin will result.

This test is known under a variety of names, depending on the aldehyde employed. Many proteins give the test without the addition of an aldehyde, owing either to contaminations or to the presence of carbohydrate units in the protein molecule. When carried out in this way the test is known as the *Liebermann* reaction. Better colors are obtained if an aldehyde is added. When formaldehyde is used the reaction carries the name of *Acree-Rosenheim*. In the *Adamkiewicz* reaction glacial acetic acid is employed because it contains glyoxalic acid, COOH—CHO. In the *Ehrlich* reaction, much used by bacteriologists to detect the presence of indole-forming bacteria, *p*-dimethylaminobenzaldehyde is present; *Cole's* test employs the unsubstituted benzaldehyde; and the *Hopkins-Cole* test uses preformed glyoxalic acid.

The Reduced Sulfur Test. In many instances a black coloration is produced when a protein or a mixture of amino acids is heated with alkali and lead acetate. The coloration is due to the formation of black lead sulfide and appears to be specific for the presence of cystine or cysteine, or rather the groups —S—S— or —SH.

Methionine does not respond to the test unless the material has first been heated with a mineral acid.

The Sullivan Cystine Reaction. Sullivan [6] devised a very specific test for cystine or cysteine. It consists essentially in adding to the cystine hydrochloride 1–2 ml. of 1 per cent sodium cyanide solution made up in 0.8 *N* NaOH, then adding 1 ml. of a fresh 0.5 per cent aqueous solution of 1,2-naphthoquinone-4-sodium sulfonate, then adding 5 ml. of a 10–20 per cent Na_2SO_3 solution made up in 0.5 *N* NaOH, and allowing the mixture to stand at 20°C. for 30 minutes. Pure red color is then developed by adding 1 ml. of a 2 per cent aqueous solution of sodium hyposulfite ($Na_2S_2O_4$). The color is specific for cystine or cysteine. A great variety of compounds were tested, but Sullivan found none that interfered. The —SH, —NH_2, and —COOH groups must all be free for the color to develop.

[6] M. X. Sullivan, *U. S. Pub. Health Repts.*, **44**, 1421, 1599 (1929).

The Molisch Test. The Molisch test is specific for carbohydrates. Many proteins contain a carbohydrate radical. In testing for such a radical, a trace of α-naphthol is added to a protein solution, and this is layered over concentrated sulfuric acid. A pink to red coloration at the interface indicates the presence of carbohydrate. This test is probably dependent on the formation of furfural from the carbohydrate.

The Sakaguchi Test. Sakaguchi [7] proposed a color reaction which he claims to be specific for the *free guanidine* group. To 5 ml. of a 1 per cent solution of a protein, or of a 1 per cent solution of an amino acid mixture, is added 2 ml. of a 15 per cent solution of sodium hydroxide, followed by 5 ml. of a 0.15 per cent solution of α-naphthol, and the entire mixture is treated with 0.3 *N* sodium hypochlorite. After standing at 2° to 4°C. for 40 minutes, the mixture is diluted and read against a standard solution in a colorimeter. The color which is developed is an intense red. Only compounds containing the free guanidine group react. Accordingly the color test should be specific for arginine and for canavanine. In a later paper Sakaguchi [8] shows that the guanidine group of arginine is free in most native proteins and also that the proteins differ widely in the rate with which arginine is set free when the proteins are hydrolyzed by enzymes or by acid or alkali.

The Diacetyl Reaction. Harden and Norris [9] observed that, if a dilute solution of protein is mixed with a small amount of 10 per cent KOH solution and then one drop of a 1 per cent solution of diacetyl (CH_3—CO—CO—CH_3) is added, a pink color with a green fluorescence is developed. The reaction depends on the presence of the arginine group in proteins, although the authors note that the exact shade of the color is somewhat dependent on the compound in which the arginine group is substituted. In bacteriology this is known as the Voges-Proskauer test and is used for testing for the presence of certain bacteria.[10]

Hessler [11] studied the correlation between the intensity of the diacetyl reaction and the known arginine content of a series of fifteen proteins. He found that he could secure more accurate results by measuring the green color developed in acid than by measuring the pink color developed in alkalies. For fourteen of the fifteen proteins he found an almost exactly quantitative relationship between the diacetyl color and the known arginine content. For gliadin, for some unknown reason, he found color development with the diacetyl reagent to be inhibited.

[7] S. Sakaguchi, *J. Biochem. Tokyo*, **5**, 25, 133 (1925).
[8] S. Sakaguchi, *J. Biochem. Tokyo*, **5**, 143, 159 (1925).
[9] A. Harden and D. Norris, *J. Physiol.*, **42**, 332 (1911).
[10] C. H. Werkman, *J. Bact.*, **20**, 121 (1930).
[11] L. E. Hessler, M.S. thesis, University of Minnesota, 1931.

QUANTITATIVE DETERMINATION OF AMINO AND CARBOXYL GROUPS IN AMINO ACIDS OR MIXTURES OF AMINO ACIDS

Van Slyke's Nitrous Acid Method. Van Slyke [12] devised apparatus whereby one can determine quantitatively with a high degree of accuracy the primary amino nitrogen which is present in a protein, an amino acid, or a mixture of amino acids. The reaction involved is

$$R—NH_2 + HNO_2 \rightarrow R—OH + N_2 + H_2O$$

where the nitrous acid is liberated by the action of acetic acid on sodium nitrite. The nitrogen gas evolved by the reaction is collected and measured. Van Slyke gives tables for the conversion of the volume of the nitrogen gas into milligrams of amino nitrogen. In 1913 Van Slyke [13] improvised a micro apparatus by which one can quantitatively determine 5 mg. or less of amino nitrogen. Koch [14] has improved the design to make for ease of manipulation.

Under the prescribed conditions, the determination is specific for primary amines in contrast to secondary and tertiary amines. Ammonium ions react to some extent; it is possible to make a correction for this provided that the concentration of ammonium ions is known, as shown by Irving, Fontaine, and Samuels.[15] Amides, such as urea, and guanidine derivatives do not react provided that strong acids are not present. Plimmer [16] has shown that in the presence of 2 *N* HCl arginine and urea will yield a considerable portion of their nitrogen in the reaction, hence care should be taken that no appreciable quantities of strong acids are present.

For all α-amino acids the time of the run is 5 minutes; if lysine is present 30 minutes are required for the complete reaction of the ϵ-group. These time periods are adequate, as shown by Dunn and Schmidt.[17]

Glycine and cystine yield values which are somewhat higher than the calculated values. Kendrick and Hanke [18] recommend that potassium iodide be included in the reagents for correct results.

The method has proved of great service not only in the rapid analysis of amino acids but also in the study of proteins. Van Slyke and Birchard [19] have shown that the ϵ-group of lysine exists free in native proteins,

[12] D. D. Van Slyke, *J. Biol. Chem.*, **9**, 185 (1911); **12**, 275 (1912).
[13] D. D. Van Slyke, *J. Biol. Chem.*, **16**, 121 (1913).
[14] F. C. Koch, *J. Biol. Chem.*, **84**, 601 (1929).
[15] G. W. Irving, Jr., T. D. Fontaine, and C. S. Samuels, *Arch. Biochem.*, **4**, 437 (1944).
[16] R. H. A. Plimmer, *J. Chem. Soc.*, **125**, 265 (1925).
[17] M. S. Dunn and C. L. A. Schmidt, *J. Biol. Chem.*, **53**, 401 (1922).
[18] A. B. Kendrick and M. E. Hanke, *J. Biol. Chem.*, **132**, 739 (1940).
[19] D. D. Van Slyke and F. J. Birchard, *J. Biol. Chem.*, **16**, 539 (1914).

and that the free amino content of proteins is approximately equal to one-half its lysine nitrogen content. However, Sanger [20] caused dinitrofluorobenzene to react with the free amino groups of insulin. On hydrolysis he found the corresponding derivative linked to two glycine and two phenylalanine residues, as well as to the ϵ-amino groups of two lysine units. This indicates that, at least in insulin, amino groups other than those of lysine are free.

The rate or extent of protein hydrolysis can be accurately followed by determining the free amino nitrogen at various intervals of time, since, when the peptide linkage is broken, a free amino group is formed. Similarly, the relative digestibility of proteins *in vitro*, *i.e.*, the rate of hydrolysis by a particular enzyme or by a mixture of enzymes, can be followed in a quantitative manner. The method affords information as to whether or not a given mixture contains proteolytic enzymes. If there is a progressive increase in free amino nitrogen, it can be taken as evidence that enzymatic action is present.

The Ninhydrin Carbon Dioxide Method. When α-amino acids are heated in solution with ninhydrin the following reaction occurs:

$$RCH(NH_2)COOH \xrightarrow{+O} CO_2 + NH_3 + RCHO$$

We have seen earlier that the liberated ammonia forms a colored compound with the reagent. Van Slyke and co-workers [21] proposed two methods for measuring the carbon dioxide evolved, from which the amino nitrogen may be calculated. Each of the known α-amino acids yields one mole of CO_2, with the exception of aspartic acid and cystine, each of which yields 2 moles of CO_2. Proline and hydroxyproline decompose to yield 1 mole of CO_2, although they do not give the color test because the ring does not open to liberate ammonia. With the exception of glutathione, peptides and proteins are not decarboxylated. The liberated CO_2 may be measured by the usual gasometric method or may be absorbed in an excess of standard barium hydroxide.

Combined with Van Slyke's nitrous acid method, this technic should serve to distinguish certain amino acids in mixtures with others. Thus proline would react to yield CO_2 but would not react with nitrous acid; mixtures of glutamic and aspartic acids could be analyzed for each component because the latter yields $2CO_2$ to $1NH_2$ group. Similarly lysine could be determined in a precipitate including the other hexone bases. Both amino groups of lysine react with nitrous acid, whereas one mole of CO_2 is liberated by ninhydrin.

[20] F. Sanger, *Biochem. J.*, **39**, 507 (1945).

[21] D. D. Van Slyke, R. T. Dillon, D. A. MacFayden, and P. Hamilton, *J. Biol. Chem.*, **141**, 627 (1941); D. D. Van Slyke, D. A. MacFayden, and P. Hamilton, *ibid.*, **141**, 671 (1941).

The following is an example of the information which can be obtained from the use of both the nitrous acid and the ninhydrin reactions. Casein was exhaustively hydrolyzed with crystallized trypsin; the product yielded 16 per cent amino-nitrogen by the nitrous acid method and practically none by the ninhydrin assay. These data indicate that the hydrolysis stopped at the peptide stage with the liberation of essentially no free amino acids. When a crude trypsin product was used, values by both technics were higher, and the ninhydrin value was 44 per cent of the value by the nitrous acid method. Evidently the crude trypsin contained peptidases in addition to the trypsin proper.

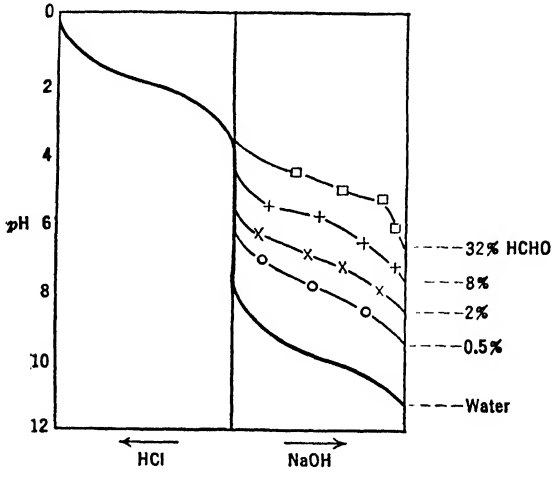


FIG. 77. Titration curve of glycine in the presence of increasing concentrations of formaldehyde. [After L. J. Harris, *Biochem. J.*, **24**, 1080 (1930).]

The Sörensen Titration. Sörensen [22] proposed a method whereby an amino acid could be titrated directly in aqueous solution with a standard alkali. Amino acids dissolve in water to form an essentially neutral solution. They cannot be titrated to completion with the use of indicators because the *p*H changes so slowly and the reaction is not complete before extreme *p*H values are attained. In the presence of formaldehyde in a concentration of 6 to 9 per cent the titration with standard alkali is complete in the range where phenolphthalein or thymolphthalein change color. The curve obtained with acids is not altered, as shown in Fig. 77. Harris (see p. 284) interprets the titration of the zwitterion as follows:

$$^{+}NH_3CH(R)COO^{-} + HCHO \rightarrow CH_2O \cdot NH_2CH(R)COO^{-} + H^{+}$$

[22] S. P. L. Sörensen, *Biochem. Z.*, **7**, 45 (1907).

Sörensen's titration yields abnormal values [23] when applied to certain amino acids. Nevertheless the variations usually offset each other in the average protein hydrolysate.

This titration method is of value not only in quantitatively determining an amino acid or a mixture of amino acids, but likewise in quantitatively determining amino acids in the presence of other organic acids. In this case, the mixture of amino acids and organic acids is titrated in aqueous solution with standard sodium hydroxide, phenolphthalein being used as an indicator, until the organic acids which are present are completely neutralized. Neutral formaldehyde is then added and the solution further titrated with standard sodium hydroxide. Sörensen's method has proved of great value in enzyme studies, inasmuch as it permits one to follow the rate at which the protein is being hydrolyzed, since the opening of the peptide linkage results in the formation of additional groups.

Foreman's Titration. Foreman [24] noted that, when amino acids or even ammonium salts are titrated in 85 per cent ethyl alcohol, the ammonium radical or the amino groups do not react basic. The acid groups with which the ammonium radical is associated, or the carboxyl group of the amino acids, can accordingly be titrated directly in 85 per cent alcohol, phenolphthalein being used as an indicator. A 0.1 *N* solution of ammonium chloride titrates in 85 per cent alcohol as though it were a 0.1 *N* solution of hydrochloric acid. It is sometimes more convenient to use Foreman's titration method than Sörensen's method. Martens [25] made a careful comparative study of the Van Slyke, Sörensen, and Foreman methods. He finds that when phenolphthalein is used as an indicator Foreman's method yields somewhat low results, Sörensen's method somewhat high results. He suggests a modification of Foreman's method: titration of the amino acid in 93 to 95 per cent alcohol and use of thymolphthalein as an indicator. Under such conditions, the carboxyl groups can be determined with essentially the same degree of accuracy as is possible for the determination of the amino groups by the Van Slyke method.

Linderstrøm-Lang's Titration. Linderstrøm-Lang [26] further modified the titration methods by titrating amino acids or proteins in 90 to 95 per cent acetone, using either standard alcoholic HCl or standard alcoholic NaOH, the HCl titrating the amino groups and the NaOH

[23] S. L. Jodidi, *J. Am. Chem. Soc.*, **48**, 751 (1926).
[24] F. W. Foreman, *Biochem. J.*, **14**, 451 (1920).
[25] R. Martens, *Bull. soc. chim. biol.*, **9**, 454 (1927).
[26] K. Linderstrøm-Lang, *Compt. rend. trav. lab. Carlsberg*, [4] **17** (1927).

titrating the carboxyl groups. Apparently acetone, like alcohol, inhibits zwitterion formation.

Titrations in Other Solvents. Many amino acids are soluble in glacial acetic acid and can be titrated with standard solutions of perchloric, hydrobromic, or sulfuric acid, with crystal violet as an indicator.[27,28] If the amino acids are insoluble in acetic acid, they may be dissolved in standard perchloric acid and back-titrated with a solution of guanidine acetate.

Most titrations in non-aqueous media involve a shift in the dissociation constants of the indicator and in that of the amino acid. In the Sörensen formol titration which is carried out in an aqueous medium, the pK_a calculated from the apparent dissociation constant is shifted, for glycine, from 9.60 in water to 5.92 in 9 per cent formaldehyde.

HYDROLYSIS OF PROTEINS

For the quantitative determination of individual amino acids, it is necessary to hydrolyze the protein into its constituent amino acids. Various methods have been proposed, *e.g.*, hydrolysis by acids, hydrolysis by alkalies, and hydrolysis by enzymes. Each has its advantages and disadvantages.

Hydrolysis by Acids. Constant-boiling (1.115 sp.g.) or approximately 20 per cent hydrochloric acid is most commonly used, although in some instances it is advantageous to use 25 per cent (by weight) sulfuric acid. Hydrolysis by 25 per cent sulfuric acid is usually somewhat slower than hydrolysis by constant-boiling hydrochloric acid, but it reaches the same completeness of hydrolysis.

Proteins differ widely in the rapidity with which hydrolysis reaches completion. In some instances, a protein may be completely hydrolyzed by boiling for 6 to 8 hours with 20 per cent hydrochloric acid. In other instances, as much as 24 hours of boiling may be required.

Various other acids have been tested. Hydrofluoric acid is unsuited for protein work, the dilute acid reacting too slowly, whereas the concentrated acid yields not simple amino acids but complex polypeptides or condensation products.

The chief drawback to the use of a strong mineral acid lies in the partial dismutation of cystine and of serine, and in the destruction of the tryptophan with the formation of humin. Hess and Sullivan [29] recommend hydrolysis with hydriodic acid to avoid the formation of humin.

[27] G. F. Nadeau and L. E. Branchen, *J. Am. Chem. Soc.*, **57**, 1363 (1935).
[28] G. Toennies and T. P. Callan, *J. Biol. Chem.*, **125**, 259 (1938).
[29] W. C. Hess and M. X. Sullivan, *Arch. Biochem.*, **3**, 53 (1943).

Similarly, Lieben [30] has found that stannous and stannic chlorides as well as titanium trichloride increase the rate of hydrolysis with hydrochloric acid.

Steinhardt and Fugitt [31] hydrolyzed wool and egg albumin at 65°C. with such acids as dodecyl sulfuric, dodecylsulfonic, picric, and *p*-diphenyl benzene sulfonic acids, as well as with orange II and β-naphthalene sulfonic acid. The values obtained were compared with those obtained with sulfuric and hydrochloric acids. The striking conclusion was that these organic acids were more effective than the mineral acids in liberating the ammonia from its amides. However, their studies do not indicate how effective these compounds may be in catalyzing the complete hydrolysis of proteins.

Formic acid causes only partial hydrolysis. Certain amino acids are hydrolyzed off, but apparently an equilibrium is reached fairly early in the process. In addition, there is the danger of forming formic derivatives with the amino group. The remaining organic acids are too weak to produce complete hydrolysis. Fodor and Epstein [32] studied the degradation of gelatin in the presence of acetic anhydride. In some instances, amino acids or amino acid derivatives were found to be acetylated. A considerable fraction of the products isolated was in the form of polypeptides, an indication that even acetic anhydride will not completely hydrolyze proteins.

Nasset and Greenberg [33] measured the rate of the hydrolysis of casein in acid solution as evidenced by the formation of amino nitrogen and found it to be a second-order reaction obeying the equation

$$K = \frac{1}{t} \times \frac{x}{100 - x} \times \frac{1}{100} \tag{154}$$

where x = the percentage of protein hydrolyzed in time t. They found hydrolysis to be proportional to the hydrogen-ion activity of the acid solutions.

Hydrolysis by Alkalies. The alkalies, sodium hydroxide, potassium hydroxide, or barium hydroxide, cause very rapid and complete hydrolysis, although the strong alkalies bring about the decomposition, especially the deamination, of certain of the amino acids. Unfortunately, alkaline hydrolysis results in the racemization of the amino acids, so that they are isolated to a large extent in the optically inactive form.

[30] F. Lieben, *J. Biol. Chem.*, **151**, 117 (1943).
[31] J. Steinhardt and C. H. Fugitt, *J. Research Natl. Bur. Standards*, **29**, 315 (1942)
[32] A. Fodor and C. Epstein, *Z. physiol. Chem.*, **171**, 222 (1927).
[33] E. S. Nasset and D. M. Greenberg, *J. Am. Chem. Soc.*, **51**, 836 (1929).

The ammonia fraction of the hydrolysate is increased by alkaline hydrolysis, which liberates 2 moles of ammonia from the guanidine nucleus of arginine.[34] Because of this, alkaline hydrolysis is rarely used. It may, however, be used to advantage in the hydrolysis of proteins in which one wishes to determine tryptophan, inasmuch as tryptophan is destroyed by acid hydrolysis but is relatively stable toward alkaline hydrolysis.

Enzymatic Hydrolysis. As noted earlier, pepsin hydrolyzes proteins only to proteoses, peptones, and polypeptides. Trypsin and the ereptic enzymes, in general, hydrolyze proteins to the constituent amino acids, but the process is slow and, in many instances, not entirely complete. However, complete hydrolysis by enzymes may be attained by use of several enzymes. Accordingly, although tryptic digestion is used to isolate tryptophan, it is not generally used for analytical purposes.

QUANTITATIVE STUDY OF THE HYDROLYTIC PRODUCTS OF PROTEINS

The methods for the quantitative study of the mixture obtained by the hydrolysis of proteins may be divided into two general groups: methods which are concerned with the identification of *groups of constituents*, and methods which are concerned with either the isolation or the quantitative estimation of *individual amino acids*.

Group Analysis. *Hausmann's Method.* Hausmann [35] proposed the characterization of a protein by distributing the nitrogen content of its hydrolytic decomposition products into three groups, *i.e.*, ammonia nitrogen, basic nitrogen, and non-basic nitrogen.

Osborne [36] modified the Hausmann method by adding a fourth fraction, *i.e.*, that of humin nitrogen (older terminology = melanin nitrogen).

THE AMMONIA NITROGEN. After acid hydrolysis, the excess of acid is removed by distillation, the solution containing the hydrolysis products is then rendered alkaline, by either magnesium oxide or preferably a suspension of calcium hydroxide, and the ammonia is distilled into standard acid, preferably under a partial vacuum. The ammonia which is formed has been generally regarded as derived from acid amide linkages, $-CONH_2$, and in many instances is referred to as the "amide nitrogen." However, a certain amount of deamination occurs under the usual conditions of acid hydrolysis, and the true amide nitrogen value can be ascertained only by hydrolyzing for relatively short periods of

[34] C. B. Croston, Ph.D. thesis, University of Minnesota, 1944.
[35] W. Hausmann, *Z. physiol. Chem.*, **27**, 95 (1899); **29**, 136 (1900).
[36] T. B. Osborne and I. F. Harris, *J. Am. Chem. Soc.*, **25**, 323 (1903).

time.[37] Under the usual conditions of hydrolysis as much as 20 per cent of the amide nitrogen of a 24-hour hydrolysate may be due to deamination. The deamination affects only the monoamino acids; the basic amino acids, arginine, histidine, and lysine, are not deaminized even by boiling for 6 weeks with constant-boiling hydrochloric acid.

The ammonia nitrogen has usually been ascribed to acid amide linkages attached to the free carboxyl group of one of the dicarboxylic acids. However, there appears to be no correlation [38] between the ammonia nitrogen of a protein hydrolysate and the dicarboxylic acids which can be isolated from that hydrolysate.

TABLE 40. PARTIAL HYDROLYSIS OF GLIADIN WITH DILUTE PHOSPHORIC ACID UNDER PRESSURE

(Analysis of gliadin: total N = 17.35%; free amino-N = unhydrolyzed 2.15% of total N, completely hydrolyzed 58.6%; ammonia-N = 24.4% of total N)

(Unpublished data—Gortner and Wiles)

Acid Concentration, normality	Temperature, °C.	Time, minutes	Ammonia-N Liberated, % of total ammonia-N	Amino-N Liberated from Protein Linkages, % of total amino-N
0.005	170	180	76.3	5.32
0.010	170	60	57.9	2.51
0.010	170	120	77.1	7.05
0.010	170	180	92.1	13.84
0.030	183	40	96.3	5.60
0.040	180	40	95.5	9.73
0.050	180	40	106.2	9.14
0.10	170	60	102.0	7.12

It is rather interesting to note that glutamine and asparagine differ widely in the ease with which ammonia is split off from the acid amide linkage. Thus, glutamine is completely hydrolyzed in 2 hours at 100° at *p*H 6.5, whereas asparagine is not affected under these conditions.[39] However, these relationships do not necessarily hold when the amino or carboxyl groups of the amino acid are combined in peptide or protein linkages.[40]

[37] R. A. Gortner and G. E. Holm, *J. Am. Chem. Soc.*, **39**, 2736 (1917).

[38] R. A. Gortner, Jr., M.S. thesis, University of Minnesota, 1934.

[39] H. B. Vickery, G. W. Pucher, H. E. Clark, A. C. Chibnall, and R. G. Westall *Biochem. J.*, **29**, 2710 (1935).

[40] A. C. Chibnall and R. G. Westall, *Biochem. J.*, **26**, 122 (1932).

Since the ammonia nitrogen is split off from the protein in the early stages of protein hydrolysis, it should theoretically be possible to prepare highly "acid proteins" by splitting off the ammonia nitrogen and leaving the peptide linkages practically unaltered. No one, however, has succeeded in completely removing the ammonia nitrogen without partially breaking down some of the peptide linkages. Table 40 shows the data resulting from such an attempt, using gliadin as the protein and hydrolyzing with dilute phosphoric acid under pressure. The work of Steinhardt and Fugitt, cited earlier, suggests that certain organic acids may be used for this purpose.

THE HUMIN NITROGEN. The acid-insoluble humin (melanin) of a protein hydrolysate is the black, amorphous material which separates when proteins are hydrolyzed with strong acids. Humin is formed by the condensation [41] of tryptophan with an aldehyde. The reaction apparently is with the indole nucleus, the condensation presumably taking place on the α-hydrogen. The colors produced in the Liebermann, Acree-Rosenheim, and Adamkiewicz reactions are due to the early stages of humin formation. The evidence presented by Gortner and his coworkers shows clearly that proteins contain either an aldehyde group or some other group which reacts as an aldehyde in condensing with tryptophan, resulting in the formation of humin. The nature of the compound to which this reactive group is attached, however, remains still to be determined. It may be that some compound similar to djenkolic acid, which breaks down to yield formaldehyde as one of the products, is present in most proteins.[42]

The humin nitrogen determination can be so conducted as to be a quantitative measure of the tryptophan content of a protein. Thus, when fibrin was hydrolyzed in the presence of increasing amounts of formaldehyde, the black, acid-insoluble humin nitrogen was increased from 9.60 mg. per 3 grams of fibrin, where no formaldehyde was added, to a maximum of 19.02 mg. of acid-insoluble humin nitrogen, where 0.10 gram of formaldehyde in the form of trioxymethylene was added. The addition of larger amounts of trioxymethylene sharply decreased the amount of acid-insoluble humin nitrogen. The 19.02 mg. of acid-insoluble humin nitrogen appears to be a practically quantitative measure of the tryptophan nitrogen which was present in the original 3 grams of fibrin.

The humin nitrogen, as obtained in an ordinary protein hydrolysate,

[41] R. A. Gortner, *et al.*, *J. Am. Chem. Soc.*, **37**, 1630 (1915); *J. Biol. Chem.*, **26**, 177 (1916); *J. Am. Chem. Soc.*, **39**, 2477 (1917); **42**, 632, 821, 2378 (1920); **45**, 550 (1923); **46**, 1224 (1924).

[42] H. A. Lillevik and W. M. Sandstrom, *J. Am. Chem. Soc.*, **63**, 1028 (1941).

bears little or no relationship to the true tryptophan nitrogen content of the protein, inasmuch as the humin formation depends on two variables, the tryptophan content and the presence of an adequate amount of aldehyde. Accordingly, the humin formed in a protein hydrolysis as ordinarily carried out, though dependent on the tryptophan content of the protein, is not a quantitative measure of that tryptophan content.

Humin formation apparently involves the interaction of the indole nucleus with three molecules of aldehyde. The first molecule of aldehyde appears to condense to form a substituted indolidene-methane. The second molecule of aldehyde forms a compound of the rosindole type. These compounds are highly colored, and certain compounds of this type have been utilized as dyestuffs. The third molecule of aldehyde condenses with the rosindole type of compound with the elimination of a molecule of water and forms the "humin." The type of structure involved still remains to be elucidated. Apparently the α-amino-propionic acid side chain of tryptophan is not involved in the reaction, since, at maximum humin formation in the presence of aldehyde, one half of the total nitrogen of the humin is still free α-amino nitrogen.

After the acid-insoluble humin has been removed from an acid hydrolysate the solution still remains black owing to the acid-soluble humin. In certain analyses this fraction is separately measured by a Kjeldahl determination after the solution has been made faintly alkaline with calcium oxide and the ammonia has been removed by distillation. Gortner and his co-workers have shown that the acid-soluble humin arises from the decomposition of part of the tyrosine.

THE BASIC NITROGEN. In the Hausmann method the basic nitrogen is determined by precipitating the diamino acids, arginine, histidine, lysine, and cystine, with phosphotungstic acid in the presence of an excess of hydrochloric acid. The phosphotungstates of these amino acids are nearly insoluble. The phosphotungstate precipitate is filtered off, washed with a dilute solution of phosphotungstic acid in dilute hydrochloric acid, and the nitrogen in the entire precipitate determined by the Kjeldahl method.

THE NON-BASIC NITROGEN. The non-basic nitrogen is determined by kjeldahling an aliquot of the filtrate from the phosphotungstate precipitate. This fraction of nitrogen represents the monoaminomonocarboxylic acids and the monoaminodicarboxylic acids.

Thus, the Hausmann method divides the nitrogen of a protein hydrolysate into four fractions. The *advantages* of the method are that it is rapid, that it requires only a small amount of protein, 0.5 gram to 1 gram, and that by this method one can obtain rather definite information as to the ratio existing between the diamino acids and the monoamino

acids comprising the protein molecule. The *disadvantages* of the method are that the method gives no clue as to the presence or absence of any individual amino acid with the possible exception of tryptophan, and that there are many proteins with somewhat similar content of basic and non-basic amino acids, such proteins being more or less indistinguishable from each other by the Hausmann technic.

Van Slyke's Method. Van Slyke [43] made use of his method for determining amino nitrogen and the ratio between total nitrogen and free amino nitrogen in certain of the amino acids, together with the Hausmann procedure, in order to estimate quantitatively certain of the amino acids in proteins. In Van Slyke's method the protein is hydrolyzed, and the *acid amide nitrogen* and *humin nitrogen* are determined essentially as in Hausmann's method.

The *basic amino acids* are precipitated with phosphotungstic acid in 5 per cent (by weight) hydrochloric acid, and the basic phosphotungstates are filtered off and washed with dilute phosphotungstic acid in dilute hydrochloric acid. This precipitate is then dissolved in very dilute alkali and decomposed by the addition of 20 per cent barium chloride solution; the insoluble barium phosphotungstate is filtered off and washed; and the filtrate containing the basic amino acids is slightly acidified and concentrated to a definite volume. This filtrate contains the arginine, histidine, lysine, and a part of the cystine which were present in the original protein. These four amino acids are then estimated as follows:

a. Total sulfur is determined on an aliquot of this solution by any of a variety of methods for determining sulfur in organic compounds, and the *cystine nitrogen* content is calculated.

b. The total nitrogen present in the basic fraction is determined on an aliquot of the solution of the bases.

c. The free amino nitrogen content is determined by Van Slyke's method in an aliquot of the solution.

d. Sufficient potassium hydroxide is added to an aliquot of the bases to make a solution containing 50 per cent of potassium hydroxide. This solution is boiled for 6 hours under such conditions that the volume remains constant and any ammonia which is evolved is collected in standard acid. By this process arginine is decomposed into urea and ornithine, the urea being further broken up into ammonia and carbon dioxide.

Consequently half of the arginine nitrogen has been evolved as ammonia. From the amount of nitrogen which is evolved as ammonia, the quantity of *arginine nitrogen* in the original solution is calculated.

[43] D. D. Van Slyke, *J. Biol. Chem.*, **10**, 15 (1911); **22**, 281 (1915); **23**, 411 (1915).

e. We now have values for cystine, arginine, total nitrogen, and amino nitrogen. The amount of *histidine* which is present is secured by a mathematical calculation, inasmuch as histidine and arginine are the only basic amino acids which contain non-amino nitrogen. Two thirds of the histidine nitrogen and three fourths of the arginine nitrogen are non-amino nitrogen (the $—NH_2$ group of the guanidine nucleus does not react with nitrous acid). Accordingly the non-amino nitrogen of the bases minus three fourths of the arginine nitrogen is equal to two thirds of the histidine nitrogen; from this we can calculate the amount of *histidine nitrogen* in the basic fraction.

f. The *lysine nitrogen* is finally obtained by subtracting from the total nitrogen, the sum of the arginine nitrogen plus the histidine nitrogen plus the cystine nitrogen.

Van Slyke further differentiates the nitrogen in the filtrate from the bases by use of his amino nitrogen apparatus into (1) amino nitrogen, and (2) non-amino nitrogen, the former being derived from the monoaminomonocarboxylic and monoaminodicarboxylic acids, the latter being derived from proline, hydroxyproline, and in part from tryptophan.

Although the essential details of the Van Slyke original method have remained unchanged, certain modifications in technic have been introduced [44] which make either for ease in manipulation or for greater accuracy in the analytical results. More recently a micro method [45] has been proposed which yields the same fractions as the original Van Slyke method, but where as little as 25 mg. of nitrogen is necessary. If the experimenter has available as much as half a gram of the protein, Cavett's modification is generally more readily applicable.

Van Slyke's method has been used very extensively in protein studies. Larmour [46] collected the various analyses which had been published up to 1928 and subjected them to statistical analyses. We are not at this point concerned with his statistical findings. His paper, however, can conveniently be used as a source of reference for an extensive series of protein analyses.

Van Slyke's method has the *advantages* of requiring a relatively small amount of protein, 3 grams or less, and of permitting the more or less quantitative determination of arginine, histidine, and lysine. The cystine value is too low, inasmuch as cystine is partly decomposed when boiled with acids, the portion which is not decomposed being either

[44] R. H. A. Plimmer and J. L. Rosedale, *Biochem. J.*, **19**, 1004, 1020 (1925); *cf.* also J. W. Cavett, *J. Biol. Chem.*, **95**, 335 (1932).

[45] J. L. Rosedale and Gladys Anne Da Silva, *Biochem. J.*, **26**, 369 (1932).

[46] R. K. Larmour, *Trans. Roy. Soc. Can.*, Section V, 349 (1928).

racemized or converted into an isomeric form, the phosphotungstate of which is appreciably soluble.[47]

In a critical study [48] of Van Slyke's method in which mixtures of known amino acids were used, the conclusion reached was that the histidine and lysine values may be somewhat in error when tryptophan and proline are present. The arginine which was added was determined within experimental error. The cystine nitrogen in three experiments was 64.5, 62.1, and 73.3 per cent of the cystine nitrogen added. The histidine values were satisfactory in the absence of tryptophan or proline but were high in the presence of either one or both of these amino acids. The lysine values were only slightly high and might well have been attributed to experimental errors.

The *disadvantages* of Van Slyke's method lie in the fact that it gives us no insight into the composition of the group of monoaminomonocarboxylic and monoaminodicarboxylic acids. In a number of instances workers have used the Van Slyke method on various biological materials which were not pure proteins, and, though valuable *comparative results* may be obtained by such technic, it should be pointed out that the "arginine nitrogen," "histidine nitrogen," "lysine nitrogen," etc., so obtained do not necessarily bear any relationship to the true arginine, histidine, lysine, etc., content of the material, because any compound which breaks down under the action of strong alkali to yield ammonia would be calculated as arginine nitrogen, the balance of the non-amino nitrogen would be calculated as histidine nitrogen, etc., and purines, pyrimidines, etc., would distribute themselves, if present, among the arginine, histidine, and lysine fractions.

Isolation and Identification of Individual Amino Acids. *Fischer's Ester Method.* In Fischer's [49] study of amino acid derivatives he observed that the ethyl esters of the monoaminomonocarboxylic and monoaminodicarboxylic acids could be distilled *in vacuo* without appreciable decomposition. It is upon this observation that he based his ester method.

The protein is hydrolyzed by hydrochloric acid; the excess of hydrochloric acid is removed by distillation; the mixture of amino acid hydrochlorides is concentrated; and the concentrated solution is saturated with gaseous hydrochloric acid. On standing at a low temperature, *glutamic acid hydrochloride* crystallizes out. This is filtered off and recrystallized from concentrated hydrochloric acid. The free glutamic acid can be obtained from this hydrochloride (any other amino acid

[47] W. F. Hoffman and R. A. Gortner, *J. Am. Chem. Soc.*, **44**, 341 (1922).
[48] R. A. Gortner and W. M. Sandstrom, *J. Am. Chem. Soc.*, **47**, 1663 (1925).
[49] E. Fischer, *Z. physiol. Chem.*, **33**, 151 (1901).

chloride where the amino acid has a low basic dissociation constant acts similarly) by suspending the hydrochloride in alcohol and adding pyridine or aniline [50] to combine with the hydrochloric acid. The free amino acid will then crystallize from the alcoholic solution.

The remaining amino acids are then converted into their ethyl esters by boiling with absolute ethyl alcohol in the presence of hydrochloric acid or zinc chloride. On concentration and cooling of the mixture of esters, *glycine ester hydrochloride* crystallizes out and can be removed by filtration. The excess of hydrochloric acid in the remaining mixture of esters is then neutralized by some appropriate technic. The preferable method is to determine accurately the amount of hydrochloric acid which is present and add an exact equivalent of sodium ethylate dissolved in absolute alcohol. The free esters are now soluble in absolute ether. The ethereal solution is dried and submitted to vacuum distillation, the following fractions being secured:

60°	(10 mm.)	= glycine, alanine, leucine, proline
60°–100°	(10 mm.)	= valine, leucine, proline
100°–130°	(0.5 mm.)	= leucine and proline
130°–180°	(0.5 mm.)	= phenylalanine, glutamic acid, aspartic acid, and serine

It will be noted that there is a considerable overlapping of amino acids in the various fractions. The residue which does not distil contains arginine, histidine, lysine, tyrosine, cystine, hydroxyproline, and various anhydrides, such as leucinimide, and other diketopiperazines formed by secondary reactions.

The ester fractions, as obtained in the vacuum distillation, are hydrolyzed by boiling with water, and the individual amino acids separated by a process of fractional crystallization.

It is obvious that such a method is far from quantitative. Osborne and Jones,[51] after many years of experience in working with Fischer's ester method, undertook a study of the method in order to ascertain the errors involved. They list the following sources of error: (1) incomplete hydrolysis, (2) loss of amino acids due to the formation of humin, (3) incomplete esterification, (4) the decomposition of the esters by hydrolysis prior to distillation, (5) unavoidable loss in separating the amino acids by fractional crystallization. To the above, we must undoubtedly add (6) the conversion of amino acid esters into diketopiper-

[50] H. C. Benedict, *J. Am. Chem. Soc.*, **51**, 2277 (1929); *cf.* also G. J. Cox, *J. Biol. Chem.*, **78**, 475 (1928).

[51] T. B. Osborne and D. Breese Jones, *Am. J. Physiol.*, **26**, 305 (1910).

azine derivatives. In this experiment, they mixed pure amino acids (omitting the boiling with hydrochloric acid for 24 hours), esterified the mixture, liberated the free esters, and distilled *in vacuo*. The various fractions were then worked up for their amino acid content. Table 41 shows that the distillable esters were only 73 per cent and the total recovery was only 66.17 per cent of the theory. This probably explains

TABLE 41. RECOVERY OF AMINO ACIDS FROM A KNOWN MIXTURE BY FISCHER'S ESTER METHOD

(Data of Osborne and Jones)

	Amino Acids Taken, grams	Amino Acids Recovered	
		Grams	Per cent
Alanine	16.00	7.34	45.88
Valine	4.00	1.64	41.00
Leucine	85.00	69.36	81.60
Proline	31.00	22.56	72.77
Phenylalanine	26.00	18.07	69.51
Aspartic acid	6.00	2.55	42.50
Glutamic acid	120.20	83.26	69.39
Tyrosine	16.00	7.97	49.81
Arginine	5.50	3.57	64.91
Histidine	2.04	0.80	39.21
Serine	2.00	0.00	0.00
Ammonia	14.31		
Total	328.05	217.12	66.17

why most of the analyses of proteins by Fischer's ester method range from 40 to 60 per cent of the weight of the protein taken.

Actually the weight of amino acids theoretically obtainable exceeds considerably the weight of the protein from which the amino acids are derived, inasmuch as the elements of water are added to each peptide linkage. A protein should yield from 110 to 120 per cent of its weight of amino acids. The isolation of 60 per cent of the amino acids in a protein analysis means, therefore, that approximately only half of the amino acids which are actually present have been accounted for.

Fischer's ester method was the only method of analysis available for many years by which the aliphatic monoaminomonocarboxylic acids could be determined. Consequently it was widely used although it re-

quired a high degree of skill and much time. Today it has only historic interest and has largely been supplanted by other methods.

Dakin's Method. Dakin,[52] in 1918, introduced a method which has advantages either in replacing or in supplementing other methods. In Dakin's method, the protein is hydrolyzed with sulfuric acid; the sulfuric acid is quantitatively removed with barium hydroxide; and the aqueous solution of amino acids concentrated until crystallization of amino acids begins. This thick, semi-crystalline mass is then extracted in a continuous liquid extractor with *n*-butyl alcohol. The amino acids themselves, with the exception of proline, are insoluble in anhydrous *n*-butyl alcohol, but the monoaminomonocarboxylic acids are slightly soluble in *n*-butyl alcohol saturated with water. The aqueous butyl alcohol distils at a lower temperature than does absolute butyl alcohol. Accordingly there is left in the receiving flask a crystalline mass of the monoaminomonocarboxylic acids. The absolute butyl alcohol containing the proline is removed from this crystalline mass, and the crystalline mass is subjected to fractional crystallization for its individual constituents, or it may be esterified and the esters distilled as in Fischer's method.

The solution which has been extracted with butyl alcohol contains the diamino acids, the dicarboxylic acids, tyrosine, and diketopiperazines. The basic amino acids can be removed by precipitation with phosphotungstic acid. Tyrosine, because of its insolubility in water, can be readily obtained, and the dicarboxylic acids are then separated by fractional crystallization.

It is rather interesting to note that, although the dicarboxylic acids cannot be extracted by butyl alcohol from a neutral aqueous solution, they are extracted by butyl alcohol from an aqueous solution having a *p*H of approximately 3.0.

Dakin's method of extraction with butyl alcohol affords in many instances a valuable procedure for the purification of an individual amino acid. Amino acids which cannot be readily obtained in a crystalline form from aqueous solution readily crystallize when extracted from the aqueous solution with *n*-butyl alcohol in a continuous extractor.

The Brazier-Schryver Method. Miss Brazier,[53] working in Schryver's laboratory on hydrolytic products of zein, introduced a radically new method for the separation of the amino acids. The protein is hydrolyzed with sulfuric acid, the acid is removed with baryta, and the ammonia is aspirated off. Any residual barium is then quantitatively removed, and the amino acids are converted into their copper salts by boiling with

[52] H. D. Dakin, *Biochem. J.*, **12**, 290 (1918); *J. Biol. Chem.*, **44**, 499 (1920).

[53] Mary A. B. Brazier, *Biochem. J.*, **24**, 1188 (1930).

copper carbonate. The mixture is then completely dehydrated in the presence of an excess of copper carbonate and then shaken in a mechanical shaker with water containing additional copper carbonate. The copper salts of leucine, phenylalanine, and aspartic acid remain as an insoluble residue.

The insoluble residue is treated with hydrogen sulfide to remove the copper, rendered alkaline with barium hydroxide, and the addition of three volumes of 95 per cent alcohol precipitates the aspartic acid as barium aspartate. The barium is removed from the filtrate of the barium aspartate with sulfuric acid, and leucine crystallizes from the concentrated solution. The filtrate from the leucine is converted into the zinc salts with freshly precipitated $Zn(OH)_2$. Zinc leucine is only slightly soluble in cold water; zinc phenylalanine is readily soluble. The zinc salts are decomposed with hydrogen sulfide to obtain the free amino acids.

The copper salts which were soluble in water are evaporated to dryness and dehydrated with acetone. The residue is then extracted with absolute methyl alcohol. The copper salts of alanine, glutamic acid, tyrosine, arginine, histidine, and glycine are insoluble in absolute methyl alcohol. The copper salts of valine, hydroxyvaline, proline, and a dipeptide, prolyl-phenylalanine, were soluble in absolute methyl alcohol.

Glutamic acid was separated from the insoluble residue as barium glutamate. Tyrosine was crystallized from the filtrate of the barium glutamate. Histidine was removed from this filtrate as the insoluble $Zn(OH)_2$—$HgCl_2$ complex. Arginine was separated as the flavianate, and the remaining alanine extracted from the residue by Dakin's butyl alcohol method.

Those copper salts which were soluble in methyl alcohol were converted into the free acids with hydrogen sulfide. Prolyl-phenylalanine crystallized from the aqueous solution. Proline was separated as the picrate or as the double compound with cadmium chloride. By reconversion through the copper and zinc salts, zinc valine was separated as insoluble in absolute ethyl alcohol, and zinc hydroxyvaline as soluble.

In Miss Brazier's analysis of zein, she accounted for 93.5 per cent of all the nitrogen in the original protein, the balance being lost in inorganic precipitates which were discarded. She recovered 87.89 per cent of the nitrogen as ammonia or as pure amino acid fractions. In almost every instance, the amount of amino acid which is reported is higher than the highest value reported in the previous literature.

In a subsequent paper from Schryver's laboratory, Damodaran [54] applied the copper-zinc salt method to the analysis of glutenin. Here

[54] M. Damodaran, *Biochem. J.*, **25**, 190 (1931).

again, in almost every instance, the yield of the isolated amino acids exceeded the highest yield reported in earlier literature. However, when one reads Damodaran's paper, one finds that the actual technic for the isolation of a particular amino acid has been varied somewhat from the specifications given by Miss Brazier. Stockelbach and Bailey,[55] in a comparative analysis of the three protein fractions in wheat gluten, also found that the procedure had to be adapted somewhat for each protein studied. Apparently, unless more specific conditions are laid down, or at least until a wide variety of proteins have been analyzed by this method, the analysis of each new protein through the series of metallic salts must be regarded almost as a research problem. Jukes [56] has suggested the removal of the dicarboxylic acids and the basic amino acids prior to the fractionation of the remaining amino acids as the copper and zinc salts, and he believes that this simplifies the procedure.

Boyd's Uramino and Hydantoin Method. When amino acids are treated with potassium cyanate, they readily form uramino or carbamino acids, and these, when treated with dilute hydrochloric acid, are easily transformed into hydantoins.

$$RCH(NH_2)COOH + HOCN \rightarrow$$

$$RCH(NH{-}CO{-}NH_2)COOH \xrightarrow{HCl} R{-}CH{-}CO$$

Uramino acid

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad NH$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad NH{-}CO$$

Hydantoin

The uramino acids and hydantoins, in contrast to the amino acids, are relatively insoluble in water and in general readily soluble in organic solvents. Boyd [57] made a study of the solubilities of the various uramino and hydantoin derivatives and proposed a method for the fractionation of the amino acids from proteins based on his solubility studies.

The proteins were hydrolyzed by boiling with hydrochloric acid, the excess of acid was removed by concentrating *in vacuo*, followed by neutralization of the hydrolysate, and the amino acids were converted into the uramino acids by boiling with potassium cyanate. The cold solution was then carefully acidified to the turning point of Congo red, and the uramino acids derived from leucine, isoleucine, and phenylalanine separated as a precipitate. These were then converted into the hydantoins and separated by differential solubility.

[55] L. S. Stockelbach and C. H. Bailey, *Cereal Chem.*, **15**, 801 (1938).
[56] T. H. Jukes, *J. Biol. Chem.*, **103**, 425 (1933).
[57] W. J. Boyd, *Biochem. J.*, **27**, 1838 (1933); **29**, 546 (1935).

The uramino acids remaining in solution were separated into two fractions based on their solubility or insolubility in alcohol at *p*H 4. The uramino acids from the monoaminomonocarboxylic acids are soluble in alcohol at *p*H 4, whereas the remaining amino acids are relatively insoluble. From those which are soluble in the alcohol, the hydantoins of leucine, isoleucine, valine, phenylalanine, alanine, proline, and hydroxyproline can be removed by their solubilities in ether or chloroform. The hydantoin of tyrosine is insoluble in ether and remains behind in practically pure crystalline form. The details of the separation by differential solubilities of the various hydantoins is too involved to be readily abstracted, but it appears as though Boyd's method may be a valuable adjunct to other methods in problems involving protein analysis.

The *advantages* of the method lie in the fact that the melting points and crystallographic properties of the uramino and hydantoin derivatives are exceedingly characteristic, so that the amino acid from which they are derived is excellently characterized by the physical properties of these derivatives. The amino acid can be regenerated by the hydrolysis of the uramino or hydantoin derivatives, the *disadvantage* of this process of securing the amino acids being that the regenerated amino acid is usually racemized.

Kossel's Separation of the Bases. Kossel and Kutscher [58] suggested a method whereby arginine, histidine, and lysine could be prepared in pure form and quantitatively determined with a fairly high degree of precision.

The protein is hydrolyzed with sulfuric acid, and the sulfuric acid later removed quantitatively with barium hydroxide. A hot, saturated solution of silver sulfate is then added to the aqueous solution of amino acids until a drop of the silver-amino acid solution turns to a brown color on the addition of barium hydroxide, indicating an excess of silver in the solution. The solution is then saturated with powdered barium hydroxide, precipitating the silver salts of *arginine* and *histidine*. *Lysine* remains in solution and can be precipitated later by phosphotungstic acid.

The precipitate containing the silver salts of arginine and histidine is filtered off, the silver removed with hydrogen sulfide, and the solution of amino acids filtered from the resulting precipitate of silver sulfide. The excess of hydrogen sulfide is removed by aeration. Hot, concentrated silver sulfate solution is again added until an excess of silver is present. At this point barium hydroxide solution is added until the solution is neutral. Histidine silver precipitates at neutrality, leaving the arginine silver in solution. The histidine silver is filtered off, decomposed with hydrogen sulfide, and the solution of the free amino acid concentrated

[58] A. Kossel and F. Kutscher, *Z. physiol. Chem.*, **31**, 165 (1900).

to a small volume. The addition of a hot, saturated alcoholic solution of picrolonic acid precipitates the histidine practically quantitatively as the histidine picrolonate.

The filtrate from the histidine, containing the arginine silver, is saturated with powdered barium hydroxide, precipitating arginine silver. This precipitate is filtered off, the silver removed with hydrogen sulfide, the excess of barium quantitatively removed with sulfuric acid, and the aqueous solution of arginine concentrated to a small volume and the picrolonate prepared as in the case of histidine.

The solution containing the lysine is freed from silver with hydrogen sulfide, acidified with sulfuric acid, and the lysine precipitated as lysine phosphotungstate. This precipitate is decomposed with baryta, the excess of barium is quantitatively removed with either carbon dioxide or an exact equivalent of sulfuric acid, the solution containing the free lysine is evaporated to a small volume, and the lysine is separated as lysine picrate.

This method has been modified in some particulars by later investigators. Vickery and Leavenworth [59] control the separation of arginine and histidine silver by hydrogen-ion concentration measurements. Histidine silver is completely precipitated at a *p*H of 7.0, whereas arginine silver remains completely in solution. They recommend double precipitation, thus insuring the absence of traces of the other amino acids, and recommend the separation of arginine as arginine flavianate.[60] Vickery [61] also recommends the use of 3,4-dichlorobenzenesulfonic acid to isolate histidine.

As already noted, Kossel's method is essentially quantitative, although the results are somewhat lower than the analyses indicated by Van Slyke's method. Its great *advantage* lies in the fact that by this method the basic amino acids can be rather readily prepared.

The Carbamate Method. Buston and Schryver,[62] in 1921, announced a new method for the separation of amino acids from the products of protein hydrolysis. The method depends on the formation of the carbamino derivatives which we have already discussed. The solution containing the amino acids is treated with an excess of barium hydroxide and saturated with carbon dioxide to form the barium carbamates. Later Kingston and Schryver [63] proposed a general scheme for the separation of the hydrolytic products of proteins, based on the carbam-

[59] H. B. Vickery and C. S. Leavenworth, *J. Biol. Chem.*, **76**, 707 (1928); *cf.* also *J. Biol. Chem.*, **79**, 377 (1928); **93**, 105 (1931); *Biochem. J.*, **26**, 1101 (1932).

[60] A. E. Pratt, *J. Biol. Chem.*, **67**, 351 (1926).

[61] H. B. Vickery, *J. Biol. Chem.*, **143**, 77 (1942).

[62] H. W. Buston and S. B. Schryver, *Biochem. J.*, **15**, 636 (1921).

[63] H. L. Kingston and S. B. Schryver, *Biochem. J.*, **18**, 1070 (1924).

ate reaction. The dicarboxylic acids can be quantitatively precipitated as the barium carbamates, when three volumes of alcohol are added to the amino acid solution which has been treated with barium hydroxide and carbon dioxide. When the amino acid solutions are sufficiently concentrated, practically all the amino acids can be precipitated in the presence of alcohol, as the barium carbamates, with the exception of proline which remains in the alcoholic mother liquor.

The carbamate method offers particular advantages for the separation of the dicarboxylic acids from the monocarboxylic acids, and it appears probable that its chief usefulness will be in connection with a study of the dicarboxylic acid fraction as shown by Foreman.[64]

Separation of the Basic Amino Acids and the Dicarboxylic Amino Acids by Electrodialysis. Foster and Schmidt [65] utilized electrodialysis to separate the basic and the dicarboxylic amino acids from the monoaminomonocarboxylic acids. The protein is hydrolyzed with sulfuric acid, the sulfuric acid is quantitatively removed with baryta, and the aqueous solution of the amino acids is placed in the center compartment of an electrodialyzer separated from the anode and the cathode chambers by collodion or parchment membranes. When carbon electrodes were used and the acidity of the protein hydrolysate was adjusted to a *p*H of 5.5, arginine, histidine, and lysine migrated almost quantitatively to the cathode compartment. At a *p*H of 7.5 only arginine and lysine migrated, the histidine remaining in the center compartment. By repeating the electrolysis on the cathode liquor, the basic amino acids were separated almost completely from the other amino acids of the protein hydrolysate. The dicarboxylic acids and proline migrated to the anode, and, when the solution in the anode compartment was again electrodialyzed, only small quantities of monoaminomonocarboxylic acids were found in the anode compartment.

A later study [66] gives explicit details for the preparation of the hexone bases by use of electrical transport.

Albanese [67] has adopted the method for the determination of the bases in 0.5 to 1 gram of protein.

Separations by Ion Exchange, by Adsorption, and by Partitions. Within the past few years several procedures have been developed for the separation of certain amino acids through the use of ion exchange materials. Because of their charge, sharp separations are possible between

[64] F. W. Foreman, *Biochem. J.*, **8**, 463 (1914).

[65] G. L. Foster and C. L. A. Schmidt, *J. Biol. Chem.*, **56**, 545 (1923); *J. Am. Chem. Soc.*, **48**, 1709 (1926).

[66] G. J. Cox, Harriette King, and C. P. Berg, *J. Biol. Chem.*, **81**, 755 (1929).

[67] A. A. Albanese, *J. Biol. Chem.*, **134**, 467 (1940).

the groups of the basic, the acidic, and the essentially neutral amino acids. Clays (fuller's earth) and particularly the zeolites have been used for this purpose; recently synthetic resins have been prepared which give considerable promise, and Myers [68] has presented a general survey of their exchange properties.

In certain cases the workers merely shook a solution containing the amino acids with the exchange materials; in other cases a chromatographic column technic was employed followed by elution. Turba [69] employed specially treated fuller's earth in a column and separated a mixture of arginine, histidine, lysine, and leucine. Wieland,[70] as well as Turba and Richter, used Brockmann's aluminum oxide to separate the dicarboxylic acids, cystine, and the neutral amino acids. Cannan [71] and Kibrick [72] used the amberlite type of polyamine-formaldehyde resins in analyses of protein hydrolysates for the dicarboxylic acids.

In contrast to ion exchange materials, activated charcoal probably removes the acids by a process of adsorption. Cassidy [73] and others report the separation of glycine, leucine, tyrosine, and phenylalanine by this method. Karrer, Keller, and Szönyi [74] employed zinc carbonate to separate the methyl esters of the N-*p*-phenylazobenzoyl derivatives of the aliphatic amino acids.

Tiselius [75] has introduced the technic of the "schlieren" method (*cf.* Tiselius' electrophoresis apparatus, p. 116) for the determination of the amounts of the amino acids adsorbed by activated charcoal. As in most of the studies reported here, no complete protein hydrolysate has been examined, although certain mixtures have been subjected to analysis with promising results.

Gordon, Martin, and Synge [76] studied the partition of acetylamino acids between such solvents as chloroform, propanol, butanol, and cyclohexane. On the basis of this preliminary work they made separations using inert silica gels as the framework to support the liquid phase. The appropriate solvents are used to "develop" the column and to elute the acids. The method has been used for the analysis of wool, gelatin, fibrin, and other proteins available only in smaller quantities.

[68] R. J. Myers, *Advances in Colloid Sci.*, **1**, 317 (1942).
[69] F. Turba, *Ber.*, **74**, 1829 (1941).
[70] T. Wieland and W. Paul, *Ber.*, **75**, 1001 (1943); **76**, 823 (1944).
[71] R. K. Cannan, *J. Biol. Chem.*, **152**, 401 (1944).
[72] A. C. Kibrick, *J. Biol. Chem.*, **152**, 411 (1944).
[73] C. S. Cleaver, R. A. Hardy, Jr., and H. G. Cassidy, *J. Am. Chem. Soc.*, **67**, 1343 (1945).
[74] P. Karrer, R. Keller, and G. Szönyi, *Helv. Chim. Acta*, **26**, 38 (1943).
[75] A. Tiselius, *Advances in Colloid Sci.*, **1**, 81 (1941).
[76] A. H. Gordon, A. J. P. Martin, and R. L. M. Synge, *Biochem. J.*, **33**, 65 (1944).

Consden, Gordon, and Martin [77] have extended the method for use in separating amino acids without previous acetylation. A sheet of filter paper is used as the supporting medium; the solvents employed are phenol, butanol, benzyl alcohol, and collidine. The two intersecting edges can in turn be dipped into two liquids to yield a two-dimensional pattern. By this means the amino acids assume different positions on the sheet as brought out by the ninhydrin color, or other suitable tests. The method is at present only qualitative; as little as 0.2 to 0.4 mg. of the protein is required for the test.

Isolation of Individual Amino Acids. The methods described above serve to separate all or certain of the amino acids present in a protein hydrolysate. At times it is desired to isolate a single amino acid, usually from an available protein in which it occurs in a fair quantity. The methods indicated below are generally applicable.

Cystine is usually isolated from a hydrolysate of wool or human hair by adjusting the *p*H to 4.4, the isoelectric point at which cystine is rather insoluble. Similarly *tyrosine* is prepared from silk fibroin or from casein by precipitation at a *p*H of 5.7. *Glutamic acid* hydrochloride can readily be isolated from a concentrated hydrolysate of wheat gluten by saturation with dry hydrogen chloride at a moderately low temperature. The free acid can be obtained by controlled neutralization or by the use of aniline. *Glycine* is at times prepared as its ethyl ester hydrochloride, which is the first step in the Fischer synthesis; however, it is readily prepared from monochloroacetic acid, a preparation which is very satisfactory in view of the fact that the compound is not optically active and hence no tedious resolution of antipodes is involved. Glycine may also be precipitated by the use of one of the specific reagents, nitranilic acid or potassium trioxalochromate. *Leucine* is often first isolated as dileucine hydrochloride from a solution saturated with sodium chloride at *p*H 1.7 to 2.8; by an adjustment of the acidity the leucine will crystallize out.

Specific precipitants have also been employed. *Proline* and *hydroxyproline* are precipitated under definite conditions by ammonium reineckate (ammonium tetrathiocyanodiammono chromite); if the two ammonia groups are replaced by aniline, rhodanilic acid, which is specific for proline, results. *Arginine* is usually isolated at the diflavianate, which is converted to arginine monohydrochloride. Similarly *histidine* has lately been prepared as indicated earlier in the description of Kossel's separation of the bases.

Tryptophan is generally isolated from an enzyme digest by precipitation with mercuric sulfate in the presence of 5 per cent sulfuric acid,

[77] A. Consden, A. H. Gordon, and A. J. P. Martin, *Biochem. J.*, **38**, 224 (1944).

which dissolves the mercury-tyrosine complex. The mercury-tryptophan is decomposed with hydrogen sulfide. An extraction with butyl alcohol (Dakin's method) will remove certain impurities before the final crystallization of the tryptophan. *Lysine* can be obtained as the picrate after the arginine and histidine have been removed from the products obtained by electrolysis, precipitation with phosphotungstic acid, or by use of Kossel's separation. Similarly *aspartic acid* can be obtained as its copper salt after the glutamic acid has been removed from a fraction containing these two acids prepared as alkaline earth carbamates.

Most of these methods are described in Morrow and Sandstrom [78] and in other manuals on preparative methods.

Determination of Specific Amino Acids. It will be recognized that the methods previously described for the group analysis of protein hydrolysates yield very little information regarding the quantities of the various amino acids present in a protein hydrolysate, except Van Slyke's nitrogen distribution which involves the determination of the four hexone bases. By the use of preparative methods, the amino acid is isolated, and from the amount obtained one knows only that that amount or more was present in the protein. Many analytical procedures have been devised to determine the concentration of amino acids in proteins. Except for the microbiological assay method, these procedures usually depend for their specificity on some functional group in the amino acids; hence the aliphatic monoaminomonocarboxylic acids are most difficult to determine.

Colorimetric and Titration Methods. Many methods of analysis have been proposed. Block and Bolling [79] have compiled the various determinations in a most useful book which should be consulted for any analytical procedure. Only a few of the more commonly used methods will be noted here. Certain of these reactions were referred to in the section dealing with the qualitative tests. Whenever a colorimetric procedure is employed the standard is usually made from an authentic sample treated in the same manner as the unknown.

Arginine is often determined by the colorimetric method of Sakaguchi; it can be assayed also by the use of the enzymes arginase and urease which together yield half of the nitrogen of arginine as ammonia. Sullivan's β-naphthoquinone sulfonic acid method is widely used for cystine (and cysteine); an iodine titration method has also been employed.

[78] C. A. Morrow and W. M. Sandstrom, *Biochemical Laboratory Methods*, 2nd ed., John Wiley & Sons, New York, 1935 (out of print).

[79] R. J. Block and D. Bolling, *The Amino Acid Composition of Proteins and Foods*, C. C. Thomas, Springfield, Ill., 1945.

Tyrosine and tryptophan are often determined on the same solution by some modification of either the Millon-Folin, or the Folin and Marenzi phenol reagent. The separation of the two is effected by the difference in the solubilities of their mercury complexes in 5 per cent sulfuric acid. Tryptophan may also be determined by the Ehrlich colorimetric reagent (see p. 318) on an alkaline protein hydrolysate. Glycine is often determined as the green derivative, soluble in chloroform, produced by *o*-phthalaldehyde.

The periodic acid or the lead tetra-acetate titration methods serve to determine serine and threonine. When both are thought to be present the resulting aldehydes are identified. Alanine may be determined as lactic acid after treatment with nitrous acid. The thiomethyl group of methionine is determined by distillation with hydriodic acid. In a definite *p*H range glutamic acid is converted to the cyclic compound, pyrrolidone carboxylic acid, which is extracted with ethyl acetate, and the amino nitrogen is determined upon opening the ring.

The Isotopic Dilution Method. This is an ingenious method due largely to Rittenberg.[80] A known quantity of the compound to be determined but containing the N^{15} isotope is added to the hydrolysate containing the unknown. The procedure of isolation and analysis is carried on in the usual manner. It is necessary to assay the final product for the proportion of N^{15} from which the concentration of the compound sought can be calculated. The method depends on the fact that the chemical properties of a compound are not altered by the introduction of an isotopic element. Thus the method is capable of great accuracy since all errors of isolation and analysis are equally distributed between the two forms. The limitations of the method lie in the ability to prepare the desired compound from N^{15} and in the fact that racemization does alter the solubility of the amino acid. The apparatus necessary to the determination of the amount of heavy nitrogen is probably the most serious deterrent to the wide use of this method.

The Solubility Method. Bergmann and his co-workers [81] have introduced the concept of the solubility product in the determination of the concentration of amino acids in protein hydrolysates. The method depends on the fact that the solubility product of the ions of a slightly soluble compound is a constant. They used a variety of aromatic sulfonic acids which would form somewhat insoluble salts with amino acids. To each of two equal aliquots of the solution to be analyzed, known quantities of the precipitant are added; this is in each case less than that amount which will give the maximum precipitation. From the weight

[80] D. Rittenberg and G. L. Foster, *J. Biol. Chem.*, **133**, 737 (1940).

[81] S. Moore, W. H. Stein, and M. Bergmann, *Chem. Revs.*, **30**, 423 (1944).

of the precipitate X, the solubility constants K_1 and K_2 may be written

$$(A - X_1)(R_1 - X_1) = K_1 \tag{155}$$

and

$$(A - X_2)(R_2 - X_2) = K_2 \tag{156}$$

where A is the amount of the amino acid present and R is that of the precipitant added. Since $K_1 = K_2$,

$$(A - X_1)(R_1 - X_1) = (A - X_2)(R_2 - X_2) \tag{157}$$

One can calculate A from the above since the X and R values are known.

The second method depends on adding a known weight of the preformed precipitate X to a known volume of the solution of A. In a parallel run a known quantity of the precipitant R is added. From similar considerations it is possible to calculate the concentration of A in the hydrolysate.

It will be noted that it is not necessary to use a specific precipitant which yields a very insoluble salt; in fact it is not desirable. In Bergmann's laboratory a number of sulfonic series have been used as reagents.

Microbiological Assays. For some years vitamin preparations have been assayed with certain organisms grown in a culture medium adequate in every respect except for the vitamin in question. This was supplied in varying quantities from the sample to be assayed. The necessary amino acids were furnished by casein acid hydrolysates supplemented with tryptophan.

The method has readily been adapted to the determination of small quantities of amino acids in proteins by making one amino acid of a hydrolysate of that protein the limiting factor in supporting growth. One method of measurement involves the determination of the rate of growth before the maximum has been attained; the other method more generally employed is to measure the total growth of the colony. This is an application of the principle that the food (nutrients) is the limiting factor in determining the maximum possible population when all other factors are adequate. When the lactic acid bacteria are used as test organisms one may measure either the turbidity of the solution or titrate the lactic acid produced. The nephelometric method requires that the suspension of the organism be uniform and that no other turbidity develops during the incubation. These conditions are difficult to obtain with certain organisms; consequently the titration method is more commonly employed. Neither of these methods can be used with the fungus *Neurospora.* It is generally grown in a liquid medium; the mycelial mat is filtered off, washed, dried, and weighed.

The two organisms most widely used are *Lactobacillus casei* and *L. arabinosus*. From runs with authentic samples of the natural antipode of each essential amino acid, growth curves for each organism are determined for purposes of reference. It should be noted that the amount of inoculum is of no great importance since the organism multiplies rapidly. What is important is the population maximum which is determined by the content of the limiting amino acid in the protein examined.

The case of *Neurospora crassa* is of interest. Normally the fungus can convert other metabolites into the leucine which it requires. However, when *N. crassa* is subjected to x-ray or ultraviolet radiations a mutant results which has lost the ability to synthesize leucine.

The following amino acids have been found necessary for the growth of one or several of the organisms; protein hydrolysates have been assayed for the content of these acids to give values which are in good agreement with those obtained by chemical methods: aspartic and glutamic acids, arginine, histidine, lysine, glycine, alanine, valine, leucine, isoleucine, phenylalanine, tryptophan, tyrosine, serine, methionine, threonine, and proline. Of the amino acids more commonly found in proteins only cystine and hydroxyproline have as yet not been found necessary to the growth of some organism. The extension of this technic is being carried on in many laboratories and by many biochemists. For more detailed information the reader is referred to the recent reviews by McMeekin and Warner [82] and by Snell.[83]

[82] T. L. McMeekin and R. C. Warner, in *Ann. Rev. Biochem.*, **15**, 119 (1946).
[83] E. E. Snell, *Advances in Protein Chem.*, **2**, 85 (1945).

CHAPTER 14

Protein Classification

Inasmuch as the exact chemical composition of nearly all the proteins is as yet unknown, it has been found convenient to divide them into classes, the basis of classification being chemical so far as possible; and where a chemical classification is not possible, solubility (or, probably better, peptization) has been made the basis. Two general classifications have been proposed, one by a joint committee of the American Physiological Society and the American Society of Biological Chemists, the other by the English Biochemical Society. They are referred to as the American classification and the English classification.

THE AMERICAN CLASSIFICATION

In 1908, the committees appointed by the American Physiological Society and the American Society of Biological Chemists recommended the classification [1] which is essentially reproduced below. In a few instances it has seemed wise to change the wording of their definitions slightly or to include additional comments.

Simple Proteins. These are the naturally occurring proteins which on being treated with enzymes or acids are hydrolyzed only into α-amino acids or their derivatives. They differ from the conjugated proteins in that the latter are not only hydrolyzed into amino acids but also yield other non-protein substances. Within the group of the simple proteins a number of subdivisions may be recognized, largely on the basis of solubility or other properties.

Albumins. The albumins are soluble in water and in dilute salt solutions and are coagulable by heat. Typical examples are ovalbumin from egg white, serum albumin, and vegetable albumins such as the leucosin from wheat. (As a matter of fact, some of the albumins which have been thoroughly studied, notably egg albumin and serum albumin, contain carbohydrate residues and should accordingly be classified strictly with the conjugated proteins under the subclass of "glycoproteins."

[1] Joint Recommendations of the Physiological and Biochemical Committees on Protein Nomenclature, *J. Biol. Chem.*, **4**, xlviii (1908).

Both the American and English classifications, however, list them as the first subclass of the simple proteins. This fact is an illustration of the difficulties of exact protein classification.)

Globulins. The globulins are simple proteins, insoluble in pure water, but soluble in dilute neutral solutions of the salts of strong bases and acids. Typical examples are the serum globulins, the myosin of muscle, and edestin from hempseed. Globulins have been isolated from a great many vegetable seeds.

Glutelins. These are simple proteins, insoluble in all neutral solvents, but readily soluble in *very* dilute acids and alkalies. Examples are the glutenin from wheat and oryzenin of rice.

Prolamines or Alcohol-Soluble Proteins. This class is defined as insoluble in water, absolute alcohol, or other neutral solvents, but soluble in relatively strong alcohol (70–80 per cent). Typical examples are zein from corn, gliadin from wheat, hordein from barley; with a single exception (the alcohol-soluble protein from milk, isolated by Osborne), the prolamines are confined to the seeds of the cereal grains.

Albuminoids. These are defined as simple proteins which possess essentially the same chemical composition as other proteins, but are characterized by a great insolubility in all neutral solvents. They are, in general, insoluble in dilute acid, alkali, water, or salt solutions. This group is in reality a heterogeneous group, containing various proteins of widely different physical properties. Apparently all the proteins which do not fit definitely into some other class are grouped with the albuminoids. Examples are the keratin from horn, hide, hoof, hair, and feathers of animals, elastin from ligaments, collagen from hide and tendons, and gelatin from hide, hoofs, bones, etc. Gelatin, though classed as an albuminoid, does not conform to the definition noted above. It is more properly a protein derived from collagen, which fact probably accounts for its inclusion in the group.

Histones. Histones are soluble in water, insoluble in very dilute ammonia, not coagulable by heat, easily soluble in very dilute acids or solutions of the fixed alkalies, and on acid hydrolysis yield a large number of amino acids, among which the basic amino acids predominate. As a general rule, histones form precipitates with solutions of other proteins. These precipitates are probably coacervates formed under conditions where the histones are positively charged but the other proteins are on the negative side of their isoelectric point. A similar precipitate which appears to be quite characteristic of the histones is formed with sodium alizarin sulfonate and is used for the characterization of the histones. Histones are basic proteins, their basicity being intermediate between the protamines and the more common proteins. Typical examples are

the globin from hemoglobin and the histones from birds' corpuscles and from the thymus gland.

Protamines. These are the simplest natural proteins. In reality they may be regarded as simply large polypeptides. The better-characterized protamines contain from 14 to 20 peptide linkages, and their molecular weights do not exceed 3,000. They are highly basic, soluble in water, soluble in ammonia, and they form true salts with mineral acids, many of the salts being crystalline. They are not coagulable by heat, and yield on hydrolysis relatively few amino acids, with the diamino acids greatly predominating. They have the property of precipitating other proteins from aqueous solution, again probably as coacervates. They form coacervates with nucleic acid, with arabic acid, and with other negatively charged lyophilic colloids. They are typically found in the ripe sperm cells and have been largely isolated from fish sperm. Examples are salmine, containing 88 per cent of arginine, from salmon sperm and sturine from sturgeon sperm.

Conjugated Proteins. These are compounds of simple proteins with some other non-protein group, the union with the non-protein molecule being in many cases other than a salt.

Nucleoproteins. These are compounds of one or more protein molecules with nucleic acid. They are probably coacervate systems. They are the proteins of the cell nuclei and apparently make up in a large measure the substance of the chromatin. Examples are nuclein and nucleohistone from tissues rich in nuclear material, such as glandular tissues and yeast.

Glycoproteins or Glucoproteins. These are proteins in which the additional group is a carbohydrate radical other than in the form of that contained in nucleic acid. The mucus-yielding proteins of tissues are particularly rich in glycoproteins. They possibly serve as a cementing substance for holding together the fibers in tendons and ligaments. The mucin secreted by the snail or by the salivary glands, the protein making up the great majority of the structure of the jellyfish, and the proteins forming the jelly surrounding fish and amphibian eggs are typical examples of these proteins.

Phosphoproteins. The prosthetic group in these proteins is *o-phosphoric* acid esterified either as the monoester or as the diester with the —OH group of the hydroxyamino acids, particularly with serine. Typical examples are casein from milk, and vitellin from egg yolk.

Chromoproteins. This group is referred to in the original classification as the *hemoglobins.* It should, however, be extended to include other proteins than the hemoglobins. The conjugated group is colored and may be hematin as in hemoglobin, cyanin as in hemocyanin, or a group

of unknown nature, such as occurs in the colored proteins of certain seaweeds where the proteins have been named phycoerythrin and phycocyan. Chromoproteins likewise occur in certain pigmented animal fibers, such as black wool and hair. The colored group is melanin. Visual purple of the retina of the eye is a chromoprotein in which the colored group is a carotenoid. Several yellow flavoproteins are known among enzyme systems; the prosthetic group is a phosphoric ester of riboflavin.

Lecithoproteins. These are the proteins of the cytoplasm and of the cell membrane area. Again, they are probably coacervate systems. The conjugated group is a phospholipid such as lecithin. Lung tissue is rich in such proteins.

Lipoproteins. This group was added by A. P. Mathews, the conjugated group being one of the higher fatty acids. Such compounds are so easily prepared artificially that their natural occurrence is deemed probable.

Derived Proteins. This group includes the various decomposition products of the naturally occurring proteins which have been produced by the action of reagents or enzymes or physical agents, such as heat, hydrogen-ion concentration. It also includes the artificially synthesized compounds. It is divided into various groups according to solubility, and to a lesser extent according to the degree of complexity.

Primary Protein Derivatives. COAGULATED PROTEINS. These are insoluble protein products produced from natural proteins by the action of heat, alcohol, or by some similar method.

PROTEANS. These are the initial product of the action of very dilute acids or, in some instances, water or enzymes, upon certain globulins. This form of derived protein is particularly characteristic of the globulins and differs physically from the globulins by a loss of solubility in dilute salt solutions. Except for the fact that they have been produced by laboratory technic from "globulin," they have all the characteristics of the naturally occurring glutelins. Examples are edestan from edestin, myosan from myosin, etc.

METAPROTEINS. These are produced by the further action of acids and alkalies upon proteins. They are, as a rule, characterized by being soluble in very weak acids or alkalies but by being insoluble in neutral solutions. Examples are acid metaprotein (acid albuminate) and alkali metaprotein (alkali albuminate).

Secondary Protein Derivatives. PROTEOSES. The partial hydrolytic decomposition products of proteins, these are soluble in water, noncoagulable by heat, and *precipitated* by saturating their solutions with ammonium sulfate.

PEPTONES. Also partial hydrolytic decomposition products of proteins, peptones are soluble in water, non-coagulable by heat, and *not precipitated* by saturating the solutions with ammonium sulfate.

PEPTIDES. These are definitely characterized compounds of two or more amino acids, the carboxyl group of one being united with the amino group of the other, with the elimination of a molecule of water. They are not heat-coagulable; they may or may not give the biuret reaction; they may be either of natural origin or synthetic.

DIKETOPIPERAZINES. These are cyclic anhydrides of two amino acids. They may be regarded as the anhydride of a dipeptide.

THE ENGLISH CLASSIFICATION

The English classification differs from the American classification only in minor details.

Simple proteins
- Protamines
- Histones
- Globulins
- Albumins
- Glutelins
- Gliadins (prolamines) (soluble in 80 per cent alcohol; insoluble in water)
- Scleroproteins (forming the skeletal structure of animals)
- Phosphoproteins

Conjugated proteins
- Chromoproteins
- Nucleoproteins
- Glucoproteins

Hydrolyzed proteins
- Metaproteins
- Albuminoses or proteoses
- Peptones
- Polypeptides

The definitions adopted for these groups are essentially those already given in the American classification. It will be noticed that the English classification places the phosphoproteins among the simple proteins.

THE CRITERIA OF PROTEIN CLASSIFICATION

A casual glance at the above systems of classification indicates definitely that the major distinction lies in physical properties. *Solubility* and *precipitability* have been emphasized at a number of points. Since the above classifications were adopted a great deal of attention has been given to the physicochemical properties of protein systems. Hardy [2] observed that untreated serum readily passed through a porous pot but that, if water, acids, or neutral salts were added, the material was largely retained by the filter. From milk whey Osborne and Wakeman [3] isolated 65 per cent of the protein as an "albumin," and some years later Palmer [4] prepared a globulin which appears to be homogeneous and which represented 60 per cent of the whey protein. These and other data raised some question as to the validity of the classification and suggested that many protein isolations may be artifacts or fractions of native protein systems.

Hoffman and Gortner,[5] in an attempt to isolate a quantity of the various proteins which had been reported to occur in wheat flour, observed that extracting the flour with 5 per cent potassium sulfate solution and with 10 per cent sodium chloride solution did not yield similar fractions, but that, instead, the amount and nature of the protein material dissolved by these two reagents were markedly different. These differences were especially noticeable in the globulin fraction. According to the definition of globulin, the two solutions should have yielded identical fractions.

Accordingly, Gortner, Hoffman, and Sinclair [6] definitely raised the question which is ignored in the definition of globulins, *i.e.*, "*What salts, and what concentrations?*" They undertook an extensive study of this question, utilizing 12 different wheat flours and studying the behavior of 22 different salts, most of them in three or more concentrations.

Table 42 shows the *average* percentages of protein extracted from this series of wheat flours by the various concentrations of the different salt solutions. In order to be certain that the effects observed were due to the salt solutions and not to variations in hydrogen-ion concentration, the 1.0 N solutions of the potassium halides were carefully adjusted to the same hydrogen-ion concentration. A lyotropic series of KF < KCl < KBr < KI was found, with extreme ranges in the percentage

[2] W. B. Hardy, *J. Physiol.*, **33**, 251 (1905).
[3] T. B. Osborne and A. J. Wakeman, *J. Biol. Chem.*, **33**, 7 (1918).
[4] A. H. Palmer, *J. Biol. Chem.*, **104**, 359 (1934).
[5] W. F. Hoffman and R. A. Gortner, *Cereal Chem.*, **4**, 221 (1927).
[6] R. A. Gortner, W. F. Hoffman, and W. B. Sinclair, *Kolloid-Z.*, **44**, 97 (1928); *Cereal Chem.*, **6**, 1 (1929).

of total protein extracted from 13.07 per cent to 63.89 per cent. *In this instance, at least, the degree of peptization can be due only to a specific influence of the anions which are present in equivalent concentrations.*

TABLE 42. AVERAGE PERCENTAGE OF TOTAL PROTEIN * EXTRACTED FROM 12 WHEAT FLOURS BY VARIOUS CONCENTRATIONS OF SALT SOLUTIONS

Salt	Concentration of Salt		
	0.5 *N*	1.0 *N*	2.0 *N*
Li acetate	25.01	22.32	
LiCl	30.08	29.32	28.60
NaCl	23.16	21.68	18.97
Na_2SO_4	20.00	18.20	17.30
Na_3 citrate	25.07	25.13	24.52
Na_2HPO_4 †	27.44	25.62	25.32
KF	.....	13.07	11.93
KCl	24.62	22.77	20.01
KBr	38.77	37.22	34.91
KI	.....	63.89	
K_2SO_4	19.68	18.59	
K_2 tartrate	26.35	24.12	
K_2CrO_4	26.92	.. ..	
$MgCl_2$	33.01	34.99	38.35
$MgBr_2$	30.17	40.11	54.38
$MgSO_4$	26.18	26.35	25.69
$CaCl_2$	34.14	36.60	32.25
$CaBr_2$	33.90	43.27	53.68
$SrCl_2$	32.59	34.50	36.47
$BaCl_2$	27.29	33.42	29.24
$ZnSO_4$	.....	16.71	
$Al_2(SO_4)_3$	.....	16.90	

* Each figure represents the average of at least 24 separate nitrogen determinations.
† Considered as a divalent salt.

The authors point out that all the various protein fractions which have been considered to be present in wheat flour and which are regarded as fixed entities can be isolated in a constant proportion, provided that a prescribed routine procedure for protein isolation is followed. If, however, that prescribed routine procedure is altered somewhat, different results are obtained. This is to be expected if one considers that *the system in question is a colloid system and that solubility is synonymous with peptization.* On the other hand, it likewise means that the fractions

which are isolated by a prescribed technic are not necessarily chemical entities but may represent only a peptized fraction.

They further point out that we know far too little about the physicochemical factors underlying protein peptization to assign a cause for the variations noted. The cause is probably complex, involving the nature of amino acid linkages, secondary valence, polar groups, degree of hydration, specific ionic effects, electrokinetic forces, etc., and until at least some of these have been evaluated it is useless to speculate about the causes involved.

The term globulin is assigned to a group of proteins separable from other protein fractions of tissues by a purely arbitrary procedure. Many globulins such as those of blood serum have been separated into euglobulins and pseudoglobulins on an equally arbitrary basis. It is to the point to note at this time that the same serum globulins have been separated into at least three fractions, designated as α-, β-, and γ-globulins, by the Tiselius electrophoretic technic (see Fig. 26).

When a globulin changes its solubility and is transformed into a protean, such as myosin to myosan, it is classified as a "derived protein," although nothing is known of the chemical change which has come about, and although the protean has practically all the properties of the class of simple proteins known as glutelins. No one can say that the glutelins are not proteans which have undergone the globulin → protean transformation in nature.

In an attempt to ascertain how generally the phenomena of peptization might be applied to the problems of protein behavior, Staker and Gortner,[7] using 0.5 *M* solutions of KF, KCl, KBr, KI, and K_2SO_4, studied most of the seeds and grains which had formed the basis of Osborne's work on the vegetable proteins. Figure 78 shows a graph of certain of the data obtained. A study of the ratios of albumin to globulin using the various technics indicated that *the albumin:globulin ratio varied widely with the particular technic employed.* In view of this observation, it is not surprising to find that one worker, using one technic, reports a high content of albumin and a low content of globulin, and another worker, using a different technic, reports exactly the opposite.

At about the same time that the above studies were in progress, the workers at the Carlsberg laboratories [8] were studying the solubility behavior of various proteins. These studies culminated in Sörensen's [9]

[7] E. V. Staker and R. A. Gortner, *J. Phys. Chem.*, **35**, 1565 (1931).

[8] S. P. L. Sörensen, *Compt. rend. trav. lab. Carlsberg*, [11] **15** (1925); K. Linderstrøm-Lang and S. Kodama, *ibid.*, [1] **16** (1925); K. Linderstrøm-Lang, *ibid.*, [9] **17** (1929); S. P. L. Sörensen and I. Sládek, *ibid.*, [14] **17** (1929); and G. Haugaard and A. H. Johnson, *ibid.*, [2] **18** (1930).

[9] S. P. L. Sörensen, *Compt. rend. trav. lab. Carlsberg*, [5] **18** (1930).

theory of "reversible-dissociable component systems." The gist of this theory can be expressed in Sörensen's own words: "Soluble proteins consist of a series of complexes or components, reversibly combined, which makes their constitution expressible by the ordinary formula $A_xB_yC_z \cdots$ A, B, C and so on each marking complete complexes, mainly polypeptides, yet in some cases also containing other groups, for example phosphorous ones, whereas the affixed indices x, y, z, and so on mark the amount to which the indicated complex is present in the entire component system. Within each complex all the atoms and atom groups are linked together by main-valences, whereas the various complexes in the whole component system are comparatively loosely and reversibly knit together by means of the residual valences which each component must be assumed to possess, and the strength and nature of which must depend on the chemical composition of the component in question as well as on its physical properties, above all on its dimensions and the resulting shape and surface. But all things considered, the linkage between the components must be supposed to be comparatively slight and of such a nature that alterations in the composition of the solution (salt content, hydrogen-ion activity, alcohol content, temperature) may give rise to reversible dissociations of the involved component systems and interchange of components between the same. When these alterations in the composition of the solution are so suited as to render possible in sufficient quantities the formation of a component system insoluble or sparingly soluble under the new conditions, such a system

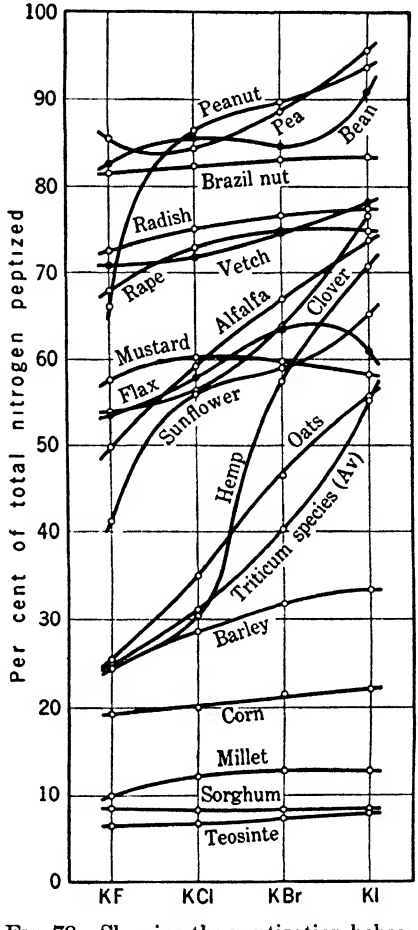


FIG. 78. Showing the peptization behavior of 0.5 M solutions of the potassium halides on the protein complex present in various seeds and grains. (Data of Staker and Gortner.)

will be formed and precipitated. In good accord with this is the fact that through suitable proceedings it has been possible to effect a reversible fractionation in the case of all hitherto investigated proteins. In the main the fractions gained possess indeed the properties of the initial material, yet both the physical properties and the chemical composition are more or less modified from fraction to fraction because of their varying contents of the different components."

Sörensen's theory differs from the peptization theory of Gortner, *et al.*, primarily in the fact that he regards solubility differences as due to the heterogeneous nature of the protein micelle and dependent on specific chemical configurations in the various peptide units, whereas the peptization theory lays emphasis upon physical surface forces and colloidal phenomena. Under Sörensen's theory a given protein aggregate can be fractionated into a series of fractions, each fraction possessing its own individual physical properties and characterized by its own specific chemical composition, the physical properties being a function of the chemical composition. A recombination of different fractions would then bring about a system possessing the same physical and chemical characteristics as were possessed by the original protein aggregate. In order to test whether or not a recombination of fractions was necessary or whether the original physical behavior of the protein aggregate was primarily determined by the colloidal state of the system, gliadin was repeatedly extracted [10] with molar potassium iodide solution until no more protein could be peptized from the residue. It was found that approximately 75 per cent of the gliadin was "dissolved" by the potassium iodide solution and approximately 25 per cent remained as a non-peptizable residue. On the basis of Sörensen's theory this should have resulted in a *chemical* fractionation of the original gliadin, and by a recombination of these fractions the properties of the original gliadin preparation should have been regained. However, the "soluble fraction" and the "insoluble fraction" were *separately* reworked (electrodialyzed, dissolved in alcohol, repeptized, redissolved, and reworked by the "standard method" for the preparation of gliadin) to secure dry white gliadin powder as in the original protein preparation. These two fractions were then subjected to peptization with potassium iodide as in the original preparation. *The peptization behavior of these separate fractions resembled the peptization behavior of the initial protein preparation.* Approximately 87 per cent of the "potassium iodide-soluble" fraction was peptized by the second series of potassium iodide extractions, and approximately 33 per cent of the reworked "potassium-iodide-insoluble" fraction was similarly peptized. Here the properties of the

[10] W. B. Sinclair and R. A. Gortner, *Cereal Chem.*, **10**, 171 (1933).

original preparation were essentially regained not by a *recombination* of polypeptide fractions as Sörensen's theory requires but by simply reworking either the "soluble" or the "insoluble" fraction by as nearly as possible the same standard technic as was used to prepare the original sample of gliadin. The dry purified protein had been brought back to approximately the same *physical* state, and the subsequent peptization behavior of the colloidal aggregates in the dry protein gel reflected the similarity in physical state. The two fractions of gliadin noted above could not be differentiated from each other by any *chemical* analyses which were applied.

There is no question that in some instances the individual fractions of a "protein aggregate" do differ from one another in chemical composition. Workers at the Carlsberg laboratories have shown that to be true in certain of their studies. Blish and Sandstedt [11] showed it for glutenin. Lustig,[12] using non-standard technic, isolated 4 "euglobulins," 4 "pseudoglobulins," and 3 "albumins" from blood serum and showed that these differ from one another in "solubility" and in chemical composition. Block [13] from an independent series of studies on blood serum reaches the following conclusions:

"Serum does not contain several independent proteins. The fractions isolated by physico-chemical methods are not pre-existent in the serum but result from the technic employed.

"The albumins and globulins obtained from cattle serum by precipitation with various neutral salts are not of constant basic amino acid composition but are simply artificial products produced by the reagents employed in their preparation.

"A better characterization of serum protein could be effected by a study of the whole protein rather than by attempting to isolate any 'pure' preparation even though it might be crystalline.

"Proteins isolated from living tissue are not necessarily present as such in that tissue during life."

Other workers have reached similar conclusions. McCalla and Rose [14] dispersed wheat gluten in a solution of sodium salicylate and precipitated a series of fractions from this solution by the gradual addition of magnesium sulfate. The various fractions were analyzed for total nitrogen, amide nitrogen, and arginine nitrogen and were found to vary systematically from fraction to fraction, the amide nitrogen increasing from

[11] M. J. Blish and R. M. Sandstedt, *Cereal Chem.*, **6**, 494 (1929).

[12] B. Lustig, *Biochem. Z.*, **225**, 247 (1930); **231**, 39, 472; **238**, 307 (1931).

[13] R. J. Block, *J. Biol. Chem.*, **103**, 261 (1933); *cf.* also *J. Biol. Chem.*, **104**, 343 347 (1934).

[14] A. G. McCalla and R. C. Rose, *Canadian J. Research*, **12**, 346 (1935).

the first to the seventh fraction, the arginine nitrogen at the same time decreasing. They reached the conclusion that there is no evidence for the existence of "gliadin" or "glutenin" as independent proteins in wheat gluten and that "the terms 'glutenin' and 'gliadin' should be discarded. . . . Gluten is a single protein complex which may be separated into a great many fractions which differ progressively and systematically in both physical and chemical properties." Rich,[15] from an independent study of wheat protein, reaches a similar conclusion: "By arbitrarily choosing the methods of isolation, five protein fractions may be obtained. If, however, the dispersing agent or the method of isolation is changed, even slightly, there is no established limit as to the *number*, *quantities*, and *character* of the protein fractions that can be isolated."

Young [16] separated the proteins of egg white by two variations of the recognized methods. He found little or no "ovaglobulin" but did prepare albumin, mucin, and mucoid fractions. Analyses of the several fractions prepared by the two technics did not agree. He concluded that the products obtained were artifacts resulting from a "mucoprotein-albumin broken by the tools we use to separate it."

One of the most striking examples of the preparation of a series of proteins by a non-conventional method is that developed by Cohn and his co-workers.[17] From human plasma they separated four major fractions by dialysis against ethanol under conditions of controlled temperature, *p*H, and ionic strength. At −2.5°C. and in a 15 per cent ethanol medium, fibrinogen was the major product; at −5°C. and an alcohol content of 20 to 25 per cent, γ-globulin; when the alcohol content had reached 30 to 40 per cent, the α- and β-globulins constituted the main portion; and, in a solution buffered to *p*H 5.5 and containing 40 per cent alcohol, albumin was recovered. Each of these fractions was reprecipitated and, in certain cases, further fractionated.

The various fractions were carefully checked by the Tiselius electrophoretic method and found to be homogeneous. The work was undertaken as a war emergency method to prepare plasma proteins and, more particularly, the albumin fraction for use in cases of shock and severe burns.[18] The detailed work is described in twenty-three papers published by Cohn and co-workers [19] from several laboratories. Without minimizing the importance of this outstanding contribution to human welfare, and to our knowledge of protein behavior, we wish at this point to cite

[15] C. E. Rich, *Cereal Chem.*, **13**, 522 (1936).
[16] E. G. Young, *J. Biol. Chem.*, **120**, 1 (1937).
[17] E. J. Cohn, *et al.*, *J. Am. Chem. Soc.*, **62**, 3386 (1940); **68**, 459 (1946).
[18] E. J. Cohn, *Am. Scientist*, **33**, 61 (1945).
[19] E. J. Cohn, *et al.*, 23 papers in *J. Clin. Invest.*, **23**, 417 (1944).

this study as an example of the fractionation of the plasma proteins (again perhaps one protein complex) by methods not included in the accepted scheme of classification.

The American system of protein classification will be employed in later discussions, but it is desirable for the student to remember that the "albumins," "globulins," etc., which will be discussed are probably laboratory artifacts arising from a protein complex and that *the properties of the protein complex as it exists in the cells and tissues of the living organism are the all-important determiners of cell behavior, and the properties of "albumins" and "globulins" as we study them in the laboratory may bear no direct relationships to the properties of proteins as they exist in living tissues.* The research worker should recognize that solubility is nothing more or less than peptization and that the proteins must be considered not alone as complex organic compounds but likewise as colloid micelles, subject to all the varied reactions of a lyophilic system. Only under such conditions will the study of protein classification and reactions characteristic of protein systems be advanced.

CHAPTER 15

Protein Structure and the Possibilities of Protein Isomerism

We have already noted that the peptide linkage has been definitely proved to be a major linkage in proteins. We have also indicated that polypeptides which have been obtained by the partial hydrolysis of proteins have been artificially synthesized, so that their structure is known. Except for a few proteins which are rich in dibasic acids, most proteins possess isoelectric points in the *p*H range 4.7 to 6.0. Few of the α-amino and the carboxyl groups are free in proteins; but these groups appear in essentially equivalent amounts as hydrolysis proceeds. These two facts can best be interpreted by recognizing that the peptide linkage is the predominant mode of combining the amino acids into chains. The reader is reminded that the positive biuret test given by proteins and polypeptides also supports the theory of peptide linkages.

This theory is not entirely adequate because it does not account for such observations as the differences in solubilities among proteins which consist of almost the same kinds and quantities of amino acids. More recently a number of workers have laid emphasis on the properties of other forms of linkage than that of the long-chain polypeptide type. We cannot discuss any of these viewpoints at length, but it seems worth while to indicate the trend of thought in this field.

Theories of Protein Structure. Critical reviews [1] of the various hypotheses which have been proposed to account for protein structure are available. Although these are possibly ultraconservative, nevertheless they bring together a number of divergent views and compare them one with another.

Kossel's "Protamine Nucleus" Hypothesis. Kossel [2] postulated from his study of the basic amino acids that the nucleus of all proteins would be found to be of the "protamine" type, *i.e.*, to be largely composed of either arginine, histidine, or lysine, or combinations of these three amino

[1] H. B. Vickery and T. B. Osborne, *Physiol. Revs.*, **8**, 393 (1928); *cf.* also H. B Vickery, *Yale J. Biol. Med.*, **4**, 595 (1932).

[2] A. Kossel, *Z. physiol. Chem.*, **22**, 176 (1896).

acids. Kossel's hypothesis was more or less ignored for a period of years. Larmour,[3] however, again called attention to this hypothesis in a statistical study of the relationship which exists between the total basic nitrogen and the arginine nitrogen or the lysine nitrogen of the various proteins, as determined by Van Slyke's method. From a statistical study of 214 analyses of proteins by Van Slyke's method, he found a coefficient of correlation of $r = +0.794 \pm 0.017$ between the total basic nitrogen and the arginine nitrogen, and a coefficient of correlation between the total basic nitrogen and the deviation of the arginine nitrogen from its most probable value, the mean, of $r = +0.211 \pm 0.044$. The coefficient of correlation between total basic nitrogen and lysine nitrogen for 213 cases was found to be $r = +0.548 \pm 0.032$, and the coefficient of correlation between total basic nitrogen and the deviation of lysine from its most probable value, the mean, was found to be $r = +0.03 \pm 0.046$. Larmour believes that the high correlations in the case of arginine are evidence in support of Kossel's hypothesis that arginine is the nucleus of the protein molecule and states, "It seems highly probable that in all proteins there is a direct and marked relationship between the arginine and the total basic nitrogen."

Block [4] finds that, although proteins may be separated into fractions which differ both in physical and chemical properties, nevertheless there is a strong tendency for the ratios between arginine, histidine, and lysine to remain relatively constant. This is most true of the keratins where the ratios are 12:1:4, respectively. He regards this as evidence "for a basic amino acid 'anlage' " existing in the proteins. Alcock [5] takes a somewhat similar view of an "urprotein" in which arginine, histidine, lysine, cystine, and tryptophan play a greater role than the other amino acids, and in a number of other papers emphasis has been laid on the possible role which the basic amino acids play in determining the structure of proteins.

Abderhalden's Anhydride Theory. As early as 1902 Fischer noted that leucinimide, a substituted diketopiperazine, could be separated from a tryptic-peptic digestion of hemoglobin, and he added that it unquestionably existed preformed in the protein molecule. Some twenty years later, Abderhalden began an extensive series of investigations on the structure of proteins, and as a result of these studies he concluded that proteins are to a very considerable extent built up of cyclic substituted diketopiperazine rings held together by the force of latent valence. The

[3] R. K. Larmour, *Trans. Roy. Soc. Canada*, Section V, 349 (1928).

[4] R. J. Block, *J. Biol. Chem.*, **103**, 261 (1933); *cf.* also *J. Biol. Chem.*, **93**, 113 (1931); **104**, 339, 343, 347 (1934); **105**, 663 (1934).

[5] R. S. Alcock, *Physiol. Revs.*, **16**, 1 (1936).

literature which Abderhalden and his students have presented is altogether too voluminous to be cited in detail. Most of the papers [6] will be found in the *Zeitschrift für physiologische Chemie.*

Abderhalden [7] pointed out that the presence of an anhydride structure possessing keto $\rightleftharpoons$ enol isomerism may well be responsible for the extreme lability of certain proteins. He noted that, even in Fischer's octadecapeptide or in his own nonadecapeptide, the polypeptide showed a great stability and was not readily altered or affected by chemical reagents which would not hydrolyze it, and he contrasted this behavior with the behavior of vital proteins. Certain of his comments are so important to the student of vital phenomena that it seems worth while to append them in the form of a rather free translation. He states, "Each conception in regard to particular structural relations in proteins, and especially of those which are concerned in living processes, must take into consideration all their reactions, their ready transformation from the natural to the denatured condition, and their greater or less lability. It is certain that proteins in protoplasm have properties of which we are at present entirely unaware. We study proteins almost invariably in a more or a less changed condition. On the one hand, we have proteins which in a certain sense are denatured in the organism, and outside of the organism undergo further changes. Thus, for example, we see the silk threads, the web of the spider, the byssus of certain mussels, etc., which are extruded as a liquid, changing to an inert solid. From a very labile form they are transformed into a very stable structure. In living processes we deal with the very reactive cell proteins which regulate in a fine degree the reactions of the cell, but when death ensues, the cell proteins coagulate and lose in a large measure their characteristic physical behavior."

Several diketopiperazines have been found in protein digests produced by acid or enzymatic action and have been identified for amino acid units by comparison with synthetic compounds. They are generally soluble in ethyl acetate which serves as the means of isolation.

Several qualitative tests are given by these anhydrides and by certain proteins or by the peptones prepared from these proteins. Picric acid and *m*-dinitrobenzene are most often used, but these two tests do not always agree when applied to a given protein. Alkaline permanganate will convert the anhydrides to oxamide; no known peptide, with the exception of glycyl-glycine, yields this product. This test would be more conclusive if it could be shown that the quantity of oxamide obtained

[6] For summaries see: E. Abderhalden, *Naturwissenschaften*, **12**, 716 (1924); *Z. physiol. Chem.*, **128**, 119 (1923); E. Abderhalden and E. Schwab, *ibid.*, **139**, 169 (1924).

[7] E. Abderhalden and E. Komm, *Z. physiol. Chem.*, **139**, 181 (1924).

from a protein exceeds that which could be accounted for by the glycine content.

Abderhalden and his co-workers have attempted to stabilize the diketopiperazine ring in silk peptone by reduction with sodium in alcohol to a piperazine ring. One such reduction was claimed to yield the piperazine corresponding to alanine and glycine with reduced tyrosine and glycine as N-substituted side chains. Proof of constitution rests upon a single analysis of the hydrochloride, which agrees moderately well with the calculated value. Later they attempted to isolate the diketopiperazine itself but the product appears to be identical with glycyl-alanyl-glycyl-tyrosine prepared and identified earlier by Fischer and Abderhalden.

The diketopiperazines of the monoaminomonocarboxylic acids are not attacked by any known proteolytic enzyme. For a time it was claimed that, if the anhydrides were made from acidic or from basic amino acids, they would slowly be hydrolyzed by trypsin and by pepsin, respectively. More careful tests have shown that this is not the case.

Some observations of Levene and Bass [8] are pertinent to the question of polypeptide *vs.* diketopiperazine structure. They note that polypeptides hydrolyze slowly and racemize slowly at low alkali concentration and that they hydrolyze slowly but racemize more rapidly at high (1.0 *N*) alkali concentration. Diketopiperazines, on the other hand, racemize rapidly at low alkali concentrations and at high alkali concentrations hydrolyze so rapidly that the amino acids of which they are composed are not racemized. Casein does not act like either polypeptides or the diketopiperazines. Levene and Bass ran the racemization curves and the amino nitrogen:total nitrogen ratios and found that the amino acids of casein were completely racemized in 1.0 *N* alkali at 125°. They concluded that either there are no diketopiperazines in the casein molecule or that, if diketopiperazines are present, the linkage is much more stable than the diketopiperazine linkages that have been studied in the laboratory.

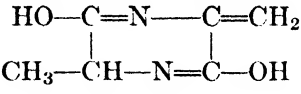
Thus the evidences adduced for and against the occurrence of diketopiperazines in proteins are not convincing, and the hypothesis still awaits more critical tests. The problem is clouded by the possibility that the isolated anhydrides may, in some cases, be secondary products arising during the process of isolation.

Bergmann's Hypothesis. Bergmann postulates a protein structure as being composed of piperazine derivatives held together by forces of latent valence. His theory is very similar to that of Abderhalden, with

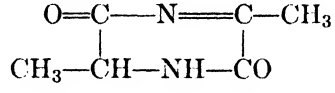
[8] P. A. Levene and L. W. Bass, *J. Biol. Chem.*, **78**, 145 (1928); **82**, 171 (1929)

the exception that Bergmann [9] has been particularly interested in the reactions of hydroxyamino acids and their effect upon the ring structure.

He suggests that a dipeptide, such as alanyl-serine, may form the usual 6-methyl-3-hydroxymethyl-2,5-diketopiperazine and that this may further lose water, the 3-hydroxymethyl group going to a methylene group The diketopiperazine then rearranges to an "iso" form:



This "iso" form forms a disodium salt and a diacetate and is capable of rearrangement to an "allo" form:



This "allo" form polymerizes so that in the presence of water it yields a colloidal hydrosol which adsorbs tannin and a number of dyestuffs, and the polymerized product in many of its reactions resembles the reactions of protein.

Bergmann points out that none of the simple polypeptides or the simple diketopiperazines adsorbs tannin in appreciable amount, whereas his polymerized "allo" form is strikingly like the proteins in that regard. Furthermore, when the "allo" form was subjected to mild hydrolysis, he isolated not the original dipeptide but rather a tetrapeptide which led him to suggest that the isolation of tetrapeptides from proteins is not necessarily proof that the tetrapeptide existed preformed in the protein molecule but may have been derived from some cyclic structure analogous to his polymerized "allo" form. He suggests that the "allo" form polymerizes through secondary valence forces which are resistant toward water or ordinary aqueous solutions.

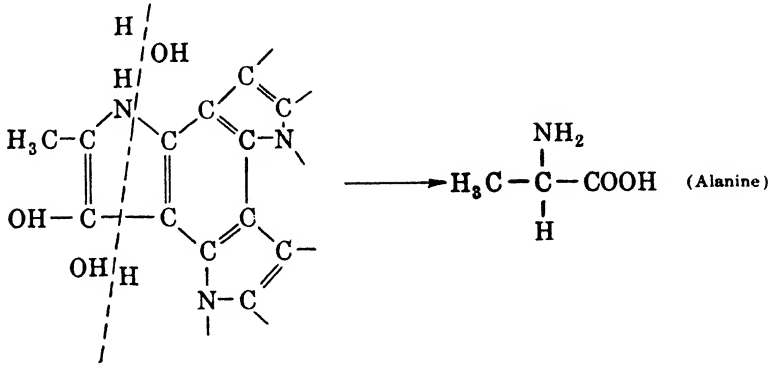
Bergmann admits that no definite proof exists for the presence of such compounds in the protein molecule but believes that the hydroxyamino acids would undergo similar reactions.

Troensegaard's Pyrrole Hypothesis. Troensegaard developed a theory of protein structure based largely upon a study of completely acetylated proteins and the hydrolytic products of such proteins. His numerous papers have recently been summarized in book form.[10] He postulates

[9] M. Bergmann, *Naturwissenschaften*, **12**, 1155 (1924); **13**, 1045 (1925); *cf.* also numerous papers in the *Annalen* and in *Zeitschrift für physiologische Chemie*, and especially *Science*, **79**, 439 (1934).

[10] N. Troensegaard, *On the Structure of the Protein Molecule*, E. Munkegaard Copenhagen, 1944.

the presence of heterocyclic units joined together through carbons; among these are rings of the pyrrole type as indicated, which would hydrolyze to give alanine:



Although Troensegaard's hypothesis has not gained any adherents, it should be recognized that he did a great deal of careful work which might well be repeated in the light of later findings. Several objections have been raised to his postulates. Thus, Vickery and Osborne have pointed out that such structures in proteins are improbable for two reasons: (1) the double bond in the condensed alanine structure should, when hydrolyzed according to Troensegaard's scheme, yield a racemic mixture of amino acids rather than a single optically active acid; and (2) the proposed ring system contains an arrangement of six carbons identical with the arrangement of carbons in the benzene ring. Such a structure should be very stable when submitted to acid hydrolysis.

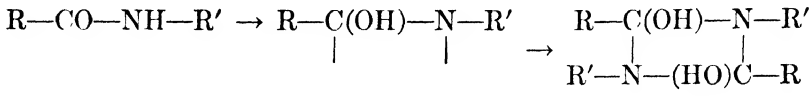
Other Heterocyclic Ring Structures. We have seen that Abderhalden has advanced the hypothesis of the presence of diketopiperazine rings in proteins and that Troensegaard has suggested that heterocyclic rings might also be present. Bergmann's hypothesis of piperazine rings has also been described because he adduced evidence that in the "allo" form such compounds possessed certain properties characteristic of proteins.

Other workers have proposed somewhat similar ring systems to account for some of the linkages within proteins. Karrer [11] pointed out that the enolization of a diketopiperazine might proceed by withdrawing the hydrogen from either the nitrogen or the adjacent carbon; these processes would give rise to a substituted dihydropyrazine or the six-membered ring with the double bonds forming C to C ties. Similarly

[11] P. Karrer and C. Gränacher, *Helv. Chim. Acta*, **7**, 763 (1924); P. Karrer and R. Widmer, *ibid.*, **8**, 203 (1925) (*cf.* also **8**, 205, 211); and P. Karrer, C. Gränacher, and A. Schlosser, *ibid.*, **6**, 1108 (1923).

polypeptides might form substituted iminazole or oxazole rings. Ssadikow and Zelinsky [12] proposed a modified diketopiperazine structure in which the cyclic nuclei are held together by primary valence linkages involving the R groups. This has been referred to as the "polypeptine" unit.

Fodor [13] has proposed that an akropeptide linkage could occur in proteins when treated with glycerol or resorcinol. This linkage results upon the enolization of the peptide linkage yielding two free valence bonds which then form a bridged complex:



This, he continues, is a labile linkage which could readily open under the influence of depolymerizing reagents, such as glycerol or resorcinol. In support of this he submits inadequate analyses of fractions described largely as tetrapeptides (2 dipeptides joined through "akro" linkage) or their polymers. Reid [14] repeated the work, using casein and resorcinol. He concluded that the reaction was essentially one of alcoholysis since each of the four fractions isolated contained resorcinol most of which behaved as a typical ester of the carboxyl group. The fractions were polypeptides ranging in molecular weight from 500 to 1,500. Reid's findings could not be interpreted as evidence of an "akropeptide" structure.

Blanchetière [15] has described some ring structures which result from peptides of the dicarboxylic acids. If the ω-carboxyl group is involved in the chain, unusual ring structures result which he believes may occur in proteins.

The Cyclol Theory. The latest theory of protein structure is that proposed by Miss Wrinch [16] based on a "cyclol" pattern which involves essentially a series "of diazine and triazine hexagons, built into one another in such a way that every triazine hexagon is surrounded by three of the diazine type and pairs of triazine hexagons are joined by a single diazine hexagon." Figure 79 shows Miss Wrinch's cyclol pattern. She observes that this pattern can be derived from the open polypeptide

[12] W. S. Ssadikow and N. D. Zelinsky, *Biochem. Z.*, **136**, 241 (1923).

[13] A. Fodor and S. Kuk, *Kolloid-Z.*, **74**, 66 (1936).

[14] T. S. Reid, PhD. thesis, University of Minnesota, 1943.

[15] M. A. Blanchetière, *Bull. soc. chim. biol.*, **6**, 854 (1924).

[16] D. M. Wrinch, *Proc. Roy. Soc.* (*London*), **A160**, 59 (1937); *Protoplasma*, **25**, 550 (1936); *Science*, **85**, 566 (1937); *cf.* also I. Langmuir, V. J. Schaefer, and D. M. Wrinch, *Science*, **85**, 76 (1937).

chain by a cyclization involving the enolization of the peptide linkage. Figure 80 shows the simplest "cyclol-6" molecule.

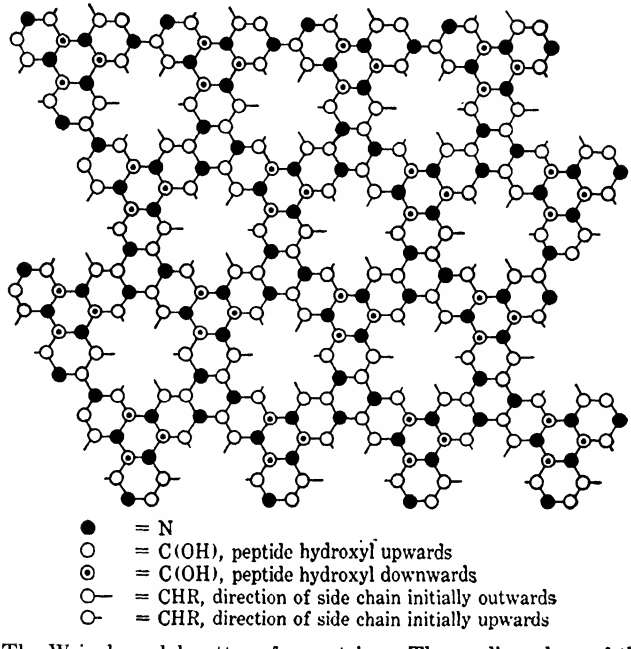


FIG. 79. The Wrinch cyclol pattern for proteins. The median plane of the lamina is the plane of the paper; the lamina has its "front" surface above and its "back" surface below the paper.

Miss Wrinch developed her theory largely from a consideration of mathematical probabilities, x-ray crystal structure data, geometrical isomerism, protein crystal symmetry, and the structure and behavior of unimolecular protein films as measured and studied by means of the Langmuir surface tension balance. She points out that the cyclol theory accounts for most of the physicochemical properties of proteins which have been investigated, including the breaking up of the Svedberg "molecules" into submultiple "molecular" units. The arrangement shown in Fig. 79 is regarded as a sheet of amino acid aggregates, polymolecular in length and breadth but monomolecular in depth. In the cyclol pattern some of the hydroxyl groups are directed downward from the cyclol plane and some are directed upward above the plane. Linkages between the cyclol sheets are accordingly formed

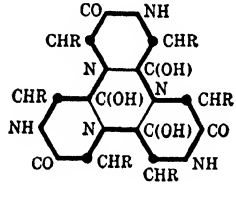


FIG. 80. A "6-cyclol." (After Wrinch.)

through the hydrogen bond or other similar forces linking electropositive groups across to electronegative groups. Thus, according to the cyclol theory, proteins would be built up of a series of lamina, and these lamina may be linked together by means of intermolecular attractions, possibly in the same way as the linkage of structural planes within clay minerals are linked. Proteins therefore can swell largely in the cross-sectional direction, but in the plane of the cyclol sheet the degree of swelling will be determined by the tensile strength of the cyclol bonds and the amount of distortion which the cyclol structure can undergo under the tensile forces which are impressed upon it.

One feature of the theory is that it permits folding into a cage structure. Because only three of the valence bonds of nitrogen are occupied, they can be shifted enough to allow for a folding at an angle of 109°. This will result in a truncated tetrahedron. The dimensions are such as to accommodate 72, 288, or 72 n^2 residues (where n is an integer). The number 288 should be borne in mind later when we consider the theories of Svedberg and of Bergmann.

Several serious objections have been raised to this hypothesis. No compounds containing triazine and diazine rings have been synthesized, and we consequently know nothing about their ease of hydrolysis with acids and with enzymes. Neurath [17] points out that there is not enough room for the R groups. In cross-sectional areas, these groups require from 11.5 to 35 square Ångstrom units (except in glycine where the R group is hydrogen); however, the surface of the 288 units is only 2,970 square Ångstrom units in area, or slightly more than 10 per R group. Pauling and Niemann [18] further argue that the bond energy values and the heats of combustion for a cyclol pattern would result in linkages less stable than that of a polypeptide pattern and, with others, question whether the conclusions drawn from x-ray diffraction patterns on insulin and pepsin are justified. No theory of protein structure has stimulated as much thought or experimentation as has the cyclol theory.

The peptide linkage is still regarded as the major type of union between amino acids in the protein molecule. With the exception of the hypotheses of Abderhalden and of Miss Wrinch, the postulates noted above have received little attention or support since they were first proposed. We shall now consider some features of the architecture of protein molecules that are predicated on the assumption that the peptide bond accounts for the major linkage in proteins.

It will be convenient to use Astbury's classification of proteins as

[17] H. Neurath, *J. Phys. Chem.*, **44**, 296 (1940).

[18] L. Pauling and C. Niemann, *J. Am. Chem. Soc.*, **61**, 1860 (1939).

fibrous or globular in structure. The fiber proteins, with the exception of myosin, are insoluble in water; they possess marked asymmetry, the longest axis being at least 30 times the two shorter axes, and they are often elastic. By contrast, the globular proteins possess less asymmetry and are generally soluble in a water or alcohol system. The shape of proteins can only be inferred from various measurements; the direct method would be to view them in an electron microscope, but this has been possible only with certain virus proteins which are seen to be rod-shaped. The interpretation of x-ray data is not sufficiently developed to give the exact dimensions of many protein units. It is accepted, however, that the methemoglobin unit [19] has these dimensions: 64 Å. long, 48 Å. wide, and 36 Å. in thickness.

Fibrous Proteins. Several examples have been studied. Silk fibroin is composed of few kinds of amino acids, of which glycine constitutes one half of the residues. Meyer and Mark [20] found that the x-ray patterns are not sharp but indicate repeated spacings at 7 Å. units, which is essentially twice the skeletal axis of 3.67 Å. found by Corey [21] for a single amino acid. Consequently their picture of the silk fibroin is that of a completely extended polypeptide chain as shown in Fig. 81. The

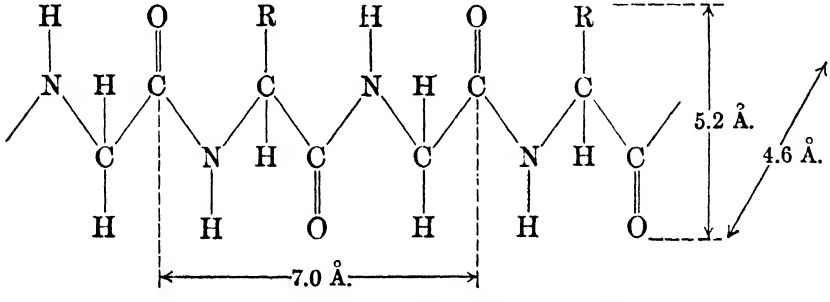


FIG. 81. Silk fibroin. (After Meyer and Mark.)

backbone distance (from one chain to another above or below it perpendicular to the plane of the paper) is 4.6 Å., and the width of the chain is 5.2 Å.

The keratins, particularly hair and wool, show the phenomenon of considerable elasticity. Astbury [22] found that, when wetted, hair can be stretched to three times its original length. This is described by saying that the α-keratin is converted to the β-keratin structure. In α-keratin

[19] J. Boyes-Watson and M. F. Perutz, *Nature*, **151**, 714 (1943).
[20] K. H. Meyer and H. Mark, *Ber.*, **61**, 1932 (1928).
[21] R. B. Corey, *Chem. Revs.*, **26**, 227 (1940).
[22] W. T. Astbury and A. Street, *Trans. Roy. Soc.* (*London*), **A230**, 75 (1931).

no regular backbone structure exists, the axis is 5.15 Å. long, and the side chain spacings extend over 9.8 Å. After α-keratin is stretched to at least 60 per cent over its original length the β-pattern is evident, with a length of 3.32 Å., width of 9.8 Å., and a backbone spacing of 4.65 Å. Thus the β-keratin structure is similar to that of silk fibroin except that the side chain spacing is greater because of larger R groups. It should be recalled that the silk fibroin length of 7 Å. involves two amino acid residues to give a repetition of pattern whereas only one unit is needed in hair with its fewer units of glycine.

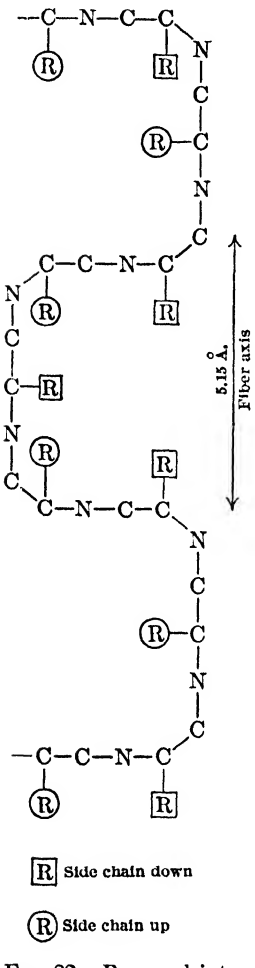


FIG. 82. Proposed intramolecular folding in the skeletal chain of α-keratin. (After Astbury and Bell.)

Astbury has revised his earlier picture of the structure of α-keratin and proposed [23] a structure shown in Figs. 82 and 83. In an earlier paper he had pictured a chain which was coiled in a spiral such that opposite ends of two amino acid units were in a position which would make possible the closing of a ring; this concept was extended by Miss Wrinch in her cyclol structure by bridging from an —NH— group to the —CO— group. Because such a structure did not allow sufficient space for the R groups, Astbury modified his structure to that shown here. In the current form, the dimensions are in keeping with the x-ray pattern, and the R groups appear alternately on one side or other of the fold. It will be noted that any three consecutive R groups on one side form a triangle, an arrangement which allows adequate space for their packing. "A typical though not necessarily universal unit of protein architecture is a sheet or ribbon of folded main chains with side chains projecting from either face." Astbury suggests that the native ovalbumin molecule consists of four such sheets superimposed, and that this picture may apply to all protein structures.

To account for the forces of attraction between the several layers, Lloyd [24] pictured these as arising between the electronegative —CO—

[23] W. T. Astbury and F. O. Bell, *Nature*, **147**, 696 (1941).
[24] D. J. Lloyd, *Biol. Revs.*, **7**, 254 (1932).

group of one and the electropositive —NH— of another layer. More recently Wrinch and Lloyd [25] and Huggins [26] have independently suggested that the linkages between these groups may also be of the hydrogen bond type.

Some such structure admirably accounts for the phenomena of protein hydration [27] and for tensile strength behavior. Fibrous proteins swell greatly in the cross-sectional direction but only elongate slightly in the longitudinal direction. This is explained by water entering in between the "backbone" chains and separating the polypeptide chains from each other, whereas in the longitudinal direction the primary valence bonds prevent any great elongation. Tensile strength decreases greatly across the swollen fiber but is not proportionally decreased in the direction of the fiber length. The x-ray diffraction patterns of the swollen fiber indicate a major displacement of the dimensions across the fiber but relatively little displacement in the direction of fiber length.

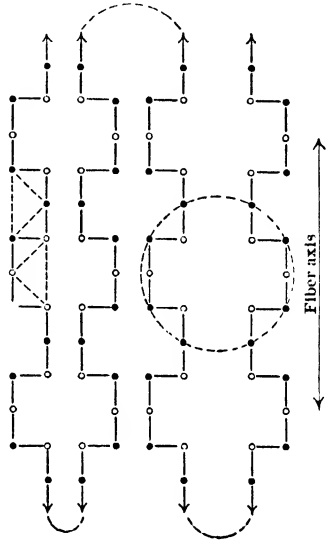


FIG. 83. Suggested arrangement of four peptide chains in α-keratin. Note the regular octagonal arrangement of the R groups. (After Astbury and Bell.)

Globular Proteins. The molecular weight of globular proteins may be determined in a variety of ways since these proteins can be brought into solution. The method most widely used is that involving the ultracentrifuge, first devised by Svedberg [28] and now quite generally used. Some of these values are given in Table 6 (p. 49), where the general theory of the method is considered. Svedberg interprets these values as closely approximating 1, 2, 4, 6, 8, 16, 24, 48, 96, 192, and 384 times the unit weight of 17,600, thus giving rise to weight classes. Bull [29] questions the validity of this conclusion on the basis of the available data; it is to be noted that other workers

[25] D. M. Wrinch and D. J. Lloyd, *Nature*, **138**, 758 (1936).

[26] M. L. Huggins, *J. Org. Chem.*, **1**, 407 (1937).

[27] D. I. Lloyd, *Biochem. J.*, **25**, 1580 (1932); *J. Soc. Chem. Ind.*, **51**, 141 (1932).

[28] The Svedberg, *Chem. Revs.*, **14**, 1 (1934). The Svedberg and I. O. Pederson, *The Ultracentrifuge*, Clarendon Press, Oxford and New York, 1944.

[29] H. B. Bull, *Advances in Enzymol.*, **1**, 1 (1941).

have used these quantized classes in their discussions of protein structure.

A second method for calculating the molecular weights of globular proteins is from osmotic pressure measurements. Bull [30] has developed a rapid and simple procedure which offers promise when used under prescribed conditions.

Calculations of molecular weights from elemental or group (amino acid) analyses will yield only minimal weights, but they can be used in conjunction with other determinations. Diffusion measurements are not adequate because the values obtained are complicated by such factors as the asymmetry of the molecule and the degree of hydration.

To return to the values obtained by use of the ultracentrifuge, there appears to be a "stability range" over which the proteins have a constant molecular weight, and sediment with a uniform velocity, whereas at *p*H's greater or less than this stability range the molecules undergo (in Svedberg's terminology) decomposition into smaller fragments which are in most instances one-half, one-quarter, etc., units of the original molecular weight. This decomposition is reversible, however, and, when the decomposed protein fragments have been reprecipitated and reworked and brought into the *p*H stability zone, the original sedimentation velocity and the original molecular weight are regained.

One additional point deserves further consideration before Svedberg's data can be accurately interpreted. His stability zone lies in the region close to the isoelectric point of the proteins, and his decomposition range lies in regions where the proteins will be ionized and accordingly possess excesses of either positive or negative charges. The decomposition may not be real but may result from the repulsion of like-charged particles which prevent sedimentation in the gravity field. Furthermore, when calculating sedimentation rates it may not be justifiable to ignore the electrical forces between particles at or close to the isoelectric point, for in reality this is the point of their maximum ionization. Here they exist as zwitterions, the negative charges balancing the positive charges within the molecule. Whether or not the zwitterion constitution will have a profound bearing on sedimentation velocity remains to be determined. Certainly it will have a profound bearing on forces which favor or inhibit aggregate formation.

Globular proteins are not necessarily spherical although their asymmetry is less than that of the fiber proteins. The problem of calculating the asymmetries of proteins is complicated by the hydration effect. Both factors increase the viscosity of a protein sol and decrease diffusion

[30] H. B. Bull, *J. Biol. Chem.*, **137**, 143 (1941).

Neurath [31] has calculated the asymmetries by the use of Perrin's equation but recognizes that the values obtained are approximations and may later be revised. Certain of his values are given in Table 43. The

TABLE 43. ASYMMETRIES CALCULATED FOR VARIOUS PROTEINS

(Data of Neurath)

Protein	Molecular Weight	Asymmetries	
		Anhydrous	Hydrated
Lactalbumin	17,400	3.6	1.5
Gliadin	27,500	11.0	7.3
Zein	40,000	30.0	21.0
Insulin	41,000	3.2	1.0
Lactoglobulin	41,500	5.0	2.8
Ovalbumin	45,160	3.6	1.2
Hemoglobin (horse)	68,000	4.2	2.6
Serum albumin (horse)	70,000	4.9	2.8
Serum globulin (horse)	167,000	7.8	4.7
Edestin	310,000	4.2	2.1
Urease	480,000	4.0	1.8
Antipneumococcus serum globulin (horse)	920,000	15.5	11.0

values reported are based on the assumptions that the hydration is 35 per cent, which is in the range often found, that the density of water of hydration is 1.00, and that the particles are prolate ellipsoids. It will be noted that hydration reduces the asymmetries in all cases.

From the above considerations it is evident that we know little about the actual architecture of globular proteins. Obviously a protein with a molecular weight in the neighborhood of 40,000 and containing around 300 units cannot exist in one long polypeptide chain; the latter must be folded over so that the chains lie in approximately parallel axes in three dimensions, as pointed out earlier in the discussion of Astbury's picture of α-keratin.

Sörensen's Reversible Dissociable Component Systems. We have discussed this theory in the preceding chapter, inasmuch as it has a profound bearing on problems of protein classification, and the student is referred to that discussion which is likewise pertinent to the present consideration of the theories of protein structure.

[31] H. Neurath, *J. Am. Chem. Soc.*, **61**, 1841 (1939).

Bergmann's Theory of Recurring Amino Acid Units. Bergmann,[32] from a study of the analytical data of the amino acid content of cattle hemoglobin, cattle fibrin, egg albumin, silk fibroin, and gelatin, has formulated the hypothesis that all simple natural proteins are constructed on similar structural principles, *i.e.*, "Every individual amino acid residue in the peptide chain of the protein molecule recurs at constant intervals . . . in a characteristic and periodic manner throughout the entire polypeptide chain. . . . The peptide structure is common to all proteins, but the various natural proteins differ from each other in that their different amino acid constituents are represented by different frequencies. Thus, the physico-chemical and the biological properties of a particular protein are based, in the last analysis, on the frequencies with which its constituent amino acid residues recur within the peptide chain."

Because the amino acid contents of most proteins are not wholly accounted for, the reasoning which Bergmann employed is illustrated from his study of beef fibrin:

(*a*) Arginine and glutamic acid are present in fibrin in the ratio of 32:72.

(*b*) This ratio may be expressed (inverted) as 9:4, 18:8, 36:16, 72:32, etc.

(*c*) Therefore, if only glutamic acid is considered, there may be 4, 8, 16, 32, etc., individual units of glutamic acid present in the protein molecule.

(*d*) The question as to which is the correct value Bergmann ascertains by the expression

$$\frac{F \times AW}{MW} = \frac{100}{\%} \tag{158}$$

where F = the "frequency" of the particular amino acid being studied
AW = the *average* molecular weight of all the amino acids resulting from a protein hydrolysis
MW = the molecular weight of the particular amino acid whose "frequency" is being determined
$\%$ = the percentage of the particular amino acid in the hydrolysate.

(*e*) In the instance of glutamic acid which is being considered, "frequencies" of 4, 8, 16 yield "average molecular weights" of 260, 130, 65, respectively. The first value is too high, the last value too low to be

[32] M. Bergmann, *J. Biol. Chem.*, **110**, 471 (1935); M. Bergmann and C. Niemann, *ibid.*, **115**, 77 (1936); **118**, 301 (1937); **122**, 577 (1938); M. Bergmann, *Chem. Revs.*, **22**, 423 (1938).

possible. Therefore, *glutamic acid forms one-eighth of the amino acid residues in fibrin.*

(*f*) It is obvious that when this line of reasoning is extended to *all* the amino acids we shall arive at a certain *minimum* number of residues for each amino acid *all of which must, of course, be expressed as whole numbers,* for there can be no partial molecules of an amino acid in the polypeptide chain.

(*g*) The actual *molecular weight* of the protein must be either the molecular weight of the polypeptide containing the *minimum* number of individual amino acid residues as ascertained in (*f*) or must be some simple whole-number multiple of this minimum size.

(*h*) Bergmann then goes further and *postulates* that the "frequencies" represent also "position numbers" and that the amino acids recur at *constant intervals* throughout the entire polypeptide chain, these constant intervals being expressed by their "frequencies" as determined by the F in the expression in (*d*). Thus, in the case of glutamic acid already cited, it will recur in the fibrin molecule as every eighth amino acid residue.

In fibrin Bergmann finds the following frequencies: glutamic acid 8, lysine 12, arginine 18, aspartic acid 18, proline 18, tryptophan 32, histidine 48, methionine 48, and cystine 64, with a *calculated* minimum polypeptide chain length of 576 amino acid residues. In egg albumin the frequencies found are: glutamic acid 8, aspartic acid 18, methionine 24, lysine 24, arginine 24, tyrosine 36, histidine 72, and cysteine 72. These data yield a "calculated" molecule of 288 amino acid residues and a "minimum molecular weight" of 35,700, which is in agreement with Sörensen's earlier value of 34,000 but is considerably less than the more recent value obtained by the ultracentrifuge (44,000) and that found from osmotic pressure measurements by Bull,[30] who calculated a value of 45,160.

The calculations of the average molecular weights involve some uncertainty. Part of this is due to the fact that for the ratios of the frequencies the nearest whole numbers are employed, and part to the fact that no frequencies can be calculated for those amino acids for which no analyses are available. From this it is seen that the number of amino acid residues may be in error, as the molecular weights will be. The whole calculation can be no more accurate than the data upon which it is built. Bull [33] has calculated the number of residues from the average residue weights of the amino acids and from molecular weights of the proteins determined by physical methods. His calculations are reproduced in Table 44.

[33] H. B. Bull, *Advances in Enzymol.*, **1**, 1 (1941).

TABLE 44. COMPARISON OF BERGMANN'S VALUES WITH THE NUMBER OF AMINO ACID RESIDUES CALCULATED FROM THE MOLECULAR WEIGHTS OF VARIOUS PROTEINS AND THE AVERAGE RESIDUE WEIGHT OF THE CONSTITUENT AMINO ACIDS

(Data of Bull)

Protein	Molecular Weight *	Average Residue Weight	Number of Amino Acid Residues	
			Calculated	Bergmann's values
Lactalbumin	17,400	120	145	144
Ovalbumin	43,960	124	352	288
Insulin	41,000	124	330	288
Hemoglobin (horse)	65,000	119	552	576
Edestin	310,000	119	2,610	2,304

* In proteins containing other than amino acid residues, the values are those obtained upon deducting the weight of the prosthetic group.

According to Bergmann, each protein is made up of $2^n \times 3^m$ amino acid units, where n and m are integers. For a protein of 288 units, the value would represent $2^5 \times 3^2$ units. The next higher group would contain 576 units, for which n would equal 6. These classes of protein size are similar to those obtained by Svedberg.

The recurrence of amino acids in the chain at regular intervals may be illustrated with silk fibroin. Glycine, alanine, and tyrosine have frequencies of 2, 4, and 16, respectively. Thus a segment of the silk fibroin would contain the amino acids spaced as shown below, where G, A, T, and X represent, respectively, glycine, alanine, tyrosine, and any other amino acid:

-G-A-G-X-G-A-G-X-G-A-G-T-G-A-G-X-G-A-G-X-G-A-G-X-G-A-G-T-G-

Bergmann admits that the regular spacing of amino acid units at definite intervals is only a theory but notes that it has some support in the case of gelatin from which polypeptides containing prolyl-glycine have been isolated in considerable amounts and where the space-relationship theory demands —P—G— juxtapositions. In any event, Bergmann's theory is challenging and will unquestionably stimulate thought and research.

The Possibilities of Protein Isomerism. In polypeptides and proteins isomerism may arise from the kinds of amino acids as well as from the numbers of each kind and their positions. This is a problem of

permutations and the number of possible isomers is given by $n!$ (factorial n), where n is the number of different amino acids each of which occurs only once. When the same amino acid occurs several times, the formula used below applies.

If it is assumed that a protein is made up of 20 amino acids, and *that each amino acid occurs only once in the protein chain,* there would still remain the possibility of having 20! or approximately 24×10^{17} different compounds, each containing the same amino acids in identical proportions and differing only in space relationships.

According to modern viewpoint, proteins have a very much greater molecular weight than would be found for such a "polypeptide." Egg albumin is made up of several hundred molecules of individual amino acids. Regardless of how these are linked together, it must be obvious from the above that there are limitless theoretical possibilities to the number of compounds which could be synthesized from the known list of amino acids.

Leathes [34] calls attention to these possibilities in a very striking way. He states, "Supposing it were a chain of only 50 links, a very simple case; if all the links were different the number of possible permutations is denoted by the innocent-looking symbol 50!. If, instead of all being different, one kind of link recurred ten times, the number would be reduced to 50!/10!. If, in addition, there were 4 that recurred four times and 10 that recurred twice, it would be further reduced to

$$\frac{50!}{10! \times (4!)^4 \times (2!)^{10}}$$

It would now consist of a chain of only 50 links, of which there were only 19 different kinds, and the number of different arrangements of its parts would be about 10^{48}. Astronomy deals with big figures. Light, it is said, takes 300,000 years to travel from one end of the Milky Way to the other; this distance expressed in Ångström units, 10,000,000 of which go to a millimeter, would be less than 10^{32}. So far are we from knowing the structure of protein molecules. So far are we from knowing what variations in disposition of the parts in such a molecule may not occur without our being within a measurable distance of detecting them. For if the number of possible varieties of a protein whose molecular weight is known, and known to be exceptionally small, and which contains the several amino acids in a known proportion, is as great as this, the number that is possible when that proportion may be changed is practically incalculable, each change in proportion being capable of a number of new arrangements that could be calculated, as was done for our hypo-

[34] J. B. Leathes, *Science*, **64**, 387 (1926).

thetical case. . . . The peculiar thing about the chemistry of living matter is not that the reactions that are characteristic in it are novel, but that in the rough and tumble of ordinary liquid systems their occurrence is almost infinitely improbable. Where there is life circumstances exist which make them the rule. . . . Forces which determine the relative positions of adjacent foreign molecules and so affect their behavior are something to which there is no analogy in the growth of crystals in a saturated solution."

Leathes' calculations do not take into consideration the possibility of various linkages between the various amino acids, keto $\rightleftharpoons$ enol isomerism, cyclic structures, such as have been discussed in the preceding part of this chapter, etc., all of which would introduce further possibilities in calculating the number of possible isomers which could be formed from a given number of amino acids.

Miss Wrinch discusses the molecular structure of chromosomes. The genes, which are regarded as the genetic units, are located in the chromosomes. She considers a pattern of protamine combined with nucleic acid, and (assuming that there are only six types of constituents and that five-sevenths of the constituents belong to one type as is the case with the protamine, clupeine) she notes that, if the protein pattern extended for a distance of 40 μ, the total number of possible patterns would approximate the inconceivable figure of $10^{50,000}$. Beside a figure of this magnitude even astronomical distances fade into insignificance.

As we shall see later, the biological reactions of the proteins enable us to differentiate rather sharply, in most instances, between the proteins which are characteristic of one species and those which are foreign to that species, and it is upon these observations that the entire modern structure of immunology has been built. Some workers point out that the species specificity which is demanded by the modern theories of immunology, and which is shown in immunological reactions, is almost infinite and therefore surprising. The author believes with Leathes that, considering the infinite possibilities of the linkages of amino acids in proteins, it is much more surprising that nature should ever synthesize twice in succession proteins which are even remotely alike.

Perhaps no more striking illustration of the exactness with which vital reactions are regulated in the living protoplasm can be given than the fact that, through untold generations, a given organism reproduces the same structural configurations in the proteins which characterize that organism. The fact that each species has a different set of proteins is not the surprising thing. The surprising thing is that nature is able to control the synthesis of proteins within a single species, so that the same proteins are synthesized by all members of the species.

CHAPTER 16

Characteristic Properties of Protein Systems

As already indicated, the reactions of protein systems must be not only those which are characteristic of the organic groups making up the protein molecule but also those which would be expected of lyophilic colloid systems. The literature in this field is so voluminous that the properties of only a few proteins can be mentioned and these very briefly and inadequately. It is hoped, however, that it will be possible to consider in a casual way some of the more important phenomena which are involved.

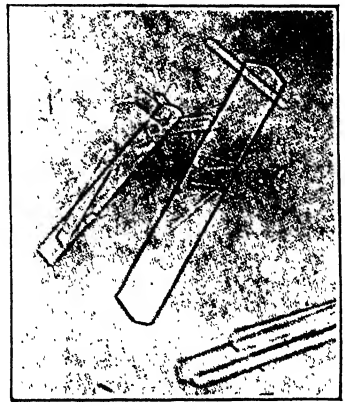


FIG. 84. Crystals of egg albumin. (After Sörensen.)

The Albumins. Egg albumin and serum albumin are the two proteins of this class which have been most extensively investigated. Both can be obtained in crystalline form. Accordingly, they are usually regarded as chemical entities. Although the fact that the albumins could be obtained in crystalline form has been known for many years, the first precise study of the crystallization process was made by Sörensen [1] in 1917. His technic for egg albumin involves fractionation with ammonium sulfate and dilute sulfuric acid.[2] Sörensen found that three crystallizations were sufficient to free the egg albumin from the other proteins of egg white. The initial crystallization was best carried out at a *p*H of 4.6, the second at 4.65, and the third at 4.70. Figure 84 shows the form of the crystals. These contained water of crystallization equivalent to 0.22 gram of water per gram of anhydrous albumin. Only at a *p*H of 4.886 did the crystals contain no surplus of either ammonia or sulfuric acid; Fig. 85 shows the excess of either at varying hydrogen-ion con-

[1] S. P. L. Sörensen, *Compt. rend. trav. lab. Carlsberg*, **12** (1917); [2] **15** (1923).

[2] C. A. Morrow and W. M. Sandstrom, *Biochemical Laboratory Methods*, 2nd ed., p. 90, John Wiley & Sons, New York, 1935 (out of print).

centrations. The crystals do not contain ammonium sulfate as an integral component. The best-defined crystals are prepared at a *p*H of 4.58 where the hydrated albumin contains one equivalent of sulfuric acid per 125 equivalents of protein nitrogen. Cole [3] has modified the procedure somewhat, but he employs ammonium sulfate, whereas Kekwick and Cannan [4] take advantage of the great solubility of sodium sulfate as the precipitating agent.

In his study of the egg albumin-ammonium sulfate-water systems, Sörensen applied the phase rule of Gibbs to the crystallization process.

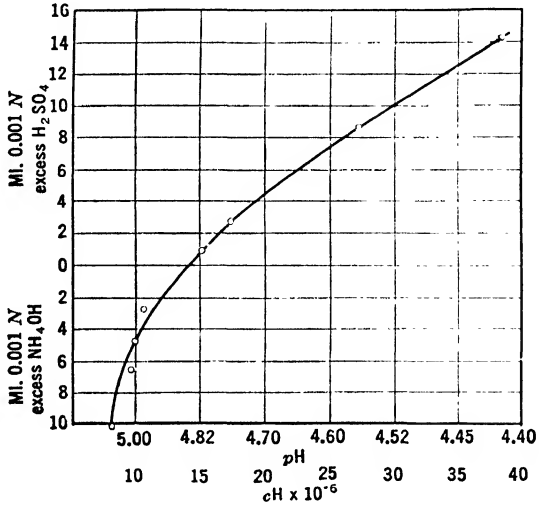


FIG. 85. Showing the excess acid (H_2SO_4) or alkali (NH_4OH) per milligram equivalents of protein nitrogen present in crystalline egg albumin in equilibrium with solutions having different hydrogen-ion concentrations. (Data of Sörensen.)

He points out that other workers have attacked this problem but that in each instance they were working with egg albumin purified by precipitation and not by repeated crystallization. Sörensen finds the Gibbs phase rule to hold rigidly in an egg albumin-salt-water system, provided that the hydrogen ion is considered as one of the components; previous workers failed to secure agreement with the phase rule because they regarded the system as a 3-component system, whereas in reality it is a 4-component system in which equilibrium is reached only at a constant temperature, a constant concentration of ammonium sulfate, and a constant hydrogen-ion concentration. When these three factors are held constant, the content of egg albumin in the mother liquor which is in

[3] A. G. Cole, *Proc. Soc. Exptl. Biol. Med.*, **30**, 1162 (1933).
[4] R. A. Kekwick and R. K. Cannan, *Biochem. J.*, **30**, 232 (1936).

equilibrium with the crystals will always be the same within experimental error. The state of equilibrium is therefore independent of the initial concentration of protein in the solution, so that from a concentrated albumin solution abundant crystals will form, whereas from a dilute solution only a small amount of crystals may be expected.

Inasmuch as he was working with a material having a very low diffusion coefficient, equilibrium was very slowly reached, more than 13 days being required in some experiments. Earlier workers had not waited sufficiently long for equilibrium to be attained.

Sörensen investigated certain other factors influencing crystallization and found, as might be expected, that the crystallization begins sooner and proceeds more rapidly as the concentration of ammonium sulfate is increased. When minimal amounts of ammonium sulfate were present, equilibrium was not reached even after 13 days. The final equilibrium in all cases, however, was the same, irrespective of the concentration of ammonium sulfate, *i.e.*, the crystals in each instance had the same chemical composition.

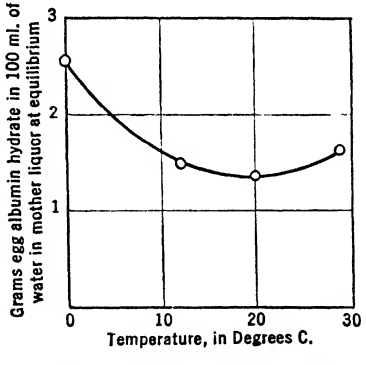


FIG. 86. The effect of temperature on the crystallization equilibrium of egg albumin solutions. (Data of Sörensen.)

He found that the amount of solid phase was lower at 0°C. than at 12° or 29°. At approximately 20°, there is an optimum temperature at which the smallest amount of hydrated egg albumin is present in solution in equilibrium with the crystals. Sörensen's data are presented in Fig. 86; it will be seen that the solubility at 0° is nearly 100 per cent greater than at 20°.

The optimum *p*H appears to be approximately at 4.58. A decrease in hydrogen-ion concentration from the optimum only causes much less albumin to crystallize. The crystals which do form are normal in shape and character. An increase in hydrogen-ion concentration causes only a slight increase in the solublity of the egg albumin hydrate crystals but does cause a decided alteration in both form and character of the crystals. The crystals contain much more sulfuric acid than normal. At a *p*H 5.22 to 5.30 the crystal form changes from small needle-shaped crystals to prisms. His crystallization curves in relation to *p*H are shown in Fig. 87.

Sörensen's fifth paper is a study of the osmotic pressure of egg albumin measured by the direct method. He finds that an egg albumin solution

of given composition will invariably show one and the same osmotic pressure, and concludes that egg albumin has a molecular weight of about 34,000 and that it exists in solution in the form of simple molecules. This value is lower than that obtained later and noted in the discussion of globular proteins (p. 371). In his paper on reversible dissociable component systems, he concludes that this protein also undergoes a dissociation but that the tendency in this system is very slight.

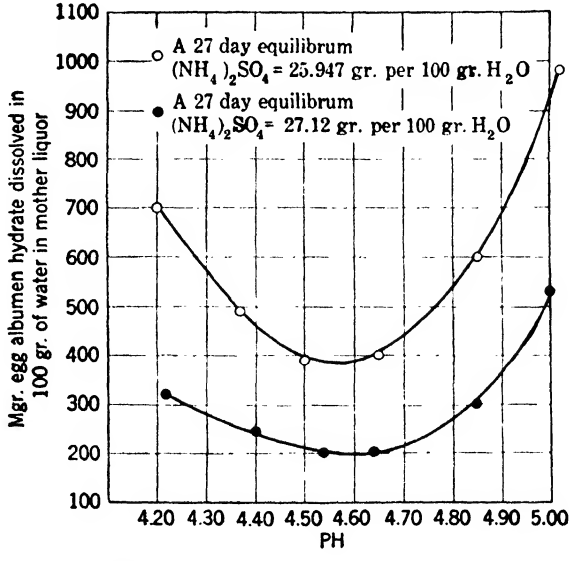


FIG. 87. Showing the effect of hydrogen-ion concentration on the "solubility" of crystalline egg albumin in ammonium sulfate solutions. (Data of Sörensen.)

Longsworth, Cannan, and MacInnes [5] were able by the Tiselius electrophoretic apparatus to separate two albumins and two conalbumins from egg white; however, it is not clear whether these four albumin fractions include more material than that obtained by Sörensen's crystallization process.

Egg albumin dialyzes slowly through a fairly porous collodion membrane. Lactalbumin dialyzes through a collodion membrane much more readily than egg albumin does. Svedberg in his studies of lactalbumin concludes that it is a heterogeneous protein made up of particles of relatively small molecular weight. No study similar to Sörensen's study of egg albumin has been carried out with lactalbumin. The fact that an egg albumin solution is monomolecular does not necessitate the abandon-

[5] L. G. Longsworth, R. K. Cannan, and D. A. MacInnes, *J. Am. Chem. Soc.*, **62**, 2580 (1940).

ing of the viewpoint that egg albumin solutions show colloidal properties. We have already pointed out that the colloidal realm is defined on the basis of the size of the micelle, irrespective of whether the micelle represents a single molecule or an aggregate of molecules. If the molecule of egg albumin hydrate is sufficiently large to form a particle having a diameter greater than 1 mμ, that particle must possess surface forces which are characteristic of the lyophilic colloids. Ultrafiltration experiments indicate very clearly that the egg albumin molecule lies well within the boundaries of the colloid realm. Consequently, although Sörensen has shown that egg albumin solutions are molecularly dispersed solutions and that they obey the physicochemical laws, such as the phase rule, nevertheless we cannot abandon the colloid viewpoint in dealing with even egg albumin systems.

We have already noted that human serum albumin may be prepared in the fractionation of whole plasma by dialysis against alcohol. Cohn and co-workers [6] have made similar fractionations using ammonium sulfate and McMeekin [7] has prepared in the crystalline state serum albumin containing 5.5 per cent carbohydrate.

As already noted, egg albumin contains a carbohydrate radical. In 1927, Fränkel and Jellinek [8] isolated from egg albumin a disaccharide composed of glucosamine and mannose, in which the union joining the two sugar radicals is through the amino group, inasmuch as the glucosamine-mannose compound does not react to yield nitrogen in the Van Slyke apparatus. Incidentally, they point out that this is the first instance in which mannose has been found in a material of animal origin. Later Rimington,[9] in investigating the carbohydrate complex of the blood serum proteins, isolated a trisaccharide, glucosamine-dimannose. This trisaccharide was non-reducing prior to acid hydrolysis. Serum globulin contained 3.7 per cent of the trisaccharide. Rimington's compound was isolated by the alkaline hydrolysis of the protein. The surprising fact about his trisaccharide is that it was optically inactive. Levene and Mori [10] investigated the carbohydrate in egg albumin and ovomucoid and found it to be a trisaccharide containing one glucosamine residue and two mannose residues. Ovomucoid contained 5.1 per cent, ovalbumin 0.26 per cent. They suggest that the carbohydrate in the three-times-recrystallized ovalbumin may represent contamina-

[6] E. J. Cohn, T. L. McMeekin, J. L. Oncley, J. M. Newell, and W. L. Hughes, *J. Am. Chem. Soc.*, **62**, 3386 (1940).

[7] T. L. McMeekin, *J. Am. Chem. Soc.*, **62**, 3393 (1940).

[8] S. Fränkel and C. Jellinek, *Biochem. Z.*, **185**, 392 (1927).

[9] C. Rimington, *Nature*, **126**, 882 (1930); *Biochem. J.*, **25**, 1062 (1931).

[10] P. A. Levene and T. Mori, *J. Biol. Chem.*, **84**, 49 (1929).

tion with ovomucoid. In a second paper, Levene and Rothen [11] from measurements of the coefficient of diffusion conclude that the carbohydrate molecule is composed of four trisaccharide units, each having a molecular weight of about 500. In his last paper Levene [12] showed that a disaccharide which was a mannose glucosaminide could be prepared from the trisaccharide.

The Globulins. Proteins which have been designated as globulins have been prepared from a great variety of vegetable sources, as well as from muscle tissue, blood serum, and other biological fluids. As has already been indicated, these proteins have been characterized as soluble in dilute salt solutions and insoluble in pure water. The general method of preparation is to extract the tissue with approximately 10 per cent sodium chloride solution and to dialyze the extract. The proteins which are precipitated will be regarded as globulins, whereas the albumins remain in solution. A modification of this method is to add to the extract an equal volume of saturated ammonium sulfate solution, the globulins precipitating in half-saturated ammonium sulfate, the albumins remaining in solution. The precipitated protein is then redissolved and reprecipitated by half saturation with ammonium sulfate, dialyzed free of ammonium sulfate, dried, and analyzed.

A number of the crude globulins so prepared have been fractionated by adding different amounts of ammonium sulfate to the solution and separating the various fractions which are precipitated at the various concentrations of ammonium sulfate. In the isolation of the globulins by "dissolving" them in neutral salt solution, we are dealing with nothing more nor less than a peptization process, such as is indicated in Fig. 88, where an increase in salt concentration throughout a certain range alters the surface forces in such a manner as to favor dispersion, following which a further increase in salt concentration causes the dehydration or the salting-out of the protein micelles. In this respect, the peptization of protein, as diagrammed in Fig. 88, would differ from the peptization of the silver halide, as diagrammed in Fig. 4, only in the amount of salt required to bring about maximal dispersion.

Subsequent to his classic study of egg albumin Sörensen [13] took up the investigation of serum globulin and concluded that the fractions which have been designated as euglobulin and pseudoglobulin are not capable of being separated as homogeneous entities but that they represent simply arbitrary fractions, the amount and character of the fraction depending on the technic employed for their isolation. We have already

[11] P. A. Levene and A. Rothen, *J. Biol. Chem.*, **84**, 63 (1929).
[12] P. A. Levene, *J. Biol. Chem.*, **140**, 279 (1941).
[13] S. P. L. Sörensen, *Compt. rend. trav. lab. Carlsberg*, [11] **15** (1925).

seen that Lustig, using non-standard technic, separated four euglobulins and four pseudoglobulins and that Block expressed doubts on the homogeneity of any of the protein fractions which are supposed to characterize the proteins of blood serum. It will be recalled that serum globulins may be fractionated by means of alcohol or ammonium sulfate to yield α-, β-, and γ-globulins. These fractions can also be prepared in a Tiselius electrophoretic cell.

Other animal globulins include the fibrinogen of blood and the myosin of muscle. The soft tissues of the body contain varying quantities of

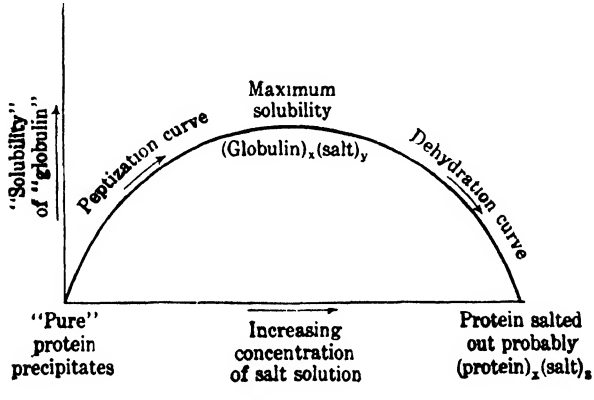


FIG. 88. A diagrammatic representation of the interrelationships of certain of the factors concerned in protein "solubility."

albumins and globulins. The higher values are found in those animals which have been fed high-protein diets; this fact is used to explain the source for the rapid regeneration of blood proteins when these animals are bled as in the Whipple [14] plasmapheresis experiments (see p. 453).

The major fraction of whey proteins is now recognized to be β-lactoglobulin, so classified because it is insoluble in water but readily soluble in very dilute salt solutions.[15] It is the major fraction of the commercial "lactalbumin." Briggs [16] has found this protein suitable for studies on electroviscous effects because it is a "strong, internally well knit micelle, incapable of changing either in size, or shape, under the relatively weak osmotic forces of its gegenions."

The vegetable globulins were extensively investigated by Osborne [17] and his students, and his book can be conveniently consulted for specific

[14] S. C. Madden and G. H. Whipple, *Physiol. Rev.*, **20**, 194 (1940).

[15] A. H. Palmer, *J. Biol. Chem.*, **104**, 359 (1934).

[16] D. R. Briggs and M. Hanig, *J. Phys. Chem.*, **48**, 1 (1944).

[17] T. B. Osborne, *The Vegetable Proteins*, 2nd ed., Longmans, Green and Company New York, 1924.

references. Since his death one of his former associates, D. Breese Jones, has been active in publishing in this field.

There are a few vegetable proteins which appear to be homogeneous and which fall in the globulin group. Among these are excelsin from the Brazil nut and pomelin of the orange seed. Referring to Fig. 78, it will be seen that solutions of the various alkali halide salts peptize the Brazil nut protein to almost the same degree, and the same is true for pomelin [18] from the seed of the citrus fruits. Both excelsin and pomelin are crystalline proteins and may be regarded provisionally as true vegetable globulins. Svedberg finds their molecular weights to be in the order of 212,000. Edestin, with a molecular weight of 312,000, has been widely used for protein studies; because the sale of Indian hemp seed has been forbidden or restricted, Vickery [19] suggests the preparation of a very similar protein from the seeds of the squash and pumpkin.

Aside from their characteristic peptization behavior toward neutral salt solutions, the chemical properties of the globulins are not greatly different from the chemical properties of many other proteins. Certain of the globulins can be obtained in crystalline form, but here again no exact study of the crystallization process has been made. Certain of the globulins, as already noted, pass into insoluble modifications, the proteans, on standing in contact with pure water or in the presence of hydrogen ions. Little is known about the reason for this transformation, which cannot be regarded as a characteristic property of all globulins, inasmuch as it is not shown by certain compounds classified as globulins.

The Prolamines. The prolamines or proteins peptized by dilute alcohol are the characteristic proteins of the cereal grains. The name prolamine was given to this group of proteins by Osborne, because, in general, they yield on hydrolysis very large amounts of ammonia (acid amide nitrogen) and relatively large amounts of proline. Gliadin of wheat and rye, zein of maize, and hordein of barley have been most extensively studied. All the cereal grains with the exception of rice contain prolamines. A small amount of an alcohol-soluble protein can be isolated from rice, but it does not show the characteristics of the true prolamines.

The literature of the prolamines, their chemical composition, and certain of their physicochemical properties have been reviewed in a comparative study.[20] There is a great similarity in certain of the physico-

[18] L. K. Rotha and F. Saunders, *J. Am. Chem. Soc.*, **54**, 342 (1932).

[19] H. B. Vickery, E. L. Smith, R. B. Hubbell, and L. S. Nolan, *J. Biol. Chem.*, **140**, 613 (1941).

[20] W. F. Hoffman and R. A. Gortner, Colloid Symposium Monograph, Vol. II, p. 209, Chemical Catalog Co., New York, 1925.

chemical properties of the proteins of this group, whereas in other properties there is a gradation from gliadin which stands at one end of the series to the prolamines of sorghum and kafir which stand at the other end of the series. Gliadin is peptized by cold 60 to 70 per cent alcohol, whereas the prolamines of kafir and sorghum are not peptized by the cold alcohol but only by hot alcohol. The prolamines are characterized by an almost complete absence of buffering action toward solutions of acid and alkali.

Hoffman and Gortner divide the prolamines into two groups, a wheat group containing the prolamines from *Triticum vulgare*, *T. spelta*, *T. durum*, *T. dicoccum*, *T. monococcum*, and *Secale cereale*, and a corn group containing the prolamines from *Zea mays*, *Euchlaena mexicana*, *Andropogon sorghum*, and *Sorghum vulgare*. The proteins within these two groups have many similar properties, and Lewis and Wells [21] have shown that within the corn group and within the wheat group the prolamines show immunological relationships but that the two groups are distinct when tested by the biological method.

Haugaard and Johnson [22] carried out an extensive study on the fractionation of gliadin and conclude that it is a reversible dissociable component system. We have already discussed the peptization of the proteins of wheat flour and have noted that the fractions which we secure by laboratory technics are probably not preexistent in the wheat berry, or, if they are present, they are intimately associated with a larger protein complex, the physicochemical properties of which are not necessarily the additive properties of the fractions which we can isolate.

Mention has been made of the alcohol-soluble protein isolated from milk by Osborne and Wakeman. Although this protein is apparently a distinct protein and is peptized by alcohol, nevertheless it does not contain the large amount of ammonia nitrogen characteristic of most of the prolamines. It is doubtful, therefore, whether it should be classed as a true prolamine, inasmuch as its peptizability by alcohol appears to be the only property which relates it to this class of proteins.

The Glutelins. These proteins are characterized by not being peptized by neutral solvents but by being peptized by dilute acids and alkalies.

Larmour [23] has collected the literature in regard to these proteins and has added certain new glutelins to the known list. They are found in greatest amount in the cereal grains. There is a progressive change in the chemical analyses of the glutelins in regard to the ammonia nitrogen,

[21] J. H. Lewis and H. G. Wells, *J. Biol. Chem.*, **66**, 37 (1925).
[22] G. Haugaard and A. H. Johnson, *Compt. rend. trav. lab. Carlsberg*, [2] **18** (1930).
[23] R. K. Larmour, *J. Agr. Research*, **35**, 1091 (1927).

arginine nitrogen, and total basic nitrogen content, ranging from glutenin from wheat at one end of the series to oryzenin from rice at the other end of the series, as shown in Table 45.

TABLE 45. VALUES OF ARGININE NITROGEN OF THE VARIOUS GLUTELINS ARRANGED IN ORDER OF MAGNITUDE WITH CORRESPONDING VALUES OF TOTAL BASIC NITROGEN AND AMMONIA NITROGEN

Glutelin	Arginine N, per cent	Total Basic N, per cent	Ammonia N, per cent
Glutenin (Blish)	9.27	18.94	16.5
Zeanin	9.45	18.84	11.32
Glutenin (Larmour)	10.90	18.80	14.78
Duro-glutenin	11.00	21.20	13.25
Hordenin	11.08	20.25	11.38
Monococco-glutenin	11.86	20.16	10.78
Dicocco-glutenin	13.03	24.60	11.06
Spelta-glutenin	13.43	24.44	8.06
Secalinin (average)	13.77	25.69	9.24
Teozeanin	14.15	26.84	10.16
Avenin	15.48	24.37	11.87
Oryzenin	17.95	27.59	8.07

The glutelins of rice varieties and subvarieties show significant differences similar to those found by Larmour for the wheat varieties. All the more recent evidence points to the group of glutelins as non-homogeneous protein systems capable of fractionation, the characteristics of the fraction depending on the technic employed.

The Albuminoids. This is a heterogeneous class, generally resistant to chemical reagents and to peptic and tryptic enzymes, and includes such proteins as silk, wool, hair, skin, nails, horn, quills, connective tissues, and bone matrix. Certain of these proteins appear to be made up of relatively few amino acids. Albuminoids do not occur in the vegetable kingdom.

Raw silk can be peptized into two fractions, silk gelatin and silk fibroin, both being classed with the albuminoids. Silk gelatin is peptized by water under slight pressure, or by dilute alkaline solutions, and amounts to 15 to 20 per cent of the weight of the raw silk. Silk fibroin is characterized by its high content of three amino acids, glycine, alanine, and tyrosine, these three accounting for 60 to 70 per cent of the weight of the silk fibroin. Silk, likewise, contains a relatively high percentage

of serine (1 to 3 per cent has been actually isolated). Silk gelatin, in contrast to silk fibroin, is extremely low in its glycine content, somewhat higher in its serine content (5.4 to 6.6 per cent has been isolated), but the analyses of silk gelatin account for only 20 to 40 per cent of the amino acids which are present, so that, in general, little can be said in regard to its amino acid content.

Certain of the albuminoids are of special interest. Thus, *spongin*, the skeletal protein of sponges and coral, contains no tyrosine but instead 3,5-diodotyrosine or gorgoic acid. The iodine content of spongin ranges from 1 to 1.5 per cent. It is interesting to note that the ancient Greeks considered ground sponges to be specific for goiter. This view, along with many viewpoints of the ancients, was held up to ridicule until modern investigations have shown that certain forms of goiter are due to lack of iodine. The iodine present in 3,5-diodotyrosine is effective in the prevention of goiter, and dried ground sponges have again been listed as an official remedy in the French Pharmacopœia.

The protein of the coral, *Primnoa lepadifera*, contains 3,5-dibromotyrosine which is one of the few organic compounds of biological origin containing bromine.

Many of the keratins, notably human hair and wool, contain a very high percentage of cystine, human hair containing as high as 14 per cent of this amino acid. Accordingly, human hair is generally used as the protein for cystine isolation, although wool, which contains considerably less cystine, can be used if necessary. Marston[24] notes that cystine may be the limiting amino acid in wool production. When he supplemented a pasture diet with blood meal which contained 2.7 per cent of cystine, there was a 30 per cent increase in the weight of the wool which was clipped.

Block and Vickery,[25] from a study of several keratins, propose a new definition, *i.e.*: "A keratin is a protein which is resistant to digestion by pepsin and trypsin, which is insoluble in dilute acids and alkalies, in water and in organic solvents, and which, on acid hydrolysis, yields such quantities of histidine, lysine, and arginine that the molecular ratios of these amino acids are respectively approximately 1:4:12." This definition was based on the study of seven keratins in which the ratios ranged from 1:3:10 to 1:5:16. In the second paper human finger nails and cattle horn keratins gave ratios of 1:5:15 and 1:5:13, respectively. These are the original studies which later led Block to the conclusion that there was a basic amino acid "anlage" in proteins.

[24] H. R. Marston, *Australian J. Exptl. Biol. Med. Sci.*, **9**, 235 (1932); *cf.* also *J. Agr. Science*, **25**, 103, 113 (1935).

[25] R. J. Block and H. B. Vickery, *J. Biol. Chem.*, **93**, 113 (1931); R. J. Block, *J. Biol. Chem.*, **104**, 339 (1934).

Wilkerson [26] studied the isoelectric point of the keratin from human hair, skin, and nails. He found essentially identical isoelectric points, *p*H = 3.67, 3.70, 3.78, for these three keratins, and the electrokinetic behavior of the three proteins gave curves which were essentially superimposable.

Routh [27] has shown that prolonged grinding renders wool and other keratins partially soluble and that certain fractions are attacked by enzymes. The process does lead to a partial oxidation of the sulfur to the sulfate ion.

The proteins of connective tissue are largely albuminoids with some glycoproteins. The three main types of albuminoids are elastin, collagen, and reticulin. Elastin is characterized by the fact that it is slightly hydrolyzed by pepsin and also by trypsin, it swells slightly, and it constitutes the main portion of the yellow connective fibers. It consists chiefly of glycine, alanine, and leucine. Collagen, on the other hand, is found in the white fibers, is slowly hydrolyzed by pepsin but not by trypsin below 40°C., but when heated in the presence of water is readily converted to gelatin. Reticulin is present in the glandular tissues and resembles collagen.

Gelatin, obtained by the pressure cooking of hides, scraps, bones, and the like, is interesting because it is one protein which is not antigenic, *i.e.*, it can be injected into the blood stream without setting up any immune reactions. For this reason it was used in World War I, with some success, to replace the plasma proteins lost through bleeding. It contains comparatively large quantities of glycine, proline, and hydroxyproline, very little tyrosine, and no cystine or tryptophan.

Since gelatin is derived from collagen, it should probably be classed as a derived protein but is usually included in the group of albuminoids. Probably gelatin has been made the basis of more physicochemical studies than any other protein because of its ready availability and the fact that its solutions undergo sol $\rightleftharpoons$ gel transformations with changes in temperature. In many respects it is a unique protein, and its sols and gels show remarkable properties. For example, when in the sol state it shows a specific optical rotation of approximately 160°, with practically no change in optical rotation when in the sol state (40°C.) over the range *p*H 1–12. On gelling, the specific optical rotation increases over 200 per cent, with marked maxima between *p*H 1–2, *p*H 4–5, *p*H 7–9, and a marked minimum between *p*H 2–3.[28] Most gels are not sufficiently

[26] V. A. Wilkerson, *J. Biol. Chem.*, **107**, 377 (1934); **112**, 329 (1935).

[27] J. I. Routh and H. B. Lewis, *J. Biol. Chem.*, **124**, 725 (1938); **135**, 175 (1940).

[28] E. O. Kraemer, Colloid Symposium Monograph, Vol. IV, p. 102, Chemical Catalog Co., New York, 1926.

optically clear to permit of polarimetric studies. Accordingly there has been doubt as to whether the enormous increase in optical rotation exhibited by the gelling of gelatin is a characteristic of all gels or is specific for gelatin gels. Gortner and MacDonald [29] have shown that zein gels (in glycol monomethyl ether) have the same specific optical rotation as the sols from which the gels were derived. Accordingly the change of optical rotation of gelatin in passing from the sol to gel state would appear to be a unique property of a gelatin system and probably is associated with binding of water by the protein micelles.

Gelatin sols and gels show enormous changes in viscosity with changes in hydrogen-ion concentration, and these viscosity changes in a large degree parallel the changes in optical rotation.[30] Again the phenomena appear to be associated with the binding of water on the polar groups associated with the asymmetric carbon atom.

Gelatin systems show marked hysteresis, and accordingly the physicochemical properties of the system vary with time and are in a very considerable measure dependent on the source of the gelatin used and the methods employed in the preparation of the sols and gels. Ultracentrifugal studies have shown gelatin to be non-homogeneous in particle size, and, because of this non-homogeneity and the great variation which exists in the commercial preparations, it is not surprising that the literature contains many conflicting observations on the physicochemical properties of gelatin systems.

The Histones. These are basic proteins, characterized by a relatively high proportion of diamino acids and by the fact that at least some of them form salts with the strong mineral acids. Their nitrogen content ranges in the neighborhood of 17 to 20 per cent. They combine readily with nuclein or nucleic acid, probably as coacervate systems, and are the characteristic proteins of glandular tissues. It has been suggested that the histones represent the union of a protamine with some other protein. The histones which have been most extensively investigated are the globins from the erythrocytes of blood, particularly from the blood of birds, where the histone occurs combined with the non-protein radical, hemin, and the histone of the thymus gland, where it occurs combined with nucleic acid. Histones also occur in the unripe sperm of fish and sometimes in the ripe sperm of some vertebrates and some invertebrates. They are, in general, characterized by a high arginine content, except vertebrate globin which is comparatively rich in histidine.

The Protamines. The protamines are the characteristic proteins of ripe sperm cells and are apparently the proteins associated with nucleic

[29] R. A. Gortner and R. T. MacDonald, *Cereal Chem.*, **21**, 335 (1944).

[30] R. H. Bogue and M. T. O'Connell, *J. Am. Chem. Soc.*, **47**, 1694 (1925).

acid in *chromatin*. It has been considered that protamines occur only in animal cells, but they have been found in pine and palm pollen and in *Lycopodium* spores. Here again we are probably dealing with coacervate systems, inasmuch as these proteins are so basic that in the free state they will blue litmus paper and have an isoelectric point well on the alkaline side of neutrality. They are therefore positively charged in the *p*H range where most proteins carry a negative charge; this fact accounts for the great ease with which they form coacervates with proteins, nucleic acid, arabic acid, and any other negatively charged lyophilic colloid. The nitrogen content is high, ranging from 25 to 30 per cent of the weight of the protein. They are sulfur-free. They are easily soluble in water, insoluble in alcohol and ether, and do not coagulate, and do not diffuse through parchment membrane. With mineral acids they form stable compounds, many of which are crystalline, and the free protamine will absorb carbon dioxide from the air, forming a more or less stable protamine carbonate. Double salts of constant composition are formed with platinic chloride and auric chloride. They are not digested by pepsin but are hydrolyzed by trypsin and ereptic enzmyes.

Kossel [31] and his associates have made the most extensive investigations of the protamines. He classifies the protamines into the *monoprotamines*, the *diprotamines*, and the *triprotamines*. The three classes differ in that they contain one, two, or three of the acids arginine, histidine, and lysine; arginine is always present.

The two protamines which have been most completely characterized are salmine and clupeine, which have been reinvestigated.[32]

Salmine is the protamine from ripe salmon sperm. All evidence indicates that it is a polypeptide containing 21 amino acid residues and 20 peptide linkages. It contains 14 molecules of arginine, 3 molecules of proline, 3 molecules of serine, and 1 molecule of valine. It contains no free amino group, proline being at the beginning of the polypeptide chain, so that what would have been normally a free amino group is the imino group of proline. At the carboxyl end of the polypeptide chain there are apparently 2 molecules of arginine. It will be noted that there are 2 arginine residues for each monoamino acid residue, and Waldschmidt-Leitz suggests that it is composed of a series of tripeptides linked together possibly in the following arrangement: M-D-D-M-D-D-M-D-D-M-D-D-M-D-D-M-D-D-M-D-D-M-D-D, where M represents a monoamino acid residue, and D an arginine residue. Such a structure

[31] A. Kossel, *The Protamines and Histones*, translated by W. V. Thorpe, Longmans, Green and Company, New York, 1928.

[32] E. Waldschmidt-Leitz, F. Ziegler, A. Schäffner, and L. Weil, *Z. physiol. Chem.* **197**, 219 (1931).

would have a molecular weight of 2,855, which agrees well with sedimentation velocity measurements carried out in Svedberg's laboratory and with the other physicochemical properties of salmine solutions.

Waldschmidt-Leitz isolated from pancreas extract a new specific enzyme, protaminase, which acts exclusively on protamines to split off arginine when the carboxyl group of arginine is free. When this protaminase acts upon salmine, the 2 terminal arginine molecules are split off leaving a residue, *salman*, containing 7 monoamino acids and 12 arginine residues. It is on the basis of this observation that Waldschmidt-Leitz concludes that not more than 2 arginine residues are conjugated together in the structure.

Clupeine is the protamine of herring sperm. It contains 15 amino acid residues and 14 peptide linkages. The amino acid composition is 10 molecules of arginine, 2 molecules of serine, and 1 each of proline, alanine, and valine, with the molecular weight of 2,021. Again the general arrangement of the formula noted for salmine appears to apply to clupeine, with proline at one end of the chain, thus providing an imino rather than an amino group and with 2 residues of arginine at the carboxyl end of the chain. Protaminase splits off these 2 molecules of arginine leaving a tridecapeptide, clupean, corresponding to the salman of salmine.

Rasmussen, working in Sörensen's laboratories, expresses doubt about the homogeneity of clupeine as ordinarily prepared. In a series of fractionations he secured preparations in which the arginine nitrogen ranged from 87.2 per cent to 91.7 per cent of the total nitrogen and concludes that the clupeine molecule cannot be built up exclusively of a series of tripeptides each containing 2 molecules of arginine to 1 molecule of a monoamino acid, since such a system would have 88.9 per cent of the total nitrogen as arginine nitrogen. In a later study [33] dealing with the electrometric titration of clupeine, the belief is expressed that it has a molecular weight lying between 4,000 and 4,100. This would necessitate doubling the molecule proposed by Waldschmidt-Leitz.

Scombrine, from mackerel sperm, is probably the simplest protamine, apparently containing only three amino acids, arginine 88.8 per cent of the total nitrogen, proline 3.8 per cent of the total nitrogen, the remainder being alanine, although there is a possibility that some other amino acid may be present in the alanine fraction. Kossel notes that the absence of serine distinguishes scombrine from salmine and clupeine.

Percine is apparently the characteristic protamine of the yellow perch, *Perca flavescens*, and the pike perch, *Stizostedium vitreum*. It contains

[33] K. E. Rasmussen and K. Linderstrøm-Lang, *Compt. rend. trav. lab. Carlsberg* [10] **20** (1935).

78.1 per cent of its total nitrogen as arginine N, 5.6 per cent as histidine N, and the balance as monoamino N. The monoamino acids have not been identified.

Sturine, from sturgeon sperm, contains all three of the diamino acids in the proportions of arginine 58.2 per cent, histidine 12.9 per cent, lysine 12.9 per cent, the three totaling 84 per cent of the protein molecule. The remaining amino acids are apparently exclusively leucine and alanine.

When protamines are injected intravenously [34] they cause a rapid fall in blood pressure and decreased respiration, death ensuing even at low dosages. The fatal dose of clupeine, injected intravenously, lies between 0.15 and 0.18 gram for a 10-kg. dog. The lethal dosage of sturine lies between 0.20 and 0.25 gram for a dog of similar weight. With a nonlethal dose, the initial reaction is over after a 30-minute period, but a second dose causes the same symptoms, indicating that no appreciable degree of immunity is conferred. The intravenous injection causes a marked delay in blood-coagulation time, which may be as much as 36 hours. The same effect is produced on blood coagulation *in vitro*. Leucocytes are rapidly destroyed by the intravenous injection of protamines. A sublethal injection of sturine reduced the leucocyte count from a normal count of 28,281 to 1,875 after a single injection and to 781 after a second injection. The diamino acids alone did not produce the above physiological symptoms, nor did a mixture of all the hydrolytic products of the protamine. Thompson accordingly concludes that the toxic action is due to the specific groupings in the molecule.

Hagedorn [35] and his associates have combined protamines and zinc with insulin to form a zinc-protamine-insulin for use in the treatment of diabetes. They note that this compound is relatively insoluble, it is absorbed slowly, and the insulin thus becomes available to the patient in small amounts over a relatively long period of time. When the free insulin is injected, there is a sharp reduction in blood sugar followed a few hours later by a similar rapid rise. With the injection of protamine insulinate the sharp reduction is avoided and the risk of hypoglycemia is thus greatly reduced. The reduction of blood sugar is slow and regular, and the regain of blood sugar likewise is slow and regular. Larger quantities of insulin can be injected, and thus the spacing between injections can be prolonged. The compound of protamine with insulin is

[34] W. H. Thompson, *Z. physiol. Chem.*, **29**, 1 (1900).

[35] H. C. Hagedorn, B. N. Jensen, N. B. Krarup, and I. Wodstrup, *J. Am. Med. Assoc.*, **106**, 177 (1936); *cf.* also H. F. Root, P. White, A. Marble, and E. H. Stotz, *ibid.*, **106**, 180 (1936).

probably a coacervate and contains about 1 part of protamine to 10 parts of insulin. One unit of insulin approximates 0.04 mg. Accordingly with an injection of even 25 units of insulin in the form of protamine insulinate there would be an insignificant amount of protamine injected; this probably accounts for the fact that no untoward symptoms have been observed which could be attributed to the protamine. Hagedorn found that the protamine insulinates, where clupeine, scombrine, or salmine was used, were too insoluble for best results. The clupeine insulin had its minimum solubility at *p*H 6.3, the scombrine and salmine compounds at *p*H 7.3. At these minimum solubilities, the solubility of the compound was of the order of that of barium sulfate. He accordingly used cyclopterine from *Salmo iridius*, which gave a compound with the desired solubility. Since the original observations of Hagedorn, protamine insulinate has been widely adopted in medical practice.

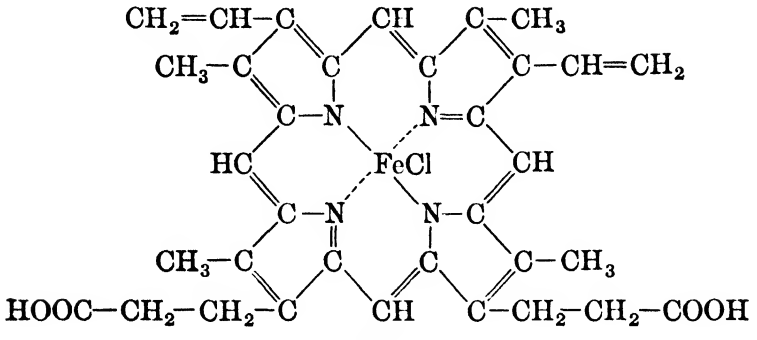
The Chromoproteins. *Hemoglobins.* The hemoglobins, found in the red blood cells of mammals, are the best characterized of the chromoproteins, and were among the first proteins to be obtained in crystal form. The prosthetic group, *heme*, constitutes about 6 per cent of the molecule in which it is tied to a histone, *globin*. The exact nature of the bond is not known but it is often pictured as between the iron of the heme group and the imidazole ring of histidine in the globin. The linkage may be broken by warming gently with acetic acid and sodium chloride.

The iron content of hemoglobins from different animals varies from 0.33 to 0.51 per cent according to the species, and on the basis of the iron content hemoglobin would have a minimum molecular weight of 16,000 to 17,000. Svedberg, using the ultracentrifuge, found a molecular weight of about 68,000, and this value has been confirmed by other technics. Apparently, therefore, four heme units are present in the hemoglobin molecule.

When the prosthetic group is removed in the presence of acid and chloride ion, *hemin* is formed; this is the chloride of heme and can readily be converted to its hydroxide (hematin) by alkalies. The iron of these derivatives has been converted from the ferrous to the ferric state; this is the type of change involved when hemoglobin is oxidized to methemoglobin by a reagent such as potassium ferricyanide. By contrast it should be noted that oxyhemoglobin still contains ferrous iron; the molecule is oxygenated, and the oxygen is in a loose linkage from which it is readily released at the lower oxygen tensions of the body tissues.

Hans Fischer, in a series of brilliant researches which led to the award of the Nobel prize in 1931, finally succeeded in synthesizing the prosthetic group of hemoglobin, so that its structure is definitely known

The synthetic product obtained, hemin, is identical in all respects with the natural product and has the formula [36]



Hemin

The compound is a ferric porphyrin, but when treated with a suitable reducing agent the corresponding ferroporphyrin (heme) results which combines readily with undenatured globin to form a compound apparently identical with natural hemoglobin.[37]

The porphyrins are substituted porphines, in this case the 1,3,5,8-tetramethyl-2,4-divinyl-6,7-di-β-propionic acid derivative. The porphines may be said to contain four pyrrole nuclei connected by methine (—CH=) bridges. It will be noted that the double bonds in the formula above are not in the positions usually found in pyrrole and that various authors present the graphic formula in slightly different forms. This is due to the fact the ring system possesses the property of resonance and the double bonds are not fixed in position. In the formula above, the iron is linked to alternate nuclei; some picture the iron as attached to two adjacent nuclei, but the writer knows of no definite proof for either structure.

Upon reduction with hydriodic acid in the presence of phosphorus triiodide and acetic acid the heme or hematin yields mesoporphyrin IX with a melting point of 212–213°. The changes produced by the reduction are the removal of the iron and the conversion of the two vinyl to ethyl radicals. The same compound results when such enzyme components as catalase, peroxidase, cytochromes *b* and *c*, and cytochrome oxidase are reduced. Although hemoglobin possesses practically no catalytic properties, its prosthetic group is rather similar to the coenzyme or prosthetic group of several oxidizing enzyme systems. By various chemical reagents heme can be converted into a series of por-

[36] H. Fischer and K. Zeile, *Ann.*, **468**, 98 (1929).

[37] M. L. Anson and A. E. Mirsky, *Physiol. Revs.*, **10**, 506 (1930).

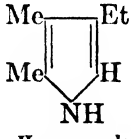
phyrins, some of which are shown in Table 46.[38] In the animal body heme gives rise to the bile pigments.

TABLE 46. THE MORE IMPORTANT PORPHYRINS DERIVED FROM HEMIN

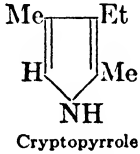
		Isomerides		Side Chains
		Possible	Synthesized (by H. Fischer)	
Hemin	$C_{34}H_{32}O_4N_4FeCl$	15	2	4Me 2X * 2CH:CH_2
Protoporphyrin	$C_{34}H_{34}O_4N_4$	15	2	4Me 2X 2CH:CH_2
Hematoporphyrin	$C_{34}H_{38}O_6N_4$	15	2	4Me 2X 2CH(OH)Me
Mesoporphyrin	$C_{34}H_{38}O_4N_4$	15	12	4Me 2X 2Et
Aetioporphyrin	$C_{32}H_{38}N_4$	4	4	4Me .. 4Et
Deuteroporphyrin	$C_{30}H_{30}O_4N_4$	15	3	4Me 2X ...
Deuterohemin	$C_{30}H_{28}O_4N_4FeCl$	15	2	4Me 2X ...

* X = —CH_2CH_2COOH.

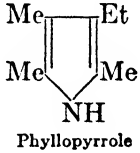
By drastic reduction with hydriodic acid, hemin gives rise to four hemopyrroles having the following structures:



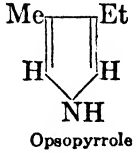
Hemopyrrole



Cryptopyrrole



Phyllopyrrole



Opsopyrrole

If one compares the formula of hemin with that of chlorophyll (p. 852), it will be noted that there is a striking similarity, indicating that the most important pigment of the plant kingdom and the most important pigment of the higher animals have much in common and are actually modifications of the same structure.

The globin of horse origin has an isoelectric point of *p*H 7.5. It is characterized by a high content of basic acids: 7.6 per cent histidine, 3.6 per cent arginine, and 8.1 per cent lysine. We have already indicated that the hemoglobins from various species differ in their crystalline forms as shown by Reichert and Brown [39]; by combining the heme from one

[38] From H. King, *Ann. Repts. Prog. Chem., Chem. Soc., London*, **29**, 209 (1932).

[39] E. T. Reichert and A. P. Brown, Carnegie Inst. Washington, Pub., No. 116 (1909).

source with the globin from a different species it has been shown that the crystal form depends on the species source of the globin [40] and that *all hemes are identical.* By replacing the globin with nitrogen compounds, like hydrocyanic acid, pyridine or nicotine, a series of *hemochromogens* have been prepared by Anson and Mirsky.[37]

The main function of hemoglobin is to carry oxygen to the tissues. In the alveoli of the lungs the oxygen tension is about 106 mm. of mercury; this is sufficient to convert over 90 per cent of the hemoglobin to oxyhemoglobin, the iron of which unites with 1.36 ml. of oxygen per gram of hemoglobin. The oxygen is held in a loose linkage and consequently is readily released at the lower oxygen pressures existent in the tissues. It is now generally recognized that the hemoglobin of venous blood also serves to carry back to the lungs some of the carbon dioxide formed by the tissues, apparently as a carbamino derivative of hemoglobin.

A different form of hemoglobin is found in tissues, particularly muscle, from which it gets its name. Muscle hemoglobin [41] has only one iron porphyrin unit per molecule in contrast to blood hemoglobin which has four; its molecular weight is in the range of 17,500. At the oxygen pressure of about 40 mm. which is found in the tissues, the blood carrier gives up part of its oxygen to the muscle hemoglobin, which in turn readily yields oxygen to the tissues at an oxygen tension of 5 mm., at which point the tissue oxidases are active. In other words, the muscle hemoglobin acts as an intermediate between the blood and the tissue to transfer and to "store" oxygen.

Warburg [42] early paid particular attention to the role of iron in the process of oxidation. He noted that heme-like compounds were present in all plant and animal cells, including yeasts and bacteria, but that these compounds were either free or combined with proteins other than globin. His early observations conform with our current knowledge of the respiratory systems of living cells. What was originally thought to be a single "respiratory enzyme" has turned out to be a complex system of enzymes, *i.e.*, the various cytochromes, cytochrome oxidase, catalase, and peroxidase, all of which have been shown to be iron-porphyrin compounds of the heme type combined with a protein other than globin. Cytochrome oxidase is apparently identical with the respiratory enzyme studied by Warburg.

The synthesis of heme and accordingly the synthesis of hemoglobin in the animal body is not as yet fully elucidated. Synthesis apparently

[40] A. Herzog, *Biochem. Z.*, **264**, 412 (1933).
[41] G. A. Millikan, *Physiol. Revs.*, **19**, 503 (1939).
[42] O. Warburg, *Science*, **68**, 437 (1928).

takes place primarily in the red bone marrow, in the liver, and in the spleen. On a diet deficient in iron, nutritional anemia develops. Rather interestingly, iron is not the only mineral element essential in the biological synthesis of hemoglobin, for both copper and manganese are in some way involved in the vital synthetic processes, and in the absence of copper and manganese nutritional anemia still appears. Still more recent observations indicate that cobalt is likewise an essential element concerned in the synthesis of hemoglobin. Neither copper, manganese, nor cobalt appear in the final product, and their role in the synthetic mechanism remains to be elucidated.

We have already noted that the prosthetic group in hemoglobin is closely allied to the porphyrin of chlorophyll. *Copper and manganese are likewise essential for the development of the chlorophyll molecule.*[43] So far as the author is aware, cobalt has not as yet been demonstrated to be essential for chlorophyll formation.

Chlorocruorin is a green oxygen-containing pigment which can be obtained in crystalline form from the blood of polychaete worms. The prosthetic group appears to be a porphyrin, but it is not heme. The red pigment occurring in the blood of a number of worms is known as *hemoerythrin* and appears to have a porphyrin residue but does not yield the mammalian hemin. *Erythrocruorins*, the red respiratory pigments of a number of the lower forms, such as *Daphnia* and the earthworm, together with the chlorocruorins, show much similarity to the hemoglobin of mammalian blood. The porphyrin contains iron, and the ratio of oxygen to iron in the oxygen-carrying capacity of the blood is the same as hemoglobin. Svedberg has measured the particle weight of a number of erythrocruorins; in some species the particle weight appears to be as low as 219,100; in others, as high as 2,850,000.

Metallo-Proteins. Ferritin is an iron-protein complex which serves to store functional iron, especially in the liver, spleen, and bone marrow. Granick [44] and others have prepared it from gland extracts from which it is readily precipitated by the addition of cadmium sulfate. The iron appears to exist in the form of colloidal iron oxide and phosphate to the extent of 23 per cent iron. The linkage appears to be a weak tie, for the apoferritin (the protein without the metallic prosthetic group) is readily removed and gives the same x-ray diffraction pattern as does the fer-

[43] A. L. Sommer, *Plant Physiol.*, **6**, 339 (1931); J. Oserkowsky and H. E. Thomas, *Science*, **78**, 315 (1933); O. S. Orth, G. C. Wickwire, and W. E. Burge, *Science*, **79**, 33 (1934).

[44] P. F. Hahn, S. Granick, W. F. Bale, and L. Michaelis, *J. Biol. Chem.*, **150**, 407 (1943); *cf. ibid.*, **146**, 451 (1942) and **155**, 661 (1944).

ritin.[45] Rothen [46] reports a molecular weight of 465,000 for the apoferritin.

Several types of copper proteins have been described. One large group consists of the hemocyanins, found in the blood of the squid, octopus, lobster, oyster, and the mollusks. For some time the hemocyanins were thought to contain porphyrins, but the latest studies fail to substantiate this. Copper is present in the range from 0.33 to 0.38 per cent, and a Van Slyke nitrogen distribution [47] indicates that 31.3 per cent of the nitrogen is present as the hexone bases. Conant [48] and others isolated what they believed to be the prosthetic group as a "complex cupric salt of a polypeptide composed of 3 molecules of serine, 1 of leucine, 1 of tyrosine, and some sulfur-containing compound." The hemocyanins investigated by Svedberg have molecular weights ranging from 2,000,000 to 5,000,000.

The oxygen-carrying capacity of hemocyanin is only approximately one-fourth that of hemoglobin. The fact that the respiratory pigment in these lower forms of animal life contains copper rather than iron and that the copper-containing compound is less efficient than hemoglobin in its vital functions, combined with the fact that copper is essential for the formation of hemoglobin, leads almost inevitably to the speculation whether the role of copper in the formation of hemoglobin may not be a vestigial remnant of a vital mechanism carried over in the processes of organic evolution. In the early stages of embryonic growth of the pig so much copper is present in the embryo that the ash is colored a characteristic blue. No studies have as yet been made to see whether or not hemocyanin may be present in the very early stages of mammalian embryonic growth.

Among the plant oxidases it has been shown that laccase, tyrosinase, and ascorbic acid oxidases contain copper as an essential part of the enzyme.[49, 50, 51] Dawson and Mallette [52] have surveyed the general field of copper proteins and discussed the role of hemocuprein and hepatocuprein and other copper derivatives reported present in the animal body.

Several other mineral elements appear to be associated with the protein of certain enzymes: magnesium in carboxylase; zinc in insulin, uricase, and carbonic anhydrase; and manganese and iron in arginase.[52] In

[45] I. Fankuchen, *J. Biol. Chem.*, **150**, 57 (1943).
[46] A. Rothen, *J. Biol. Chem.*, **152**, 679 (1944).
[47] C. L. Alsberg and E. D. Clark, *J. Biol. Chem.*, **8**, 1 (1910).
[48] J. B. Conant, F. Dersch, and W. E. Maydans, *J. Biol. Chem.*, **107**, 755 (1934).
[49] D. Keilin and T. Mann, *Biochem. J.*, **34**, 1163 (1940).
[50] J. M. Nelson and C. R. Dawson, *Advances in Enzymol.*, **4**, 99 (1944).
[51] W. H. Powers, S. Lewis, and C. R. Dawson, *J. Gen. Physiol.*, **27**, 167 (1944).
[52] C. R. Dawson and M. F. Mallette, *Advances in Protein Chem.*, **2**, 179 (1945).

most cases the linkages are not known. Several of these conjugated proteins are not colored, but it seemed desirable to note them at this point.

Manganese was thought to be conjugated in "pinnaglobin," a respiratory pigment of a tropical mussel, but Henze [53] has shown that the metal was not part of a protein. A similar situation obtains in the case of certain tunicates which were reported to contain vanadium as part of a respiratory protein. Webb [54] showed that the product did contain the metal but it was neither a protein nor a porphyrin derivative.

Pigmented Plant Proteins. Certain sea algae contain the chromoproteins *phycoerythrin* and *phycocyan.* These have been studied by a number of workers. Kylin [55] prepared them in crystalline form, identified them as of protein nature, studied their behavior toward various reagents, including enzymes, acids, and alkalies, and their reactions toward heat and light.

Later Kitasato [56] isolated a reddish violet, water-soluble pigment to which he gives the formula $C_{48}H_{120}O_{18}N_{12}$ from phycoerythrin. This is further hydrolyzed by acids to yield a blue pigment, insoluble in water, soluble in organic solvents, and having a composition, $C_{30}H_{54}O_{12}N_6 \cdot 1.5\ H_2O$. He notes that phycocyan yields, on peptic digestion, the final blue pigment directly, the intermediate compound not being obtained.

Svedberg studied the light absorption of these proteins and their minimal particle weight as determined in the ultracentrifuge. The particle weight ranged from 106,000 to 208,000, depending on the biological source of the protein. The chemical nature of the colored group still remains for future investigation.

Melanoproteins. Black wool by treatment with extremely dilute alkalies [57] yields a material of protein nature which is intensely black but which on acid hydrolysis yields about 90 per cent of its weight in amino acids. This material is apparently a conjugated protein in which melanin is the associated chromogen. Inasmuch as no similar compound could be isolated from non-pigmented wool, and since certain other pigmented structures, such as human hair, black rabbit fur, and black feathers, did not yield a similar product, it seems reasonable to believe that the pigment of black wool exists in the form of *melanoproteins* rather than in the form of melanin which is not associated with the protein molecule. Brown horse hair and brown cattle hair yield similar products. The protein, as isolated, appears to be relatively unaltered, inasmuch as it

[53] M. Henze, *Z. physiol. Chem.*, **162**, 136 (1926).
[54] D. A. Webb, Jr., *J. Exptl. Biol.*, **16**, 499 (1939).
[55] H. Kylin, *Z. physiol. Chem.*, **69**, 169 (1910).
[56] Z. Kitasato, *Acta Phytochim. Japan*, **2**, 75 (1926).
[57] R. A. Gortner, *J. Biol. Chem.*, **8**, 341 (1910); *Biochem. Bull.*, **1**, 207 (1911).

still will undergo heat coagulation after it has been separated and purified by dialysis.

Flavoproteins. Warburg and Christian [58] prepared the first of a series of flavoproteins (the "old yellow enzyme") involved in the transfer of hydrogen in anaerobic respiration. To date ten enzymes [59] of this general composition are known. The riboflavin is tied to phosphoric acid and constitutes a nucleotide. In certain cases the riboflavin nucleotide is combined with adenylic acid through their respective phosphoric acid groups, thus forming a dinucleotide. In either form the riboflavin nucleotide acts as a hydrogen transport medium when combined with the appropriate protein of an enzyme system (see p. 996).

Carotenoid Proteins. The visual purple of the retina of the eye appears to be a chromoprotein in which the prosthetic group is a carotenoid, retinene. Retinene is closely related to vitamin A, inasmuch as in the absence of vitamin A the normal amount of visual purple fails to develop and a nutritional type of "night blindness" [60] ensues. Visual purple, when acted upon by light, decomposes into visual yellow, which consists of the mixture of a protein and retinene; in the dark, visual purple is in part regenerated. In this cycle vitamin A is expended and must be replenished from the diet. Visual violet (iodopsin),[61] is another carotenoid protein of the cones in the retina. Ovoverdin, found in lobster eggs, has been shown to be a protein with a carotenoid grouping.

The Glycoproteins. We have already noted that egg albumin contains a carbohydrate radical. The amount of carbohydrate which is present, however, is extremely small in comparison with the amount of carbohydrate in the group known as the mucins and the mucoids. The mucin contained in the submaxillary glands has been reported to contain 23.5 per cent, salivary mucin to contain as high as 36.9 per cent, and tracheal mucin from 34 to 37 per cent, of carbohydrate.

Meyer [62] has surveyed the mucoids and the glycoproteins and proposed a classification of these compounds on the basis of their chemical compositions. Certain groupings are not conjugated proteins, but of the latter it is noted that most contain an acetylated hexosamine together with uronic acid and sulfuric acid as the prosthetic group. These have been studied particularly by Levene [63] and his co-workers. He

[58] O. Warburg and W. Christian, *Biochem. Z.*, **254**, 438 (1932).

[59] J. B. Sumner and G. F. Somers, *Chemistry and Methods of Enzymes*, 2nd ed., Academic Press, Inc., New York, 1947.

[60] G. Wald, *J. Gen. Physiol.*, **17**, 351, 781 (1936).

[61] S. Hecht and J. Mandelbaum, *Science*, **88**, 219 (1938).

[62] K. Meyer, *Advances in Protein Chem.*, **2**, 249 (1945).

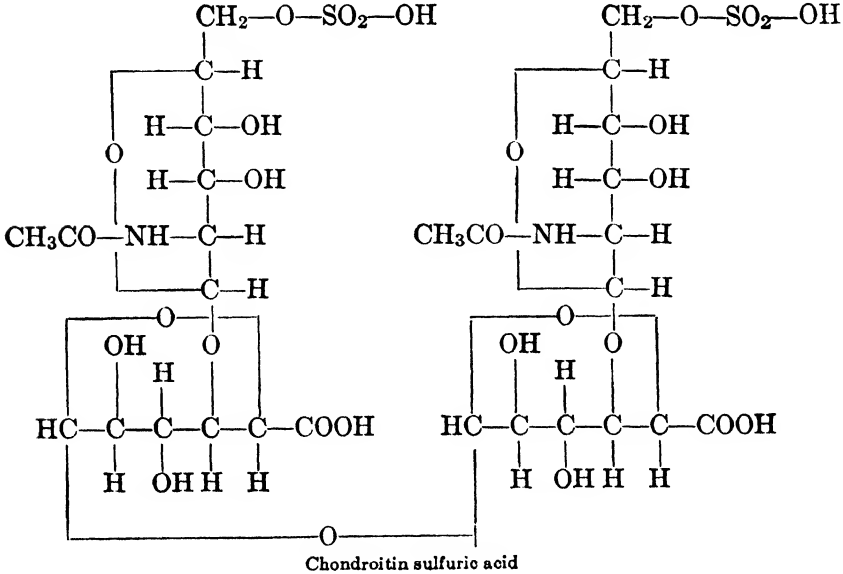
[63] P. A. Levene, *Hexosamines, Their Derivatives, and Mucins and Mucoids*, The Rockefeller Institute for Medical Research Monograph 18, 1922.

distinguishes two types of groups associated with the protein, *i.e.*, chrondroitin sulfuric acid and mucoitin sulfuric acid.

Chondroitin sulfuric acid on hydrolysis yields a reducing disaccharide, *chondrosin*, composed of 1 molecule of chondrosamine (aminogalactose) combined with a molecule of glucuronic acid. Both the amino group and the carboxyl group are free, the linkage apparently being from the carbonyl group of the chrondrosamine to the fourth carbon of the glucuronic acid. Other hydrolytic products are glucuronic acid, acetic acid, and sulfuric acid.

Chondroitin sulfuric acid contains 2 chondrosamine residues in which the amino group is acetylated, and the hydroxyl group on the sixth carbon atom is esterified with sulfuric acid. These acetylated and esterified chondrosamine residues are linked through glycosidal linkages to a disaccharide acid composed of 2 molecules of glucuronic acid as shown in the following graphic formula. Provisionally the linkage from the chondrosamine residue is to the fourth carbon of the glucuronic acid.

The decomposition products of *mucoitin sulfuric acid* are mucosin, a disaccharide composed of glucuronic acid, and glucosamine. *Mucosin* apparently has a structure very similar to chondrosin, excepting that the amino sugar is derived from glucose instead of galactose.



Chondroitin sulfuric acid

These amino sugar residues likewise occur in the acetylated form and are linked with a diglucuronic acid residue through glycosidal linkages. The sulfuric acid ester residue is much less stable in mucoitin sulfuric

acid, so presumably the esterification is on a different hydroxyl group from that in chondroitin sulfuric acid.

When the sulfuric acid radical is removed from the chondroitin sulfuric acid or from the mucoitin sulfuric acid, substances are formed which are non-reducing, relatively inert chemically, and in many respects resemble gum acacia (gum arabic) in properties. These compounds are highly hydrophilic and probably account for the very hydrophilic properties of the mucins and mucoids with which they are associated in the conjugated proteins.

Chondroitin sulfuric acid has found a use in medicine in the treatment of migraine. Alvarez [64] cites Crandall and Roberts as having observed that relief was afforded in about 50 per cent of the cases where chondroitin sulfuric acid was administered.

Heparin, the anticoagulant from the liver, is a glycoprotein. Charles and Todd [65] have prepared the barium salt of the prosthetic group and conclude, from its properties, that it is of the mucoitin type except that five sulfuric acid residues are present on each tetrasaccharide unit.

Hyaluronic acid [66] appears to be the prosthetic group of certain products found in a large variety of places. It is of the uronic acid-acetylglucosamine type, extremely viscous and capable of forming a mucin clot with soluble proteins. An enzyme, hyaluronidase, has been reported in testes, malignant tissues, snake and spider venom, and in leeches, which is capable of hydrolyzing the hyaluronic acid and thus reducing its clotting properties. The system is also referred to as a "spreading agent" because normal skin is rendered more susceptible to penetration by large molecules such as occur in vaccines when a preparation containing the enzyme is first spread on the surface of skin.[67]

The Phosphoproteins. *Casein* of milk and *vitellin* of egg yolk are the two most important phosphoproteins. Casein, because of its ease of preparation, has been most extensively investigated. The chemistry of casein, including its uses in industry and the arts has been adequately discussed by Sutermeister and co-authors,[68] and in the American Chemical Society Monograph dealing with dairy science.[69]

[64] W. C. Alvarez, *Proc. Staff Meetings Mayo Clinic*, **9**, 22 (1934).

[65] A. F. Charles and A. R. Todd, *Biochem. J.*, **34**, 112 (1940).

[66] K. Meyer and E. Chaffee, *J. Biol. Chem.*, **138**, 491 (1941).

[67] J. H. Humphrey, *Biochem. J.*, **37**, 177, 460 (1943); F. Duran-Reynals, *Bact. Rev.*, **6**, 197 (1942).

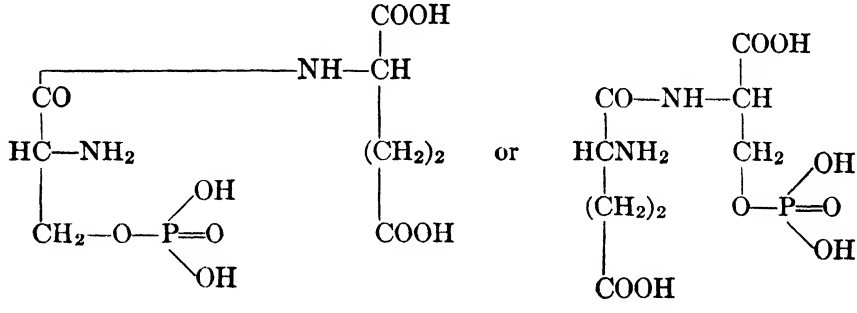
[68] Edwin Sutermeister and Frederick L. Browne, *Casein and Its Industrial Applications*, 2nd ed., Reinhold Publishing Corp., New York, 1939. Note in particular Chap. 2.

[69] Associates of Lore A. Rogers, *Fundamentals of Dairy Science*, 2nd ed., American Chemical Society Monograph Series, Reinhold Publishing Corp., New York, 1935.

Casein is the characteristic protein of the milk of all species of mammals, even the egg-laying monotreme or spiny anteater, *Echidna aculeata multiaculeata*, which secretes in its mammary glands a protein having the characteristics of casein.[70] Similarly, vitellin appears to be the characteristic protein of egg yolk, where it actually exists as a lecithoprotein (p. 416), although the various vitellins have not been investigated so extensively as have the caseins.

The caseins are acted upon by a special enzyme, rennin, which converts them into another protein, *paracasein*, having somewhat different properties. Whether rennin actually brings about a hydrolytic cleavage resulting in a paracasein with a smaller micellar weight than the original casein or whether we are dealing with a special type of denaturation is still uncertain.

The prosthetic group of the phosphoproteins is orthophosphoric acid. Levene and Hill [71] and Schmidt [72] almost simultaneously but independently reported the isolation of a phosphodipeptide, glutamyl-serine phosphate, which Levene notes possesses one or the other of the following alternative formulas:



More recently several similar phosphopeptones have been isolated from a tryptic digest of casein. Rimington's [73] nonapeptide contains 3 phosphate residues, 4 of serine, and 5 of glutamic acid; Lowndes, Macara, and Plimmer [74] isolated an octapeptide containing 2 phosphate nuclei, 2 serine, 2 glutamic acid, and 3 residues of what is probably isoleucine.

The fact that phosphoric acid is esterified on the —OH group of serine was proved by the isolation of serine phosphoric acid.[75] Rimington's paper indicates that the phosphoric acid may be esterified both as the

[70] H. R. Marston, *Australian J. Exptl. Biol. Med. Sci.*, **3**, 217 (1926).
[71] P. A. Levene and D. W. Hill, *J. Biol. Chem.*, **101**, 711 (1933).
[72] G. Schmidt, *Z. physiol. Chem.*, **223**, 86 (1934).
[73] C. Rimington, *Biochem. J.*, **35**, 321 (1941).
[74] J. Lowndes, T. J. R. Macara, and R. H. A. Plimmer, *Biochem. J.*, **35**, 315 (1941).
[75] F. Lipmann, *Biochem. Z.*, **262**, 3 (1933); *cf.* also *ibid.*, **262**, 9 (1933).

monophosphate and as the diphosphate esters. Through the differential action of bone phosphatase, which does not hydrolyze diphosphoric acid esters, and kidney phosphatase, which hydrolyzes both the mono- and diesters, he concludes that the phosphoric acid in casein exists in a ratio of 2 monoesters to 1 diester. If that is true, phosphoric acid may, when in the diester form, act as a bridge between two polypeptide chains and link such chains together in the larger casein molecule.

Casein, in comparison with most proteins, is a relatively strong acid, and its titration curve with sodium hydroxide shows a sharp inflection between *p*H 6.0 and 7.5 (*cf.* Fig. 17). It is in the same range that the second hydrogen of phosphoric acid is neutralized, so that presumably a very considerable part of the alkali-binding capacity of casein is attributable to the secondary hydrogen of the phosphoric acid. An extensive study of casein has been conducted in Sörensen's laboratories with the conclusion that casein is a reversible dissociable component system,[76] and this point of view is borne out by sedimentation studies in Svedberg's laboratories where the sols were found to be heterodisperse, the particle weight varying from 75,000–100,000 up to 375,000, depending on the fraction which was studied and the technic employed in the preparation of the protein sols.

The Nucleoproteins. The nucleoproteins consist of a protein group, usually a *histone* or *protamine*, combined with *nucleic acids*. The nucleic acids have been extensively studied, particularly by Levene,[77] and a summary of the more recent literature is presented by Greenstein.[78]

The nucleic acid may usually be separated from the protein moiety by salts or by cautious addition of acids or alkalies. The ease of separation suggests that the linkages are essentially salt-like although it is recognized that the proportions of the two components are not constant. The two nucleic acids which have been most widely studied are derived from yeast and from calf thymus glands. The former is taken as typical of plant materials, and the latter of animal sources; however, there are exceptions to this rule.

Upon hydrolysis with acid, nucleic acids yield four molecules each of orthophosphoric acid, of a sugar, and of certain bases. This broad statement is based upon the assumption of a tetranucleotide structure, which will be considered later. In yeast the sugar is D-ribose, and the bases consist of two purines, *adenine* and *guanine*, and two pyrimidines, *cytosine* and *uracil*. From thymus nucleic acid similar products are obtained ex-

[76] K. Kondo, *Compt. rend. trav. lab. Carlsberg*, [8] **15** (1925); K. Linderstrøm-Lang and S. Kodama, *ibid.*, [1] **16** (1925); K. Linderstrøm-Lang, *ibid.*, [9] **17** (1929).

[77] P. A. Levene and L. W. Bass, *Nucleic Acids*, American Chemical Society Monograph 56, Chemical Catalog Co., New York, 1931.

[78] J. P. Greenstein, *Advances in Protein Chem.*, **1**, 210 (1944).

cept that the sugar is 2-desoxy-D-ribose and the pyrimidines are *cytosine* and *thymine*. On the basis of Bial's test for pentoses and the similarity of the bases, the following contain acids of the yeast type or ribose-nucleic acids: wheat embryo, tobacco mosaic virus, pancreas β-nucleoproteins, and chick embryo β-nucleoproteins. By analogous tests the thymus type or desoxyribose-nucleic acids are found in the pancreas, liver, spleen, and other sources. Thus it is seen that the chemical distinctions between nucleic acids of animal and of plant origins are not sharply drawn. In their study of tuberculinic acid, Johnson and Coghill [79] found a new pyrimidine, 5-methyl cytosine, which is unusual. It is coming to be recognized that both types of nucleic acids, at least as far as the sugars are concerned, may be found in the same cell. Thus, the desoxyribose type has been found in the nucleus, whereas the ribose type is found in the cytoplasm and in the nucleolus of the same cell.[80]

D-Ribose was identified by Levene and Jacobs in a series of papers during the years 1908–1911. Rather interestingly, neither this pentose nor desoxyribose was known to occur in nature until they were discovered as decomposition products of nucleic acid. Fischer, during his studies on the carbohydrates, had synthesized D-ribose in 1901 and desoxyribose in 1913. We have, thus, examples of synthetic products prepared first in a chemical laboratory which after a number of years were found to be identical with what are probably the most important carbohydrates occurring in nature. An interesting philosophical discussion could be built around this question of this "most important carbohydrate." We derive our energy from glucose. The cell nuclei contain either D-ribose or 2-desoxy-D-ribose. Is the engine more important than the fuel which runs it, or is the fuel more important than the engine?

Levene and Tipson [81] have shown that, in nucleic acids, both ribose and desoxyribose occur in the furanose structure. Suzuki, Odake, and Mori [82] isolated a thiopentose from a nucleoside in yeast, which fact Levene and Sobotka [83] have verified. It is not known whether this sugar occurs in nucleic acid, and the observation seems to have been overlooked.

Guanine and adenine are the only naturally occurring purines in the nucleic acids. Xanthine and hypoxanthine are derived from these by the action of acids or enzymes. Purine, the mother substance of the

[79] T. B. Johnson and R. D. Coghill, *J. Am. Chem. Soc.*, **47**, 2838 (1925).

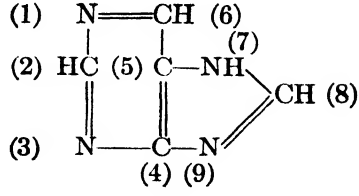
[80] T. Caspersson and J. Schultz, *Proc. Natl. Acad. Sci.*, **26**, 507 (1940); T. S. Painter and N. Taylor, *ibid.*, **28**, 311 (1942); R. Feulgen, M. Behrens, and S. Mahidassan, *Z. physiol. Chem.*, **246**, 203 (1937).

[81] P. A. Levene and R. S. Tipson, *J. Biol. Chem.*, **109**, 623 (1935).

[82] U. Suzuki, S. Odake, and T. Mori, *Biochem. Z.*, **154**, 278 (1924).

[83] P. A. Levene and H. Sobotka, *J. Biol. Chem.*, **65**, 551 (1925).

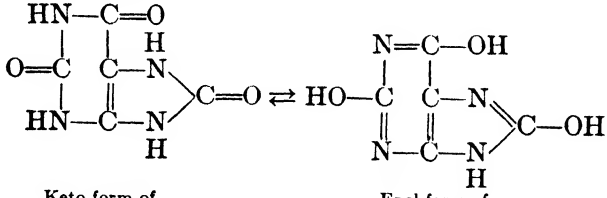
purine bases, does not occur in nature but has been synthesized and has the following structure:



The various derivatives with which we are concerned are as follows:

Guanine, 2-amino-6-oxypurine.
Adenine, 6-aminopurine.
Guanine + enzyme (guanase) → 2,6-dioxypurine, *xanthine*.
Adenine + enzyme (adenase) → 6-oxypurine, *hypoxanthine*.
Hypoxanthine by oxidation goes to *xanthine*.
Xanthine + oxygen + enzyme → 2,6,8-trioxypurine, *uric acid*.

The oxygenated purine or pyrimidine compounds undergo keto ⇄ enol isomerism and accordingly may exist in either the enol form or the keto form, or as an equilibrium mixture of the two forms:

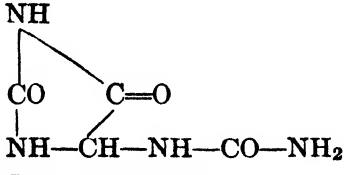


Keto form of uric acid

Enol form of uric acid

Rose [84] has reviewed the changes which take place in the various purine and pyrimidine derivatives under the influence of various reagents and enzymes, and has discussed the fate of the various purine derivatives in man and in animals. The question whether or not animals are able to synthesize the purine derivatives needed for nucleic acid formation has not been satisfactorily answered.

Uric acid is the primary end product of purine metabolism in man and in the anthropoid ape. Monkeys excrete only traces. In the lower animals it is further oxidized to allantoin:

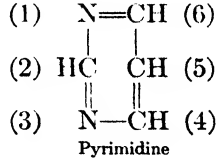


[84] W. C. Rose, *Physiol. Revs.*, **3**, 544 (1923).

which in some cases is excreted as such, in others is further oxidized to urea in the liver. When the liver is removed, the purines are excreted as uric acid. Allantoin excretion decreases after liver removal (dog).[85] The enzymes adenase and guanase are *deaminases*, the further oxidation to uric acid being produced by *oxidases*.

Besides the purines noted above, some others are important from a phytochemical standpoint. These are *caffeine*, 1,3,7-trimethylxanthine, *theobromine*, 3,7-dimethylxanthine, and 1,3,7,9-tetramethyl-2,6,8-trioxypurine (tetramethyl uric acid), which has been found to occur in tea. Theobromine occurs in cocoa beans to the extent of 1.5 to 2.4 per cent and in chocolate to the extent of 1 to 2 per cent. Caffeine occurs in tea, coffee, kola nut, cocoa bean, etc. The coffee bean contains about 1 per cent, tea leaves 1.5 to 2.5 per cent. It is a diuretic and cerebral stimulant and has a pronounced stimulating action on the heart.

The pyrimidine compounds found in nucleic acids are derivatives of the pyrimidine nucleus:



Pyrimidine

The most important derivatives from our standpoint are

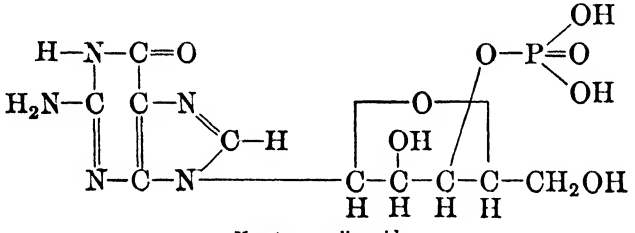
Cytosine, 2-oxy-6-aminopyrimidine
Thymine, 2,6-dioxy-5-methylpyrimidine
Uracil, 2,6-dioxypyrimidine
5-*Methylcytosine*, 2-oxy-5-methyl-6-aminopyrimidine

In recent years a synthetic pyrimidine compound, *alloxan* (2,4,5,6 tetraoxypyrimidine), has proved useful in the experimental production and study of diabetes (see p. 963). Whether this substance is a factor in human diabetes remains to be elucidated.

The structure of nucleic acid has been arrived at by a study of the mononucleotides, more often called the *nucleotides*, compounds containing a single molecule each of phosphoric acid, sugar, and a purine or pyrimidine base. Typical nucleotides are adenylic, guanylic, and inosinic acids, containing adenine, guanine, and hypoxanthine, respectively. Inosinic acid does not occur as such in nucleic acid but results from the deamination of adenylic acid. Attention has also been paid to the *nucleosides* which consist of the sugar and the base. In yeast, and in most other nucleic acids, the linkage between the base and the sugar is a glycosidal tie to nitrogen of position-9 in the case of the purines and to

[85] J. L. Bollman and F. C. Mann, *Proc. Staff Meeting Mayo Clinic*, **7**, 631 (1932).

nitrogen of position-3 in the pyrimidines. The glycosidal linkage is evidenced by the rate of hydrolysis, which is similar to that of other glycosides. The position on the purine ring was decided by considering the possibilities presented by the several nitrogens; since all purine nucleotides in yeast appear to be similar, only positions-7 and -9 could be considered. Seven- and 9-methyl purines are known from their syntheses and, of these, 9-methyl purine gives an absorption curve similar to that of the purine nucleotides and nucleosides. Hence the linkage to ribose is at position-9, and similar studies fix the point of attachment on the pyrimidine nucleus at position-3. The phosphoric acid in yeast is esterified to carbon-3 of the sugar; this was shown by Levene and Harris [86] who reduced the sugar ester to its alcohol, which was found to be optically inactive. These and other arguments are summarized in a lecture by Gulland.[87] The structure of yeast guanylic acid is therefore that of 9-guanine-3-phosphoribofuranoside:



Yeast guanylic acid

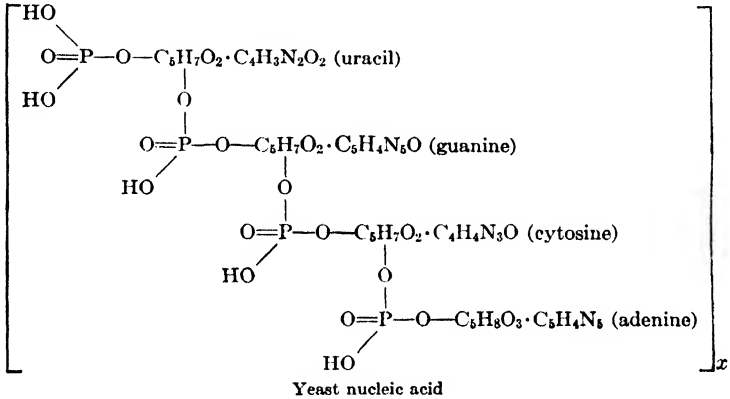
It should be noted that the above is true only for yeast nucleic acid. In thymus nucleic acid the phosphoric acid is attached to carbon-5 (the primary alcoholic group) of the sugar. In free muscle adenylic acid and adenosine, which play important roles in anaerobic respiration, the sugar is tied to position-7 of adenine; again the phosphoric acid is esterified to carbons-3 or -5, depending on the source.

Before leaving the field of the nucleotides it should be noted that in recent work on anaerobic respiration several compounds similar to the nucleotides have been reported. *Niacin* is readily converted *in vivo* to a sugar-phosphoric acid derivative, and *riboflavin* is similarly converted to a phosphate ester on the ribityl group. Some authors consider these compounds nucleotides; others call them pseudonucleotides. Obviously they differ from the nucleotides of nucleic acid only in the nature of the base involved.

The traditional concept of the structure of nucleic acid considers it to be a tetranucleotide, each unit of which differs from the others in the

[86] P. A. Levene and S. A. Harris, *J. Biol. Chem.*, **98**, 9 (1932).
[87] J. M. Gulland, *J. Chem. Soc.* (*London*), 1722 (1938).

nature of the base. The linkages between nucleotides involves the sugar of one nucleotide and the phosphoric acid of the next; this has been arrived at by the isolation of several fragments.[88] It will be recalled that, in yeast, phosphoric acid is esterified on the third carbon of ribose within the nucleotide; the link between two nucleotides involves esterification on the second carbon atom of ribose, as illustrated below. Obviously this union between nucleotides could not take place in the thymus-nucleic acid-type which contain 2-desoxyribose; here the phosphoric acid is tied to carbon-5 of the sugar within the nucleotide and the cross linkage involves esterification on the third carbon of the sugar.



Yeast nucleic acid

It will be noted that such a tetranucleotide would contain five equivalents of ionizable acid whereas titrations reveal approximately four per tetranucleotide. This finding can be reconciled with a polymeric structure wherein only one end nucleotide would carry five ionizable hydrogens. Molecular weight determinations establish the facts that the nucleic acids are in polymeric forms and that the size of these polymers varies with the method of preparation. Loring [89] reports the value of x in the formula for yeast nucleic acid to be 15; for virus nucleic acid Cohen and Stanley [90] give a value of over 200; also Tennent and Vilbrandt [91] find that thymus nucleic acid is larger, with a value of x ranging from 500 to 1,000. Recently an enzyme named *ribonuclease* has been prepared in crystalline form by Kunitz [92] from beef pancreas. This enzyme depolymerizes polynucleic acids of the ribose type.

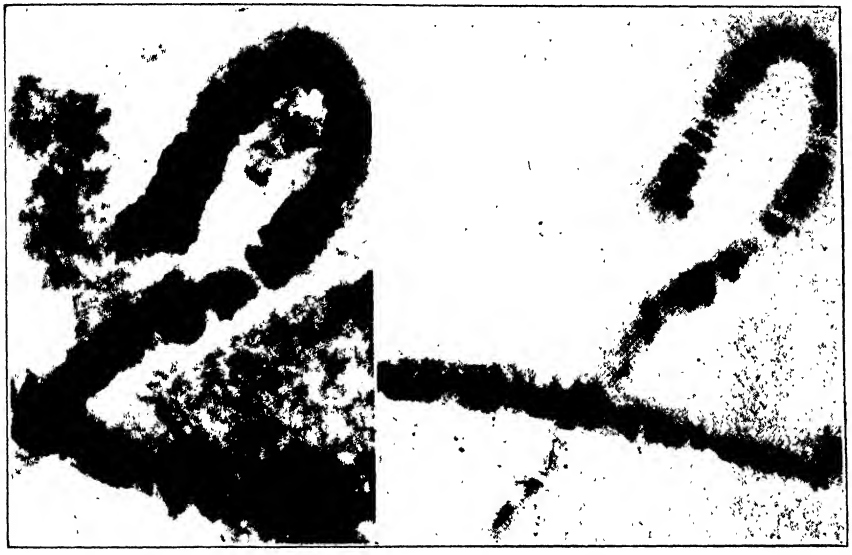
[88] P. A. Levene and W. A. Jacobs, *J. Biol. Chem.*, **12**, 411 (1912).
[89] H. S. Loring, *J. Biol. Chem.*, **128**, lxi (1939).
[90] S. S. Cohen and W. M. Stanley, *J. Biol. Chem.*, **144**, 589 (1942).
[91] H. G. Tennent and C. F. Vilbrandt, *J. Am. Chem. Soc.*, **65**, 424 (1943).
[92] M. Kunitz, *J. Gen. Physiol.*, **24**, 15 (1940).

As already noted, *chromatin* appears to be almost exclusively composed of protamine nucleate, a protamine combined with nucleic acid. The available evidence points to either a true salt formation or a coacervate system. Chromatin is the chemical substance from which *chromosomes* are derived. The nucleic acids form approximately 40 per cent of the solid components of the chromosomes, and, as Leathes notes, if we consider that into these chromosomes "are packed from the beginning all that preordains, if not our fate and fortunes, at least our bodily characteristics down to the color of our eyelashes, it becomes a question whether the virtues of nucleic acids may not rival those of amino acid chains in their vital importance." Leathes notes that, on the basis of 40 per cent of the nucleic acids in the chromatin, there are approximately half a million molecules of nucleic acid in a sperm cell.

Both Miss Wrinch [93] and Caspersson [94] have discussed the molecular structure of chromosomes. Miss Wrinch's paper is largely a theoretical discussion of the available literature. Caspersson's paper, on the other hand, is an extensive experimental study. Both reach essentially the same conclusions. Caspersson studied the structure of the chromosomes by making use of the ultraviolet absorption spectra. Nucleic acid has a great absorption in the ultraviolet. Proteins likewise have marked absorption in the same region, but Caspersson was able to digest out the proteins, leaving what he believes to be the nucleic acid residues, and thus to demonstrate that nucleic acid is deposited in the chromosomes in bands which correspond to the bands which the cytologist has observed in many of the chromosomes and which appear to be associated with the genes of the geneticist. Figure 89(*a*) shows parts of two of the giant salivary gland chromosomes of *Chironomus* larvae as photographed by Caspersson, showing dark ultraviolet absorption bands. Figure 89(*b*) shows the same preparation after the protein had been digested out in a 24-hour proteolytic digestion. The dark bands in Fig. 89(*b*) occupy essentially the same positions as in Fig. 89(*a*), but the absorbing protein material between the areas where the nucleic acid is concentrated has been digested out so that Caspersson believes that the bands shown in Fig. 89(*b*) demonstrate the localization of nucleic acid within the chromosome structure. Miss Wrinch postulates the structure of chromosomes as being made up of a protein (protamine) warp and a nucleic acid fill. The nucleic acid is linked to the *basic* groups of the protamine but is not attached to the monoaminomonocarboxylic acid residues. There would

[93] D. M. Wrinch, *Protoplasma*, **25**, 550 (1936).
[94] T. Caspersson, *Skand. Arch. Physiol.*, 73 (Supplement No. 8) (1936).

thus be a banding of nucleic acid surrounding the parallel chains of protamines in those areas where the dibasic amino acids are concentrated. This would account for the banded structure of the chromosomes. The coacervate viewpoint would lead to the same picture; the negatively charged anions of the nucleic acid would be attracted to the positively charged areas of the protamine so that localized concentrations of nucleic



(*a*) (*b*)

FIG. 89. (*a*) Chromosomes from salivary glands of *Chironomus* larvae photographed with ultraviolet microscope technic. The dark bands are due to an intense absorption of ultraviolet light by protamine-nucleate. (*b*) Same preparation as (*a*) after the protein portion of the chromosomes had been solubilized by digestion with proteolytic enzymes for 24 hours. Note that nucleic acid is unaffected by proteolytic enzymes and remains in a banded structure. (After Caspersson.)

acid would occur along the protamine structure. The only difference between this picture and a true salt formation would be that in the coacervate the water shells would prevent the positively charged hydrated cations from completely neutralizing the hydrated negatively charged anionic residues. Thus, the positive and negative charges would still retain their identities, the structure being held together by electrostatic attractions but the charged areas being held apart by the repulsion of the oriented water shells.

Virus Proteins. Within the past fifteen years great progress has been made in the study of this class of proteins. The literature is voluminous;

the reader is referred to recent reviews by Bawden,[95] Lauffer,[96] Pirie,[97] and Stanley and others.[98]

The filterable viruses possess the ability to reproduce under favorable conditions; so far this has been achieved only in the living select host and not by *in vitro* culture. The production of the additional virus material is at the expense of the host and, in plants, may vary from 0.05 to 2 grams per liter of the infected sap. The viruses may be isolated upon fractionation with salts or with alcohol but are most often separated by ultracentrifugation. Some of the purest preparations have been obtained by recrystallization of the ultracentrifugate from ammonium sulfate systems.

Among the plant viruses most widely studied are the tobacco mosaic virus and the tomato bushy stunt virus. The latter is very constant in its properties, but the former may be modified by chemical treatment to alter its chemical properties and its infectivity upon related plants.[99]

The most carefully purified virus preparations from plant sources are nucleoproteins. In general the animal viruses appear to be more complex because their preparations contain lipid and enzyme components as well as nucleoproteins. The pentose type of nucleic acid occurs in plant viruses and in the agents responsible for chick sarcoma and for equine encephalomyelitis; the desoxypentose type is present in the viruses of influenza, vaccinia, and rabbit papilloma. The nucleic acid content varies from 5 to 40 per cent of the total protein content in the several infective agents. The tobacco mosaic virus, which has been widely studied, contains 5.8 per cent of nucleic acid. Its protein is well characterized as to amino acid content. However, the various strains of this virus act upon different hosts, and the proteins differ in their amino acid composition.[100] Like all nucleic acids and nucleoproteins, the viruses show a strong absorption band in the region of 2,600 Å.

The particle size of various viruses has been calculated from their sedimentation constants in the ultracentrifuge. The largest reported

[95] F. C. Bawden, *Plant Viruses and Plant Diseases*, Chronica Botanica, Waltham, Mass., 1943.

[96] M. A. Lauffer, *Chem. Revs.*, **31**, 561 (1942); also M. A. Lauffer and W. M. Stanley, *ibid.*, **24**, 303 (1939).

[97] N. W. Pirie, *Advances in Enzymol.*, **5**, 1 (1945).

[98] W. M. Stanley, *Physiol. Revs.*, **19**, 524 (1939); also W. M. Stanley and C. A. Knight, *Cold Spring Harbor Symposia Quant. Biol.*, **9**, 255 (1941).

[99] G. A. Miller and W. M. Stanley, *J. Biol. Chem.*, **141**, 905 (1941); **146**, 331 (1942).

[100] C. A. Knight, *J. Am. Chem. Soc.*, **64**, 2734 (1942); *J. Biol. Chem.*, **147**, 663 (1943).

molecular weight is that for the tobacco mosaic, found by Lauffer [101] to be 40 million. This is of the same order of magnitude as the value of 69 million obtained by Neurath and Saum [102] from diffusion measurements. The smallest reported value for any purified virus is that of the alfalfa mosaic with a weight of 2 million.[103] The tobacco mosaic virus is rod-shaped with the long axis 40 times the other two axes of 150 Å. By contrast, the tomato bushy stunt virus is essentially spherical in shape with a diameter of 260 Å. and a molecular weight of about 7 million.

There exist certain points of similarity between the viruses and genes. Both are nucleoproteins and can perpetuate themselves, given favorable conditions. Both may be altered by external conditions. The genes are caused to change under certain radiations; viruses may similarly be altered, but a more striking case exists here because they have also been altered *in vitro* by chemical reagents.

In view of the fact that certain cells have been stripped of much of their cytoplasm leaving mainly the "naked nuclei" (consisting largely of the chromatin which comprises the genes) and yet are able to carry on vital functions, Gortner [104] among others has questioned whether a similar condition may not obtain in the case of the known virus preparations. He points out that the methods employed in their isolation and purification are such that vital function would persist even if much of the cytoplasmic material is removed. In other words, are the viruses now described merely essential parts of organisms or are they molecules in the accepted meaning of the word? Those who insist that the best virus preparations are chemical entities point to the constant properties on repeated purification. Many regard the viruses as an unusual class of matter which bridges the gap between living (organized material) and non-living matter. The problem is an interesting one; perhaps it may be regarded as largely semantic.

The Lecithoproteins. The literature on this class of proteins is somewhat confused; some workers feel that the lecithoproteins are definite compounds, others that they represent mixtures of proteins with lecithin or other lipids in varying proportions. Within recent years, however, there has been a trend to view the combinations between proteins and lipids which occur in tissues as conjugated proteins.

It seems probable that in many instances the proteins existing in the tissues may be associated with fats, lipids, or fatty acids. However, in

[101] M. A. Lauffer, *J. Physical Chem.*, **44**, 1137 (1940).
[102] H. Neurath and A. M. Saum, *J. Biol. Chem.*, **126**, 435 (1938).
[103] M. A. Lauffer and A. F. Ross, *J. Am. Chem. Soc.*, **62**, 3296 (1940).
[104] R. A. Gortner, *Science*, **87**, 529 (1938).

our method of purification of the protein we remove these groups by fat solvents. This is perhaps simply a restatement of the remarks already credited to Abderhalden where he notes that the proteins prepared and purified for chemical study may in many instances bear little resemblance in their physical properties to the proteins as they actually exist in the cells and tissues. It is a common observation that, in a fat determination, ether fails to remove all the lipids unless some alcohol is added; this suggests that part of the fatty material is bound to the protein. Perhaps it would be better to designate this group as *lipoproteins* since lipids other than lecithin may be involved.

A recent review by Chargaff [105] considers these compounds as they occur in the cytoplasm, plastids, membranes, and mitochondria of cells, in egg yolk, in bacteria and viruses, in blood serum, and particularly as the thromboplastic protein (tissue fibrinogen).

Chargaff and others [106] have prepared the lipovitellin of egg yolk. Their product was constant in composition and after repeated ether extraction contained 18.8 per cent of the total yolk lipids. The lipid content of the product was 23 per cent, of which the phospholipids constituted the major part.

Mills [107] had earlier prepared an extract from lung tissue which he called tissue fibrinogen and which served to convert the pro-enzyme, prothrombin, to thrombin. The latter is the enzyme which acts upon soluble fibrinogen to produce the fibrin clot. Chargaff and co-workers [108] have obtained the most potent preparation by fractional ultracentrifugation, in which several of the lipid hydrolytic components have been identified. The protein component free of lipids had no thromboplastic property. The conjugated protein had a molecular weight of 167 million, and as little as 0.008 γ would produce a clot. They also observed that the anticoagulant heparin would disrupt the complex with the liberation of the lipid component and the formation of a heparin-protein.

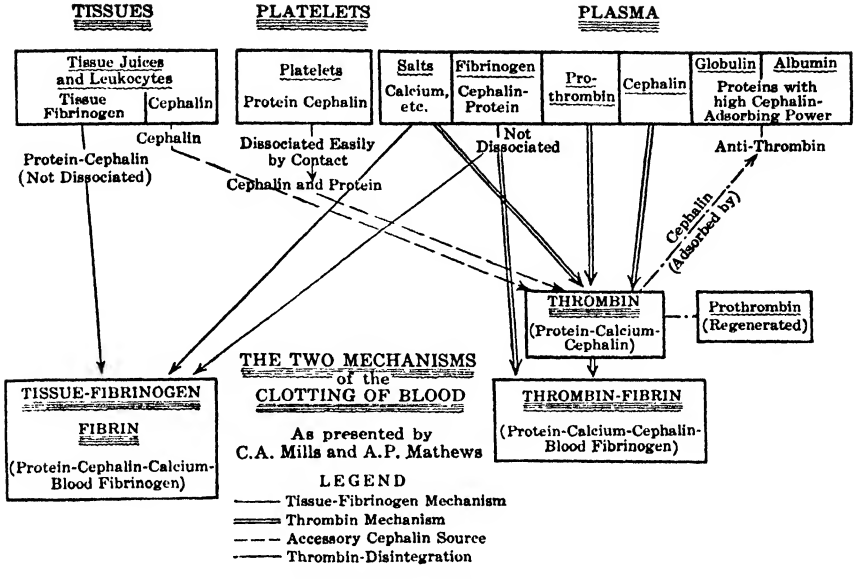
Mills and Mathews have studied the mechanism of the clotting of blood and find that, in both the normal clotting process and the clotting process induced by tissue fibrinogen, there is the interaction of both protein and phospholipid. The following chart prepared by Mills and Mathews indicates the two mechanisms of blood clotting.

[105] E. Chargaff, *Advances in Protein Chem.*, **1**, 1 (1944).

[106] E. Chargaff, *J. Biol. Chem.*, **142**, 505 (1942); also E. Chargaff, M. Ziff, and D. Rittenberg, *ibid.*, **144**, 343 (1942).

[107] C. A. Mills, *J. Biol. Chem.*, **40**, 425 (1919); **76**, 651 (1926).

[108] S. S. Cohen and E. Chargaff, *J. Biol. Chem.*, **139**, 741 (1941); also E. Chargaff, D. H. Moore, and A. Bendich, *ibid.*, **145**, 593 (1942).



Derived Proteins. Derived proteins, as already noted, are proteins which have been chemically altered by manipulation subsequent to their isolation. They include such groups as the coagulated proteins, the halogenated proteins, the nitrated proteins, the formaldehyde proteins, deaminized proteins, racemized proteins, and proteans. The literature in this field is so voluminous that mention can be made of only a few of the reactions involved. The coagulated proteins constitute one type of denatured proteins which are considered in a separate chapter.

Halogenated Proteins. When the halogens, chlorine, bromine, or iodine, are introduced into the protein, the question always arises as to whether or not the method of halogenation involves a partial hydrolysis of the protein. In most instances, this undoubtedly occurs. In general, substitution takes place on the aromatic nuclei, notably the benzene rings of tryptophan, tyrosine, and to some extent that of phenylalanine. The same may be said for the nitrated proteins. The Millon and the Adamkiewicz reactions, as a rule, are negative, the characteristic hydrolytic products being the halogenated or nitrated aromatic radicals, particularly that of tyrosine, although the nitrated proteins yield nitroarginine.

From iodized casein Reinecke and Turner [109] isolated 0.4 per cent of good-quality thyroxine of which one fourth was obtained as crystalline L-thyroxine. The mechanism of the formation of thyroxine from some

[109] E. P. Reinecke and C. W. Turner, *J. Biol. Chem.*, **149**, 555 (1943).

component of casein has never been explained, although it has been made *in vitro* from L-tyrosine.[110]

Formaldehyde Proteins. It will be recalled that, in the Sörensen formol method, titration is carried out at room temperature in a 6 to 8 per cent solution of formaldehyde. The reaction is rapid, and, since the linkage to the amino group is weak, it is readily reversible upon dilution. However, when proteins are subjected to low concentrations of formaldehyde for longer periods of time more permanent linkages result. Thus bacterial toxins may be converted to toxoids (see p. 471) by 0.15 per cent formaldehyde upon incubation for several weeks at about 38°C. The result is a product which has lost its virulence but is still antigenic. In virus proteins Ross and Stanley [111] observed an inactivation which could in part be restored by prolonged dialysis at *p*H 3.

Proteins unite with formaldehyde to form compounds having very different properties from the original protein. Thus, egg albumin treated with formaldehyde is no longer coagulable by heat. Most of the work on the formaldehyde proteins has been carried out with casein, and the manufacture of artificial ivory, galalith, or casein plastics, involving the union of formaldehyde and casein, has resulted in the building up of a large chemical industry, the more or less chemically reactive casein being converted into a relatively inert horn-like mass. A discussion of this process has been given by Brother.[112] Little is known in regard to the chemical reactions involved. Yeast and yeast residues, or the protein of the soybean, have been used to replace casein and form similar "artificial ivories."

Formaldehyde has been widely used in the arts to render proteins more insoluble, as in certain tanning practices. Several methods have been proposed to measure the amount bound to the protein. The whole problem of the action of formaldehyde on amino acids and proteins is reviewed by French and Edsall.[113]

Deaminized Proteins. Deaminized proteins, prepared by the action of nitrous acid on the protein, have been rather extensively studied. Dunn and Lewis [114] have shown that deamination of casein leads to a product containing no free amino group and that the hydrolysate contains no lysine in the phosphotungstic acid precipitate, but the filtrate from the bases is increased by an amount of amino nitrogen equivalent

[110] C. R. Harington and R. V. P. Rivers, *Biochem. J.*, **39**, 157 (1945).

[111] A. F. Ross and W. M. Stanley, *J. Gen. Physiol.*, **22**, 105 (1938).

[112] G. H. Brother, in Sutermeister and Browne, *Casein and Its Industrial Applications*, Reinhold Publishing Corp., New York, 1939.

[113] D. French and J. T. Edsall, *Advances in Protein Chem.*, **2**, 278 (1945).

[114] M. S. Dunn and H. B. Lewis, *J. Biol. Chem.*, **49**, 327, 343 (1921).

to the α-amino group of lysine. These observations are in keeping with the postulate that the ϵ-amino group of lysine is free in proteins and hence subject to deamination.

Sandstrom [115] found that several proteins behaved in a similar manner. He also showed that the acid-binding capacity of proteins was decreased upon deamination.

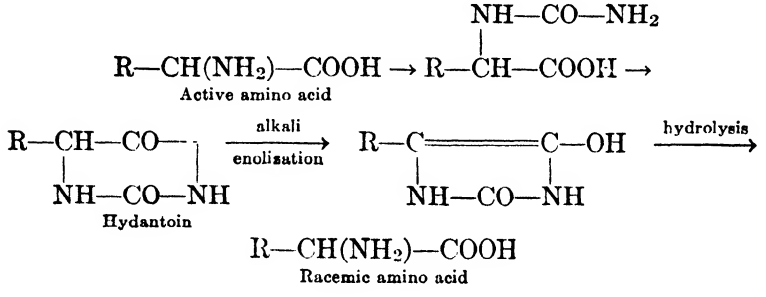
Deaminized proteins when fed to experimental animals produce a nutritional anemia. This is a toxic anemia [116] rather than a deficiency anemia, since supplementing the diet with good protein does not relieve the symptoms. Alcoholic sodium hydroxide extracts the toxic factor, leaving a relatively non-toxic residue. The nature of the toxin has not been elucidated.

Proteans. The change of certain globulins to an insoluble modification, known as a protean, has already been noted. Edestin of hempseed, excelsin of the Brazil nut, and myosin of the muscle fibers are readily transformed from a protein which is easily peptized by 10 per cent sodium chloride to proteans which are no longer peptized. Here, again, we do not know the nature of the chemical change involved. Osborne [117] concluded that this transformation involves hydrolysis and that the edestan may be regarded as the first stage in the hydrolysis of edestin. He found that the transformation is accelerated by the presence of acids and that various acids produced in a given time very different amounts of the edestan. The figures which he gives indicate that the hydrogen ion is the active catalyst, although his work was done prior to the general use of hydrogen-ion control. The chemical analysis of edestan is within experimental error of the chemical analysis of the edestin from which it was derived.

Racemized Proteins. Dakin [118] observed that, when proteins are dissolved in dilute alkali, there is a progressive fall in the specific optical rotation with time. The reaction is usually carried out for at least 10 days at a temperature of 38°C. in 0.2 to 0.5 *N* sodium hydroxide. In a typical case the specific rotation changed from $-80°$ to $-20°$, after which it was essentially constant. The reader will note that this is not racemization in the sense in which the term is normally understood. Dakin observed a similar change in the rotation of hydantoins prepared from optically active amino acids. When hydrolyzed they yielded racemic rather than active amino acids. He explains the change as a keto $\rightleftharpoons$ enol isomerism, the hydrogen migrating from the α-carbon

[115] W. M. Sandstrom, *J. Phys. Chem.*, **34**, 1071 (1930).
[116] M. I. Smith and E. F. Stohlman, *U. S. Pub. Health Repts.*, **51**, 772 (1936).
[117] T. B. Osborne, *Z. physiol. Chem.*, **33**, 225 (1901).
[118] H. D. Dakin, *J. Biol. Chem.*, **13**, 357 (1912).

atom to the carbonyl group, forming a double bond, as shown in the accompanying structural formulas. Later when hydrolysis occurs and the amino acids are regenerated, the double bond no longer exists, but, since either one of the unions of the double bonds may later become attached to hydrogen, we would expect to find not the active but equal amounts of D- and L-acids to be formed.



In working with gelatin, Dakin found that a similar racemization occurred and that when the gelatin was later hydrolyzed, certain of the amino acids were optically inactive, whereas others were optically active. The inactive amino acids were leucine, aspartic acid, arginine, histidine, and phenylalanine. Part of the alanine was active; part was inactive. All the proline, glutamic acid, and lysine which he isolated was optically active. He believes that the reason that certain amino acids were racemized, whereas others were not, depends on the position of the amino acid in the protein molecule, and notes that only those amino acids, where both the $—NH_2$ group and the —COOH are combined in the protein molecule, undergo racemization. According to this viewpoint, the inactive amino acids are situated somewhere in the interior of the chain making up the protein molecule, whereas those which are not racemized are on the ends of the peptide chain.

Later, Dakin and Dale [119] compared the albumins of hens' eggs and ducks' eggs in regard to racemization and to acids which they yielded on hydrolysis. Only three amino acids were found to show characteristic differences. Leucine from racemized albumin from hens' eggs was almost completely racemic, whereas leucine isolated from the racemized duck albumin was almost completely active. Aspartic acid from the duck albumin was completely inactive, whereas that from hen albumin was partly active. Histidine from hen albumin was completely inactive, from duck albumin almost completely active. They believe that these observations indicate structural differences in the two albumins.

[119] H. D. Dakin and H. H. Dale, *Biochem. J.*, **13**, 248 (1919).

Woodman [120] employed the racemization phenomenon in a second manner; he compared the rates of change in optical activity as a function of time when the temperature and the alkalinity were kept constant. Since the two proteins traced the same curve he concluded that the globulins of cow and ox serum and of colostrum are identical, that the albumin of milk is identical with the albumin of colostrum, but that lactalbumin and serum albumin are not identical.

Dakin and Dudley [121] studied racemized casein, and Dudley and Woodman [122] repeated the work of Dakin and Dudley, comparing casein from cow's milk and casein from the milk of sheep. In casein from cow's milk, all the amino acids isolated were racemized, with the exception of proline and part of the valine and leucine. In casein from sheep's milk, lysine, proline, and tyrosine were all optically active, glutamic acid, leucine, and valine partly active. They, therefore, reached the conclusion that cow and sheep caseins contained the amino acids linked in different positions.

Dakin and Dudley [123] attempted to isolate tryptophan from racemized casein by tryptic digestion and found that neither trypsin, erepsin, nor pepsin would attack the racemized protein; neither was it digested *in vivo*, and, when fed to dogs, it was eliminated unchanged in the feces. The racemic casein was not attacked by putrefactive bacteria during a 10-day interval, but after that was very slowly attacked. They point out that all artificial polypeptides which have been synthesized have been subjected at some stage or other to alkaline conditions, and suggest that possibly the failure of enzymes to attack certain of these synthetic products may be due to some change similar to the change which has taken place in protein racemization. They also note that, in the separation of proteins from biological materials, precautions must be taken not to subject them to the presence of alkalies for any extended time.

Ten Broeck [124] prepared racemic egg albumin by allowing egg albumin to stand at 37° for 3 weeks in contact with 0.5 *N* sodium hydroxide. He states that the product which was isolated differed chemically from egg albumin only in rotatory power. He found that this racemic protein had no immunological power. It neither sensitized nor intoxicated; no antibodies were formed. These observations support Vaughan's idea (see p. 423) that a splitting of the protein must take place before im-

[120] H. E. Woodman, *Biochem. J.*, **15**, 187 (1921).
[121] H. D. Dakin and H. W. Dudley, *J. Biol. Chem.*, **15**, 263 (1913).
[122] H. W. Dudley and H. E. Woodman, *Biochem. J.*, **9**, 97 (1915).
[123] H. D. Dakin and H. W. Dudley, *J. Biol. Chem.*, **15**, 271 (1913).
[124] C. Ten Broeck, *J. Biol. Chem.*, **17**, 369 (1914).

munological reactions occur. Obviously, if racemic proteins are not attacked by enzymes, no splitting can take place.

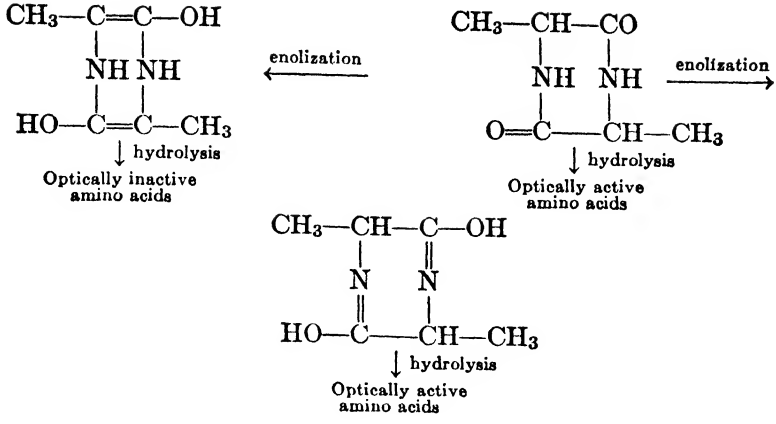
The above discussion and the papers cited would lead one to believe that a racemic protein is merely a protein in which certain shifts in linkages have taken place with a concomitant change in optical rotation, and that racemic proteins are somewhat analogous to coagulated proteins and proteans. However, in most instances, the racemic product which is isolated amounts to only a fraction of the weight of protein which is originally taken. Thus, 20 to 30 grams of racemic casein can be isolated if 100 grams of casein have been acted upon by the 0.5 *N* sodium hydroxide. A large part of the casein has undergone deep-seated hydrolysis, and one of the hydrolytic products, *i.e.*, the racemized "protein," is still capable of being readily isolated. The question arises, therefore, whether the analysis of this protein fraction can be compared with that of the original protein in drawing valid conclusions. Csonka and Horn [125] show that rapid hydrolysis of the protein takes place in 0.5 *N* sodium hydroxide at room temperature and that the digestion products are optically active in the sodium hydroxide solution. However, when they are later boiled with 20 per cent sodium hydroxide to effect a complete hydrolysis, optical activity disappears. Csonka and Horn suggest the discontinuance of the name "racemic protein."

Levene and Pfaltz [126] investigated the action of alkali on diketopiperazines and polypeptides. They find that the dipeptide, D-alanyl-D-alanine, is not racemized by alkali. The anhydride, however, changes from an optical rotation of $+17.5°$ to $-16.0°$, at which point 50 per cent of the nitrogen is amino nitrogen, *i.e.*, hydrolysis of the polypeptide has taken place. They note that the diketopiperazine ring must remain in contact with the alkali for an appreciable period of time before it hydrolyzes, in order to form intermediate products which on hydrolysis yield optically inactive acids. It will be noted that there are two enol forms of the dimethyldiketopiperazines, as shown in the diagram on p. 423, only one of which will yield the optically inactive amino acids.

Levene and Pfaltz note that even when the tripeptide, glycyl-L-alanyl-glycine, and the tetrapeptide, glycyl-glycyl-L-alanyl-glycine, were allowed to stand in contact with alkali, no racemization occurred during the first 48 hours, but, on long standing, there was a slight decrease in optical rotation, the maximum racemization never exceeding 10 per cent, and even in these cases hydrolysis reached 80 per cent or more. They note that there is no way of determining whether racemization precedes or follows the hydrolysis, and point out that perhaps the race-

[125] F. A. Csonka and M. J. Horn, *J. Biol. Chem.*, **93**, 677 (1931).

[126] P. A. Levene and M. H. Pfaltz, *J. Biol. Chem.*, **63**, 661 (1925); **68**, 277 (1926).



mization of proteins may be the racemization of diketopiperazine compounds rather than the long peptide chains which Dakin postulates. In considering the diketopiperazine theory of protein structure, we have already noted the later studies of Levene and Bass on the racemization of polypeptides, diketopiperazines, and proteins where they found that casein did not behave like either the polypeptides or the diketopiperazines. The rate of racemization was more nearly what one would expect for diketopiperazines. The rate of hydrolysis was more nearly what one would expect for polypeptides. They conclude that in casein either the diketopiperazine rings are unusually stable or the order of the linkage favors racemization.

Acid and Alkali Albumins. These are perhaps best typified by Paal's [127] protalbinic and lysalbinic acids which have already been noted as examples of protective colloids. We should perhaps add to this group the products prepared by Vaughan's method [128] of partially hydrolyzing proteins in alcoholic solution with sodium hydroxide. In this way, proteins can be separated into two fractions, an alcohol-soluble portion and an alcohol-insoluble residue. The alcohol-soluble portion was found by Vaughan to be exceedingly poisonous when injected intravenously, exceeding in its toxic action even such drugs as strychnine. No adequate explanation for the high degree of toxicity has been offered. Vaughan notes, however, that gelatin appears to be the only protein which does not yield the toxic fraction. We have already noted that gelatin is deficient in a number of amino acids, those containing aromatic nuclei, and that it does not induce the usual immunological reactions. Vaughan believes the immunological reactions are, at least

[127] C. Paal, *Ber.*, **35**, 2195 (1902).
[128] V. C. Vaughan and F. G. Novy, *Cellular Toxins*, Lee Brothers, 1902.

in part, induced by a hydrolytic cleavage in the animal body, more or less similar to the hydrolytic cleavage which takes place in an alcoholic sodium hydroxide solution, and that the toxins liberated in the anaphylactic reaction are similar, if not identical, to his protein poisons. Similar protein poisons [129] have been prepared from the proteins of the tubercle bacillus.

Relatively little work from the chemical standpoint has been done with these products. Miss Wheeler [130] has shown that they are complex and has isolated a number of amino acids from the various preparations.

[129] B. White and O. T. Avery, *J. Med. Research*, **26**, 317 (1912).

[130] S. M. Wheeler, *J. Biol. Chem.*, **6**, 509 (1909).

CHAPTER 17

Reactions of Proteins with Acids and Bases

From the preceding discussions of the chemical nature of the proteins, we have seen that the protein molecule possesses both acidic and basic groups and accordingly must be looked upon as an amphoteric substance which exists in the zwitterion state at its isoelectric point. We have noted that proteins in solution show the characteristic properties of colloidal sols and also that they may be either positively or negatively charged, depending on the hydrogen-ion concentration of the medium. Furthermore, changes in the medium may bring about molecular rearrangements, such as a shift in the keto $\rightleftharpoons$ enol isomerism, whereby additional reactive groups may come into play.

It is not surprising, therefore, to find that proteins undergo reactions with acids, with bases, and with certain salts. The binding of acids and bases has been studied by a great many workers, and a voluminous literature has resulted. At this time mention can be made of only a few of the more important considerations.

In spite of the large amount of study which has been carried out in this field the workers are not in entire agreement as to the interpretations put upon the experimental results, but they may in general be classified into two schools. One group of workers interprets the reactions between proteins and acids or bases as *stoichiometrical* in character, with colloidal reactions playing no appreciable role. In other words, the primary valence forces of the ionogenic groups present in the protein molecule determine the amounts of acid and of alkali bound by that protein. This school of thought received a great impetus from the work of Loeb [1] and has been perpetuated through the activities of Schmidt,[2] Cohn and Edsall,[3] Hitchcock, Wilson, and Cannan,[4] to mention only a few of the outstanding workers.

[1] J. Loeb, *Proteins and the Theory of Colloidal Behavior*, McGraw-Hill Book Co., New York, 1922.

[2] C. L. A. Schmidt, *Chemistry of the Amino Acids and Proteins*, 2nd ed., with addendum, Charles C. Thomas, Springfield, Illinois, and Baltimore, 1944.

[3] E. J. Cohn and J. T. Edsall, *Proteins, Amino Acids and Peptides*, Reinhold Publishing Corp., New York, 1943.

[4] R. K. Cannan, *Chem. Revs.*, **30**, 395 (1942).

A second group of workers include those who, though admitting that the forces of primary valences play a role, nevertheless view these reactions as those of a lyophilic *colloidal* system and support the view that the action of ions upon proteins is probably due largely to the surface forces of *adsorption*. This group insists that, at least under certain conditions, the surface forces outweigh or tend to modify the forces of primary valence. Gortner and Bancroft supported this viewpoint for many years.

We have already noted in the consideration of the Donnan equilibrium that the establishment of such an equilibrium is independent of the nature of the process which causes the formation of a non-diffusible ion. Similarly, it is immaterial in many instances whether a charged particle is formed by the process of ionization or by the colloidal process of adsorption. In either instance the charged micelle will behave as an ion, and as such will obey the physicochemical laws characteristic of an ionized system. This is one reason why it is difficult to decide between the colloidal and the strictly stoichiometrical viewpoint in regard to the action of acids and bases upon proteins.

In an earlier discussion it was pointed out that the proteins possess a buffering capacity (*cf.* Fig. 17) which results from the fact that a part of the acid or base in solution is "bound" by the protein. The earlier literature has been reviewed by Hoffman and Gortner [5] in a study of the prolamines, with casein and fibrin included for purposes of comparison. The earlier workers used various types of technics, certain of which are only briefly cited since they are not much used today.

One method involved the *direct precipitation of a protein compound.* Thus casein was titrated with a solution of calcium hydroxide to a definite indicator; then the soluble calcium-casein complex was precipitated by adding alcohol. Depending on the indicator selected, caseinates containing various amounts of calcium could be prepared. Obviously this method has serious limitations.

A second method involved *dissolving water-insoluble compounds in protein solutions.* Casein dissolves freshly precipitated calcium carbonate when the two are ground in water. Similarly, insoluble alkaloids or bases, such as strychnine and freshly precipitated copper hydroxide, can be brought into solution by means of proteins. Edestin will dissolve almost 35 per cent of its weight of copper in this manner. Again this method is of no practical value, and the question arises here whether these are the reactions characteristic of the hydrogen ions formed by the dissociation of —COOH groups or whether they are, in part at least, the

[5] W. F. Hoffman and R. A. Gortner, Colloid Symposium Monograph, Vol. 2, p. 209, 1925.

reactions characteristic of peptization. Still a third possibility is the formation of complex metallic compounds in which a chelate type of linkage binds the metal to the nitrogen groups of the peptide linkages or other amino acid residues. We do not have a clear-cut effect of the hydrogen ion, particularly in view of such solubility effects as have been noted for salt solutions by Kruyt and Robinson (p. 219). The protein sols may act as peptizing agents and undoubtedly will act as protective colloids, thus favoring peptization, so that at least part of the reactions between proteins and insoluble substances may well be attributed to the colloidal behavior of the protein system.

A third method which has been proposed is to study *precipitates formed by the interaction of two soluble substances*, one containing a protein radical. Thus, protein dispersed in hydrochloric acid, when added to a metallic phosphotungstate, precipitates a protein phosphotungstate complex. The latter is possibly a chemical compound. Equally possibly it is an adsorption complex. Again we have a method where we cannot sharply separate the colloidal and the stoichiometrical reactions.

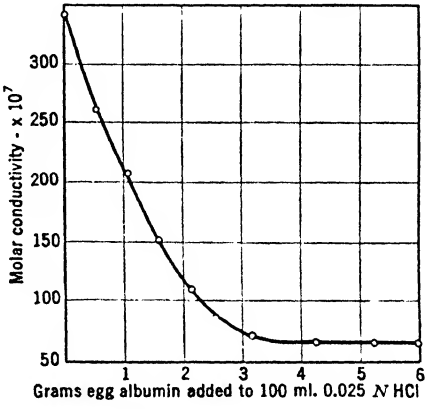


FIG. 90. Showing the change in the electrical conductivity of a dilute hydrochloric acid solution to which increasing amounts of protein (egg albumin) were added. (Data of Sjöqvist.)

Electrical conductivity studies were first used by Sjöqvist [6] when he added increasing amounts of protein to a constant amount of acid until a constant electrical conductivity was reached. Figure 90 shows certain of his results.

Hoffman and Gortner report similar studies for the prolamines and casein and fibrin. They were unable, however, to correlate their findings in any definite way with the chemical composition of the proteins studied.

More recently a number of conductivity studies have been carried out by Schmidt [7] and co-workers. These studies have shown that, in sodium caseinate, the current is carried by the cations of the alkali metal and the casein anions and that ion transference experiments give about

[6] J. Sjöqvist, *Skand. Arch. Physiol.*, **5**, 277 (1895); **6**, 255 (1895).

[7] For literature see: S. Miyamoto and C. L. A. Schmidt, *J. Biol. Chem.*, **99**, 335 (1933).

the same mobility for the casein ion as is indicated by conductivity measurements. When, however, the alkaline-earths are combined with the casein, abnormal values for the transference numbers are obtained, indicating that a considerable portion of the alkaline-earth metals were bound to the protein to yield negatively charged complex ions of the protein and the alkaline-earth residues. We know that in blood serum a considerable part of the calcium is "bound" to the protein, so that it is held in a non-diffusible, non-ultrafilterable form and behaves as if it were actually a part of the protein micellar complex. Presumably other proteins behave similarly in forming these complex ions. Even in solutions of the sodium salts of aspartic and glutamic acids, there is evidence [8] that a few ionic aggregates are present. In proteins there is no definite evidence that the ionic micelles are stoichiometrical compounds.

The cryoscopic method was first employed by Bugarszky and Liebermann,[9] who found that the addition of 6.4 grams of egg albumin to 100 ml. of 0.5 *N* hydrochloric acid reduced the depression of the freezing point of the acid solution to approximately 50 per cent of its initial value.

Barnett [10] made a comparative study of acid and alkali binding, using electrical conductivity, cryoscopic, and potentiometric technic, and reached the conclusion that the cryoscopic technic is the least valuable, owing to the small depressions contributed by the protein ions or micelles. In a number of instances the depression obtained was not appreciably greater than the experimental error of the method.

Another method is the *catalytic effect of hydrogen and hydroxyl ions.* Some reaction is selected which is catalyzed by one of these ions, for example, the inversion of sucrose by hydrogen ions [11] or the saponification of an ester by hydroxyl ions.[12] When a protein is added to such a system, the rate of catalysis is changed, and it is concluded that hydrogen ions have been removed from that system. Here, again, the method gives no indication of the mechanism whereby hydrogen or hydroxyl ions are removed from the solution.

The method which has been most generally employed in recent years is the *potentiometric method,* involving the measurement of the change in hydrogen- or hydroxyl-ion concentration in acid or alkali solutions to which proteins have been added. This method was first used by Bugarszky and Liebermann,[9] who concluded that chlorine ions, as well as hy-

[8] W. M. Hoskins, M. Randall, and C. L. A. Schmidt, *J. Biol. Chem.*, **88**, 215 (1930).
[9] S. Bugarszky and L. Liebermann, *Pflügers Arch. ges. Physiol.*, **72**, 51 (1898).
[10] H. M. Barnett, M. S. Thesis, University of Minnesota, 1927.
[11] F. A. Hoffman, *Z. klin. Med.*, **10**, 793 (1889); **11**, 521 (1890).
[12] R. Wintgen and K. Krüger, *Kolloid-Z.*, **28**, 81 (1921).

drogen ions, are bound by the protein. During the period 1900 to 1925 various workers [13] proposed equations by means of which the amount of acid or alkali bound by the protein could be calculated from hydrogen-ion-concentration measurements. All these equations, however, included the dissociation constant of the acid or alkali as calculated from conductivity measurements.

More recently, workers have calculated the amount of acid or base bound by the protein from potentiometric measurements made on the aqueous systems both with and without the protein. It is recognized that this procedure involves the assumption that the activity of the hydrogen ion is not greatly changed by the presence of protein ions. Hoffman and Gortner [5] used the equation

$$n = N - \frac{(H^+)}{\alpha'} \tag{159}$$

where n = the normality of acid or alkali bound
N = the original normality of the acid or alkali
(H^+) = the hydrogen-ion concentration of the protein-acid or protein-alkali solution at equilibrium
α' = the degree of ionization of the acid or alkali as determined by potentiometric methods.

Cohn criticized this formula and proposed a new formula which in turn involves several assumptions. The derivation of his formula is given in a consideration of the base-binding capacity of casein.[14] Using the fundamental equation for deriving pH, *i.e.*,

$$\frac{\text{emf. (observed)} - E\text{ (calomel electrode)}}{0.00019837T} = \log\frac{1}{H^+} = p\text{H} \tag{160}$$

Cohn converts this into pOH by the equation

$$p\text{H} + p\text{OH} = p\text{K}_w \tag{161}$$

where

$$p\text{OH} = \log\frac{1}{(OH^-)} \tag{162}$$

[13] L. Blaŝel and J. Matula, *Biochem. Z.*, **58**, 417 (1913); W. Pauli and A. Spitzer, quoted by Pauli, *Colloid Chemistry of the Proteins*, Pt. I, p. 93 (translated by P. C. L. Thorne), J. and A. Churchill, London, 1922; D. J. Lloyd and C. Mayes, *Proc. Roy. Soc. London*, **B93**, 69 (1922).

[14] E. J. Cohn and R. E. L. Berggren, *J. Gen. Physiol.*, **7**, 45 (1924).

Cohn then introduces the activity coefficient γ in place of the dissociation coefficient α' of equation (159), the activity coefficient being determined by the ratio

$$\frac{(OH^-)}{(NaOH)} = \gamma \tag{163}$$

Combining equations (162) and (163), we have the equation

$$p\text{OH} = \log \frac{1}{\gamma(\text{NaOH})} = p\text{NaOH} + p\gamma \tag{164}$$

Table 47 shows the activity coefficients [15] γ and the logarithms of the reciprocals of the activity coefficients $p\gamma$ for varying concentrations of hydrochloric acid and sodium hydroxide.

TABLE 47. THE ACTIVITY COEFFICIENT γ (LEWIS AND RANDALL) AND THE LOGARITHMS OF THE RECIPROCAL OF THE ACTIVITY COEFFICIENT $p\gamma$ FOR VARIOUS CONCENTRATIONS OF HYDROCHLORIC ACID AND SODIUM HYDROXIDE

Molarity	HCl		NaOH	
	Activity coefficient, γ	$\log 1/\gamma$, $p\gamma$	Activity coefficient, γ	$\log 1/\gamma$, $p\gamma$
0.001	0.98	0.009	0.98	0.009
0.002	0.97	0.013	0.97	0.013
0.005	0.95	0.022	0.95	0.022
0.010	0.92	0.036	0.92	0.036
0.020	0.90	0.046	0.89	0.051
0.025	....	(0.048)	....	(0.055)
0.030	....	(0.050)	....	(0.059)
0.040	....	(0.053)	....	(0.065)
0.050	0.88	0.056	0.85	0.071
0.100	0.84	0.076	0.81	0.091

In using equation (164) one must assume (*a*) that the sodium-protein compound is completely dissociated, (*b*) that the sodium ions from the sodium-protein compound have the same activity as sodium ions in a sodium hydroxide solution, and (*c*) that the protein ion does not influence this activity, or rather that the influence of the protein ion on

[15] G. N. Lewis and M. Randall, *Thermodynamics and the Free Energy of Chemical Substances*, McGraw-Hill Book Co., New York, 1923.

the activity of the sodium ion is the same as is the influence of a hydroxyl ion. Similarly, if equation (164) were used to study acid binding, one would have to assume (*a*) the complete dissociation of the so-called "protein chloride" with the same activity coefficient for the chloride ions of the "protein chloride" as for the chloride ions in an equivalent concentration of hydrochloric acid, and (*b*) no adsorption of the acid in the molecular state which can give rise to hydrogen ions in the solution. The data on gliadin-acetic acid cited on p. 435 show that this assumption is probably unjustifiable. Various workers have indicated that either there is not complete dissociation of the "sodium proteinates" and "protein chlorides" or both anions and cations are adsorbed by the protein. If such a viewpoint is correct, this method of calculation introduces errors which may be as serious as the errors introduced by the equations which have been previously employed.

Using the above equations, Cohn concludes that casein has a maximum base-binding capacity of approximately 0.0014 mole of sodium hydroxide per gram of casein which had never been exposed to greater alkalinities than those which exist in nature, whereas casein which has been prepared by more drastic treatment has a maximum base-binding capacity of 0.0018 mole of sodium hydroxide per gram, 1 mole of sodium hydroxide, therefore, combining with 735 grams of unaltered casein or with 535 grams of casein somewhat altered.

Cohn's studies of "maximum" base-binding capacity involved concentrations of sodium hydroxide ranging between 0.03 *N* and 0.05 *N*. Hoffman and Gortner in their studies covered the range from 0.0005 *N* to 0.50 *N*. It seemed desirable, therefore, to repeat [16] the work, using the technic and method of calculation exactly as suggested by Cohn. In this study a succinic acid-sodium hydroxide system was compared with casein-sodium hydroxide and paracasein-sodium hydroxide systems. In the succinic acid-sodium hydroxide system the experimental and theoretical maximum base-binding capacities coincided. Neither the casein- nor paracasein-sodium hydroxide systems showed maximum base-binding capacities, but it was found that the amount of base bound by the casein was dependent on the equilibrium hydrogen-ion concentration. Even alanine showed a varying base-binding capacity in the presence of different normalities of sodium hydroxide. Incidentally this study also included the effect of neutral salts on the protein-base-binding reaction. The "salt" effect was very much more marked for the protein-base systems than is characteristic of simple acid-base systems, indicating that the reactions of a protein-base system cannot be cal-

[16] A. D. Robinson, R. A. Gortner, and L. S. Palmer, *J. Phys. Chem.*, **36**, 1857 (1932).

culated by applying equations derived from the reactions of simple acid-base systems.

The conclusion was drawn from this study that equation (164) is inadequate for the calculation of base binding, or the "sodium proteinate" is not completely dissociated at the higher concentrations of alkali, or sodium hydroxide is "adsorbed" on the ionic micelles according to a typical adsorption isotherm.

Certain observations of Hitchcock [17] are in agreement with the above studies. He studied acid binding of gelatin and edestin at various concentrations of hydrochloric acid, using both the hydrogen electrode and the chloride electrode. In the case of gelatin he concludes that in 0.1 N hydrochloric acid the protein combines with a maximum of 9.4×10^{-4} equivalent of H^+ and 1.7×10^{-4} equivalent of Cl^- per gram protein. In the case of edestin the corresponding values are 13.4×10^{-4} equivalent of H^+ and 3.9×10^{-4} equivalent of Cl^- per gram protein. *It will*

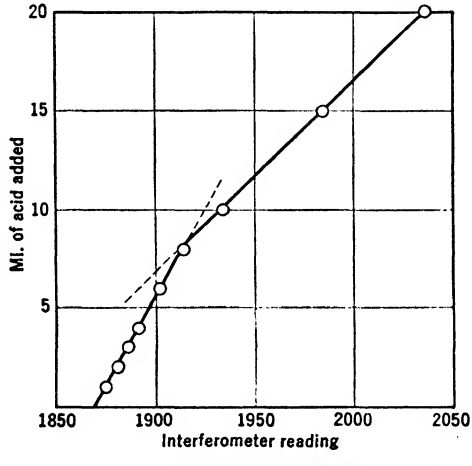


FIG. 91. The titration curve of 0.881 gram gelatin with 0.1004 N HCl, volume 200 ml., as measured by readings in a liquid interferometer. (Data of Thomas and Mayer.)

be noted that in both of these systems Cl^-, as well as H^+, is bound, thus indicating that either the "protein chloride" is not completely dissociated or that hydrochloric acid is in part adsorbed on the protein.

Thomas and Mayer [18] used a Zeiss-Löwe interferometer and found that the refractive indices of mixed solutions of gelatin and of hydrochloric acid are not additive, whereas the refraction of mixed solutions of "gelatin chloride" and of hydrochloric acid are additive. Accordingly, when

[17] D. I. Hitchcock, *J. Gen. Physiol.*, **12**, 495 (1929); *ibid.*, **14**, 99 (1930).

[18] A. W. Thomas and C. W. Mayer, *Proc. Soc. Exptl. Biol. Med.*, **25**, 667 (1928).

a gelatin sol is titrated with an acid, there is an abrupt change in the slope of the curve at some particular acid concentration, which they believe indicates the amount of acid which must be added to gelatin to form the "gelatin chloride." Their curve is shown in Fig. 91. Here again the technic is not capable of distinguishing between stoichiometrical compound formation and an adsorption reaction.

It will be realized that the curves obtained when studying the reactions of varying amounts of acids or of bases upon proteins are in reality dissociation curves; this follows from a consideration of the zwitterion state of proteins. Earlier it was pointed out that, in general, two interpretations have been placed on the results obtained. We should briefly consider the arguments advanced for each.

The Stoichiometrical Argument. When the amounts of acid or of alkali bound by a definite quantity of protein are plotted as functions of

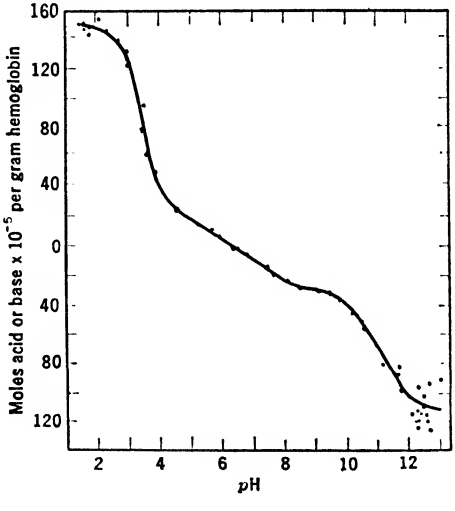


FIG. 92. The titration curve of horse carboxyhemoglobin. (Data of Cohn, Green, and Blanchard.)

the *equilibrium* *p*H, one obtains a curve similar to that shown in Fig. 92, in which carboxyhemoglobin was studied by Cohn, Green, and Blanchard.[19] However, the results obtained upon a variety of proteins differ in the positions and the extent of the inflections; this is generally true over a *p*H range of 2 to 12. From these inflections, which represent ionizable groups, Cohn and his co-workers calculated that those at *p*H 3.7, 4.0, and 4.8 are due to carboxyl groups; those at 5.7 and 7.5 to histidine; and those at 10.8 and 11.6 to lysine and arginine, and possibly

[19] E. J. Cohn, A. A. Green, and M. H. Blanchard, *J. Am. Chem. Soc.*, **59**, 509 (1937).

tyrosine. These values compare reasonably well with the limited analytical data on the amino acid content of hemoglobin.

Wyman [20] has studied the influence of variations in temperature upon the points of inflection, from which it was possible to identify the particular grouping responsible for each of these inflections and to estimate the value of the intrinsic dissociation constants (designated as pK_0') for these groups; the term *intrinsic* represents the isolated effect of the ionogenic group in the absence of effects due to other constituents. Table 48 indicates the estimated values of these groups in proteins.

TABLE 48. IONOGENIC GROUPS OF PROTEINS

(Data of Cannan)

Group	*Amino Acid*	pK_0'
α-Carboxyl		3.5
ω-Carboxyl	Aspartic acid / Glutamic acid	4.0
Imidazole	Histidine	7.0
α-Amino		8.0
ε-Amino	Lysine	10.0
Phenolic	Tyrosine	10.0
Sulfhydryl	Cysteine	10.0
Guanadino	Arginine	12.5

Similar studies have been reported by Cannan and co-workers for egg albumin [21] and β-lactoglobulin.[22] In both cases the values for histidine and arginine as determined independently agree well with those estimated from the curves, but with lysine the isolation values are considerably less than those obtained by titrations. For these and other proteins, the analyses of the dicarboxylic acids are not adequate for a comparison with the titration data. The examples cited are typical of a large body of data which can be interpreted on the basis of the back-titration of ionogenic groups in proteins and thus lead to the conclusion that such "binding" is stoichiometrical in nature.

Many workers point to a maximum binding capacity similar to that illustrated in Fig. 92, in which 148×10^{-5} equivalents of acid and 113×10^{-5} of base are fixed. The horizontal zones at the two ends of the curve indicate the maximum values. However, as Cannan points out, with many proteins the values in the extreme alkaline region do not give a plateau. Partial hydrolysis or some denaturation of the protein may influence the values obtained at the higher *p*H.

[20] J. Wyman, Jr., *J. Biol. Chem.*, **121**, 1 (1939).

[21] R. K. Cannan, A. C. Kibrick, and A. H. Palmer, *Ann. N. Y. Acad. Sci.*, **41**, 243 (1941).

[22] R. K. Cannan, A. H. Palmer, and A. C. Kibrick, *J. Biol. Chem.*, **142**, 803 (1942).

The Adsorption Argument. Gortner argued that the ionogenic groups in proteins are present largely as zwitterions and hence are in a salt-like state. As a consequence one should not expect very weak bases to displace the hydrogen from the ammonium group of the zwitterion, nor should one expect very weak acids to repress the ionization of the carboxyl group. Glycine apparently does not form a compound with ammonia, and most amino acids do not form salts with the fatty acids, *e.g.*, acetates with acetic acid. This is what one would expect from their zwitterion constitution. There seems accordingly no *a priori* reason why proteins, if the amino and carboxyl groups are in the zwitterion form, should be expected to bind stoichiometrically appreciable quantities of either ammonia or the weaker organic acids. With the stronger acids, such as hydrochloric acid, or the stronger bases, such as the fixed alkalies, one would anticipate a destruction of the zwitterion structure and a stoichiometric binding of acid or base equivalent to the number of zwitterions which were affected. However, gliadin does unite [23] with acetic acid to form a "compound" in which 1 gram of gliadin binds a constant amount of 15.4 mg. of acetic acid. This acetic acid is so strongly bound that it cannot be removed by any amount of washing with water, or by exhaustive extraction with alcohol or ether. The gliadin-acetic acid complex in solution has a *p*H of 3.99 as contrasted with a *p*H of 6.50 for the gliadin sol after electrodialysis, indicating that *the acetic acid was not neutralized by basic groups present in the protein.* The acid can be completely removed by electrodialysis. This would appear to be a clear-cut case of the formation of a "gliadin acetate," if it were not for the fact that *purified α-cellulose adsorbs approximately the same amounts of acetic acid and holds it equally tenaciously.*[24] In the α-cellulose there are no basic groups to react stoichiometrically with the acetic acid, so that there the mechanism must be an interfacial adsorption. It seems probable that the "gliadin acetate" is likewise an adsorption complex where the adsorption of the acetic acid is molecular rather than ionic. We have already seen [25] that the adsorption of the fatty acids on charcoal is likewise a molecular rather than an ionic adsorption; therefore gliadin apparently behaves in acetic acid solution somewhat similarly to charcoal.

Thus, while recognizing that proteins do react essentially stoichiometrically with strong acids in the range between the *p*H of their isoelectric point and 2.5 and with strong bases up to a *p*H of 10.5, Gortner

[23] W. B. Sinclair and R. A. Gortner, *Cereal Chem.*, **10**, 171 (1933).

[24] R. A. Gortner and J. J. McNair, *Ind. Eng. Chem.*, **25**, 505 (1933).

[25] V. L. Frampton and R. A. Gortner, *J. Phys. Chem.*, **41**, 567 (1937); *cf.* also E. R. Linner and R. A. Gortner, *ibid.*, **39**, 35 (1935).

contended that surface adsorption may slightly alter the numerical values. However, he adduced considerable evidence that the binding of acids and of alkalies at the concentrations beyond 2.5 and 10.5 is by a process of adsorption.

At these higher concentrations proteins of varied compositions bind approximately the same amount of acid or base. Admittedly, the experimental errors are greater in these ranges. However, when the logarithmic values are employed the constants a and b of the Freundlich adsorption isotherm can be calculated for each protein by the method of least squares. This gives a weighted average of the various readings in any one titration curve. The striking finding is that the lines thus obtained are essentially similar.

It is known that the type of adsorption described by the Freundlich isotherm possesses a negative temperature coefficient, *i.e.*, at equilibrium less adsorption takes place at the higher temperatures. From a comparative study of a large number of prolamines, Hoffman and Gortner [5] found that the binding was essentially in the ratio of 3:2:1 when the determinations were made at 15°, 25°, and 35°C., respectively.

Finally, when one considers the amount of alkali removed from a solution which was initially 0.5 N, more base is bound than could be accounted for on the assumption that every nitrogen atom could come into play.

From the above considerations it now appears reasonably certain that both chemical and adsorption forces play a role in the fixing of alkali and acid by proteins, although the two phenomena take place under different conditions. In his survey article Cannan [4] presents a more extended theoretical consideration of the dissociation curves to which the reader is referred.

The most recent technic is that introduced by Bancroft and Barnett [26] and used by Belden [27] and by Czarnetzky and Schmidt.[28] To avoid a liquid medium, gaseous hydrogen chloride or ammonia was employed with proteins. Schmidt extended the study to amino acids, to which he also added gaseous carbon dioxide and hydrogen sulfide. The amount of gas which "combines" with the protein and the corresponding equilibrium pressure are measured. If the protein forms a solid "ammonium proteinate" or a solid "protein chloride," the equilibrium pressure of the gas must remain constant until all the protein has been converted into the protein compound. If ammonia or hydrogen chloride is adsorbed by the protein, there will be a smooth combination curve characteristic of adsorption.

[26] W. D. Bancroft and C. E. Barnett, *J. Phys. Chem.*, **34**, 449 (1930); **34**, 753, 1217, 1930, 2433 (1930); **36**, 1285, 2299 (1932).

[27] B. C. Belden, *J. Phys. Chem.*, **35**, 2164 (1931).

[28] E. J. Czarnetzky and C. L. A. Schmidt, *J. Biol. Chem.*, **105**, 301 (1934).

When the phase rule

$$F = C - p + 2 \tag{165}$$

is applied to this system of protein plus vapor, there will be three phases —protein, ammonium proteinate (or protein chloride), and ammonia (or hydrogen chloride)—with one degree of freedom. When the experiments are carried out at constant temperature, the degree of freedom is removed, and the system becomes invariant. In such a case the pressure must remain constant, irrespective of the amount of gaseous component in the two solid phases, *i.e.*, until all the protein has been converted into the protein compound. When this point is reached, only two phases and two degrees of freedom are present, so that the pressure will change at constant temperature. If the protein adsorbs the gas, the system is bivariant, since it consists of only two phases, and at constant temperature the pressure will vary continuously with the concentration of the gas in the solid phase. These conditions are shown diagrammatically in Fig. 93. In *A*, we have the gaseous titration of a solid capable

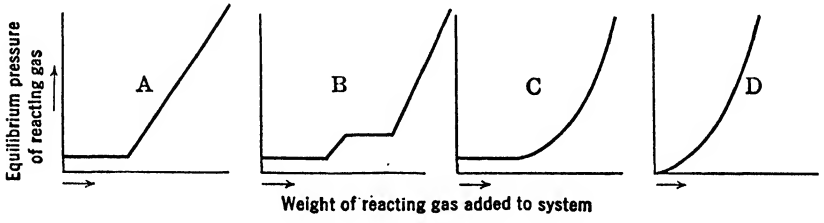


FIG. 93. Showing diagrammatically the phase rule behavior of a gas reacting on a solid where: (*A*) the solid is monoreactive and where the reaction is purely stoichiometric; (*B*) the solid is direactive and the reaction is purely stoichiometric; (*C*) the solid is monoreactive and the product formed by the stoichiometric reaction adsorbs the gas; and (*D*) the interaction between the solid and the gas is wholly one of adsorption.

of combining with one molecule of an acid (or base) to form a definite monobasic compound. In *B*, the solid combines with two molecules of an acid (or base) to form a dibasic compound. In *C*, a monobasic stoichiometrical reaction precedes additional adsorption of the acid (or the base) on the original compound which was formed; and, in *D*, the solid adsorbs the acid (or the base) without definite compound formation.

As the result of his phase rule studies Bancroft concludes that casein, zein, arachin, fibrin, and gliadin readily adsorb ammonia, but that there is no evidence for the formation of any chemical compound. Casein, arachin, fibrin, gliadin, and edestin formed definite compounds with hydrogen chloride, but the amount of hydrogen chloride which combined with these proteins was different from the amount found by previous investigators working with protein solutions and using potentiometric

technics. In the dry state much larger quantities of acid combined. Bancroft found that zein did not form a chemical compound with hydrogen chloride. All the protein chlorides adsorbed additional hydrogen chloride after compound formation had been completed (Fig. 93*C*). Zein adsorbed hydrogen chloride without compound formation (Fig. 93*D*).

Czarnetzky and Schmidt do not agree with Bancroft's findings with respect to compound formation between zein and hydrochloric acid. Neither do they agree to the absence of compound formation in the case of casein, gelatin, edestin, and zein in the presence of ammonia. They report definite ammonium compounds. They also report a definite zein-hydrochloric acid compound but note that the amount of hydrochloric acid which is bound is exceedingly small. It is difficult to account for the divergence in the two series of data. The acid-binding capacity of casein and gelatin is of the same order of magnitude in Bancroft's studies and in those of Czarnetzky and Schmidt. In the case of edestin the acid-binding capacity in Bancroft's experiments is approximately twice that found by Czarnetzky and Schmidt, and the marked discrepancy in the two series of data in the case of zein has already been noted. Both groups of workers report that glycine forms no compound with ammonia. This would be anticipated if glycine is already a substituted ammonium salt existing in the zwitterion form. If the proteins used were in the zwitterion form, one would anticipate that they would not form compounds with ammonia. If in some way the zwitterion form had been destroyed, the compound formation with ammonia would be expected. It may be that the discrepancy in the two series of data is a reflection of the methods by which the proteins were originally prepared. In any event, Czarnetzky and Schmidt note that, "beyond the region of the ammonia or hydrogen chloride-protein titration curves where chemical combination has taken place, the curves are generally not, as might be expected, straight lines. This is probably due to the fact that the relation between the solid phase and the gas is somewhat analogous to that which exists between other solids having large surfaces and certain gases." In other words, although they do not use the colloid chemist's terminology, *adsorption of ammonia or hydrogen chloride takes place on the surface of the ammonium proteinate or the protein chloride which is initially formed, and this is what Bancroft observed in his studies.* These experiments to which the phase rule can be rigidly applied, therefore, bear out Gortner's original contention that *chemical combination of proteins with acids or bases to form stoichiometrical compounds is complicated by colloidal adsorption reactions.*

CHAPTER 18

Protein Denaturation [1]

A comprehensive and excellent review of protein denaturation has been published by Neurath, Greenstein, Putnam, and Erickson.[2] The more chemical aspects of the problem have been reviewed by Anson.[3] The purpose of the present short chapter is to outline the subject without attempting to cover it in detail.

In dealing with proteins one is constantly impressed with the ease with which proteins undergo changes. The rather ill-defined term of denaturation is applied to changes in the native protein brought about by relatively mild agents. Neurath, *et al.*,[2] have defined protein denaturation as *any non-proteolytic modification of the unique structure of native proteins, giving rise to definite changes in chemical, physical, or biological properties.* It is understood that denaturation does not include the addition of protons to, or their removal from, the native protein.

The usual changes which native proteins undergo when exposed to denaturing agents are: (1) decreased solubility; (2) increased digestibility by proteolytic enzymes; (3) exposure of oxidizing and reducing groups, notably the sulfhydryl (—SH) groups; (4) loss of enzymatic properties if the protein is an enzyme; (5) modification of the specific immunological properties; and (6) decreased diffusion coefficient and increased intrinsic viscosity of the protein. Proteins undergoing denaturation do not always show all the changes listed above. There is, however, a definite tendency for these changes to be associated together.

The most generally accepted theory of protein denaturation is that of Wu,[4] who proposed that the phenomenon be looked upon as a change from the unique and specific structure of the native protein to the much more randomly arranged denatured form. The peptide chains in the

[1] This chapter contributed by Dr. H. B. Bull, Professor of Chemistry in Northwestern University Medical School; formerly on the staff of the Division of Agricultural Biochemistry, University of Minnesota.

[2] H. Neurath, J. P. Greenstein, F. W. Putnam, and J. O. Erickson, *Chem. Revs.*, **34**, 157 (1944).

[3] M. L. Anson, *Advances in Protein Chem.*, **2**, 361 (1945).

[4] H. Wu, *Chinese J. Physiol.*, **5**, 321 (1931); see also A. E. Mirsky and L. Pauling, *Proc. Natl. Acad. Sci.*, **22**, 439 (1936).

native molecule exist in a highly ordered structure, and when the protein molecule is denatured this precise folding of the peptide chains is destroyed. An analogy is the conversion of a crystalline material to an amorphous one. As a broad picture little objection can be raised against this interpretation of protein denaturation, and indeed the theory of Wu could in a sense serve as a definition of denaturation.

Denaturation can be brought about by a variety of agents such as heat, concentrated acids and bases, alcohol, urea, guanidine salts, alkyl sulfates, certain hydrotropic ions, surface action, high pressures, x-rays and ultraviolet radiation.

Denaturation by Heat. Undoubtedly, the action of heat on proteins has been studied in greater detail than any other denaturing agent. Characteristically, when a protein solution at or near its isoelectric point is heated, the protein coagulates. Usually the temperature at which coagulation becomes visible is almost as sharp as a melting point. Actually, the sharpness of the coagulation point is something of an illusion because the temperature at which this change occurs is dependent on the length of exposure of the protein to the elevated temperatures. Few proteins will stand temperatures above 60°C. for any length of time without undergoing considerable denaturation.

The sequence of changes associated with heat coagulation can, under appropriate conditions, be made especially clear cut and be shown to consist of three stepwise reactions. The first is that of *denaturation* proper, which is an intramolecular rearrangement whereby certain chemical groups which were not detectable in the native protein are rendered so in the denatured product. Denaturation is necessary, but not sufficient, for the production of a coagulated protein. The second reaction involves *flocculation* of the denatured protein. The third and last step, *coagulation*, results in the formation of an insoluble coagulum due to polymerization of the denatured protein molecules.

It is a very simple matter to demonstrate these three reactions. If a solution of egg albumin in a salt-free medium is shifted from its isoelectric condition by the addition of acid or base and heated, the solution remains perfectly clear. The thiol groups have become exposed, as can be shown by the nitroprusside test, and there is considerable increase in the viscosity of the solution. The protein has been denatured. If the clear solution of heat-denatured protein is brought to its isoelectric point, flocculation occurs. The flocculum is, however, completely soluble upon the addition of acid or base. But if one waits too long before adding acid or base or if the flocculated protein is heated, the flocculum is found to be so insoluble that nothing short of hydrolysis of the protein will render it soluble. Coagulation has taken place.

Beginning with the work of Chick and Martin [5] there have been a number of studies on the kinetics of heat denaturation. In general, these studies seem to indicate that heat denaturation follows the kinetics of a first-order reaction. If the change of the velocity constant of the reaction with increasing temperature is known, it is possible to apply the Arrhenius equation for calculating the energy of activation required for the denaturation reaction to occur. As a consequence of the very large temperature coefficient of denaturation, very great energies of activation (in the order of 100,000 calories per mole) are calculated. Steinhardt [6] made a very careful study of heat denaturation of pepsin. He concluded that of the 63,500 calories of total energy of activation, a large fraction, 45,200 calories, was needed to ionize 5 protons from the pepsin molecule. Accordingly, the only form of pepsin which will heat denature has to have a negative charge on it. The rest of the energy of activation (18,300 calories) is involved in activating some non-electrostatic bonds in the pepsin molecule. Although no other protein has received the careful attention that pepsin has, there is every reason to believe that other proteins would behave in a similar manner.

The studies on the energy of activation of heat denaturation of proteins lead to the conclusion that there are two types of linkages broken upon denaturation: (1) linkages which are *p*H sensitive and (2) linkages which are largely insensitive to *p*H changes. The first type of linkage is undoubtedly electrostatic in nature (salt linkages) and arises from interaction between charged groups, such as carboxyl and amino groups, in the protein molecule. The second type of linkage broken is probably hydrogen bonds between neighboring peptide chains. The breaking of these bonds results in the displacement of the peptide chains relative to each other.

Subsequent to the disruption of the arrangement of the peptide chains in heat denaturation there results a natural tendency for the denatured molecules to polymerize. Cooper and Neurath,[7] in a study of the influence of heat treatment on solutions of crystalline horse serum albumin, concluded that heating tends to produce polymerization of the protein even though the heating was conducted at such a *p*H as to obtain a completely soluble product. MacPherson, Heidelberger, and Moore [8] have shown that egg albumin in a salt-free medium at *p*H 6.8 and 7.3 yields

[5] H. Chick and C. J. Martin, *J. Physiol.*, **40**, 404 (1910); **43**, 1 (1911–12); **45**, 61, 261 (1912–13).

[6] J. Steinhardt, *Kgl. Danske Videnskab. Selskabs. Mat. fys. Medd.*, **14**, 11 (1937).

[7] G. R. Cooper and H. Neurath, *J. Phys. Chem.*, **47**, 383 (1943).

[8] C. F. C. MacPherson, M. Heidelberger, and D. H. Moore, *J. Am. Chem. Soc.*, **67**, 578 (1945).

clear solutions when heated. However, on the basis of diffusion and ultracentrifugal measurements they concluded that the egg albumin had undergone extensive polymerization and had a molecular weight of about 900,000. Briggs and Hull,[9] investigating the heat denaturation of β-lactoglobulin at *p*H 7, found that they could separate the polymerization into two reactions. The first reaction was initiated at temperatures above 65°C. and resulted in a polymer containing four molecules of β-lactoglobulin per particle. The second process, which can take place only after the first reaction has occurred and proceeds at temperatures below 65°C., involves extensive polymerization and gives rise to particles of very large molecular weight.

It is not clear why the extent of polymerization should be limited at intermediate *p*H values since, as we have noted, in the isoelectric condition the polymerization reaction goes to completion and a coagulum results.

In the light of the above results, the increase in viscosity and decrease in diffusion constants following denaturation can be understood. Polymerization undoubtedly accounts for most of the effects noted.

When a protein is denatured by heat it becomes possible to detect sulfhydryl (—SH) groups in the denatured product.[10] In some proteins, such as the crystalline lens protein and myosin, a certain amount of sulfhydryl can be detected in the native protein, and on denaturation there is an increase in the amount which can be detected. Other proteins, such as egg albumin, have no detectable sulfhydryl in the native condition, and these groups make their appearance upon denaturation.

The classic test for the sulfhydryl group is the red color which it gives with sodium nitroprusside in an alkaline solution. Frequently the technic used to estimate the sulfhydryl quantitatively is to titrate the denatured protein with an appropriate oxidizing agent such as potassium ferricyanide, using sodium nitroprusside as an indicator to detect the end point.

In egg albumin there is a marked tendency for the sulfhydryl groups of heat-denatured protein to be considerably less than the sulfhydryl content observed in the denaturation by urea or by guanidine salts.[11] Anson,[12] however, reports that if excessive quantities of the oxidizing reagents are used the amount of sulfhydryl detected in heat-denatured egg albumin is the same as when this protein is denatured by other agents.

[9] D. R. Briggs and R. Hull, *J. Am. Chem. Soc.*, **67**, 2007 (1945).
[10] A. E. Mirsky and M. L. Anson, *J. Gen. Physiol.*, **18**, 307 (1935).
[11] J. P. Greenstein, *J. Biol. Chem.*, **125**, 501 (1938).
[12] M. L. Anson, *J. Gen. Physiol.*, **25**, 355 (1942).

Other groups, such as disulfide and tyrosine, likewise become detectable upon denaturation.

Undoubtedly, the increased digestibility of denatured protein by proteolytic enzymes, as contrasted with the relative if not absolute indigestibility of native proteins, is a reflection of the appearance of "detectable" peptide bonds in the denatured protein.

Attempts have been made to explain the liberation of sulfhydryl as well as the other groups upon denaturation, but no theory is entirely adequate. In general, there are two ways of looking at this problem. According to one theory, the sulfhydryl and the other groups considered above are involved in some kind of chemical bonding in the native molecule; upon denaturation this bonding is ruptured, and the sulfhydryl groups become liberated and detectable. The other theory holds that, although these groups are not bonded, they are inaccessible by reason of their position in the native molecule, *i.e.*, they are buried in the interior of the protein molecule, and only upon denaturation with the consequent disruption of the internal structure of the protein molecule do they become reactive and detectable. Evidence seems to favor the last point of view.

Of considerable interest is the observation by Luck and co-workers [13] that human and bovine serum albumins are stabilized against heat denaturation, as well as against the action of urea, by the anions of fatty acids. In general, the longer the hydrocarbon chain of the fatty acid, the greater is the protection conferred on the serum albumin. The actual concentrations needed for protection are small, most of the protection being achieved by 0.1 *M* solutions of the fatty acids in 25 per cent solutions of the albumin. Many derivatives of carboxylic, sulfonic, sulfuric, and phosphoric acids as well as other types of compounds were tested. The essential feature for protection appears to be the presence of a large non-polar (hydrocarbon) group and a strongly polar group combined in the same molecule. It appears that the polar group must be anionic in nature, as cationic detergents render the serum albumin more susceptible to denaturation.

The stabilizing action of the anions of the fatty acids and related compounds on serum albumin probably can be visualized this way: Undoubtedly, the anionic group of the fatty acid attaches itself to the positively charged groups on the surface of the protein molecule. The hydrocarbon tails of the fatty acid molecules extend outward into the

[13] G. A. Ballou, P. D. Boyer, J. M. Luck, and F. G. Lum, *J. Biol. Chem.*, **153**, 589 (1944); P. D. Boyer, F. G. Lum, G. A. Ballou, J. M. Luck and R. G. Rice, *J. Biol. Chem.*, **162**, 181 (1946); P. D. Boyer, G. A. Ballou, and J. M. Luck, *J. Biol. Chem.*, **162**, 199 (1946).

aqueous solution. These tails pack together on the surface of the protein molecule, and any disruption of the arrangement of the peptide chains within the protein molecule must involve the separation of the tails of the fatty acids from each other. The arrangement of the peptide chains is thus stabilized by the energy of interaction of the hydrocarbon tails; the longer the fatty acid tails, the greater the energy of interaction and the greater the stabilization effect. This speaks for a very regular and close packing of the charged groups on the surface of the serum albumin molecule. Some proteins, such as serum globulin, are not stabilized by fatty acids. It is probable in these proteins that the packing of the charged groups on the surface of the protein molecule is not sufficiently close or regular to permit the interaction of the hydrocarbon tails to take place.

Johnson and Campbell [14] have reported that hydrostatic pressures in the range of 10,000 lb. per sq. in. inhibit the precipitation of human serum globulin as well as of egg albumin at 65°C. An increase in volume during the course of denaturation of several proteins was observed by Heymann [15] and, accordingly, it can be understood why hydrostatic pressure would tend to prevent heat denaturation of a protein. On the other hand, several workers have found that pressures of around 100,000 lb. per sq. in. will cause proteins to denature at room temperature.[16-18]

Urea Denaturation. Hopkins [19] was the first to describe the denaturation of proteins by urea and to report that urea in sufficient concentrations renders the sulfhydryl groups detectable in egg albumin. To produce denaturation the concentration of urea must be quite large, *i.e.*, in excess of about 4 *M*. In general, urea produces a marked *increase in the intrinsic viscosity* of a protein. Another common effect of urea is to *induce protein molecules to split.* For example, the molecular weight of urea-denatured horse hemoglobin is one-half that of the native protein, whereas edestin, excelsin, and amandin have their molecular weights decreased to about one-sixth and tobacco mosaic virus to about one-twelfth that of the native protein by the action of urea. Urea produces no change in the molecular weight of egg albumin, serum albumin, gliadin, and pepsin.

Urea is an excellent *solvent* for proteins, and its action appears to be

[14] F. H. Johnson and D. H. Campbell, *J. Cellular Comp. Physiol.*, **26**, 43 (1945).

[15] E. Heymann, *Biochem. J.*, **30**, 127 (1936).

[16] P. W. Bridgman and J. B. Conant, *Proc. Natl. Acad. Sci.*, **15**, 680 (1929).

[17] R. B. Dow and J. E. Matthews, Jr., *Phil. Mag.*, **27**, 637 (1939).

[18] R. B. Dow, J. E. Matthews, Jr., and W. T. S. Thorp, *Am. J. Physiol.*, **131**, 382 (1940).

[19] F. G. Hopkins, *Nature*, **126**, 328, 383 (1930).

one of dispersal. The peptide chains tend to be pushed apart, and the protein molecule unfolds, resulting in a large increase in the intrinsic viscosity and exposure of sulfhydryl groups and a splitting of the molecule if the peptide chains are not attached to each other through covalent bonds. There is a question, however, whether the unfolding of the protein gives rise to a more or to a less asymmetric structure. Although it is undoubtedly true that urea-denatured protein can be pulled into extended fibers by applying the appropriate mechanical stress, this does not mean that the denatured protein in solution exists in fiber form. An extended peptide chain is an improbable configuration; the natural tendency is for the peptide chains of the denatured protein to assume a random arrangement. A loosely and randomly folded arrangement is the more probable structure.

The action on proteins of guanidine salts, as well as of certain hydrotropic ions such as thiocyanate, resembles very closely that of urea.

It has been claimed by several workers that it is possible to reverse protein denaturation and to recover the native protein. For example, if an acid solution of trypsin is heated to denature the protein, upon cooling all the trypsin can be recovered in an active native condition.[20] Reversal of the denaturation of serum albumin proteins has been reported.

Neurath and co-workers[21] have studied the denaturation of serum globulin by urea and by guanidine hydrochloride. It was found that when the urea or the guanidine was removed by dialysis a certain fraction of these proteins returned to a condition which resembled rather closely that of the native protein as far as the viscosity, solubility, etc., were concerned. The properties of the reversed protein were not, however, identical with those of the native protein. For example, the reversed protein shows a greater degree of digestibility by trypsin than does the native.[22] The viscosity and the solubilities likewise are not identical with those of the native protein. The question of the complete reversal of denaturation is still unsettled although it is true that the protein molecules can be made to return to a condition which certainly resembles very closely that of the native state.

Denaturation by Synthetic Detergents. There are a great many synthetic detergents on the market. Typically, these are characterized by having a strongly polar group such as sulfate attached to a large alkyl or aromatic residue.

[20] J. H. Northrop, *J. Gen. Physiol.*, **16**, 333 (1932).

[21] H. Neurath, G. R. Cooper, and J. O. Erickson, *J. Biol. Chem.*, **142**, 249 (1942); **142**, 265 (1942); *J. Phys. Chem.*, **46**, 203 (1942).

[22] F. Bernheim, H. Neurath, and J. O. Erickson, *J. Biol. Chem.*, **144**, 259 (1942).

Bull and Neurath [23] were the first to report that an alkyl sulfate, such as lauryl sulfate, in very low concentrations will precipitate isoelectric egg albumin. Anson [24] showed that such detergents cause the exposure of sulfhydryl groups in egg albumin. The action of synthetic detergents has since been studied by a number of workers.[25] In general, when a synthetic sulfated detergent is added to an isoelectric protein solution, a precipitation of the protein results; the precipitated protein will dissolve in an excess of the detergent.

An especially interesting feature about the interaction between detergents and proteins is that complexes containing definite stoichiometric ratios of detergent to protein can be recognized. Both serum albumin and egg albumin yield two complexes each. The first complex corresponds to a ratio of detergent to protein which titrates one-half the total positive groups of the protein molecule by the anionic detergent, lauryl sulfate. The second complex involves the titration of all the positive groups. If more detergent is added, it is bound to the protein to form a soluble product with an indefinite number of detergent molecules.

Surface Denaturation. Ramsden [26] showed that, when solutions of egg albumin, egg globulin, vitellin, serum albumin, serum globulin, fibrinogen, lactalbumin, myosinogen, and potato proteins are shaken, insoluble precipitates of these proteins are formed. He suggested that the coagula are due to the rolling or breaking up of the protein film at the air:solution surface. He was also able to show that the process is independent of the nature of the gas present.

The surface denaturation of egg albumin, oxyhemoglobin, and methemoglobin was studied in some detail by Wu and Ling,[27] who shook solutions of these proteins in glass bottles in a shaking machine. They found the rate of surface coagulation to be independent of time and concentration but strongly influenced by the rate of shaking and by the size of the bottles. Mirsky [28] has shown that sulfhydryl groups can be detected in surface-denatured egg albumin.

Surface denaturation of protein can in some respects be more con-

[23] H. B. Bull and H. Neurath, *J. Biol. Chem.*, **118**, 163 (1937).

[24] M. L. Anson, *Science*, **90**, 256 (1939).

[25] G. L. Miller and K. J. I. Anderson, *J. Biol. Chem.*, **144**, 475 (1942); H. P. Lundgren, D. W. Elam, and R. A. O'Connell, *J. Biol. Chem.*, **149**, 183 (1943); F. Putnam and H. Neurath, *J. Biol. Chem.*, **15**, 263 (1943); **159**, 195 (1945); *J. Am. Chem. Soc.*, **66**, 692, 1992 (1944); H. B. Bull, *J. Am. Chem. Soc.*, **67**, 10 (1945).

[26] W. Ramsden, *Arch. Physiol.*, p. 1517 (1894); *Proc. Roy. Soc. London*, **B72**, 156 (1903); *Z. physik. Chem.*, **47**, 336 (1904).

[27] H. Wu and S. M. Ling, *Chinese J. Physiol.*, **1**, 407 (1927).

[28] A. E. Mirsky, *J. Gen. Physiol.*, **24**, 725 (1941).

veniently studied by placing a drop of dilute protein solution on a clean water surface and allowing the protein to spread on the surface. If the experiment has been properly done, all the protein will remain on the surface, and none will go into the underlying water phase. The surface film can be compressed on a Langmuir surface trough and the film pressure measured as a function of the area of the film. Protein films exhibit remarkable similarity in dimensions. They all occupy about 0.8 sq. mm. per mg. of protein in the maximally compressed state at their critical pressures. The thickness of such films is about 10 Å. and corresponds to a thickness of one peptide chain.

Several proteins form "gaseous" films below about 1 dyne per centimeter film pressure. The molecular weight of the molecules in such a film can be calculated.[29]

In order for a substance to spread on water it must contain exposed polar and non-polar groups. In a spread film the hydrophilic groups are oriented into the water phase, the hydrophobic groups pointing upward. The lower surface of the film is thus hydrophilic and wet by water, whereas the surface exposed to the air is hydrophobic or "oily." It has been proposed by several workers that native proteins are made up of layers of peptide chains, and the films formed on a water surface probably preexist in the native molecule.[30]

[29] H. B. Bull, *J. Am. Chem. Soc.*, **67**, 4 (1945).

[30] W. T. Astbury, *Nature*, **137**, 803 (1936); L. Pauling, *J. Am. Chem. Soc.*, **62**, 2643 (1940); D. G. Dervichean, *J. Chem. Phys.*, **11**, 236 (1943); K. J. Palmer, *J. Phys. Chem.*, **48**, 12 (1944).

CHAPTER 19

Metabolism of Proteins

Man (and, as far as we know, every vertebrate) is absolutely dependent on the plant kingdom for certain of the amino acids which he needs to synthesize into the proteins characteristic of his own tissues. Whether or not all forms of animal life below the vertebrates are similarly dependent is still an open question. *These essential amino acid molecules in our bodies are derived, directly or indirectly, from plant proteins where they are produced from their inorganic constituents by the action of photosynthesis.* To be sure, some of these amino acids in our body proteins may have been derived from the proteins of animals used by us as food, but, in the last analysis, they were obtained originally from plants.

Essential Amino Acids. In a series of brilliant researches Rose [1] showed by feeding known mixtures of amino acids that *ten amino acids derived from food sources are necessary for growth of the white rat. These ten amino acids are lysine, valine, tryptophan, histidine, phenylalanine, leucine, isoleucine, threonine, arginine, and methionine.* The omission of any one of these amino acids, except arginine, caused failure in growth and finally death. When arginine was omitted, however, growth was suboptimal, suggesting some, but inadequate, synthesis by the growing rat. All other amino acids were found to be unessential under his feeding conditions.

Other investigators have subsequently claimed that arginine [2] and lysine, leucine, phenylalanine, and histidine [3] are not required for maintenance of nitrogen equilibrium in the adult rat. The dog [4] requires the same amino acids as the growing rat except that arginine is non-essential for the adult. On the other hand, glycine, which is readily synthesized by the rat, is needed by the growing chick for creatine synthesis.[5]

[1] W. C. Rose, *Physiol. Revs.*, **18**, 109 (1938).

[2] R. C. Corley, P. A. Wolf, and E. K. Nielsen, *J. Biol. Chem.*, **123**, xxvi (1938).

[3] E. W. Burroughs, H. S. Burroughs, and H. H. Mitchell, *J. Nutrition*, **19**, 363 (1940).

[4] W. C. Rose and E. E. Rice, *Science*, **90**, 186 (1939).

[5] H. J. Almquist and E. Mecchi, *J. Biol. Chem.*, **135**, 355 (1940).

In recent years Rose [6] has attacked the much more difficult problem of the amino acid requirements of man. Using mixtures of pure amino acids as the sole source of dietary nitrogen, he has demonstrated that young adult men can be maintained in nitrogen balance on the ten amino acids essential for the growing rat. Furthermore, *histidine and arginine could be removed without upsetting this balance.* It should be recognized, however, that he was observing but one criterion and for intervals of only a few days, so one cannot state with certainty that the eight remaining amino acids (lysine, valine, tryptophan, leucine, isoleucine, phenylalanine, threonine and methionine) would sustain reproduction, lactation, and a long life span. In fact it has been claimed [7] that, after nine days on an arginine-deficient diet, the number of spermatozoa in the seminal fluid of young men was reduced to one-tenth the normal level, indicating that atrophy of the spermatogenic tissue may have been supplying arginine needed to maintain nitrogen equilibrium.

Additional studies are under way in Rose's laboratory to ascertain the amount of each amino acid which must be present in the diet to sustain nitrogen balance in man.

The utter dependence of the vertebrates on the plant kingdom perhaps cannot be better expressed than by the statement that the vertebrate cannot even attach an $-NH_2$ group to the ϵ-carbon atom of α-amino caproic acid to form α-ϵ-diaminocaproic acid or lysine, in spite of the fact that the animals were starving to death because there was no lysine in the diet and were deliberately given [8] an adequate amount of α-amino caproic acid. Similar experiments have been conducted with other synthetic materials with essentially similar results, except that in some instances closely allied derivatives having the same carbon skeleton can replace an "essential" amino acid. Thus DL-β-4-imidazolelactic acid [9] and imadazole pyruvic acid,[10] when fed with diets deficient in histidine, caused a resumption of growth, indicating that an amino group had apparently been inserted in place of the α-hydroxy or keto groups. Methylimidazole, hydroxymethylimidazole, imidazolecarboxylic acid, β-imidazolepropionic acid, and β-imidazoleacrylic acid, however, were ineffective.

Tyrosine apparently is formed in the animal organism from phenylalanine, but the animal organism cannot form phenylalanine from tyro-

[6] W. C. Rose, *et al.*, *J. Biol. Chem.*, **146**, 683 (1942); **148**, 457 (1943); also personal communication (1947).

[7] L. E. Holt, Jr., A. A. Albanese, L. B. Shettles, C. Kajdi, and D. M. Wangerin, *Federation Proc.*, **1**, 116 (1942).

[8] H. B. Lewis and L. E. Root, *J. Biol. Chem.*, **43**, 79 (1920).

[9] G. J. Cox and W. C. Rose, *J. Biol. Chem.*, **68**, 781 (1926).

[10] B. Harrow and C. P. Sherwin, *J. Biol. Chem.*, **70**, 683 (1926).

sine. Cystine, methionine, and homocystine (the next higher homolog, prepared synthetically by removing the methyl group from the sulfur of methionine, after which two molecules of the homocysteine combine to form homocystine), can apparently be used more or less interchangeably as one of the essential amino acids, although a minimal amount of methionine must be present and cannot be replaced by cystine. The organic chemistry involved in the problems of the interchangeability of these three amino acids is discussed later in this chapter.

The modern ideas of protein nutrition are vastly different from those annunciated by Liebig. The old idea involved the presence of an ample supply of protein, carbohydrate, fat, and mineral matter. Later, physiologists introduced the qualification that a minimal amount of energy was required, and accordingly the calorific equivalent of the diet was added. The modern viewpoint, largely brought about by the pioneer researches of Osborne and Mendel, and McCollum and his co-workers, has shown that *protein quantity* is not an adequate criterion of diet, and that *protein quality* must be taken into consideration, *i.e.*, the question must be answered, "Does the protein which is eaten contain the essential amino acids in adequate amounts for normal growth and maintenance?" To use Dr. McCollum's illustration, if the body needs "Peter-Piper-picked-a-peck-of-pickled-peppers" protein, it will not secure normal nutrition but will starve to death on a "Theosophilus-Thistle-the-successful-thistle-sifter" protein.

We cannot go into the role played by the individual amino acids, or into the experiments which prove them to be essential. Any modern textbook on nutrition will be found to cover these points. All that can be emphasized is the striking specificity of the biological processes involved both in growth and maintenance, as illustrated by the fact that, if one essential amino acid is lacking from the diet, the entire growth mechanism may break down completely and the animal may starve to death.

Digestion and Absorption. The whole viewpoint of the mechanism of protein digestion and absorption within the animal body has changed within the last three or four decades almost as radically as have the ideas of what constitutes normal nutrition. The older workers supported the view that the proteins were digested by pepsin and hydrochloric acid in the stomach to proteoses and peptones, and that these were further digested to peptides and amino acids by the trypsin and "erepsin" of the pancreas and small intestine. It was then believed that the peptides were absorbed, as such, and were promptly resynthesized into protein in the intestinal wall, passing into and being transported by the blood stream as proteins.

Van Slyke and Meyer [11] reported that *amino nitrogen* is present in the non-protein nitrogen fraction of the blood, and that the non-protein amino nitrogen in the blood increases with a meat diet, thus suggesting that free amino acids are transported, as such, in the blood stream. However, the presence of amino nitrogen does not definitely prove the presence of the simple amino acids, for it may have been derived from relatively simple peptides or from other compounds.

We have already noted the vividiffusion apparatus of Abel (p. 57), and, so far as the author is aware, Abel should be credited with the first absolute proof that amino acids are transported, as such, in the blood stream. Shortly after the preliminary publication of Abel,[12] Abderhalden [13] reported the isolation of a considerable quantity of amino acids from blood, thus confirming Abel's observations by an independent method. These series of observations changed the earlier theories to the modern belief that the amino acids are transported as such in the blood stream to the various cells and tissues.

Gastric digestion by pepsin reduces proteins only to the stage of proteoses and peptones. Probably no simple amino acids are formed by gastric digestion, and no appreciable absorption takes place from the stomach. We can, therefore, look upon gastric digestion of protein as a preliminary hydrolysis and hydration, rendering the split products more susceptible to enzyme action after the food has passed into the duodenum.

In intestinal digestion, the proteins are attacked by trypsin, chymotrypsin, and carboxypeptidase in the uppermost portion of the intestines and by ereptic enzymes (peptidases) somewhat lower down. The tryptic enzymes degrade the proteoses and peptones to polypeptides, which are then broken down to amino acids by the peptidases. The amino acids then diffuse through the intestinal wall into the blood stream and are transported to the various cells and tissues of the body.

At any one time there is an extremely small concentration of amino acids in the blood stream. It has been calculated that at the height of digestion the rate of blood flow through the portal vein in a 9.5-kg. dog is approximately 9,000 ml. per hr., or at a rate of 150 ml. per min. Protein is only slowly digested and absorbed. Pflüger has stated that the absorption of 1.14 grams of protein per kilogram of body weight per hour is a good absorption for a human being. Accordingly if such absorption of protein were to take place in the dog noted above, the con-

[11] D. D. Van Slyke and G. M. Meyer, *J. Biol. Chem.*, **12**, 399 (1912).

[12] J. J. Abel, L. G. Rowntree, and B. B. Turner, *Trans. Assoc. Am. Physicians*, **28**, 51 (1913).

[13] E. Abderhalden, *Z. physiol. Chem.*, **88**, 478 (1913).

centration of amino acids in the blood would not exceed 0.12 per cent. It is not surprising, therefore, that, by the earlier and somewhat crude methods, the presence of amino acids in the blood stream was overlooked and that only the more refined technic of the vividiffusion or the Van Slyke apparatus was able to detect the presence of appreciable quantities of amino acids. Abderhalden, in his isolation of amino acids from the blood stream, worked up approximately 100 liters of blood and identified proline, leucine, valine, aspartic and glutamic acids, glycine, arginine, histidine, and lysine. No amino acid was found in amounts exceeding 0.40 gram.

It should be noted at this point that the earlier workers were justified in concluding that amino acids were not present, partly because of the inadequacy of their methods and partly because certain other nitrogenous compounds, *e.g.*, urea, creatine, and uric acid, are always present in the blood serum from which the proteins have been removed by appropriate technic. Some of these constituents occur in much larger quantities than the amino acids, and accordingly, unless very delicate technics are employed, it is impossible to identify such traces of amino acids as are normally present during the absorption of protein.

It is perhaps pertinent to ask whether or not polypeptides, proteoses, peptones, and perhaps even proteins may not at times pass from the digestive tract into the circulation. The author believes that this is not only a possibility but a probability. Certain individuals, particularly infants, develop pathological conditions which can be traced to the inclusion of some particular protein in the diet, in that they become exceedingly hypersensitive to some of these proteins. Egg albumin appears to be one of the proteins most commonly producing such hypersensitivity, although many cases have been noted where the proteins of the legumes, particularly peas, are involved. We have already noted that the egg albumin molecules are so small that they will dialyze slowly through a fairly porous collodion membrane. It seems probable, therefore, that conditions may arise whereby the membranes separating the intestinal contents from the blood stream become sufficiently permeable to permit the passage of small amounts of foreign proteins, thus giving rise to protein intoxication. Aside from its pathological significance this does not appear to be an important problem in protein assimilation.

Dynamic State of Body Proteins. No adequate theory has been advanced to account for the synthesis of the vital proteins in the various cells and tissues from the heterogeneous mixture of amino acids which are transported in the blood stream. Whether peptide-bond synthesis by proteolytic enzymes is the main pathway is uncertain, since known conditions in the tissues seem to favor hydrolysis. We know only that

the vital process involved is one of rigid selection and that a given cell or tissue invariably synthesizes proteins containing certain specific amino acids linked in a certain specific way. We have no clue to the mechanism whereby this selective action is brought about.

In recent years, however, convincing evidence has been put forward which has modified older concepts of the relatively static nature of organ proteins. In the main, two types of biochemical "tools" have been used: (*a*) the technic of *plasmapheresis*, mainly by Whipple and his associates, and (*b*) the use of *isotopes* of C, N, and S as tracer substances for studying protein metabolism.

Continuous Interchange of Plasma and Tissue Proteins. In an extensive series of investigations on the origin and utilization of plasma protein, Whipple and co-workers [14] have demonstrated that it is an integral part of a "*balanced system of body proteins.*" By maintaining dogs on a protein-free or low-protein basal diet and repeatedly withdrawing about one-fourth the total blood and reinjecting the washed red corpuscles suspended in physiological salt solution (a technic termed *plasmapheresis*) the plasma protein can be gradually depleted. Then the effects of fasting, the feeding of various proteins, or the parenteral administration of proteins, protein hydrolysates, or amino acid mixtures can be assessed in relation to their effects on plasma protein regeneration.

Such studies have shown conclusively that the normal dog has appreciable *reserve stores* of plasma protein-building materials in his tissues, probably sufficient to replenish one or two times the amount of plasma protein normally present in the circulating blood. The amount of such storage materials varies widely between individual animals and is largely influenced by the preceding nutritional history. Thus in ten dogs weighing 10–14 kg. this protein reserve amounted to 24 to 68 grams. Two to six weeks of plasmapheresis are required to deplete these stores and bring the animals to a steady state of hypoproteinemia; after such depletion the body can provide very little new plasma protein (about 2 grams per week).

These reserve protein stores differ from those of carbohydrate and fat in that they are not physically demonstrable entities such as one observes in the intracellular deposits of glycogen and fat. The liver, kidneys, and intestines are important storage sites, but a great mass is also stored in the carcass tissues. The liver of a rat, for example, may lose 40 per cent of its total protein during a week of fasting by the animal,[15] yet careful chemical analysis of the liver protein before and after

[14] For good reviews of the literature see: S. C. Madden and G. H. Whipple, *Physiol. Revs.*, **20**, 194 (1940); also G. H. Whipple, *Am. J. Med. Sci.*, **203**, 477 (1942).

[15] T. Addis, L. J. Poo, and W. Lew, *J. Biol. Chem.*, **116**, 343 (1936).

storage showed no differences in composition.[16] Apparently little if any of this reserve protein is stored as plasma protein or in a form more readily utilizable than orally ingested protein.

It is apparent also that the storage protein is not converted into plasma protein at a constant rate. During a seven-day fast, half of the liver protein which is lost by the rat disappears in the first two days,[17] and after severe rapid depletion of plasma protein (by plasmapheresis) the fasting dog regenerates new plasma protein much more rapidly during the first few hours than subsequently. This suggests that the reserve protein consists of a *labile* portion which is mobilized quickly and

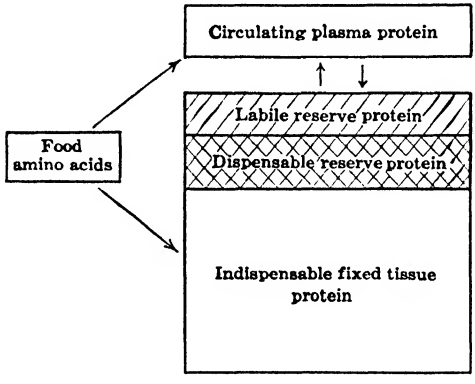


FIG. 94. Diagrammatic representation of the equilibrium between the circulating plasma protein and the reserve stores of tissue protein.

a further *dispensable* portion which is made available more slowly. Both of these portions may be depleted without interfering with body function; under such conditions animals can be kept hypoproteinemic for relatively long periods and be normal in all other respects. Aside from the reserve store there is the larger mass of indispensable or "fixed" tissue protein which cannot be withdrawn without injury to the tissue. The plasma protein:tissue protein equilibrium may be represented schematically as in Fig. 94.

A most amazing observation[18] was that *dogs on a protein-free diet could be kept in nitrogen balance for considerable periods, and probably indefinitely, by the intravenous administration of plasma protein as such.* The injected protein rapidly left the blood and entered the tissues.

[16] J. M. Luck, *J. Biol. Chem.*, **115**, 491 (1936).

[17] T. Addis, L. J. Poo, and W. Lew, *J. Biol. Chem.*, **115**, 117 (1936).

[18] F. S. Daft, F. S. Robscheit-Robbins, and G. H. Whipple, *J. Biol. Chem.*, **123**, 87 (1938).

Howland and Hawkins,[19] using phlorizinized dogs, demonstrated that after such injections neither protein nor sugar appeared in the urine. When such dogs were *fed* plasma protein, sugar was eliminated. It is clear, then, that *the metabolism of the protein by vein is different from that of protein by mouth.* The evidence strongly suggests that *the injected protein passed only through a large aggregate stage, rather than through amino acid stages, before incorporation into the tissues.*

Madden and Whipple [14] have summarized this new concept of plasma protein metabolism in relation to tissue metabolism as follows:

> The ultimate source of construction materials is food protein furnishing amino acids absorbed from the intestinal tract and synthesized in the liver cells (and elsewhere) into plasma proteins. This same influx of amino acids, and/or the plasma proteins, support protein formation throughout the body. A part of the body protein forms a *reserve against adversity* in the sense that it can be circumspectly depleted without apparent injury to the body. The supplies of amino acids coming from outside the body and the demand for protein materials within the body are in a constant state of balance with these protein reserve stores, a *dynamic equilibrium.*
>
> If the body need for protein to build new cell protein, new plasma protein, or new hemoglobin is greater than the exogenous supply, this reserve store of protein will be drawn upon; if the body need is less, the store will be repleted. The maximum and minimum limits of these stores are an individual characteristic of each organ, tissue, and fluid, and are determined by factors not yet understood.
>
> When the body demands protein the plasma contributes its share, and evidence from plasma injection experiments indicates that plasma protein molecules can be accepted as such by body cells and recast into specific cell protein without loss of nitrogen.
>
> It seems to us probable that new plasma protein flows largely from the liver although some globulins may be formed elsewhere.

Transamination. In 1937 Braunstein and Kritzmann [20] upset long-established concepts of protein metabolism when they discovered that *intermolecular transfer of amino groups* occurred in various animal tissues. The following reversible reaction was found to proceed in muscle, heart, brain, liver, and kidney:

$$\text{L-glutamic acid} + \text{pyruvic acid} \rightleftharpoons \alpha\text{-ketoglutaric acid} + \text{L-alanine}$$

In this mechanism aspartic acid can replace glutamic acid, in which case oxalacetic acid takes the place of α-ketoglutaric acid in the equilibrium. Keto acids other than pyruvic, such as α-ketobutyric and oxalacetic

[19] J. W. Howland and W. B. Hawkins, *J. Biol. Chem.*, **123**, 99 (1938).

[20] A. E. Braunstein and M. G. Kritzmann, *Biokhimiya*, **2**, 242, 859 (1937); *Nature*, **140**, 503 (1937).

acid, were also able to accept the amino group of glutamic acid; correspondingly, it was demonstrated that sixteen different amino acids were able to produce glutamic acid from α-ketoglutaric acid in the presence of muscle tissue. Since the rate of reaction was very high, it appeared that this mechanism of transamination probably plays an important role in cellular metabolism.

Subsequent investigations have revealed that the enzyme *transaminase*, which catalyzes this reaction, is not only widely distributed in animal tissues, but is also present in plants [21] and microorganisms,[22] where the rate of transamination is even higher than in most animal tissues [23] (Table 49).

TABLE 49. Q_{TN} * VALUES OF *Escherichia coli*, ANIMAL, AND PLANT TISSUES

(Data of Lichstein and Cohen)

Tissue	Q_{TN}
Escherichia coli	3,900
Oat seedlings (96 hr)	5,650
Brain (rat)	2,800
Liver (rat)	2,200
Kidney (rat)	1,750
Heart muscle (rat)	3,330
Purified transaminase (from beef heart muscle)	10,300

* Substrates: L-glutamic acid plus oxalacetic acid.

$$Q_{TN} = \frac{\mu\text{l. aspartic acid formed}}{\text{mg. N} \times \text{hr.}}$$

The role played by the dicarboxylic amino acids in transamination assumes even greater importance in view of the discovery by Euler, *et al.*,[24] of a highly specific enzyme in animal tissues (principally the liver and kidneys) which reversibly catalyzes the formation of glutamic acid from ammonia and α-ketoglutaric acid. It seems highly probable, therefore, that aspartic and glutamic acids play a key role in protein metabolism.

Metabolism Studies Using Isotopes. The lability of the proteins and amino acids in plant and animal tissues has been further strikingly demonstrated through the development of procedures for preparing

[21] A. I. Virtanen and T. Laine, *Biochem. J.*, **33**, 412 (1939).
[22] E. Adler, *et al.*, *Z. physiol. Chem.*, **255**, 14, 27 (1938).
[23] H. C. Lichstein and P. P. Cohen, *J. Biol. Chem.*, **157**, 85 (1945).
[24] H. von Euler, E. Adler, G. Günther, and N. B. Das, *Z. physiol. Chem.*, **254**, 61 (1938).

compounds "labeled" with isotopes of hydrogen, carbon, nitrogen, or sulfur and tracing their paths of intermediary metabolism.[25] Thus deuterium can be incorporated into a carbon chain by hydrogenation of a double bond with deuterium gas, or sometimes by treating a compound with hot concentrated D_2SO_4. The stable heavy isotope of nitrogen, N^{15}, can be introduced into an amino acid by catalytic reduction of the corresponding α-keto acid in the presence of $N^{15}H_3$. After administration of the "tagged" compounds to the organism, various substances can be isolated from the tissues or excreta and analyzed for their isotope content.

For precise determination of the stable heavy isotopes, use of the mass spectrometer has proved most satisfactory. In deuterium studies the substance is burned and the heavy water content determined; with N^{15} studies, the organic substance is digested by the Kjeldahl procedure, and nitrogen gas is formed by adding alkaline hypobromite to the resulting ammonium sulfate. Accurate determinations of the N^{14}:N^{15} ratio can be made with the mass spectrometer, as little as ½ cc. of gas being used. With radioactive isotopes, such as S^{35} and C^{14}, devices such as the Geiger counter are employed for analysis.

To cover adequately all the numerous recent reports on protein metabolism which have utilized this "tracer" technic would take us beyond the scope of this book. Only a short summary of some of the information which such studies have uncovered will be presented here.

It has been repeatedly demonstrated that when ammonium salts or amino acids containing N^{15} are fed to adult animals in energy and nitrogen balance, about half of the labeled nitrogen can be detected in the body proteins, the rest being excreted largely in the urine as urea and ammonia. No conclusion can be reached other than that chemical reactions of the tissue proteins occur without changes in composition or the formation of additional proteins. Furthermore, not all the introduced isotope appears in the tissues as the compound which was fed, since many other amino acids isolated from the tissues contain appreciable quantities of N^{15}. Table 50 illustrates the distribution of isotopic nitrogen in various substances isolated from a rat fed tyrosine containing N^{15} for a ten-day period.[26]

Schoenheimer, Ratner, and Rittenberg calculated that only about one fourth of the N^{15} retained in the body proteins was present as tyrosine. The liver took up roughly three times as much of the isotope

[25] R. Schoenheimer and D. Rittenberg, *Physiol. Revs.*, **20**, 218 (1940); also R. Schoenheimer, *The Dynamic State of Body Constituents*, Harvard University Press, Cambridge, Mass., 1942.

[26] R. Schoenheimer, S. Ratner, and D. Rittenberg, *J. Biol. Chem.*, **127**, 333 (1939).

as did the carcass. The isotopic nitrogen in the arginine was all in the guanido group, whereas that in histidine was exclusively in the α-amino group. Of the amino acids isolated, *only lysine failed to incorporate* N^{15}. This corresponds to results of an earlier experiment [27] in which mice received heavy water; of the ten amino acids isolated, only lysine contained no added deuterium. These data confirm the view that *only plant-synthesized lysine is utilized in building animal protein.*

TABLE 50. N^{15} CONCENTRATION IN NITROGEN OF SUBSTANCES ISOLATED FROM RAT GIVEN DL-TYROSINE N^{15}

(Data of Schoenheimer, Ratner, and Rittenberg)

Substance	N^{15} *Atom per cent excess* *
Liver	
Total protein	0.039
Tyrosine	0.316
Arginine	0.030
NH_3 from arginine	0.058
Glutamic and aspartic acid	0.056
Lysine	0.003
Histidine	0.015
Amide-N	0.044
Non-protein-N	0.038
Carcass	
Total protein	0.012
Tyrosine	0.100
Glutamic and aspartic acid	0.010
Amide-N	0.012
Non-protein-N	0.015

* Limit of error = approximately 0.003.

By incorporating both deuterium (on the carbon chain) and N^{15} into a given amino acid, it has been proved that dietary amino acids can be introduced as such into the proteins of adult, non-growing animals. This could be accomplished only by rapid, intermittent opening and closing of peptide bonds. The appearance of N^{15} in amino acids other than that fed indicates a continuous process of deamination and reamination of amino acids in the various body proteins. This concept fits well with the observations of Braunstein and Kritzmann [20] and Euler, *et al.*,[24] regarding transamination in living systems. The studies with N^{15} also lend weight to the important role of glutamic and aspartic

[27] G. L. Foster, D. Rittenberg, and R. Schoenheimer, *J. Biol. Chem.*, **125**, 13 (1938).

acids in the utilization of ammonia and the transfer of nitrogen from one carbon chain to another; when isotopic nitrogen is administered, either as an ammonium salt or in amino acid combination, the N^{15} concentration in these dicarboxylic acids is always greater than in any other amino acid except the one originally given.

Evidence has been put forth, also, which indicates that in the living organism there is a continuous conversion of carbon chains from one amino acid into another. Thus ornithine can be converted into proline, glutamic acid, and arginine [28] even when these end products are amply supplied in the diet. This has not been demonstrated, however, for any of the truly *essential* amino acids, the carbon skeletons of which apparently must be preformed by the plant.

It is clear that the once widely accepted Folin theory of exogenous and endogenous metabolism has undergone drastic revision. According to this theory the organ proteins are essentially static after growth is complete, and dietary protein is almost completely catabolized and excreted, only a small amount being utilized to replace the slow, constant tissue "wear and tear" (endogenous) losses. Today, however, one cannot escape the conclusion that the continuous, rapid metabolism of cell proteins is a characteristic of all living matter.

Transmethylation. This phenomenon, as the name suggests, involves the intermolecular transfer of methyl groups within the organism. It was first postulated by du Vigneaud, *et al.*,[29] who showed that homocystine plus either choline or betaine could replace methionine in the diet of the rat. Conversely it was ascertained, by using deuterium, that methionine is able to furnish methyl groups to choline and creatine. Borsook and Dubnoff [30] were the first to show, however, that methionine is a specific methylating agent for creatine synthesis in the liver (for details of the mechanism, see p. 496). The transmethylation process is apparently continuous and automatic.

The lability, *in vivo*, of the methyl group of methionine has provided a key to the mechanism of its conversion into cystine by the living organism. That the sulfur in methionine can be converted to cystine sulfur has been conclusively demonstrated by feeding methionine containing S^{35} to rats and isolating cystine containing radioactive sulfur from the skin and hair.[31] Also studies with serine labeled with N^{15} strongly suggest that the carbon chain of this amino acid can be con-

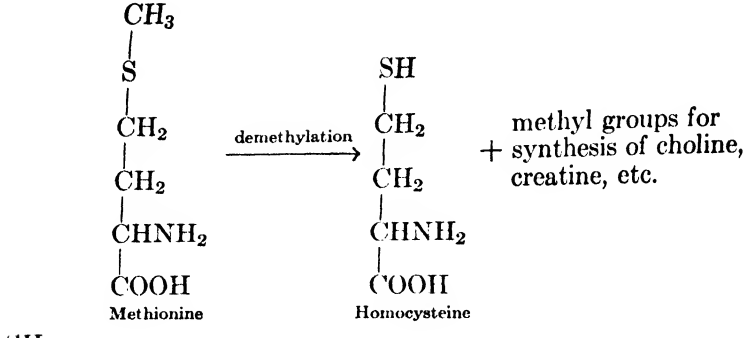
[28] M. Roloff, S. Ratner, and R. Schoenheimer, *J. Biol. Chem.*, **136**, 561 (1940); R. F. Clutton, R. Schoenheimer, and D. Rittenberg, *J. Biol. Chem.*, **132**, 227 (1940).

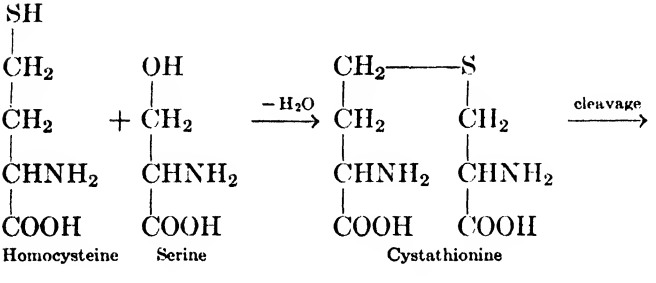
[29] V. du Vigneaud, *et al.*, *J. Biol. Chem.*, **131**, 57 (1939); **134**, 787 (1940).

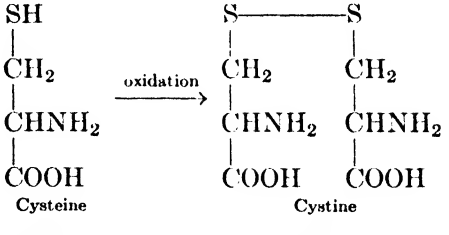
[30] H. Borsook and J. W. Dubnoff, *J. Biol. Chem.*, **132**, 559 (1940).

[31] H. Tarver and C. L. A. Schmidt, *J. Biol. Chem.*, **130**, 67 (1939).

verted to cystine.[32] The currently accepted mechanism for *in vivo* cystine synthesis from methionine may be schematically represented as follows:







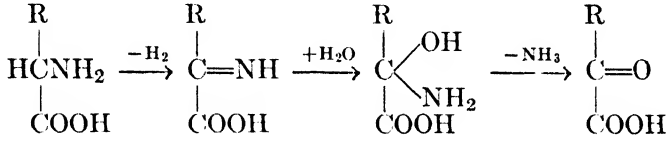
In view of the above, the "essential" nature of methionine, in contrast to cystine, is apparently due to (*a*) its importance as a source of labile methyl groups, and (*b*) the fact that cystine can be formed from methionine, but methionine cannot be formed from cystine. Some of the methionine "requirement" can be replaced by cystine, however, by sparing methionine that would otherwise have been diverted to tissue cystine formation.

The Degradation of Amino Acids in the Body. The body possesses a very efficient mechanism for maintaining a low concentration of amino acids in the blood stream even when protein digestion and absorption are at a maximum. To a considerable extent the free amino

[32] D. Stetten, *J. Biol. Chem.*, **144**, 501 (1942).

acids can be temporarily stored by the liver and other tissues and gradually incorporated into plasma or tissue proteins. On the other hand, amino acids are constantly being deaminized and broken down, chiefly in the liver, the nitrogenous portion being eliminated as urea by the kidneys.

The principal mechanism of deamination of amino acids in the animal body is an oxidative deamination giving rise to an α-ketonic acid and ammonia as end products. This mechanism, originally proposed by Knoop,[33] apparently proceeds in three stages:



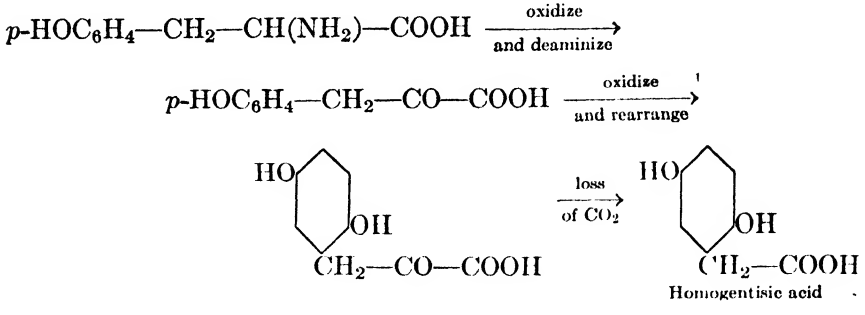
The first step is catalyzed by dehydrogenases and requires an accompanying hydrogen acceptor. Nearly all dehydrogenation reactions are accompanied by utilization of oxygen. The deamination process is no exception to this rule, for it fails under anaerobic conditions.

α-Ketonic acids corresponding to many of the amino acids have been isolated as would be required by this theory. Knoop observed that the sequence of reactions noted above was reversible and that *α-amino acids could be synthesized in the animal organism by perfusing the liver (dog) with the corresponding α-keto acid.* Thus, tyrosine and tyrosine derivatives were isolated by perfusing with defibrinated blood to which the ammonium salt of *p*-hydroxyphenylpyruvic acid had been added. Similarly alanine, phenylalanine, leucine, α-amino-*n*-butyric acid, and norleucine were prepared by perfusing with the ammonium salts of the corresponding α-keto acid. In the last two compounds, the author notes that they are not naturally occurring amino acids; since that observation both of these compounds have been reported to occur in proteins. Still more interesting, *the amino acids which were synthesized by perfusing with the ammonium salt of the α-keto acid were optically active.*

Amino acids containing aromatic nuclei apparently undergo the α-ketonic oxidation, after which the rings open, permitting complete oxidation of the residue to carbon dioxide and water. The intermediary catabolic processes differ greatly for individual amino acids. However, the phenols, which are normal constituents of the urine, and indole, skatole, etc., appear to have their origin in the benzene and indole rings of the amino acids by action of the intestinal bacteria.

[33] F. Knoop, *Z. physiol. Chem.*, **67**, 489 (1910); *cf.* also G. Embden and E. Schmitz, *Biochem. Z.*, **29**, 423 (1910); **38**, 393 (1912); K. Kondo, *ibid.*, **38**, 407 (1912).

Inborn Errors of Metabolism. Under certain conditions the mode of oxidation of certain of the amino acids appears to be altered within the animal body, tyrosine, for example, giving rise to *homogentisic acid*, probably according to the following scheme:



The homogentisic acid appears in the urine in certain diseases, such as melanouria and *alkaptonuria*, the urine being colorless when voided and rapidly darkening to an intense black liquid from which black particles are precipitated.

The presence of homogentisic acid in the urine should not be taken necessarily as an indication of a pathological condition. This mechanism for the oxidation of tyrosine and phenylalanine appears to be hereditary, is more prevalent in males than in females, and may persist throughout life without any evidence of harmful effects. An increased consumption of proteins containing tyrosine or phenylalanine will increase the amount of homogentisic acid excreted. Rather surprisingly *p*-hydroxyphenyllactic acid when ingested does not affect the amount of homogentisic acid excreted; the corresponding keto acid is excreted as homogentisic acid, indicating that hydrolytic deamination probably does not take place normally to any great extent.

In 1939 Sealock and Silberstein [34] discovered that guinea pigs deficient in vitamin C excreted homogentisic acid when fed half a gram of tyrosine daily, but that 5 grams daily of the vitamin promptly stopped this excretion. Similar results were noted with two human subjects, suggesting that vitamin C is concerned with the oxidation of the phenyl nucleus in aromatic amino acids.

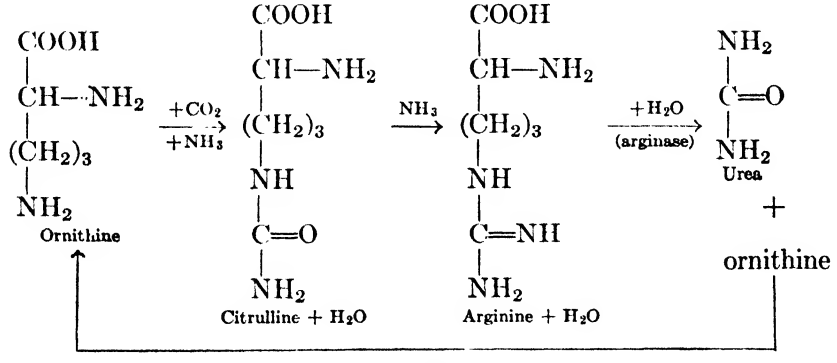
The sulfur of cystine, on oxidation in the body, yields sulfuric acid and sulfates. Again, in this instance, *cystinuria* may occur, the cystine being eliminated as such in the urine. This again need not necessarily be taken as indication of a pathological condition, inasmuch as it is

[34] R. R. Sealock and H. E. Silberstein, *Science*, **90**, 517 (1939); *J. Biol. Chem.*, **135**, 251 (1940).

hereditary, affecting males and females about equally. The only untoward symptom which may arise is the aggregation of the cystine crystals into kidney stones or urinary concretions. Cystine is so insoluble that there are a number of instances on record where such concretions have been formed. Thus, Tennant [35] reported a surgical case in which 15 stones, having a total weight of 73 grams, were removed from a kidney. These stones contained [36] 93 per cent of cystine.

Urea Formation. For many years it was assumed that urea arose in the animal body through the condensation of ammonia and carbon dioxide to form ammonium carbonate, followed by the loss of water to yield the intermediate ammonium carbamate and finally urea. An alternative hypothesis was the formation of ammonium cyanate which rearranged to form urea. The difficulty with either of these theories was that, in the mammal, urea synthesis appears to reside almost wholly in the liver, and all efforts have failed to link urease, the enzyme which decomposes urea into ammonia and carbon dioxide, with the synthetic mechanism.

In 1932, Krebs and Henseleit [37] demonstrated a new mechanism for the synthesis of urea in the mammalian liver which has received almost universal acceptance. This involves a cycle in which ornithine combines with one molecule of ammonia and one molecule of carbon dioxide with the loss of a molecule of water to form citrulline. The citrulline then combines with an additional molecule of ammonia with the loss of a molecule of water to form arginine, which is subsequently hydrolyzed by the enzyme *arginase* to form one molecule of ornithine and one molecule of urea. The ornithine which is thus regenerated is available for again starting a new cycle:



[35] C. E. Tennant, *J. Am. Med. Assoc.*, **80**, 305 (1923).
[36] R. A. Gortner and W. F. Hoffman, *Proc. Soc. Exptl. Biol. Med.*, **23**, 691 (1926).
[37] H. A. Krebs and K. Henseleit, *Z. physiol. Chem.*, **210**, 33 (1932).

Krebs and Henseleit incubated ammonium carbonate with macerated tissue from seventeen mammalian organs, and in their experiments liver was the only organ which converted the ammonium carbonate into urea. They found that both ornithine and citrulline had remarkable catalytic activity in favoring urea synthesis. Later Manderscheid [38] found that the same mechanism holds for amphibia and reptiles, so far as these animals form urea. Subsequent studies have largely confirmed this mechanism for urea synthesis *in vivo*. In birds, where the end product is uric acid, the failure of the mechanism appears to be due to the absence or limiting amounts of arginase in the liver.

The end products of protein metabolism, allantoin and urea, have taken on an added interest since the observations of Robinson.[39] He was investigating the reasons for the beneficial effects in the healing of chronic ulcers which have been cleaned by maggot therapy. During World War I it was observed that certain suppurative wounds which had become infected with blow-fly maggots healed much more rapidly than similar wounds in which the maggots were absent. This gave rise to the use of maggots in cleaning suppurative wounds, the belief being that the only function of the maggots was to clean the wound by devouring dead tissue. However, it was observed that if the maggots were removed from the wound after they had completely cleaned it, the healing process slowed down. On the other hand, if the maggots were left in contact with the cleaned wound, the healing was remarkably accelerated. Inasmuch as no secretions of the blow-fly maggot were known, Robinson investigated the excretory products. One of these he found to be allantoin, and, when wet dressings saturated with a solution of allantoin were placed in contact with the clean wound, healing was remarkably accelerated. Later, after crystallizing out the allantoin from the excretory products of the maggots, Robinson found the residual solution still to have healing activity. The residual nitrogenous constituents were largely urea, and, when it was tested as a healing agent, urea likewise was found to stimulate the healing process and to promote the development of granulation tissue. Many suppurative wounds and chronic ulcers of long standing have been healed by means of either allantoin or urea. Ammonia, secreted by surgical maggots or spontaneously produced from urea through the action of the enzyme *urease*, also acts as a cell proliferant and stimulates healing of purulent wounds.

[38] H. Manderscheid, *Biochem. Z.*, **263**, 245 (1933).

[39] W. Robinson, *J. Bone and Joint Surgery*, **17**, 267 (1935); *Am. J. Surgery*, **33**, 192 (1936); **47**, 111 (1940); **49**, 319 (1940); *cf.* also H. G. Holder and E. M. MacKay, *J. Am. Med. Assoc.*, **108**, 1167 (1937).

Macalister [40] had earlier noted the healing properties of allantoin and had suggested its use in the treatment of chronic ulcers. His studies grew out of the observation that the peasants of Europe used the macerated root of comfrey (*Symphytum officinale*) as an application for suppurative wounds, and the peasantry believed that the root of the comfrey had great healing power. Macalister made an extract of the comfrey roots and crystallized allantoin from the extract. He then tested the allantoin and demonstrated its healing properties.

Urea in the form of urine has been used as a healing agent by almost every primitive people. Lawrence of Arabia in his *Pillars of Wisdom* notes that it is generally used by the Bedouins. In the light of Robinson's studies there appears to be no question that the primitive practices had merit. Today synthetic urea can be purchased at a negligible cost, and, since the solutions are non-irritating to the most tender tissue and are non-toxic, it can be used without fear of untoward effects.

Formation of Carbohydrates and Fats from Amino Acids. Carbohydrates and fats may be formed from proteins under the normal processes of metabolism. In studies of the factors influencing sugar excretion by diabetics, it had long been suspected that protein might give rise to carbohydrates.

Lusk administered amino acids to phlorizinized dogs and found that certain of the amino acids yielded sugar, whereas others did not. Those containing 2, 3, 4, and 5 carbon atoms in a straight chain were more or less completely converted into glucose. Glycine and alanine could be quantitatively converted into glucose, whereas only three of the carbon atoms of aspartic and glutamic acids were converted into glucose, three fourths of the carbon of the aspartic acid and three fifths of the carbon of the glutamic acid appearing in the form of glucose.

Dakin has shown that serine, proline, arginine, and ornithine will yield at least a part of their carbon in the form of glucose, the arginine going first to ornithine. Lysine is the only straight-chain amino acid tested which failed to yield sugar. Apparently the aminopropionic acid side chain of tyrosine and phenylalanine is not available for glucogenesis, since glucose does not arise when these amino acids are fed.

We now recognize that common intermediates result from the intermediary metabolism of protein and carbohydrate, thus providing a rational basis for the interconversion of these foodstuffs. For example, alanine, glutamic acid, and aspartic acid on oxidative deamination yield pyruvic, α-ketoglutaric and oxalacetic acids, respectively, all of which are components of the tricarboxylic acid cycle of carbohydrate metabo-

[40] C. J. Macalister, *Brit. Med. J.*, **1**, 10 (1912).

lism (see Fig. 107, p. 693). About half of the amino acids in protein are regarded as potential carbohydrate formers.

The formation of fats from amino acids has proved a more difficult problem for investigation than the formation of sugars. There is no question, however, that fat can be formed in the animal body from protein. We know that carbohydrates are readily converted into fats; therefore at least those amino acids which can give rise to glucose in the body can likewise give rise to fats. Possibly certain other of the amino acids may give rise to fats through side reactions. Thus leucine, tyrosine, and phenylalanine, although they do not directly give rise to glucose, do appear at least in part as acetoacetic acid, which should be convertible into fatty acids.

The interrelationships between carbohydrate, fat, and protein metabolism are more fully discussed in Chapter 26 (p. 698).

Specific Dynamic Action of Proteins. One additional property of proteins and amino acids in metabolism deserves mention, *i.e.*, the phenomenon which Rubner called the specific dynamic action of proteins. It had been noted by many workers that increased metabolism, as measured in a respiration calorimeter, resulted when proteins were fed. If an amount of sugar or fat equivalent to a given number of calories is fed to a fasting dog, the heat, as measured in a respiration calorimeter, is only slightly greater (4–6 per cent) than that which would be produced by the combustion of the sugar or fat which was fed. If, however, protein having an equivalent calorific value replaces the sugar or fat, the heat, as measured in a respiration calorimeter, exceeds by approximately 30 per cent the calorific value of the protein which was fed.

This phenomenon of excess heat production following the feeding of protein and of amino acids was extensively investigated by Lusk and Benedict, who showed that the excess heat is not due to mechanical heat from increased intestinal activity or peristaltic action. Lusk suggested that the amino acids behave as cell stimulants, raising the metabolic power of the tissue cells, the excess energy being derived from carbohydrates and fats which are burned in amounts greater than the normal level. He suggested that perhaps the stimulant is the keto- or the hydroxy-acids formed as intermediate metabolic products.

Borsook [41] has reviewed critically the literature on the specific dynamic action of protein and amino acids. He concludes that the suggestion of Lusk of a cell stimulant is probably unnecessary and in any event contributes only a minor fraction to the excess calories. Borsook notes that there are two factors which account for the increase in metabolism after the ingestion of protein or amino acids. One of these factors

[41] H. Borsook, *Biol. Revs., Cambridge Phil. Soc.*, **11**, 147 (1936).

is approximately constant and includes the increased oxygen consumption necessary for the oxidative deamination of the amino acids. One molecule of oxygen is used for each molecule of nitrogen deaminized. An increase in metabolism of 4 calories per gram of nitrogen, therefore, arises from oxidative deamination. In the synthesis of urea from ammonia and carbon dioxide we again have an evolution amounting to a maximum of 4 calories per gram of nitrogen. In order to excrete urea the kidney has to do work, and this work results in the production of heat. Borsook notes that this may amount to as much as 2 calories per gram of nitrogen. Thus, simply *handling the nitrogen* causes an increase in metabolism. Borsook's second source of increased metabolism concerns the *utilization of the carbon residue* after deamination. If the carbon residue is synthesized into glycogen, there will be an evolution of heat. He points out a number of possible and probable exothermic reactions which could adequately account for the increased metabolism which has been observed, and furthermore points out that the excess heat production which has been noted by various workers is not a constant but a variable and depends on the conditions of the experiment and particularly on the physiological state of the experimental animal.

Perhaps certain of the amino acids or their metabolic residues do act as cell stimulants. Tyrosine, when injected into the blood stream, causes a very appreciable increase in metabolic rate over that observed when even larger amounts of tyrosine are fed. Similarly the ingestion of 2,4-dinitrophenol has been found to have an enormous effect on metabolic rate. This compound, administered at a 20 mg. per kg. level, will increase the metabolic rate by as much as 1,000 per cent. The amino acid thyroxine, isolated by Kendall from the protein of the thyroid gland, apparently regulates in a large measure the metabolism of the body. The merest trace of thyroxine markedly raises the metabolic rate and maintains it at a high level for a considerable period of time. Perhaps other amino acids reflect in a small measure the extraordinary effects of thyroxine.

Forbes has repeatedly put forward evidence indicating that in a normal *mixed diet* the protein content does not dominate the heat production. Recently he has reported the dynamic effects on rats of beef protein, lard, and dextrose determined separately and in various dietary combinations.[42] Whereas beef protein alone had double the dynamic action of lard, when the two were combined so that lard supplied about 44 per cent of the total calories, heat production was even less than for lard alone and only half that calculated from the values for the individual components. It seems evident, therefore, that the dynamic

[42] E. B. Forbes and R. W. Swift, *Science*, **99**, 476 (1944).

effects of mixed diets are not the additive effects of their components, and that at times fat may be more important than protein in determining the dynamic action of food mixtures.

Enzymatic Synthesis of Proteins. Bourquelot has shown that the reaction between glycosides and glycoside-splitting enzymes is a reversible reaction according to the following scheme:

$$\text{Glycoside} + \text{enzyme} \rightleftharpoons \text{enzyme} + \text{sugar} + x$$

where x represents a non-sugar radical. Accordingly, in dilute solutions the reaction is driven toward the right, resulting in the hydrolysis of the glycoside, whereas in concentrated solutions of the components plus the enzyme, the reaction is driven toward the left and the glycoside is resynthesized.

A number of workers [43] have attempted to synthesize proteins by means of a reversed enzyme reaction, adding the enzyme to a concentrated solution of protein-split products. The only synthesis which is generally accepted is that by which Taylor appears to have synthesized the protamine *salmine*.

Taylor [44] digested the protamine salmine with trypsin, converted the amino acids into carbonates, destroyed the trypsin by heat, and concentrated the amino acids from 400 grams of the protamine to the point of crystallization. He then added 300 ml. of a glycerol extract from the liver of the clam *Schizothaerus nuttalii*, added toluene as a preservative, and set the mixture aside. The mixture gradually became opalescent, then cloudy, and finally a precipitate formed. At the end of 5 months, 2 grams of salmine sulfate was isolated from the mixture. A duplicate mixture, omitting the enzyme, yielded no trace of the protamine.

A very considerable number of workers have reported the synthesis of proteins particularly in the presence of pepsin. These proteins are generally referred to as *plasteins*. There appears to be no question that an actual synthesis does take place, but it is clear that these proteins are not analogous to the natural proteins. They are certainly much more complex than the proteose mixture from which they were derived, and Wasteneys and Borsook believe that they do represent the products of a reversed equilibrium more or less similar to that noted above for the glycosides. However, they state that "peptic synthesis in concentrated peptic digests cannot be considered as an *in vitro* reproduction of biological processes; at best it is only a model." With this viewpoint Alcock [43] agrees.

[43] For literature and general references see: H. Wasteneys and H. Borsook, *Physiol Revs.*, **10**, 110 (1930); and R. S. Alcock, *ibid.*, **16**, 1 (1936).

[44] A. E. Taylor, *J. Biol. Chem.*, **3**, 87 (1907).

CHAPTER 20

Biological Reactions of the Proteins

No discussion of the importance of the proteins in biological processes would be complete without at least a casual mention of the important role which they play in the problems of immunity. Wells,[1] Landsteiner,[2] Marrack,[3] and Ratner [4] have summarized the more important literature with particular reference to the chemical problems involved. Accordingly, in the following pages only a very brief outline of these questions will be considered.

Foreign proteins (*antigens*) introduced into a living animal, by subcutaneous, intraperitoneal, or intravenous injection, by inhalation of a protein "dust," etc., give rise to the formation of specific substances (*antibodies*) in the blood serum of the animal, the presence of which may be detected by subsequent reactions of the animal or the blood serum of the animal. The reactions which follow the introduction of foreign proteins may be classified into four groups: anaphylaxis, the precipitin reaction, hemolysis, and complement fixation.

Anaphylaxis. When a small quantity of a protein (a *sensitizing* dose) is injected into the blood stream of an animal and an appropriate time interval (7 to 30 days or more, depending on the initial dosage, a larger dosage requiring a longer time) is allowed to elapse, a second injection of the same protein (the *intoxicating* dose) will cause the animal to undergo a severe shock, death often ensuing within a few minutes. The physiological reaction which occurs is known as the anaphylactic shock. The amount of protein necessary for the sensitizing dose may be exceedingly small, as little as 0.00000005 gram of egg albumin being sufficient. The initial injection has sensitized the animal to this particular

[1] H. G. Wells, *The Chemical Aspects of Immunity*, rev. ed., Chemical Catalog Co., New York, 1929.

[2] K. Landsteiner, *The Specificity of Serological Reactions*, Charles C. Thomas, Springfield, Ill., and Baltimore, 1936.

[3] J. R. Marrack, *The Chemistry of Antigens and Antibodies*, Special Rept. Ser., No. 194, Privy Council, Medical Research Council, H. M. Stationery Office, London (1934).

[4] B. Ratner, *Allergy, Anaphylaxis, and Immunotherapy*, Williams and Wilkins Co., Baltimore, 1943.

foreign protein. The second injection brings about the anaphylactic reaction.

Dragstedt [5] has summarized the literature bearing on the mechanism of anaphylactic shock. He concludes that the liberation in the tissues of heparin, histamine, and possibly choline during the reaction can account for all the prominent symptoms observed. He regards anaphylaxis as an autointoxication by substances normally present in various fixed or circulating tissue cells, these substances being freed by changes in cellular permeability induced by the antigen-antibody reaction. Others, however, disagree with this conclusion, pointing out that these agents may result from, rather than cause, the shock mechanism. Furthermore, the physiological pathology of anaphylaxis differs markedly in various animal species; in some animals the blood histamine is definitely reduced during shock. Thus the true mechanism is far from being completely understood.

The Precipitin Reaction. When an initial injection of a foreign protein into an animal is followed by other injections at 3- to 4-day intervals, the dosage being gradually increased until an appreciable quantity, 0.25 gram or more of the foreign protein, has been injected in a series of 5 to 6 up to 20 or 30 injections, the blood serum of the animal acquires the property of precipitating the particular foreign protein which was injected, when the immune serum is added to a solution of that protein *in vitro*. The delicacy of the test depends somewhat on the nature of the protein injected and the potency of the immune sera used. The test is often used to detect human blood in criminal cases. Ordinary chemical tests for blood have a sensitivity which will detect dilutions of blood not much greater than 1:1,000. The precipitin reaction has been found to be positive at a dilution of 1:50,000 for blood, and at a dilution of 1:1,000,000 for egg albumin. The "agglutination" reactions of bacteria are probably due to the precipitin reaction of proteins which are on the surface of the bacteria.

Since precipitins (or agglutinins) are found in both blood and tissues, they can promote the rapid removal of the foreign protein (or bacteria) from the blood, or immobilize it near its port of entry, where it can be set upon by phagocytes.

Hemolysis. When red blood cells of an animal are injected into the blood stream of an animal of a different species through a series of rather shortly spaced injections, the blood serum of the injected animal acquires the ability to dissolve the foreign blood corpuscles when tests are made *in vitro*. Here, again, the reaction has been employed in criminal cases to ascertain whether or not the cells in a blood stain were human cor-

[5] C. A. Dragstedt, *Physiol. Revs.*, **21**, 563 (1941).

puscles. Similarly, bacteriolysins may be formed which will disintegrate specific bacteria. Natural or acquired immunity may, at least in part, be due to the presence of such bacteriolysins.

Complement Fixation. This reaction depends on the fact that immune sera contain at least two distinct substances, both of which are required for the production of an immuno reaction. These are (*a*) the relatively heat-stable antibody, and (*b*) the complement which is destroyed at 55°C. Complement occurs in fresh, normal serum (in the globulin fraction) and can be supplied from that source.

Five biological reagents are necessary to carry out the complement fixation test.

Reagent 1 is a suspension of red blood cells (*e.g.*, sheep cells).

Reagent 2 is the immune serum A (*e.g.*, the serum of a patient who is suspected of having typhoid fever). The complement of this immune serum has been previously destroyed by heating the serum to 55°C.

Reagent 3 is a supply of fresh, normal serum to act as a source of complement.

Reagent 4 is the hemolytic serum B (*e.g.*, serum from rabbits which have been immunized to sheep corpuscles). This hemolytic serum B has had the complement destroyed by heating to 55°C.

Reagent 5 is a suspension of the bacteria which cause typhoid fever.

The test is carried out by taking the immune serum A (reagent 2), adding complement (reagent 3), and then adding the bacteria suspension (*the antigen*, reagent 5). If reagent 2 is in reality a typhoid immune serum, combination will occur between the typhoid bacteria, the typhoid antibody, and the complement (reaction I), which will remove all the complement from the solution and "fix" (adsorb?) it on the agglutinated or precipitated bacteria. However, it may not be possible for us to detect this reaction. It is necessary, accordingly, to determine whether or not reaction I has taken place. It is here that reagents 1 and 4 are used. To the original mixture of typhoid bacteria, typhoid antibody, and complement are now added reagents 1 and 4. *If hemolysis occurs, reaction I did not take place,* and complement is still present in the solution, as evidenced by the fact that the red cells were broken down by the hemolytic serum B. *If hemolysis does not take place, reaction I has already occurred, and reagent 2 was in reality serum from a patient with typhoid fever,* all the complement having been used up in reaction I. A modification of this reaction, but essentially the same as far as the technic is concerned, is the Wassermann test for syphilis.

Toxin, Antitoxin, and Toxoid. Toxins are the poisonous substances elaborated by bacteria, which are responsible for the diseases which they produce. Certain of these are *endotoxins,* which are retained within the

bacterial cells and liberated as these cells are broken down. The toxins of pneumococcus, streptococcus, and anthrax are of this type. *Exotoxins*, such as are produced by the organisms responsible for scarlet fever, diphtheria, and tetanus, are given off into the surrounding medium. The poisons in the venom of certain snakes and the poisons of spiders, scorpions, etc., also are exotoxins.

Evidence has been mounting to show that bacterial toxins are protein in nature; certain of them, such as diphtheria and tetanus, have been isolated in a practically pure state.

When exotoxins are introduced into an animal they stimulate the formation of specific antibodies, or *antitoxins*. Thus, after recovery from a disease such as diphtheria, the continued presence in the blood of the antitoxin formed during the bacterial invasion confers future immunity (partial or complete) on the animal. Endotoxins either do not stimulate antibody formation or give rise to serum elements of only weak neutralizing power.

Active immunity to a particular toxin sometimes can be achieved by administering the corresponding *toxoid* to an animal. When a toxin is treated with formaldehyde and heat, changes occur which render the product non-toxic, yet it usually has only slightly less antigenic value than the toxin. Toxoids produced in this way are widely used in inoculations against diphtheria and tetanus.

Origin of Antibodies. It has been recognized for more than a decade that antibodies are closely associated with the proteins in blood serum. Today it is widely accepted that they are *modified serum proteins* formed in response to stimulation by a foreign non-diffusible colloid (usually protein). In 1937, Tiselius,[6] using his newly developed electrophoresis method, found that the rabbit antibody to ovalbumin was all in the γ-globulin fraction, and that immunized animals had a greater amount of this fraction than normal rabbits. Subsequent studies by various workers have shown that antibodies are confined largely to two blood globulin fractions, *i.e.*, γ-globulin and a T-globulin, which migrates electrophoretically between the β and γ fractions.

Evidence obtained through diffusion or viscosity measurements and the ultracentrifuge indicates that antibodies have molecular weights of about 150,000 and 900,000. The former molecular weight is similar to that of normal serum globulins, and all human antibodies apparently fall in this category.

Cannon, Chase, and Wissler[7] have demonstrated the *importance of*

[6] A. Tiselius, *Biochem. J.*, **31**, 1464 (1937).

[7] P. R. Cannon, W. E. Chase, and R. W. Wissler, *J. Immunol.*, **47**, 133 (1943); see also, P. R. Cannon, *Advances in Protein Chem.*, **2**, 135 (1945).

dietary protein in antibody formation. They depleted the protein reserves of rabbits by maintaining them on protein-deficient diets (sometimes accompanied by plasmapheresis), after which they were subjected to antigenic stimulation by *Eberthella typhosa* or *Salmonella paratyphi* vaccines. In following the agglutinin titers of the sera in comparison with those of correspondingly vaccinated controls, they noted a marked lessening in the ability of the protein-depleted animals to produce antibodies.

Two theories have existed for many years concerning the site of antibody formation, *i.e.*, the reticulo-endothelial theory and the lymphocytic theory. Ehrlich and Harris [8] recently have discussed these theories in the light of newer observations and conclude that in all probability several types of phagocytic cells, *i.e.*, the lymphocytes, granulocytes, and macrophages, play an essential role in antibody production.

Haptens. Landsteiner [9] coined the name hapten, about 1921, to designate *substances which, when injected alone, do not give rise to the formation of antibodies in the blood serum but which, when combined with a protein and injected, give rise to immune sera specific for the chemical grouping of the compound (non-protein) which was present in the protein-hapten combination. In many instances the uncombined hapten reacts in vitro with the antibody so formed.* He accordingly notes that there are two systems of species specificity in the animal kingdom, the specificity of the proteins and the specificity of non-protein residues which may react as haptens, and that these specific non-protein substances may occur in the alcoholic extracts of cells and tissues, giving rise to antibodies. In the bacteria, in particular, *specific polysaccharides* have been isolated which are characteristic of particular bacterial types. These specific polysaccharides react as haptens, inasmuch as they do not, when injected in the pure state, give rise to the formation of antibodies but do *in vitro* react with specific immune serum characteristic of the particular type of bacteria containing the specific polysaccharide.

Our knowledge of the chemistry of immunity has had an amazing development during the last two decades. This has been largely, but not exclusively, due to the work of Landsteiner and his co-workers and the work of Michael Heidelberger and his co-workers, although many others have been and are now active in the field. This development has been brought about by the demonstrated formation of specific antibodies reacting with specific haptens. The antibody formation was induced by combining a chemical compound possessing a definitely known structure with a protein and then injecting this derived protein into the

[8] W. E. Ehrlich and T. N. Harris, *Science*, **101**, 28 (1945).
[9] K. Landsteiner, *Biochem. Z.*, **119**, 294 (1921).

experimental animal. *The immune serum so produced is specific, by and large, for the specific grouping of the non-protein constituent combined in the derived protein.* The developments in this field are covered by Chapter V of Landsteiner and Chapter III of Marrack, and the student interested in biological reactions is urged to read these chapters.

The original observations that non-protein constituents may give rise to antibody formation may be traced back to Obermayer and Pick,[10] but it was not until the observations of Pauly,[11] followed by the work of Landsteiner and Prasek, and Landsteiner and Lampl,[12] that the phenomenon was recognized as general. In these later studies *acyl* groups were introduced into the protein, and the immune sera were found to react with acyl-substituted proteins, being more or less specific for the particular acyl group which had been introduced and relatively nonspecific for the original protein. The next advance was the formation of *azoproteins*, where proteins were coupled with a diazonium compound to form azo derivatives. *These azoproteins gave rise to immune sera specific for the diazonium group which had been introduced into the protein*, so that it became possible to study immuno reactions characteristic of particular and of known chemical groupings. Conclusive proof has developed that the diazonium compounds couple with the tyrosine and histidine residues in the protein and that the immuno reactions of the resulting azoprotein reflect the specific configuration of the chemical residues linked to the azo group. Table 51 shows the cross-reactions of immune sera for a series of azoproteins differing somewhat in chemical constitution with a series of antigens consisting of azoproteins containing different hapten residues. It will be noted that the immune serum for the azoprotein containing *p*-aminobenzoic acid as a hapten reacted strongly with its own azoprotein, very weakly with an azoprotein containing *m*-aminobenzoic acid, but that there was no reaction with *o*-aminobenzoic acid or with a considerable number of closely related compounds even including *p*-aminophenylarsenic acid. Similarly when *m*-aminobenzoic acid is diazotized and linked to protein, and this protein injected into an animal to form an immune serum, the immune serum will react with the azoprotein which was initially injected but will not react with other azoproteins derived by the coupling of the diazonium salts of the following compounds with protein: aniline, *o*-aminobenzoic acid, amino-*m*-toluic acid, amino-*o*-toluic acid, 3-amino-

[10] F. Obermayer and E. P. Pick, *Wien. klin. Wochschr.*, **16**, 659 (1903); **17**, 265 (1904); **19**, 327 (1906).

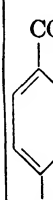
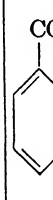
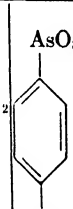
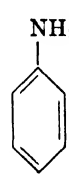
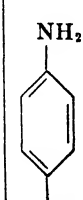
[11] H. Pauly, *Z. physiol. Chem.*, **42**, 508 (1904); **94**, 284 (1915).

[12] K. Landsteiner and E. Prasek, *Biochem. Z.*, **61**, 191 (1914); K. Landsteiner and H. Lampl, *Z. Immunitätsforsch.*, **26**, 258, 293 (1917); *Biochem. Z.*, **86**, 343 (1918).

4-methylbenzoic acid, 4-chloro-3-aminobenzoic acid, 4-bromo-3-aminobenzoic acid, *p*-aminobenzoic acid, *o*-aminocinnamic acid, *m*-aminocinnamic acid, *p*-aminocinnamic acid, *o*-aminobenzene sulfonic acid, 4-bromoaniline-2-sulfonic acid, *m*-aminobenzene sulfonic acid, and 4-aminotoluene-2-sulfonic acid. Thus, the introduced group induces a remarkable specificity.

TABLE 51. SHOWING CROSS-REACTIONS BETWEEN IMMUNE SERA FOR AZOPROTEINS CONTAINING CERTAIN GROUPINGS AND AZOPROTEIN ANTIGENS CONTAINING VARIOUS HAPTENS

[Landsteiner and van der Scheer, *J. Exptl. Med.*, 45, 1045 (1927)]

Azoprotein Antigen Containing the Residue from Diazotized	Immune Sera for Azoprotein Containing the Residue from Diazotized					
						
p-Aminobenzoic acid	+++	0	0	0	0	0
m-Aminobenzoic acid	±	0	0	0	0	0
o-Aminobenzoic acid	0	++	0	0	0	0
p-Aminophenylarsenic acid	0	0	++++	0	0	0
Sulfanilic acid	0	0	0	0	0	0
o-Aminocinnamic acid	0	0	0	0	0	0
Aniline	0	±	0	++±	+	++
p-Nitroaniline	0	±	0	+	++	+±
p-Toluidine	0	....	....	+±	+	++±
m-Toluidine	0	±	±	++	+	+±

Perhaps a still more pronounced illustration of the specificity of the induced hapten group occurs in the experiments of Landsteiner and van der Scheer,[13] where D-, L-, and *meso*-tartaric acids were introduced into proteins. It will be noted that there is a certain cross-reaction between the sera from the active forms of the tartaric acids and those from the *meso* form. This Landsteiner explains by the fact that the *meso* form contains carbon atoms having both D- and L-configuration, and, although the reaction is not so intense with the *meso* form, nevertheless it does indicate similarity of chemical configuration. From these and

[13] K. Landsteiner and J. van der Scheer, *J. Exptl. Med.*, **50**, 407 (1929).

other observations in the literature, it becomes evident that *the stereoisomeric configuration around an asymmetric carbon atom can be detected by serological reactions.* A β-glucoside hapten and a β-galactose hapten are immunologically distinct. Even α- and β-glucoside haptens, when linked in a protein, can be distinguished from each other serologically.

TABLE 52. SHOWING THE REACTIONS OF IMMUNE SERA (WHERE D-, L-, AND *meso*-TARTARIC ACIDS WERE INTRODUCED INTO THE IMMUNIZING AZOPROTEIN) WITH THE CORRESPONDING PROTEIN-HAPTEN ANTIGENS. A DEMONSTRATION OF THE EFFECT OF STEREOISOMERIC CONFIGURATION ON IMMUNOSPECIFICITY. (ANTIGEN CONCENTRATION: FIRST COLUMN = 0.05%, SECOND COLUMN = 0.01%)

(From Landsteiner, by permission. Courtesy Charles C. Thomas, publishers)

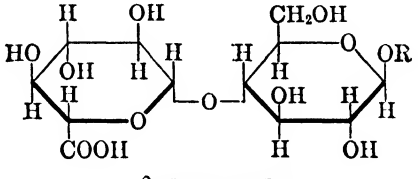
Immune Sera from	Antigen from					
	L-Tartaric Acid		D-Tartaric Acid		*meso*-Tartaric Acid	
L-Tartaric acid	+++	++±	±	0	+	±
D-Tartaric acid	0	0	+++	++±	+	±
meso-Tartaric acid	±	±	0	0	+++	+++

The serological differentiation appears to be as pronounced as the corresponding α- and β-glycosidal enzyme relationships which have been known for so long. When polypeptides are used as haptens and introduced into protein residues, Landsteiner found that "the specificity is determined chiefly by the amino acid carrying the free carboxyl group, to a lesser degree by the second amino acid" from the carboxyl end of the polypeptide chain. Landsteiner notes that those haptens which contain a free carboxyl, sulfonic, or arsenic acid grouping give rise to a much greater antigen specificity than those in which the substituting groups are neutral. Apparently an acidic residue increases the specificity of the reaction.

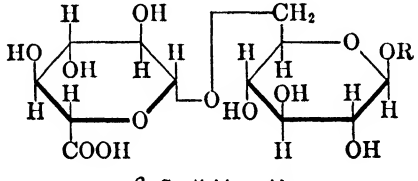
Goebel [14] has more recently prepared artificial antigens which evoke immunity to three types of pneumococci. Cellobiuronic acid is known to be a constituent of the specific polysaccharide of type III pneumococcus; gentiobiuronic acid has the same composition and differs from cellobiuronic acid only in the *position*, not in the *configuration*, of the glucuronosidic linkage. By combining the diazonium salts of the *p*-aminobenzyl aldobionides with horse serum globulin, artificial antigens

[14] W. F. Goebel, *J. Exptl. Med.*, **69**, 353 (1939); *Science*, **91**, 20 (1940).

were prepared. When injected into rabbits, the antigen containing the cellobiuronide gave rise to antibodies which agglutinated type III pneumococci and conferred passive immunity to types III and VIII virulent pneumococci. The gentiobiuronide antigen, on the other hand, had no such effects, proving that "*the position of the glucuronosidic linkage determines the immunological specificity* of antigens containing these acids



β-Cellobiuronide



β-Gentiobiuronide

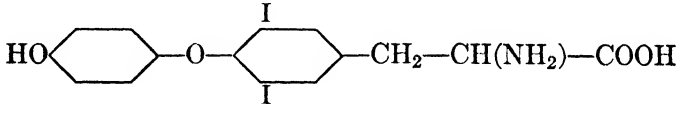
in so far as their capacity to evoke antibacterial immunity to Types III, VIII pneumococci is concerned." Goebel noted further that both synthetic antigens gave rise to antibodies which conferred passive immunity to type II pneumococcus, and clearly demonstrated that this was attributable to the glucuronic acid component common to both antigens.

We have already noted that the original biological specificity of the protein into which the hapten was introduced is, to a very considerable extent, lost and that the hapten-protein compound acquires a new specificity. In his discussion of this question Landsteiner points out that in many instances the original specificity of the protein is retained to some extent and can be detected by very careful technics, but that the specificity for the introduced hapten in general overshadows in intensity the original specificity of the protein. Hopkins and Wormall,[15] using phenylisocyanate or bromophenylisocyanate to form the phenylureidoprotein derivatives, found that the original protein specificity was lost and that the immune sera from the ureido derivatives would interact with all sorts of ureidoproteins including even gelatin. Similarly xanthoproteins lose their protein specificity, and the immune serum for one xanthoprotein will react with other nitrated proteins regardless of

[15] S. J. Hopkins and A. Wormall, *Biochem. J.*, **27**, 740, 1706 (1933); **28**, 228 (1934).

whether they are derived from plant or animal sources. Halogenated proteins similarly give rise to immune sera specific for halogenated proteins. The iodoprotein specificity apparently resides in the 3,5-diiodotyrosine residue, so that any protein containing this residue would react with the immune serum from any other protein containing this residue. Methylated and acetylated proteins also lose part or all of the original protein specificity and acquire new specificities for the introduced groups.

As previously stated, simple haptens do not form precipitates with the corresponding immune sera *in vitro*; precipitates are formed only with the immune sera and the protein-hapten complex. Landsteiner, however, observed that, *if the immune serum was mixed with the simple hapten and then mixed with the protein coupled with the hapten, no precipitate formed.* He designated this as *the phenomenon of specific inhibition*, and an enormous literature is being built up in this field. It is a phenomenon very similar and probably closely related to the phenomenon of complement fixation. Apparently the soluble hapten couples with or neutralizes the immune serum, but, since no precipitate forms, there is no visible evidence of the reaction or coupling because the resulting product is still soluble. However, the original coupling with the hapten prevents a later coupling with the protein-hapten complex. This allows for a study (*in vitro*) of the serum reactions of simple substances of known constitution which are themselves devoid of antigenic power but which can combine specifically with antibodies. In the halogenated proteins where 3,5-dihalogenated tyrosine is the specific hapten, it has been shown that thyroxine is effective as a specific inhibitor but that diiodothyronine



does not inhibit. Apparently the two iodines must be *ortho* to the —OH group. In the phenylureido compounds, the ϵ-amino group of lysine was found to be the point of attachment for the grouping, for phenylureidolysine acted as a hapten in the specific inhibition reaction of ureidoproteins.

Landsteiner goes further and suggests that hay fever, allergies of various kinds, and acquired hypersensitivity to chemicals which is often observed in the manufacture of dyestuffs, etc., as well as hypersensitivity to formaldehyde and to specific drugs, may in many instances be due to compounds of small molecular weight becoming in some way attached to a protein molecule and thus having induced the initial sensitization of the individual. In his experiments, in certain instances

when the animal has been sensitized to a protein coupled with a hapten it has been possible to induce an anaphylactic shock by an appropriate dosage of the pure uncoupled hapten. It seems as though any compound containing labile hydrogen, halogen, or other labile atomic groupings might, if it secures entrance to the blood stream, spontaneously couple with some of the serum proteins, thus giving rise to sensitization. In certain systems the specific inhibition reaction has been applied to the serum of hypersensitive individuals, and haptens of low molecular weight have been found to be responsible for the hypersensitivity.

There is some evidence that even coupling with a protein may not be necessary, for workers have reported the production of antibodies where the hapten was adsorbed on hydrous aluminum oxide or on charcoal. It may be that the compound circulating in the blood stream must either be in the colloidal state or have its kinetic energy greatly reduced. It seems possible that a reduction in kinetic energy might be advantageous in inducing specific orientation on the surface of tissue cells or possibly ingestion by phagocytes or tissue cells as a preliminary to antibody formation.

Nature of the Antibody-Antigen Combination. Marrack reaches the conclusion that the combination of antibody with antigen is an adsorption reaction which follows the Freundlich adsorption isotherm. When small amounts of antibody are present practically all the antibody is fixed by the antigen. With increasing amounts of antibody, holding the antigen constant, a part of the antibody apparently remains uncombined, the amount remaining uncombined being greater as the concentration of the antibody increases. If the logarithms of the amount bound are plotted against the logarithms of the amount remaining free, a straight line results, as would be anticipated from an adsorption isotherm.

Hydrogen-ion concentration has a marked effect on antibody-antigen combination. In one case which Marrack cites, combination was constant and maximum in the range of *p*H 5.55 to *p*H 8.0. The combination was markedly decreased, and the reaction was much slower at *p*H 4.5 and at *p*H 9.5 and was completely inhibited at a *p*H 3.0 or less or *p*H 10.0 or greater.

Salt effects are likewise prominent among factors affecting the antibody-antigen combination. The anions show specific lyotropic series effects either in inhibiting the combination or in dispersing the precipitate which forms, with CNS^- standing at the dispersing end of the series and F^- at the coagulating end. Marrack further suggests that a secondary reaction more or less similar to the heat denaturation of proteins may follow the primary adsorption reaction.

Pauling, Campbell, and Pressman [16] *point out that the properties of antigen-antibody systems suggest that the components are linked by* rather weak interactions such as electronic van der Waals attraction, the attraction of electronic dipoles or multipoles, and formation of hydrogen bonds. Since they regard the molecules of antigen and antibody as complementary in structure, close contact between the two would be established over a sufficiently large area to permit these relatively weak forces to form a strong bond. The evidence points to the reacting molecules being largely multivalent, *i.e.*, capable of forming two or more antigen-antibody linkages; therefore a continuation of these processes may build up a large complex with alternate antigen and antibody molecules, thus forming a precipitate.

Specificity of Immuno Reactions. For a number of years immunologists have been in general agreement that *protein specificity depends on the structure of the protein molecule and not necessarily on the biological origin of the protein*, although the great majority of proteins of different species are different both chemically and immunologically. In certain instances, however, immunological identity, or at least close similarity, has been observed in proteins isolated from different biological sources. Thus, the globulin of the squash seed appears to be immunologically identical with the globulin from the seeds of the cantaloupe, and there is an immunological similarity between ovomucoid from hens' eggs and a mucoid-like protein present in the seeds of the sweet almond. The caseins from the milk of animals of different species show very close biological relationships. The same is true of the vitellins of the egg yolk from various species of animals. Wells notes that "casein from the milk of an animal of any given species shows a closer biologic relationship to the casein of another species than it does to either the whey proteins or the serum proteins of its own species," and "egg yolk proteins from even such widely different species as fish and turtle may give precipitin reactions with the antiserum for hen egg yolk proteins."

Keratins generally show very little species specificity.[17] Thus pigeons sensitized to ox lens protein are also sensitized to that of the dog. Whereas there is little immunological difference between the lens protein of sheep and swine, there is a considerable difference between these mammalian proteins and chicken lens protein and a still greater dif-

[16] L. Pauling, D. H. Campbell, and D. Pressman, *Physiol. Revs.*, **23**, 203 (1943).

[17] Uhlenhuth, *Festschrift*, Robert Koch, Jena, 1903 (cited by Landsteiner in *The Specificity of Serological Reactions*, Charles C. Thomas, Springfield, Ill., and Baltimore, 1936); H. Hoffmann, *Z. Immunitätsforsch.*, **71**, 171 (1931); L. Markin and P. Kyes, *J. Infectious Diseases*, **65**, 156 (1939); E. E. Ecker and L. Pillemer, *J. Exptl. Med.*, **71**, 585 (1940).

ference between chicken and fish lens proteins. Pillemer and Ecker [18] have demonstrated that the species specificity of keratins can be increased by first reducing the —S—S— bonds to —SH groups, indicating that the state of oxidation $\rightleftharpoons$ reduction is a factor in specificity.

From the observation of proteins deliberately coupled with haptens, it has been absolutely demonstrated that immuno reactions are specific for particular groupings, and that substances containing very closely related, but nevertheless slightly different, chemical groupings will react with the immune bodies to a greater or less extent. As the degree of relationship diverges, there is a lessening in the intensity of the serological reaction until finally the serological reaction is undetectable. Similarly in proteins the species specificity is not absolute but varies with the biological relationships of the organisms. Closely related species apparently have proteins which are not absolutely identical but are closely related. As the degree of relationship becomes less, the differentiation of the protein becomes greater until finally there are no cross-reactions between the immune sera. Manwaring [19] points out that proteins may be 99 per cent identical, but that the immunological tests may be measuring the 1 per cent which is non-identical, and that the immuno reactions are for specific groupings rather than for the entire protein molecule.

Is serological specificity dependent on the *chemical* nature of groupings or on purely *structural* relationships? During the past ten years strong evidence has been gathered in support of the latter conception—the "lock and key" complementariness originally proposed by Paul Ehrlich. The work of Landsteiner, Heidelberger, and others on the specificity of haptenic groupings can be satisfactorily explained on this basis. The results of Goebel, cited above, also support this view. The more recent studies of Pauling, *et al.*,[20] now leave little doubt about the correctness of the structural explanation of immunological specificity.

Pauling regards antibodies as serum globulins which, under the stimulus of the antigen, have been "folded" to fit specific groups on the surface of the antigen molecule, *i.e., a portion of the antibody molecule structurally mirrors these antigen groups.* He has ascertained that the two parts fit together within less than 1 Å. Thus a methyl group (van der Waals radius 2.0 Å.) can replace a chlorine atom (radius 1.8 Å.) in a hapten without appreciably altering the antibody-antigen reaction, but the reaction is interfered with when a methyl group replaces a hydrogen

[18] L. Pillemer and E. E. Ecker, *Science*, **88**, 16 (1938).

[19] W. H. Manwaring, *Science*, **74**, 324 (1931).

[20] L. Pauling, *Chem. Eng. News*, **24**, 1064, 1375 (1946); *cf.* also numerous papers by Pauling, *et al.*, in *Journal of American Chemical Society* since 1940.

atom (1.2 Å.). Accordingly, haptenic groupings of different chemical nature but closely similar in size and shape will behave more or less the same immunologically; others containing closely analogous chemical groupings, perhaps differing only in the position of a linkage or in the configuration of one carbon atom, may be immunologically distinct.

Phylogenetic Relationships as Determined by Serological Reactions. A number of workers have used serological studies to trace phylogenetic relationships in both the plant and the animal kingdoms. Nuttall [21] studied more than 16,000 precipitin reactions quantitatively with the blood of more than 900 species of animals and found that antihuman precipitating serum gave comparative volumes of precipitate when tested against the blood sera of various orders of primates, as shown in Table 53. Landsteiner and Miller [22] have added to these observations.

TABLE 53. THE INTENSITY OF THE PRECIPITIN REACTION BETWEEN THE SERUM OF AN ANIMAL IMMUNIZED AGAINST HUMAN BLOOD PROTEINS AND EQUIVALENT AMOUNTS OF BLOOD SERA FROM VARIOUS ORDERS OF PRIMATES

Blood Serum from	*Number of Individuals Tested*	*Intensity * of Precipitin Reaction*
Human	34	100
Simiidae (anthropoids) (3 species)	8	100
Cercopithecidae (common monkeys of the Old World)	36	92
Cebidae (capuchins and spider monkeys of the New World)	13	78
Hapalidae (marmosets)	4	50
Lemuridae (lemurs)	2	0

* 100 indicates a reaction essentially identical in intensity with that of the original serum used for immunizing.

Martin and Cotner [23] studied the reaction of fourteen genera and twenty species of moths in six subfamilies of the *Phalaenidae* and concluded that the serological reactions were very useful in the determination of phylogenetic relationships. Similarly Wilhelmi [24] differentiated between various species of helminths in this manner and obtained results in good agreement with the morphological classification. He found the lipid-free (protein) fractions to be more *species specific* than the lipid-containing antigens, but the latter were widely *group* specific.

[21] G. H. F. Nuttall, *Blood Immunity and Blood Relationships*, Cambridge University Press, 1904.

[22] K. Landsteiner and C. P. Miller, Jr., *Science*, **61**, 492 (1925).

[23] S. Martin and F. B. Cotner, *Ann. Entomol. Soc. Am.*, **27**, 372 (1934).

[24] R. W. Wilhelmi, *Biol. Bull.*, **79**, 64 (1940).

Irwin and Cole,[25] in studying the immunogenetic relationships of the erythrocytes of the common pigeon and doves and hybrids between the various species, found that the erythrocytes possess many substances in common but that *each species possesses specific immunodetectable compounds peculiar to the species*, that these can be traced through the F_1

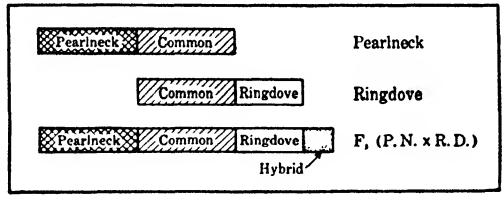


FIG. 95. Showing diagrammatically the immunologically specific components of the erythrocytes of "Pearlneck" and "Ring dove" and the F_1 hybrids. (After Irwin and Cole.)

hybrids, and that they segregate in the backcross generations. Thus, there are species specific "Pearlneck" substances, species specific "Ring dove" substances, etc. Figure 95 shows this diagrammatically for the Pearlneck dove, Ring dove, and the F_1 hybrid. In the F_1 hybrid a new immunologically specific "hybrid" substance likewise appeared. The chemical nature of these species specific substances has not been ascertained.

More recently Irwin and Crumley [26] have pointed out that "*any species of pigeon of the New World is more similar to any other species of*

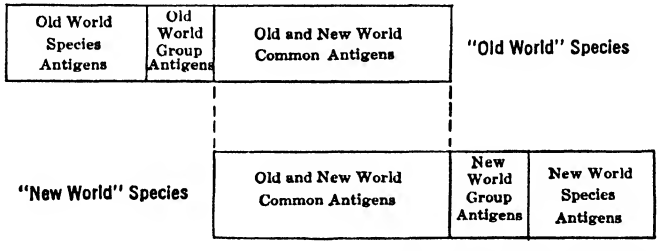


FIG. 96. Interrelationship between antigens in Old and New World species of pigeons. (After Irwin and Crumley.)

the New World than it is to any Old World species." The same holds true for the Old World pigeons with respect to the New World species (Fig. 96).

[25] M. R. Irwin and L. J. Cole, *Proc. Soc. Exptl. Biol. Med.*, **29**, 850 (1932); *J. Exptl Zool.*, **73**, 85, 285, 309 (1936).

[26] M. R. Irwin and R. W. Crumley, *Am. Naturalist*, **74**, 222 (1940).

In the plant kingdom we have the studies of Magnus,[27] Zade,[28] Mez,[29] and Moritz.[30] Certain of these studies have dealt with the genetic relationships of the *Leguminosae* and the *Gramineae*. Others, particularly those of Mez and Moritz, have been directed primarily toward phylogenetic problems. Mez has surveyed almost the entire plant kingdom and constructed a detailed chart showing phylogenetic relationships of the plant kingdom based upon serological reactions. The phylogenetic relationships shown in his chart are in most instances similar to or identical with those which have been generally accepted by taxonomists. In some instances the serum-diagnostic method of Mez indicates relationships different from those usually accepted. The author, however, has discussed this chart with several taxonomic botanists and in each instance has received the assurance that the chart may well express the true phylogenetic relationships. It would accordingly seem as if serum diagnosis may be used in questions of disputed phylogeny. Moritz in his studies has extended and confirmed the observations of Mez and concludes that the technics which Mez used were adequate.

We have already noted that Lewis and Wells found immunological relationships between the prolamines of *Triticum vulgare*, *T. dicoccum*, *T. monococcum*, and *T. spelta*. They similarly found immunological relationships between the prolamines of the corn group, *Zea mays*, *Andropogon sorghum*, and *Euchlaena mexicana* Schrad, but no relationship between the "wheat" group and the "corn" group.

In a later study [31] the globulins from the Georgia velvet bean and the Adzuki, Mung, navy, and lima beans were compared immunologically. The Georgia velvet bean globulins were distinct from those of the four other varieties. The lima and navy bean globulins were more or less alike but distinguishable from each other. The globulins from the Adzuki and Mung beans were similar or closely alike but were apparently not closely related to those of the other globulins tested. The Georgia velvet bean belongs to the genus *Stizolobia*, all the others to the genus *Phaseolus*.

[27] W. Magnus, *Landw. Jahrb.*, **38** (Erg. Bd. V), 207 (1909); *Ber. deut. botan. Ges.*, **26a**, 532 (1908).

[28] A. Zade, *Z. Pflanzensüch.*, **2**, 101 (1914); *Fuhlingslandw. Ztg.*, **61**, 807 (1912).

[29] C. Mez and H. Ziegenspeck, *Botan. Arch.*, **12**, 163 (1925); *cf.* also numerous papers in *Botanisches Archiv* under the general title, "Sero-diagnostische Untersuchungen," and C. Mez, *Proc. Intern. Congr. Plant Sci.*, 1402 (1926).

[30] O. Moritz, *Biol. Zentr.*, **48**, 431 (1928); *Beih. Botan. Centralbl.*, **46**, 114 (1929); *Planta*, **7**, 759 (1929); *Biol. Zentr.*, **51**, 290 (1931); *Planta*, **15**, 647 (1932); *Ber. deut. botan. Ges.*, **51**, 52 (1933); *Beitr. Biol. Pflanz.*, **22**, 51 (1934).

[31] H. G. Wells, J. H. Lewis, and D. B. Jones, *J. Infectious Diseases*, **40**, 326 (1927).

Human Blood Specific Substances. Normal sera of the higher animals usually contain substances which agglutinate or hemolyze the red cells of individuals of the same species. These have been designated as isoagglutinins or isohemolysins. Three distinct sets of substances have been found in human blood.

A-B Blood Groups. Nearly a half century ago Landsteiner [32] discovered that man can be classified on the basis of four distinct blood groups. In human blood there are two isoagglutinins (α and β) in the serum and two agglutinable substances (A and B) in the red corpuscles. The α- and β-isoagglutinins are inherited in the Mendelian fashion as genetic dominants. Table 54 shows the blood groups in man, the iso-

TABLE 54. IMMUNO REACTIONS CHARACTERISTIC OF THE FOUR BLOOD GROUPS IN MAN

(From Landsteiner, by permission. Courtesy Charles C. Thomas, publisher)

Blood Group	Isoagglutinins in Serum	Reactions with Erythrocytes of Group			
		O	A	B	AB
O	α and β	0	+	+	+
A	β	0	0	+	+
B	α	0	+	0	+
AB	.....	0	0	0	0

agglutinins in the serum, and the reaction with erythrocytes from the various groups. The A and B group specific substances are found in nearly all body cells and to a greater or lesser extent in body fluids and gland secretions (*e.g.*, saliva and gastric juice).

The anthropoid apes contain isoagglutinins and agglutinable substances indistinguishable from those in human blood. Chimpanzees, in contrast to the four groups in man, have only two blood groups, corresponding to the O and A groups in man.

Racial differences in man are markedly shown by differences in blood groupings. This was first demonstrated by the Hirszfelds.[33] Only group O occurs in the full-blooded American Indian. Agglutinable substance A is much more common than agglutinable substance B in

[32] K. Landsteiner, *Centr. Bakt. Parasitenk.*, **27**, 361 (1900); *cf.* also *Science*, **73**, 403 (1931).

[33] L. Hirszfeld and H. Hirszfeld, *Przeglad Epidemjol.*, **1**, (Heft 2), 1 (1920).

Europeans and Australian blacks. In the Hindus the ratio is reversed (Englishmen 46.4 per cent A, 10.2 per cent B; Hindus 49.7 per cent B, 27.5 per cent A). Blood grouping has even been used to trace race relationships in Egyptian mummies and likewise in cases of disputed parentage. Races of animals show distinctive blood groupings within the species.

M-N Blood Types. Landsteiner and Levine [34] first reported the presence of these factors in human red cells. An individual may possess M, N, or MN, but never are both factors lacking. Normally human blood does not have anti-M or anti-N agglutinins, so these factors are not important in blood transfusions. The agglutinins can be produced in the blood of a rabbit, however, by injecting human red corpuscles.

Since these blood types are genetically determined, they are sometimes useful in solving cases of disputed parentage; this would also be true of the other blood factors.

Rh Factor. In 1940 Landsteiner and Wiener [35] discovered this substance by using the serum from rabbits which had received blood from a rhesus monkey. Subsequently it was found to exist in the corpuscles of about 85 per cent of the human white population. It is inherited as a genetic dominant.

Aside from its obvious significance in repeated blood transfusions, it is believed to be responsible for many miscarriages, stillbirths, and to some extent the condition known as *erythroblastosis fetalis* in the newborn child. If the mother is Rh-negative and the fetus is Rh-positive (from the father), the fetal antigens may enter the maternal blood stream and stimulate antibody formation. The antibodies might then diffuse back through the placenta and react with the antigen in the fetus. Apparently, however, this takes place in only a small percentage of cases where this relationship exists.

Chemically Induced and Directed Genetic Transformations. It has been the dream of the geneticist for years to acquire the technic of inducing by experimental technics changes in an organism which are heritable and which will be passed on to the offspring. Isolated reports of success which have appeared from time to time have in general failed to be confirmed. Therefore, the experiments of Dawson and Sia [36] and of Alloway [37] take on added significance, inasmuch as these workers not only have apparently effected the permanent transformation of one

[34] K. Landsteiner and P. Levine, *Proc. Soc. Exptl. Biol. Med.*, **24**, 600, 941 (1927); *J. Exptl. Med.*, **47**, 757 (1928).

[35] K. Landsteiner and A. S. Wiener, *Proc. Soc. Exptl. Biol. Med.*, **43**, 223 (1940).

[36] M. H. Dawson and R. H. P. Sia, *J. Exptl. Med.*, **54**, 681, 701 (1931).

[37] J. L. Alloway, *J. Exptl. Med.*, **55**, 91 (1932); **57**, 265 (1933).

type of organism into another heritable type of organism, but also these have been *directed* changes in which the conditions of the experiment determine the direction which the inheritable modification will take, and in addition cause the production within the new type of organism of specific chemical compounds chemically and immunologically distinct from the chemical compounds which the original organism could elaborate. In Alloway's experiments in particular this was accomplished by the growing of pneumococci of one type in a culture medium to which had been added a sterile filtered extract of a different type organism. Thus, culture of type R Pneumococci were grown in culture media to which were added extracts of type I-S, type II-S, and type III-S Pneumococci, and under such conditions *the type R Pneumococci changed to the particular type characteristic of the material from which the active extract was derived.* The active extract could be clarified with charcoal and when so clarified was water-clear, yet the filtrate contained practically all the activity. The active substance could be precipitated from this filtrate with alcohol of 70–100 per cent concentration or with acetone. No loss in potency resulted from precipitation. Filtrates from the precipitate were inactive. The potent factor was thermolabile but was occasionally active after 10 minutes' exposure to 90°C. The extract passed through both N and W types of Berkefeld filters and retained its activity. The action of the extracts was specific. An extract from type II Pneumococci transformed type R Pneumococci into type II forms. An extract from type III Pneumococci transformed type R into type III forms. The transformed pneumococci produced the particular capsular specific polysaccharide characteristic of the type. It appears, therefore, as though some complex chemical substance had induced a change in the hereditary properties and definitely transformed one organism into a different but a closely related organism. This, if it is confirmed, is the first successful experiment of inducing a specifically directed genetic change in an organism by a chemical compound. All changes induced by radiation have been random or chance changes. Here a chemical compound from type I organism *directs* a change in type to type I organisms. The chemical compound from type II *directs* the change to type II, etc. Alloway and Dawson and Sia vouch for the sterility of the extracts and for the absence of the derived type in the original culture of the type R organisms. Their proof is accepted by Landsteiner. Since these experiments are of such importance in their genetic implications, they should be repeated with cultures from single-cell isolates of the type R organism. If such experiments confirm the reports, we shall have the first demonstrated proof of a directed genetic change.

CHAPTER 21

Nitrogen Bases

A discussion of the nitrogen bases need not be considered a part of protein chemistry. It seems probable, however, either that the nitrogen bases represent metabolic products formed by the breaking down of amino acids in the plant or animal organism, or that amino acids enter into the synthesis of these compounds. The nitrogen bases, therefore, may be looked upon as either amino acid decomposition products or products which have been derived from amino acids through some vital process.

It will be impossible in the space at our disposal to consider all the nitrogen-containing compounds that have been isolated from plant or animal material. Only a few of the more important will be mentioned, and only a few of the reactions involved will be considered.

The Decomposition Products of Amino Acids. Proteins are hydrolyzed by fungi and bacteria to their constituent amino acids which are then acted upon by the microorganisms to yield either bases or acids. Deamination occurs with both anaerobes and aerobes, the characteristic *aerobic* reactions being either the production of a saturated acid possessing one less carbon atom than the original amino acid or the production of a hydroxy acid having the same number of carbon atoms as the original amino acid.

$$RCH(NH_2)COOH + O_2 \rightarrow RCOOH + CO_2 + NH_3$$

$$RCH(NH_2)COOH + H_2O \rightarrow RCH(OH)COOH + NH_3$$

In some instances this reaction is modified, producing instead an alcohol with the elimination of carbon dioxide and ammonia. This reaction occurs [1] when amino acids are present in a medium where yeast is actively fermenting sugars, and has been used to prepare such rare alcohols as β-hydroxyphenylethyl alcohol (tyrosol), etc.

$$\underset{\text{Tyrosine}}{p\text{-}HOC_6H_4CH_2CH(NH_2)COOH} + H_2O \rightarrow \underset{\text{Tyrosol}}{p\text{-}HOC_6H_4CH_2CH_2OH} + NH_3 + CO_2$$

[1] F. Ehrlich, *Ber.*, **44**, 139 (1911); F. Ehrlich and P. Pistschimuka, *Ber.*, **45**, 1006 (1912).

The characteristic *anaerobic* changes may be either (*a*) deamination and reduction or (*b*) decarboxylation. Deamination and reduction form a saturated acid containing the same number of carbon atoms, or deamination may take place without reduction, and an internal rearrangement will cause the formation of an unsaturated acid.[2]

$$RCH(NH_2)COOH + H_2 \rightarrow RCH_2COOH + NH_3$$

$$RCH_2CH(NH_2)COOH \rightarrow RCH{=}CHCOOH + NH_3$$

Decarboxylation results in the formation of amines and is a typical reaction of the putrefying bacteria:

$$RCH(NH_2)COOH \rightarrow RCH_2NH_2 + CO_2$$

or formic acid may be liberated, in which case reduction takes place with the formation of an amine having one less carbon atom than the original amino acid:

$$RCH(NH_2)COOH + H_2 \rightarrow RCH_2NH_2 + HCOOH$$

Some workers believe that decarboxylation always involves both the above reactions, inasmuch as carbon dioxide and formic acid are usually formed simultaneously. The precise conditions for amine and acid formation under anaerobic conditions have not been accurately worked out. It seems probable that both these reactions proceed simultaneously, the preponderance of one or the other being determined by the particular type of organism involved. In only a few instances have the products of the reaction been investigated for some specific organism.

The fundamental differences between aerobic and anaerobic putrefaction may be stated as *the elimination of nitrogen and the formation of acids by aerobic organisms and the elimination of carbon dioxide with the formation of bases by anaerobic organisms.* It is recognized that there are exceptions to this rule. By and large, however, the rule will hold.

Certain of the amines produced from amino acids by bacteria are of importance. Arginine yields urea and ornithine; the latter in turn loses carbon dioxide to form *putrescine*, $NH_2—CH_2—CH_2—CH_2—CH_2—NH_2$. Lysine loses carbon dioxide to form *cadaverine*, the next higher homolog of putrescine. These bases have similar properties. They were among the first compounds isolated from the putrefactive decomposition products of proteins and were originally classified as "ptomaines." The names are suggestive of undesirable compounds. The bases themselves are physiologically practically inert.

[2] H. Raistrick, *Biochem. J.*, **11**, 71 (1917).

Tyrosine on putrefactive decomposition loses carbon dioxide to form the base tyramine, β-(p-hydroxyphenyl)ethylamine, an active principle of ergot. Histidine similarly forms histamine, β-imidazoleethylamine; tryptophan forms tryptamine, β-indoleethylamine; and arginine forms agmatine, δ-guanidobutylamine. Tyramine, histamine, and tryptamine all have a powerful physiological action. They are violent poisons. Tyramine is more or less similar to epinephrine in that it raises the blood pressure, whereas histamine reduces arterial pressure to the shock level (30–40 cm. Hg) even when used in dosages as low as 1 mg. per kg. of body weight. Histamine has been found by Abel in the secretion of the skin of the African toad. Toad skin, according to Pliny, was a medicine. Abel showed that it contained a powerful drug. This is an additional instance where one of the old folk remedies was found to be not so foolish as it appeared to be.

Tyramine likewise occurs in the salivary gland [3] of the cephalopod *Octopus macropus*, and in the pods of the broom,[4] *Sarothamnus scoparius*. In the latter it occurs in association with hydroxytyramine which had not been previously found to occur in nature.

Although, in man and other animals, decarboxylation of amino acids to their corresponding amines is largely restricted to the action of the microflora in the large intestine, evidence is available to show that certain amino acids can be degraded in this manner within the body. Under anaerobic conditions histidine can be converted to histamine in the liver and kidney, and tyrosine can form tyramine. Both of these bases are present in small amounts in normal blood. But this reaction appears limited to a few amino acids, and even then it is minor compared to other paths of metabolism.

Koessler and Hanke [5] made extensive investigations of proteinogenous amines. The earlier laboratory method of preparation involved decarboxylation by *Escherichia coli*. However, heating an amino acid with diphenylamine [6,7] may serve as a general method for the preparation of the corresponding proteinogenous amine. Using this method, Abderhalden obtained 95 per cent of the theoretical amount of pure tyramine from tyrosine.

The odor of putrefying protein is due to indole and skatole (β-methylindole) derived from tryptophan and not to the bases noted above.

[3] M. Henze, *Z. physiol. Chem.*, **182**, 227 (1929).

[4] H. Schmalfuss and A. Heider, *Biochem. Z.*, **236**, 226 (1931).

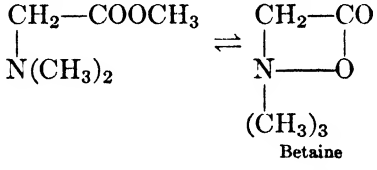
[5] K. K. Koessler and M. T. Hanke, *J. Am. Chem. Soc.*, **40**, 1716 (1918); *J. Biol. Chem.*, **39**, 497 (1919); K. K. Koessler, *Proc. Inst. Med. Chicago*, 3, 46 (1920).

[6] T. B. Johnson and P. G. Daschavsky, *J. Biol. Chem.*, **62**, 725 (1925).

[7] E. Abderhalden and F. Gebelein, *Z. physiol. Chem.*, **152**, 125 (1926).

We have already noted that certain of these bases were classified originally as "ptomaines" and were supposed to be responsible for ptomaine poisoning. Most of the bases, however, are physiologically inert, and it is very doubtful if ptomaine poisoning is ever caused by any of these decomposition products of proteins. The common cause of ptomaine poisoning is the presence of the bacterial poison, *botulinus toxin*, secreted by *Clostridium botulinum*. *Odor is no criterion of the presence or absence of botulinus toxin.* Completely rotted meat *may* be safe to eat, whereas apparently wholesome food may contain botulinus toxin, if opportunity has been afforded for the organism to develop. Botulinus toxin is destroyed by heat, which probably accounts for the relatively few cases of ptomaine poisoning which actually occur.

The Betaines. The betaines are a group of bases found in plants, which can be defined as *completely methylated amino acids.* The α-, β-, and γ-amino acids form betaines. Those from α-amino acids are known as the α-betaines, those from β-amino acids as β-betaines, etc. The simple betaine is that of glycine, which is formed by an intramolecular rearrangement of the methyl ester of dimethylamino acetic acid,



Betaine

forming an internal salt, the nitrogen atom changing from the trivalent to the pentavalent condition. It occurs in the sugar beet and passes into the molasses in the process of sugar manufacture. Young sugar beets contain up to 2.5 per cent; old ones contain about 1 per cent of betaine. It likewise occurs in the leaves of many families of plants in quantities ranging from a fraction of 1 per cent to 3.78 per cent in *Atriplex canescens.*

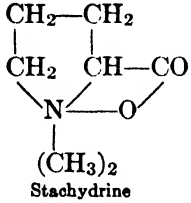
Barger [8] suggests that in plants betaines are end products of nitrogen metabolism and that they do not function in the vital processes. In view of more recent studies, however, it is possible that they may be concerned to some extent with transmethylation reactions (see p. 459). They are apparently most abundant in young or rapidly growing tissues.

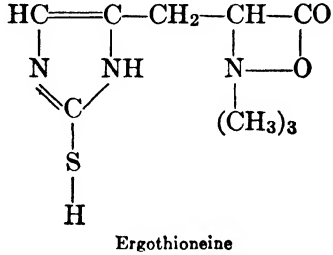
Stachydrine, the betaine of proline, occurs in the leaves of the orange tree to about 0.19 per cent of the dry weight. It is present [9] in the aqueous extract of alfalfa hay to the extent of about 0.5 per cent of the total

[8] G. Barger, *The Simpler Natural Bases*, Longmans, Green and Company, London, 1914.

[9] H. Steenbock, *J. Biol. Chem.*, **35**, 1 (1918).

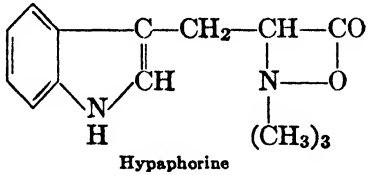
nitrogen. Its chemical properties are such that in a Van Slyke analysis for nitrogen distribution it would be calculated in the *histidine* nitrogen fraction. Its physical properties are such, however, that with Kossel's method it would appear in the *lysine* fraction. These considerations emphasize the importance of recognizing the limitations of the various methods of protein analysis. *The methods of protein analysis are valid only when the material being analyzed is a pure protein.* In the event that non-protein nitrogenous compounds are present, neither the Van Slyke method nor any other method based upon a nitrogen determination will necessarily yield accurate values for any particular amino acid.

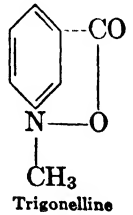

Stachydrine


Ergothioneine

Benedict, *et al.*,[10] isolated a sulfur-containing nitrogenous base from blood and later showed this to be identical with the base *ergothioneine*, isolated by Tanret [11] from ergot and shown by Barger and Ewins [12] to be the betaine of thiolhistidine. Its function in the blood stream still remains to be determined. *Trimethyl histidine*, the betaine of histidine, is found in certain edible mushrooms. Its chemical properties are such that it would appear in the lysine fraction of the Kossel separation.

Hypaphorine, the betaine of tryptophan, has been found in the seeds of *Erythrina hypaphorus*, a shade tree grown on the coffee plantations of Brazil. It has a very slight physiological action.


Hypaphorine


Trigonelline

Trigonelline, the betaine of nicotinic acid, is widely distributed in plants. It occurs in the garden pea. Also it is a normal constituent of

[10] S. R. Benedict, E. B. Newton, and J. A. Behre, *J. Biol. Chem.*, **67**, 267 (1926); E. B. Newton, S. R. Benedict, and H. D. Dakin, *Science*, **64**, 602 (1926).

[11] Ch. Tanret, *J. pharm. chim.*, **VI, 30**, 145 (1909).

[12] G. Barger and A. J. Ewins, *J. Chem. Soc.*, **99**, 2336 (1911).

urine,[13] being one of the end products of nicotinic acid metabolism. It is physiologically inert.

It seems very possible that a betaine corresponding to each of the known amino acids may well occur in nature. However, aside from their interest as naturally occurring nitrogen-containing compounds, they appear to be of little biological importance.

Choline and Allied Substances. Certain bases are formed by the bacterial decomposition of substances other than proteins. Among these are *choline* and its natural allies, *neurine* and *possibly muscarine.* These are all strong bases. β-Amino ethyl alcohol is probably the precursor of choline, which is trimethyl-β-hydroxyethylammonium hydroxide, $(CH_3)_3N(OH)CH_2CH_2OH$.

β-Amino ethyl alcohol, $NH_2CH_2CH_2OH$, or *ethanolamine*, occurs as a constituent of cephalin from various sources. Although the mechanism by which this base is formed is still uncertain, there is a possibility that it may be derived by the loss of carbon dioxide from serine.

$$HOCH_2CH(NH_2)COOH \rightarrow HOCH_2CH_2NH_2 + CO_2$$

Also dietary glycine containing N^{15} has been shown to form ethanolamine.[14]

Acetylcholine has been demonstrated to be involved in the transmission of the nerve impulse to the ganglion cells, and also to transmit the motor nerve impulses to the fibers of our voluntary muscles. The pioneer work in the demonstration of the importance of acetylcholine in nerve reactions is due to Loewi, *et al.*,[15] who demonstrated the presence of a "Vagusstoff" as liberated into the salt solution after the stimulation of the vagus nerve of the isolated frog heart. In a series of careful studies he identified his "Vagusstoff" with acetylcholine. Dale and Gaddum,[16] in 1930, extended Loewi's observation to the mammal, using dogs. They concluded that the nerve impulses were transmitted through the medium of acetylcholine. This mechanism of nerve impulse transport has been confirmed by subsequent studies. Dale [17] comments, "As I speak to you, I have every reason to suppose that the muscle fibers of my tongue and my jaws are being activated by innumerable little charges of acetylcholine, fired at them, as it were, from the endings of the nerve fibers."

[13] D. Melnick, W. D. Robinson, and H. Field, Jr., *J. Biol. Chem.*, **136**, 145 (1940).

[14] D. Stetten, Jr., *J. Biol. Chem.*, **140**, 143 (1941).

[15] O. Loewi, *et al.*, *Arch. ges. Physiol.*, **189**, 239 (1921); **193**, 201 (1922); **203**, 408 (1924); **204**, 361, 629 (1924); **206**, 123, 135 (1924); **208**, 694 (1925); **210**, 550 (1926); **214**, 678, 689 (1926); **217**, 610 (1927); **225**, 721 (1930).

[16] H. H. Dale and J. H. Gaddum, *J. Physiol.*, **70**, 109 (1930).

[17] H. Dale, *Science*, **80**, 343 (1934); *cf.* correction, *ibid.*, **80**, 450 (1934).

He calculates that the transmission of a single nerve impulse to a single ganglion cell involves the liberation of approximately 3,000,000 molecules of acetylcholine.

Lecithin (a phospholipid) is a *choline glycerol phosphoric acid ester*, and inasmuch as it occurs in all cells and especially in nervous tissues, it must be of marked physiological importance. Putrefaction of lecithin yields choline as one of the products. Choline is a depressant, causing a fall in blood pressure, and is antagonized by the alkaloid atropine. Choline may occasionally occur in the free state in plants. Vinson [18] isolated a considerable quantity of choline from a fraction of the dilute sodium hydroxide extract of corn pollen. Whether or not the choline was present in the free state in the corn pollen or represented a decomposition product of the lecithin was not determined.

When putrefying bacteria act upon choline, the alcoholic side chain suffers dehydration, forming *neurine*, vinyltrimethylammonium hydroxide, $(CH_3)_3$—N(OH)—CH=CH_2. Neurine is exceedingly poisonous and may occur among the putrefaction decomposition products of meat and fish. Further decomposition results in the formation of trimethylamine. Trimethylamine occurs [19] in the spores of *Tilletia levis*, the "stinking smut" of wheat.

According to the older literature, *muscarine* is the oxidation product of choline, where the alcohol group has been oxidized to an aldehyde. A more recent investigation [20] of the chemistry of muscarine points to the probable formula

$$(CH_3)_3N(OH)—CH(CHO)CH(OH)CH_2CH_3$$

which would be a β-hydroxyvalerianic aldehyde substituted in the α-position with the trimethylammonium hydroxide radical. It would thus be a higher homolog of the oxidation product of choline. Muscarine is the poison of the "fly mushroom" or "fly agaric," *Amanita muscaria*. This mushroom was Caesar's favorite poison because it resembles so closely *Amanita caesarea* which is edible. He accordingly had *Amanita caesarea* served to himself, and *Amanita muscaria* served on certain occasions to his guests. Muscarine poisoning is antagonized by atropine.[21]

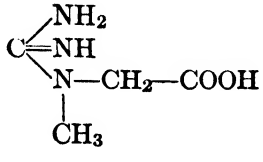
[18] C. G. Vinson, *J. Agr. Research*, **35**, 261 (1927).

[19] W. F. Hanna, H. B. Vickery, and G. W. Pucher, *J. Biol. Chem.*, **97**, 351 (1932).

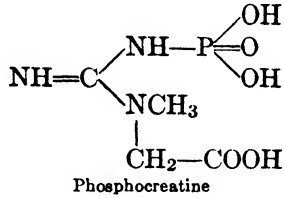
[20] F. Kögl, H. Duisberg, and H. Erxleben, *Ann.*, **489**, 156 (1931).

[21] Warning should be inserted at this point, that whereas the poison of *Amanita muscaria* is antagonized by atropine, this is not the case for the "deadly amanita," *Amanita phalloides*. *Amanita phalloides contains the amanita toxin for which there is no antidote.* The toxin is a complex, possibly protein-like substance acting as a

Creatine and Creatinine. Methylguanidoacetic acid, *creatine*, is a constituent of all vertebrate muscle and appears normally in the blood. It occurs in the urine during starvation and is an indication that muscle tissue is being broken down. Fiske and Subbarow [22] isolated a creatine phosphoric acid complex from muscle tissue and pointed out that this compound is present in considerable proportion in resting muscle but is rapidly decomposed in fatigued or dead muscle. They suggested [23] a physiological role for the creatine and phosphoric acid in the muscle, *i.e.*, that they act as a buffer to bind the lactic acid formed from sugars by muscular activity.



Creatine



Phosphocreatine

At approximately the same time that Fiske and Subbarow isolated their *phosphocreatine*, the Eggletons [24] isolated the same compound and named it *phosphagen*. Elucidation of the reactions of this compound and the role which it plays in vital processes revolutionized our previous concepts of muscle physiology.

The primary role of phosphocreatine in the phenomenon of muscle contraction is that of a rich store of readily available energy.[25] On muscle stimulation, adenosine triphosphate (ATP) breaks down to adenosine diphosphate (ADP) and phosphoric acid, with the release of considerable energy. This is followed immediately by the breakdown of phosphocreatine to creatine and phosphoric acid, liberating 120 calories per gram of phosphoric acid, or nearly six times the energy of the ATP reaction. *These two energy-rich, labile organic phosphates provide the immediate energy for muscle contraction.* Resynthesis of phosphocreatine

hemolysin, dissolving the red blood cells. It is one of the most poisonous of all plant products. One half of an *Amanita phalloides* is sufficient for a fatal dose. As a rule, the symptoms of *Amanita* poisoning do not appear for some time after the mushroom has been eaten. For those who are interested in the properties of poisonous mushrooms, see W. W. Ford and E. D. Clark, *Mycologia*, **6**, 167 (1914) (60 refs. to literature).

[22] C. H. Fiske and Y. Subbarow, *Science*, **65**, 401 (1927).

[23] C. H. Fiske and Y. Subbarow, *Science*, **67**, 169 (1928).

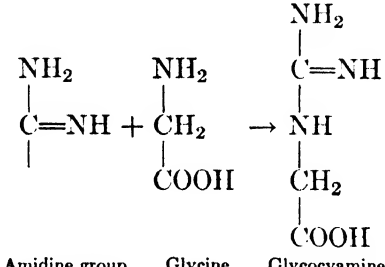
[24] P. Eggleton and G. P. Eggleton, *Biochem. J.*, **21**, 190 (1927).

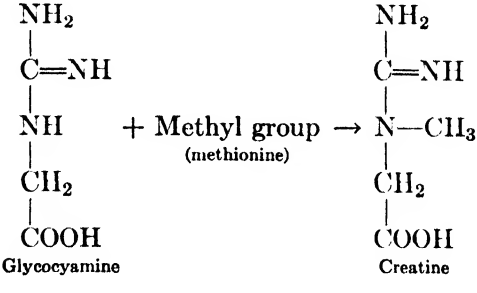
[25] The role of *phosphate-bond energy* in metabolism is described in detail in Chapter 26.

takes place during muscle recovery, the energy for resynthesis being supplied by the exothermic formation of lactic acid from muscle glycogen.

Whereas phosphocreatine plays a major role in vertebrate muscle contraction, the corresponding arginine phosphoric acid ester is the active agent in invertebrate muscle (*e.g.*, crustaceans). An echinoderm (*Stronglyocentrotus*) and a primitive chordate (*Balanoglossus*), which may represent intermediate stages in the transition from invertebrates to vertebrates, contain *both* phosphocreatine and phosphoarginine. Another simple chordate, *Amphioxus*, resembles the vertebrates in that it possesses only phosphocreatine.

Until a few years ago the exact mechanism whereby the body synthesized creatine was unknown. Now it is recognized that three amino acids are concerned in creatine formation:[26]



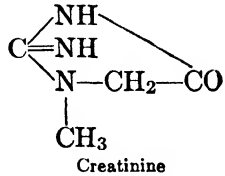


The first step occurs in the kidneys, whereas the methylation process takes place in the liver.

Creatinine, the anhydride of creatine, occurs normally in all urine. The ratio of creatine to creatinine has been extensively studied as indicating relationships to physiological conditions of the individual. A

[26] H. Borsook and J. W. Dubnoff, *J. Biol. Chem.*, **132**, 559 (1940); K. Bloch and R. Schoenheimer, *J. Biol. Chem.*, **134**, 785 (1940); **138**, 167 (1941).

discussion of such relationships, however, properly belongs in textbooks of physiological chemistry.



Creatinine

Miscellaneous Nitrogen Bases. Only a few of the nitrogen bases occurring in plants or animals can be considered. For a more adequate discussion the reader is referred to the book by Barger, already noted.

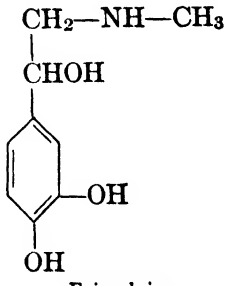
Guanidine occurs occasionally in plants. It is found in the sugar beet and passes into the molasses in the process of sugar manufacture. It is probably originally derived from arginine.

Another base, of the utmost importance, has three names which are rather common in literature, *epinephrine* (Abel), *suprarenine* (Von Fürth), and *adrenaline* (Takamine). Since this base was first isolated by Abel, it should be known as epinephrine. It is a derivative of pyrocatechol. It will be noted that the formula (p. 498) shows a marked resemblance to tyrosine. Undoubtedly epinephrine is one of the most important of all the naturally occurring bases. It is one of the active principles of the adrenal glands and regulates the blood pressure in mammals. In health, about 0.1 per cent of the gland is epinephrine. The natural product is *levo*-rotatory. The racemic mixture, containing the *dextro*- and *levo*-modifications in equivalent amounts, has a much lower physiological activity than the natural product, the *dextro*-form having only approximately 5 per cent of the physiological activity of the *levo*-form, again a striking example of biological specificity. An injection of as little as 0.0003 mg. per kg. of body weight is sufficient to produce a marked effect on blood pressure. The lethal dose (for man) is about 0.06 gram when injected intravenously. It contracts the blood vessels so that no blood can flow. Because of this marked contractile effect, epinephrine has come to be a valuable adjunct in surgery to control bleeding during a surgical operation. Together with other physiologically active constituents, epinephrine occurs in the secretions of two African toads, *Bufo regularis* and *B. arenarum*.[27]

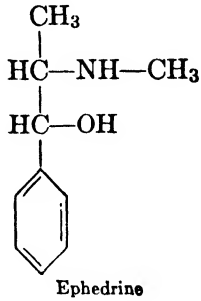
Ephedrine, another base very similar in structure to epinephrine and having the formula noted, occurs in the plant *Ephedra equisetina*, and in smaller amounts in the leaves of the yew, *Taxus baccata*. This base

[27] H. Jensen, *J. Am. Chem. Soc.*, **57**, 1765 (1935).

has many of the desirable properties of epinephrine. It raises the blood pressure and has the advantage that it can be taken orally



Epinephrine



Ephedrine

Two synthetic products combining portions of the structure of both epinephrine and ephedrine have proved of value in medicine. These are "corbasil" or "cobefrin" and "benzedrine." Corbasil is apparently β-hydroxy-β′-(3,4-dihydroxyphenyl) isopropylamine, and benzedrine is β-phenylisopropylamine. Benzedrine has a stimulating effect on the higher centers of the central nervous system and has been used with good results in cases of nervous fatigue, chronic exhaustion, psychoneuroses, and manic-depressive psychoses.

3,4-Dihydroxyphenylalanine or *dopa* is an amino acid similar in some respects to epinephrine in structure. It occurs in the Georgia velvet bean (*Stizolobium deeringianum*) and in the seeds of *Vicia faba*. It reacts more or less like epinephrine but does not have the highly marked physiological properties, although it is distinctly toxic. It has not as yet been isolated from a protein. It is apparently the *chromogen* involved in certain brown and black animal pigmentations,[28] such as in butterfly wings, *e.g.*, *Vanessa antiopa*. In *Tenebrio molitor*, the chromogen [29] appears to be 3,4-dihydroxyphenylacetic acid, which presumably arises from dopa.

Carnosine, the dipeptide β-alanyl-histidine, has already been referred to. It is one of the principal products in "beef extracts" and may be looked upon either as a dipeptide or as a base which is a constituent of muscle.

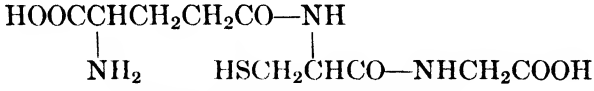
Anserine, a dipeptide of β-alanyl-N-methylhistidine, where the methyl group is attached to the nitrogen of the imidazole ring, occurs among the extractives of the flesh of the dog, cat, rabbit, white rat, goose, chicken, turkey, pigeon, and crocodile. It has not been isolated

[28] K. Hasebroek, *Fermentforschung*, **5**, 1, 297 (1921–22); **7**, 139, 143, 183 (1923–24); **8**, 197, 199, 553, 568 (1925–26).

[29] H. Schmalfuss, A. Heider, and K. Winkelmann, *Biochem. Z.*, **257**, 188 (1933).

from the muscle protein. The fact that it contains an N-methyl group may account for a part or all of the N-methyl compounds which have been reported to occur in proteins.

Glutathione, a tripeptide, glutamyl-cysteinyl-glycine, was isolated by Hopkins [30] from yeast, muscle, and mammalian liver. Such a tripeptide has twelve possible structural formulas. Kendall [31] has shown that glutathione has the formula



Glutathione (glutamyl-cysteinyl-glycine)

This appears to be the substance which is present in all cells that give the nitroprusside test. It is generally believed that glutathione acts as a hydrogen receptor and takes an active part in cell oxidations and reductions: [32]

$$2R{-}SH \rightleftharpoons R{-}S{-}S{-}R + H_2$$

Kendall and Nord [33] suggest that an unstable highly reactive oxygen addition product is formed between glutathione and oxygen, in which product the sulfur atom has a higher state of oxidation. This compound, together with the more stable —SH and —S—S— forms, make up a reversible oxidation-reduction system.

Attention has been called also to the role which glutathione may play in intermediary protein metabolism.[34] The only three amino acids which are known to be available for the process of detoxification in man are the same as those which form glutathione. It is suggested that the mechanism may be that the substance being detoxified first combines with glutathione, and then this compound breaks down, leaving a part of the original glutathione molecule associated with the toxic compound.

Numerous reports have demonstrated that the —SH group is effective in stimulating growth of epithelial tissue and in promoting granulation of wounds. It may be that a very considerable part of the biological effects of glutathione reside in the stimulation afforded tissues

[30] F. G. Hopkins, *Biochem. J.*, **15**, 286 (1921); *J. Biol. Chem.*, **84**, 269 (1929).

[31] E. C. Kendall, B. F. McKenzie, and H. L. Mason, *J. Biol. Chem.*, **84**, 657 (1929); *cf.* also *Proc. Staff Meetings Mayo Clinic*, **4**, 359 (Dec. 11, 1929).

[32] M. Dixon and J. H. Quastel, *J. Chem. Soc.*, **123**, 2943 (1923); *cf.* also J. H. Quastel, C. P. Stewart, and H. E. Tunnicliffe, *Biochem. J.*, **17**, 586 (1923); H. E. Tunnicliffe, *Biochem. J.*, **19**, 194 (1925); C. P. Stewart and H. E. Tunnicliffe, *Biochem. J.*, **19**, 207 (1925).

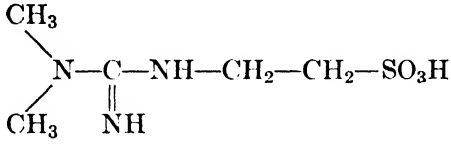
[33] E. C. Kendall and F. F. Nord, *J. Biol. Chem.*, **69**, 295 (1926).

[34] E. Brand and M. M. Harris, *Science*, **77**, 589 (1933).

by the —SH linkage. In malignant growth, spontaneous mammary carcinomas are in a large measure controlled [35] by a diet deficient in cystine or methionine. After such repression, the growth of the carcinoma is greatly stimulated by either cystine or glutathione.

Taurine, aminoethylsulfonic acid, has the formula $NH_2—CH_2—CH_2—SO_3H$. It is found in small quantities in lung and muscle tissues, but mainly it is present, combined with cholic acid, in the bile as taurocholic acid. Taurine is probably derived from cystine as a result of oxidation and decarboxylation in the liver.

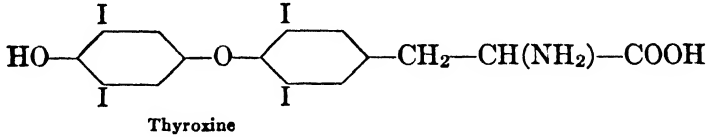
Dimethylguanyltaurine, or *asterubin*, has been isolated in small



amounts from starfish.[36] Its physiological significance is not known.

Spermine, isolated by Rosenheim,[37] has been shown to be α-δ-bis-(γ'-aminopropylamino) butane, $NH_2(CH_2)_3NH(CH_2)_4NH(CH_2)_3NH_2$. It is apparently a constant constituent of sperm, although it has been found to occur in other animal organs. A similar compound *spermidine*,[38] α-(γ'-aminopropylamino)-δ-aminobutane, $NH_2(CH_2)_3NH(CH_2)_4NH_2$, occurs in association with spermine in the various tissues. The physiological function of these bases has not been determined.

Thyroxine, the hormone of the thyroid gland, was isolated by Kendall [39] as a decomposition product of the protein of the thyroid gland.



Thyroxine

It has been synthesized [40] by known reactions. Thyroxine has a very marked physiological action, exceptionally small amounts increasing the

[35] C. Voegtlin, J. M. Johnson, and J. W. Thompson, *U. S. Pub. Health Repts.*, **51**, 1689 (1936).

[36] D. Ackerman, *Z. physiol. Chem.*, **232**, 206 (1935).

[37] O. Rosenheim, *Biochem. J.*, **18**, 1253 (1924); *cf.* also H. W. Dudley, O. Rosenheim, and W. W. Starling, *Biochem. J.*, **20**, 1082 (1926).

[38] H. W. Dudley, M. C. Rosenheim, and O. Rosenheim, *Biochem. J.*, **18**, 1263 (1924); H. W. Dudley, O. Rosenheim, and W. W. Starling, *Biochem. J.*, **21**, 97 (1927).

[39] E. C. Kendall, *J. Biol. Chem.*, **39**, 125 (1919); *Ind. Eng. Chem.*, **17**, 525 (1925).

[40] C. R. Harington, *Biochem. J.*, **20**, 293, 300 (1926); (and G. Barger) *Biochem. J.*, **21**, 169 (1927).

metabolism of the body. As little as 1 mg. of thyroxine injected intravenously may increase the metabolic rate as much as 2.5 per cent or, as Kendall puts it, "The injection of 1 milligram of thyroxine will produce an increase in the carbon dioxide output of approximately 400,000 milligrams of carbon dioxide."

After the injection of thyroxine, there is a pronounced delay in reaction, 6 to 8 hours being required before the basal metabolic rate is affected. Subsequently there is a rapid increase in metabolic rate, but the maximal response may not be reached for several days after the injection. In a number of instances the maximal response to a single injection was not reached until the eighth or tenth day after the injection. The physiological effect of most drugs is over within a relatively short period of time. Thus, the effect of epinephrine may last, at the most, only a few hours. Kendall notes that a single injection of 5 to 10 mg. of thyroxine may affect the basal metabolic rate for a period as long as 5 to 6 weeks.

When certain pathological conditions influence the function of the thyroid gland so that thyroxine is no longer synthesized, as in myxedema and cretinism, the entire physiological process undergoes alteration which is particularly noted in the mental reactions, the individual either remaining essentially an idiot (cretinism) or losing his mental faculties (myxedema). In each instance the mental faculties can be more or less completely restored and physiological processes brought back to normal by means of thyroxine. Accordingly the isolation of this hormone in a form suitable for use in medicine ranks as one of the great contributions to modern medicine.

The Alkaloids. Among the vegetable products, numerous oily or crystalline bases have been found, to which the term "alkaloid" has been applied, and, because of their physiological properties they have, for ages past, interested mankind.

When the orientals used opium or hashish, or the South American aborigines chewed the coca leaves for stimulation or the cinchona bark for fevers, they did not know that the reason they secured results lay in the alkaloidal content of the material. Only within the last few decades has the organic chemistry of certain of the alkaloids been elucidated. Owing to the complex structure of the alkaloidal molecule, a study of the chemistry of the alkaloids is one of the most difficult fields of organic chemistry, and even today the structural formulas of several important drugs are more or less uncertain.

We can only consider briefly a few of the more important alkaloids.

For those who are interested in the further development of this subject reference may be made to a number of books.[41-47]

It is difficult to formulate an exact definition of an alkaloid. The definition may be made so broad as to include all nitrogen-containing compounds or so narrow as to leave out compounds with definite alkaloidal properties. Ladenberg originally defined them as "*those naturally occurring vegetable substances of a basic character which contain at least one nitrogen atom forming a part of a heterocyclic ring.*" If we except the purine and pyrimidine bases, we can limit the definition to "basic substances found in plants and which contain a cyclic nitrogenous nucleus." Even this is too narrow, for a few compounds do not contain a "cyclic nitrogenous nucleus"; nevertheless they may have the marked physiological properties of alkaloids. The modern chemist may well take exception to the words, "naturally occurring." We have prepared synthetically medicinal substances which are *better* than the natural alkaloids, and these synthetic products are truly alkaloidal, in both chemical and physiological properties, if we accept chemical and physiological properties as criteria. Perhaps the definition of Thompson,[48] *a relatively complex organic chemical entity containing nitrogen, whose free base is alkaline in reaction and capable of neutralizing acids to form salts, whose free base is much less soluble in water than its salts, whose free base is much more soluble in ether, chloroform, benzol, etc., than its salts, and as a substance whose salts in solution release the free base upon the addition of alkalies,*" is as satisfactory as any which can be devised.

Historically, the work on alkaloids dates back to 1803 when Derosne isolated a crystalline compound from opium, which he called "opium salt." He did not, however, notice its basic character. In 1805, Sertürner, a German apothecary, isolated the material again independently, purified it, recognized its basic properties, and called it "morpheum." At the same time he separated an acid which he called "meconic" acid and expressed the view that the two were combined in opium. These observations remained unnoticed until 1817, when

[41] T. A. Henry, *The Plant Alkaloids*, P. Blakiston's Son and Co., Philadelphia, 1913.

[42] A. Pictet, *The Vegetable Alkaloids*, translated by H. C. Biddle, John Wiley & Sons, New York, 1904 (out of print).

[43] N. V. Sidgwick, *The Organic Chemistry of Nitrogen*, revised and rewritten by T. W. J. Taylor and Wilson Baker, Clarendon Press, Oxford, 1937.

[44] L. Spiegel, *Der Stickstoff*, Fr. Vieweg und Sohn, Braunschweig, 1903.

[45] E. Winterstein and G. Trier, *Die Alkaloide*, Gebrüder Borntraeger, Berlin, 1910.

[46] J. Schmidt and V. Grafe, *Alkaloide*, Urban und Schwarzenberg, Berlin, 1920.

[47] A. W. Blyth, *Poisons: Their Effects and Detection*, 3rd ed., Charles Griffin and Co., Ltd., London, 1895.

[48] M. R. Thompson, *Science*, **82**, 62 (1935).

Sertürner published a second paper, in which he further pointed out the basic character of morphine and described a number of its salts. Chemists then began to look for other similar compounds, and in 1818 Pelletier and Caventou found strychnine in *Nux vomica*, then brucine in 1819, and in 1820 they isolated quinine and cinchonine from cinchona bark. At least two or three new alkaloids have been isolated and described each year since 1820.

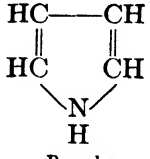
There was considerable speculation by the early chemists regarding the chemical constitution of these compounds, but the first definite clue was obtained about 1842–1846. Gerhardt, in 1842, distilled quinine, strychnine, and cinchonine with solid potassium hydroxide and obtained an oily base which he called "quinoleine"; later the name was changed to quinoline.

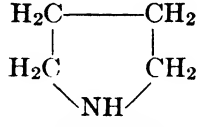
Before this, Runge (1834) had obtained a base from coal tar which he called "leucol," and Hoffman found that "quinoleine" and "leucol" were identical. Meanwhile, Andersen (1849–1851) separated pyridine from bone oil, and later this base was isolated from the alkaloids, nicotine, coniine, piperidine, etc., by distillation with zinc dust. Isoquinoline was isolated from coal tar in 1885 and was later obtained from certain of the alkaloids, hydrastine, papaverine, etc. The fact that coal-tar bases could also be obtained from alkaloids stimulated research, and, by breaking down alkaloids on the one hand and building up derivatives of the coal-tar bases on the other, points of contact were established and information regarding the alkaloidal molecule was obtained. Eventually, some of the alkaloids were synthesized, and the recent advance in this field has been in the discovery of the *potent groupings* and then the use of these new groupings in new compounds which are better than the naturally occurring alkaloids, in that *they have the same curative or medicinal properties and less of the undesirable properties.*

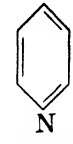
In using alkaloids in medicine, two properties of the alkaloid must be taken into consideration, *i.e.*, the minimal medicinal dose and the minimal lethal dose. The minimal medicinal dose is that dosage which will bring about the desired physiological effect. The minimal lethal dose is that dosage which may cause death. A number of alkaloids having very desirable physiological properties cannot be used or can be used only with great caution, because their physiological dose lies very close to their toxic dosage. Sometimes only three or four times the physiological dose is sufficient to cause death. The aim of the synthetic organic chemist interested in the field of the alkaloids is to ascertain those groupings which give to an alkaloidal molecule medicinal properties and those groupings which give to the alkaloid its toxic properties, with the hope of being able to synthesize new organic compounds having the desirable

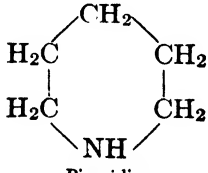
properties and lacking the undesirable properties of the naturally occurring alkaloids. The synthesis of *novocaine, procaine, butyn,* etc., has been accomplished by studying the molecular configuration of cocaine and attempting to synthesize local anesthetics having a similarity to certain parts of the cocaine molecule but lacking the toxic groupings.

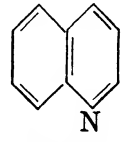
A classification of the alkaloids has been made on the basis of the heterocyclic nucleus, although more than one heterocyclic ring may be present in the molecule of a single alkaloid. The following ring structures are the characteristic structures found in naturally occurring alkaloids.

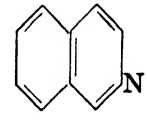

Pyrrole


Pyrrolidine


Pyridine


Piperidine


Quinoline


Isoquinoline

In most discussions of alkaloids, the *purine derivatives,* such as caffeine and theobromine, are considered to belong to the alkaloid group. We have already noted these compounds as derivatives of the purine nucleus under nucleic acids and will not include them again at this point.

Before considering the organic chemistry of the alkaloids, it is perhaps justifiable to note certain of the families of plants in which alkaloids occur. Alkaloids are very unevenly distributed throughout the various groups of the plant kingdom. In some families of plants many alkaloids are found. Other families are noted by the practical absence of alkaloids. Very few of the monocotyledonous plants contain alkaloids, these compounds being confined almost exclusively to the dicotyledonous plants.

Among the angiosperms six families are noteworthy for their alkaloidal content:

1. The *Apocynaceae,* "dogbane" (tropic and subtropic).
2. The *Leguminosae,* legumes.
3. The *Papaveraceae,* poppies.
4. The *Ranunculaceae,* buttercup and crowfoot.
5. The *Rubiaceae,* madder.
6. The *Solanaceae,* potato, tomato, nightshade.

Other families, such as the mints, roses, and orchids, sometimes, though rarely, contain alkaloids. They may occur in cell sap (opium), leaves (coca), stems, fruits (piperine, black pepper), seeds (*Nux vomica*), bark (quinine), roots (berberine in barberry roots), etc. Rarely does one alkaloid occur alone; usually two or three or more occur together; opium is known to contain at least 20, and new ones are still being isolated from opium.

The alkaloids are usually solid and crystalline, but a few, like nicotine and coniine, are liquids. They are mostly colorless, although a few are yellow. They rarely occur free in the plant tissue, but as salts of organic acids, *e.g.*, malic, citric, succinic, oxalic, tannic, quinic, meconic, or aconitic. In several groups of plants special alkaloids occur with special acids. Thus, the aconite alkaloids occur combined with aconitic acid, the opium alkaloids with meconic acid, and the cinchona alkaloids with quinic acid. Alkaloids readily form crystallizable salts with inorganic acids and are extracted from plant tissues by dilute sulfuric acid or hydrochloric acid. Certain of the alkaloids are volatile and may be steam-distilled from alkaline media; the non-volatile ones are set free by sodium hydroxide and extracted with ether, chloroform, etc., or may be absorbed on Lloyd's reagent.

McNair,[49] in a series of papers directed toward a study of plant phylogeny, has surveyed the distribution and physical properties of the acids, alcohols, glycerides, essential oils, and alkaloids, as they occur in various plant families, with particular reference to the climatic distribution of the compounds. He finds that alkaloids occur in 57 of the 295 families of angiosperms and gymnosperms. Forty-four per cent of the families are mostly tropical; 14 per cent are wholly temperate. Two of the 5 families of the gymnosperms, 5 of the 45 families of monocotyledonous plants, and 44 of the 241 families of the dicotyledonous plants contain alkaloids. The same alkaloid is seldom found in different plant families. On the other hand, a particular alkaloid may be often found in a number of members of a single plant family. When different alkaloids occur in the same plant family, the alkaloid is usually confined to a single genus. In a study of the tropical alkaloids, McNair concludes that the higher the plant is in the scale of evolutionary development, the higher will be the molecular weight of the alkaloid which the family contains. He further points out that the melting point of the tropical alkaloids is, in general, higher than the melting point of the subtropical alkaloids, and these in turn have a higher melting point than those of the temperate alkaloids. A similar observation holds for the other compounds (glyc-

[49] J. B. McNair, *Am. J. Botany*, **18**, 416 (1931); **21**, 427 (1934); *Bull. Torrey Botan. Club*, **62**, 219, 515 (1935).

erides, acids, etc.) investigated, and he suggests that chemical studies of plant constituents may be an aid in tracing evolutionary developments.

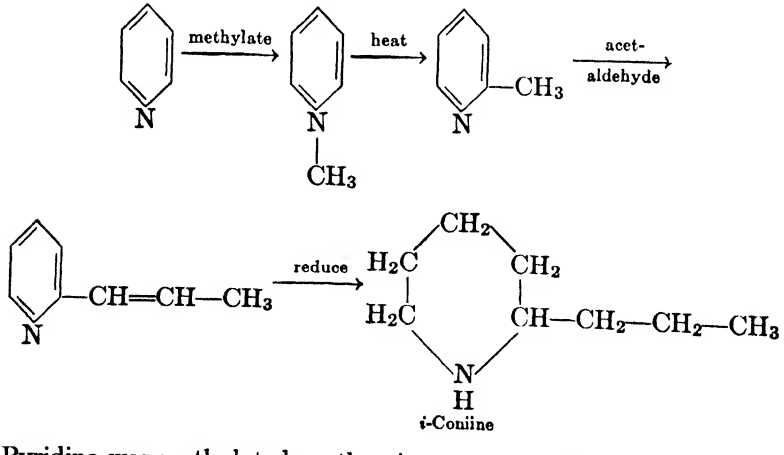
Various tests are used to detect the presence of alkaloids. Among the more common are: iodine in potassium iodide forming a yellow-brown precipitate (Wagner's reagent), platinic chloride forming a crystalline double salt which can be analyzed for platinum content, auric chloride forming a crystalline aurichloride (a similar crystalline double salt), and lead acetate yielding a precipitate.

Most alkaloids are bitter, but this is not a necessary property, for piperine, from black pepper, is tasteless. (The pungent taste of pepper is *not* due to its alkaloidal content but to an essential oil.)

Many of the alkaloids contain an asymmetric carbon atom and are accordingly optically active. Most of them are *levorotatory*. *Dextro*-tartaric acid is commonly used to resolve racemic mixtures, following the same technic as already noted for resolving amino acid mixtures.

The Pyridine Group. This group includes nicotine (tobacco), coniine (hemlock), piperine (black pepper), etc.

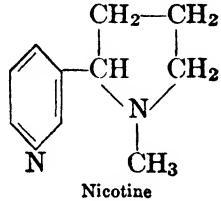
Coniine is α-*n*-propylpiperine. This was the first alkaloid to be synthesized, Ladenberg, in 1886, accomplishing the synthesis.



Pyridine was methylated on the nitrogen, forming *n*-methylpyridine, which when heated caused a migration of the methyl group to the α-position. The α-picoline so formed condensed with acetaldehyde to form α-allylpyridine, which on reduction yielded inactive coniine. The racemic mixture was separated by crystallization with *dextro*-tartaric acid. It would appear as if the synthesis could have been easily accomplished by adding a normal propyl group at the stage where the $-CH_3$ group was added to the pyridine. Ladenberg attempted to do this and found that, when the alkyl radical shifted from the nitrogen to

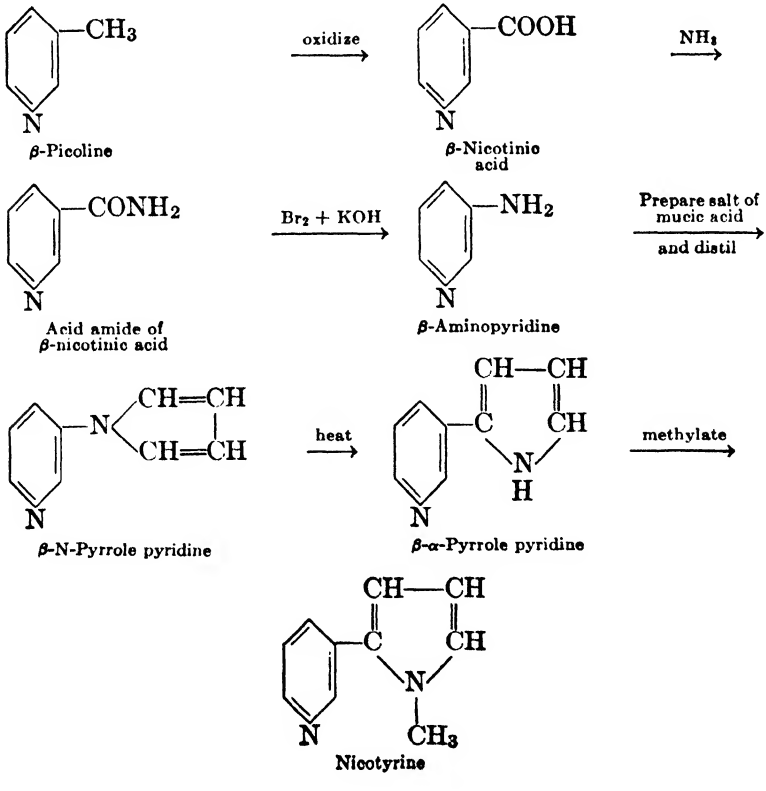
the α-carbon, it became an isopropyl group, thus resulting in the synthesis of an isomer of coniine. From the historical standpoint coniine is an interesting alkaloid, inasmuch as Socrates was supposed to have died from the effect of coniine in his drink of the deadly hemlock.

Nicotine, present in tobacco, is β-(*n*-methylpyrrolidine) pyridine. This alkaloid was synthesized by Amé Pictet during the period 1895–



Nicotine

1904, starting from β-picoline (β-methylpyridine) which is formed when glycerol, an organic nitrogen compound, and P_2O_5 are dry-distilled. The mechanism of the reaction is unknown. The various steps in Pictet's synthesis are as follows:

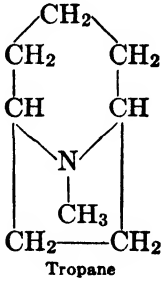


Nicotyrine treated with iodine forms the diiodo compound which reduces with zinc and potassium hydroxide to the dihydro compound. The dihydro compound with bromine gives the dibromo inactive nicotine, and the bromine can be removed from this by reducing with tin and hydrochloric acid, forming *inactive nicotine* which can be separated into its optical isomers. The trick in this synthesis was the reduction of the nicotyrine, reducing the pyrrole ring without reducing the pyridine ring. Solving the problem of reduction required nearly ten years of research.

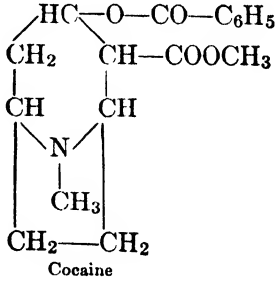
Although nicotine normally constitutes 95 per cent or more of the alkaloids of *Nicotinia tobacum*, this does not always hold true. In low-nicotine strains which have been developed by plant breeding for milder smoking, *nornicotine* may be the predominating alkaloid.[50] Structurally, nornicotine differs from nicotine in having no methyl group attached to the pyrrolidine ring. As a contact insecticide it is about as effective as nicotine, but its toxicity to animals is less than that of nicotine.

The Pyrrolidine Group. The pyrrolidine group includes relatively few alkaloids. Only one of these is of any interest to us, *i.e.*, *stachydrine* which has already been mentioned as the betaine of proline.

The Tropane Group. Compounds in the tropane group contain a ring composed of a 6-membered piperidine ring and the 5-membered pyrrolidine ring, and then can be regarded as derivatives of the heterocycle *tropane*. These alkaloids are found in the *Solanaceae* and in cocoa and are represented by *atropine, hyosine, hyoscyamine, cocaine*, etc.



Tropane

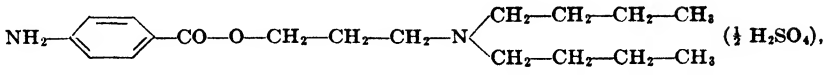


Cocaine

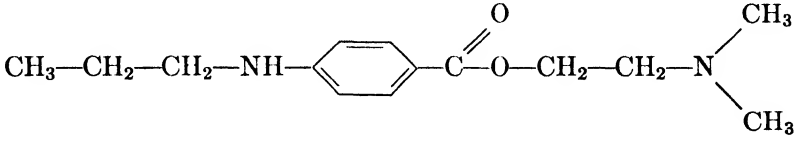
Cocaine has been synthesized by Willstätter. Attempts to synthesize compounds having the desirable properties have yielded *novocaine* $(C_2H_5)_2N—CH_2—CH_2—O—CO—C_6H_4—NH_2(p)$, which is only one-seventh as toxic as cocaine. The local anesthetic properties of cocaine lie in the benzoic acid ester and nitrogen portion. Undesirable properties

[50] L. N. Markwood, *Science*, **92**, 204 (1940).

accompany the methyl ester group. Novocaine cannot be used as a surface anesthetic. *Butyn,*



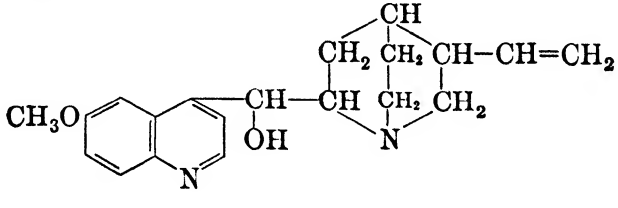
however, is fairly efficient in this regard. *Pantocaine,* another synthetic local anesthetic, and one which has had considerable use in spinal anesthesia, apparently has the structure



For injection into the body it is desirable to have a water-soluble, neutral material; therefore, in the synthetic products the p-NH_2 group is added to the benzene nucleus. The hydrochloride of the synthetic compounds containing the p-amino group is essentially neutral.

The Quinoline Group. This includes the alkaloids *quinine, cuprine,* and *cinchonine.* Of these alkaloids, only quinine has been synthesized [51]; therefore the exact structures of the others are not certainly known.

Quinine has been proved to be



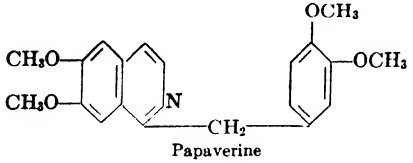
If the $-OCH_3$ on the quinoline nucleus is replaced by —OH, the compound is *cuprine;* if —H is on the quinoline nucleus, the compound is *cinchonine. Ethyl hydrocuprcine* has $-OC_2H_5$ on the quinoline nucleus, and the $-CH{=}CH_2$ group is hydrogenated to $-CH_2{-}CH_3$. It has very desirable properties as a specific for pneumococcus meningitis, but the lethal dose is dangerously close to the curative dose, so that it is used only in an emergency.

For many years quinine has been extensively used as an antimalarial drug. Prior to and during the recent world conflict intensive research was conducted on the synthesis of other even more effective antimalarials. Many compounds have been prepared and tested in this connection. Two which have been produced on a large scale and ex-

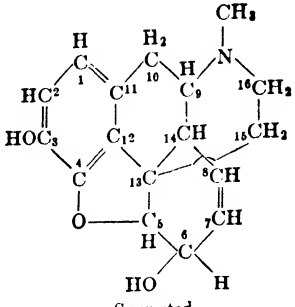
[51] R. B. Woodward and W. E. Doering, *J. Am. Chem. Soc.,* **67,** 860 (1945).

tensively used are *plasmochin*, a derivative of quinoline, and *atabrine* (or quinacrine), containing the acridine nucleus.

The Isoquinoline Group. The isoquinoline alkaloids occur mainly in the opium series.[52]



Papaverine



Suggested formula for morphine ($C_{17}H_{19}NO_3$)

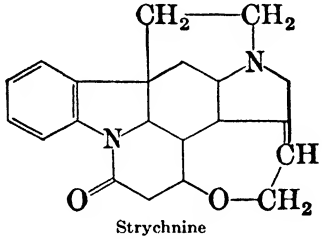
Papaverine is one of the simplest of these; *morphine* is probably one of the most complex.

Small notes that this formula for morphine appears to satisfy all the known reactions and decomposition products of morphine and that it may be accepted provisionally as indicating the structural relations of the more complex opium alkaloids. The problem of the opium alkaloids is complicated by the fact that there are a number of asymmetric carbon atoms in the molecule. Thus, in morphine there are five asymmetric carbon atoms, and Small notes that carbons -5, -6, and -9 are levorotatory and carbon-13 and carbon-14 are dextrorotatory. Therefore, the difficulties of attempting the synthesis of such a complicated structure are almost insuperable because the molecule contains four 6-membered rings and one 5-membered ring, and even if such a structure were synthesized, one would still face the problem of securing the proper space configurations on five asymmetric carbon atoms.

In discussing the phenomenon of molecular orientation and the influence of specific molecular configuration on orientation at interfaces, we have already referred to the change in physiological properties brought about by the conversion of the —OH group on the benzene ring of morphine to a carbonyl group and the simultaneous conversion of the benzene ring into a dihydrobenzene derivative. This synthetic compound, *dihydromorphinone* (dilaudid), is much more effective for the relief of pain, and the psychic and habit-forming effects of morphine are greatly reduced.

[52] L. F. Small and R. E. Lutz, *Chemistry of the Opium Alkaloids*, Supplement 103 to *U. S. Pub. Health Repts.*, Government Printing Office, Washington, D. C. (1932).

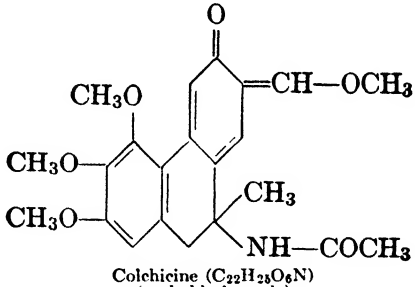
Another group of alkaloids, of which *strychnine* and *brucine* are the two best-known members, contains a highly intricate ring structure with two heterocyclic nitrogen atoms. Neither strychnine nor brucine has been successfully synthesized, but the available evidence points strongly to the following structure for strychnine:



Strychnine

Brucine is dimethoxystrychnine.

One unusual alkaloid deserves special mention. This is *colchicine*, an alkaloid obtained from *Colchicum autumnale*. This alkaloid is interesting from two standpoints. In the first place it does not contain a nitrogen in a heterocyclic ring but is in reality an acetylated amine of a substituted tetrahydrophenanthrene.



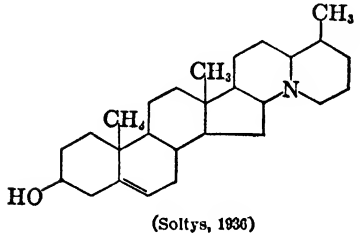
Colchicine ($C_{22}H_{25}O_6N$)
(probable formula)

Its interest from the standpoint of biology lies in the fact that Blakeslee [53] found that when plants are treated with colchicine there is a doubling of the chromosome number so that tetraploids are produced from diploids. Tetraploid plants have in each organ approximately the same number of cells as diploid plants. The doubling of the chromosomes, however, results in the approximate doubling of the cell volume. Similarly octaploid plants having four times the chromosome number of diploid plants would have cell volumes approximately four times as great as those of the diploid type. The implications of polyploidy for

[53] A. F. Blakeslee, *Compt. rend.*, **205**, 476 (1937); A. F. Blakeslee and A. G. Avery, *J. Heredity*, **28**, 393 (1937).

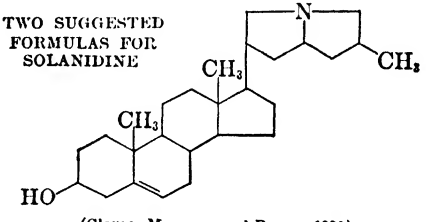
the horticulturist are obvious, for the flowers and fruits of the polyploid varieties would be much larger than those of the normal type. Colchicine was the only one of many chemical compounds tested which induced a doubling of the chromosome number.

Another compound, which may be considered as intermediate between the true alkaloids and other compounds containing a basic nitrogen group, is *solanine*. Solanine occurs in the potato, and on acid hydrolysis is reported to yield one molecule, each, of dextrose, galactose, and rhamnose, and a base, *solanidine* ($C_{27}H_{43}ON$). Solanine has been re-investigated [54] by the dehydrogenation of solanidine from potatoes. A hydrocarbon, methylcyclopentenophenanthrene or Diels hydrocarbon, was obtained, indicating that solanidine contains in part the same nucleus which is characteristic of the sterols, the sex hormones, and the digitalis glycosides and saponins. On the basis of the best available evidence, solanidine has one or the other of the following formulas:



(Soltys, 1936)

TWO SUGGESTED FORMULAS FOR SOLANIDINE



(Clemo, Morgan, and Raper, 1936)

Solanine is, therefore, possibly an alkaloidal glycoside. Solanine is obtained commercially from the juice of potato sprouts, and nearly every year workers in agricultural experiment stations receive reports of the deaths of farm animals which on checking back are found to be due to the feeding either of potato sprouts or of sprouted potatoes. Under ordinary conditions, potatoes seldom contain enough to produce toxic effects, but certain strains of potatoes contain appreciable amounts. "Sunburned" potatoes are nearly always bitter and contain an increased quantity of solanine. Death may result, if a considerable quantity of such tubers are eaten,[55] especially if they are baked and eaten "in their jackets." Potatoes normally contain about 0.024 part of solanine per 1,000, but sunburned potatoes have been found to contain as high as 0.588 part per 1,000. Approximately 70 per cent of the solanine is removed in the parings.

[54] A. Soltys and K. Wallenfels, *Ber.*, **69B**, 811 (1936).

[55] A. Bömer and H. Mattis, *Z. Nahr. Genussm.*, **45**, 288 (1923); **47**, 97 (1924); *cf.* also J. C. Rothe, *Z. Hyg. Infektionskrankh.*, **88**, 1 (1919); A. Behre and H. Ehrecke, *Chem. Ztg.*, **42**, 593 (1918).

The Origin of Alkaloids in Plants. Obviously the question of origin and purpose of such compounds as the alkaloids, in the economy of a plant, is a question which will probably always remain a subject of speculation. Inasmuch as they do not occur in all forms of plant life, it is obvious that they are not essential to the life process of a plant. The great variability in amount from season to season and from locality to locality also indicates that they may possibly be regarded as by-products of the synthetic and metabolic activities of the plant. This is the view of Pictet, who believes that the nitrogen residues from protein utilized by the plant in its metabolic processes are resynthesized into alkaloids (by certain plants which possess such synthetic power), because in this form the alkaloids are less harmful to the plant than the direct decomposition products of the proteins. Plants do not excrete their waste products, so do not synthesize any considerable quantities or urea or uric acid. The above view is obviously only a guess. We have no evidence as to the real significance of the alkaloids in the organs or tissues of a plant. It has been suggested that they are a form of protection, so that animals will not eat them, or, if they do eat them, they will be "punished" and die. Such a view, however, attributes rather too much intelligence to a plant!

The alkaloidal content may be increased or decreased by appropriate selection of strains, by plant breeding, and by fertilizing (especially nitrogen and phosphorus). Collecting at the proper time is also an important factor. Thus, the first lancings of the poppy capsules yield an opium containing a very much greater morphine content than the later lancings.[56]

The nicotine content in tobacco has been increased by plant breeding and manuring. This is desirable in certain instances in order to secure a high nicotine content for spray manufacture.

[56] H. E. Annett, *Biochem. J.*, **14**, 618 (1920); *Mem. Dept. Agr. India, Chem. Series*, **6**, 61 (1921); (and H. D. Singh) *ibid.*, **8**, 27 (1925).

III

CARBOHYDRATES AND RELATED SUBSTANCES

Für das Studium der chemischen Prozesse im Tier- und Pflanzenkörper ist nächst den Eiweisskörpern keine Gruppe von Kohlenstoffverbindungen so wichtig, wie die Kohlenhydrate, und als Nahrungsmittel nehmen sie unstreitig die erste Stelle ein.

EMIL FISCHER (1890)

CHAPTER 22

Introduction to the Chemistry of the Carbohydrates

The carbohydrates are especially prominent constituents of plants. They comprise about 75 per cent of the dry weight of most plants and serve not only as a source of available energy (sugars) but also as reserve food (starch) and as structural materials (cellulose). Carbohydrates form the direct link between the radiant energy emitted by the sun and the energy manifested by living plant and animal tissues. They are probably the first of the three great groups of food substances (proteins, fats, and carbohydrates) to be synthesized in the plant from simple inorganic substances, and in the long run most of the energy of living matter comes through them. Quantitatively, carbohydrates are the chief foodstuffs of animals and are their most important source of energy. Carbohydrates, such as lactose and glycogen, are synthesized by the animal organism. Two vitamins, ascorbic acid (vitamin C) and inositol, are structurally related to common carbohydrates, and other members of the group play unique functions in intermediary metabolism.

Biochemically, interest in the carbohydrates centers around glucose. It is the form in which carbohydrates are principally transported by plants and animals, and it is a structural unit in such important carbohydrates as sucrose, lactose; maltose, starch, glycogen, and cellulose. It is therefore not surprising that the structure of glucose has been very extensively investigated; indeed, the growth of our knowledge concerning the molecular structure and stereochemical configuration of this and other sugars is one of the most interesting chapters in the history of organic chemistry.

Plant sources of carbohydrates comprise the primary raw materials for the manufacture of starch, pulp and paper, and rayon, as well as for the preparation of many industrially important chemicals, such as ethanol, butanol, glycerol, citric acid, and lactic acid, through fermentations brought about by various microorganisms.

CHEMICAL NATURE AND BASIS OF CLASSIFICATION

The name, carbohydrate, originated from the fact that the com, pounds of this class which were first analyzed by early French chemists-

such as glucose, sucrose, and starch, could be represented by the formula $C_m(H_2O)_n$, and hence were thought to be hydrates of carbon. The French term, *hydrate de carbone*, was changed in the German to *Kohlenhydrat*, and this in turn was translated into English as carbohydrate. The name has long been known to be quite inappropriate. Aside from the fact that the hydrogen and oxygen of sugars and related substances are not present as loosely bound water (as, for example, in the compound $CuSO_4 \cdot 5H_2O$), some sugars, such as rhamnose ($C_6H_{12}O_5$) and rhamnoheptose ($C_7H_{14}O_6$) do not possess the general composition $C_m(H_2O)_n$, and yet they have the chemical properties of carbohydrates; also, compounds, such as formaldehyde, acetic acid, and lactic acid, which differ in chemical properties from the carbohydrates, have the empirical formula CH_2O. Despite the unsuitability of the name it is still retained to designate the sugars and those substances which resemble them in structure and properties.

Method of Classification. The carbohydrates are conveniently classified into two main groups according to complexity: (1) the *sugars* and (2) the *polysaccharides*. The sugars are crystalline, sweet-tasting substances which form true solutions in water and are also soluble in aqueous ethanol; the most complex contain 24 carbon atoms. In contrast, the polysaccharides are usually amorphous, are tasteless, have a high molecular weight, and are insoluble in ethanol; those which can be dispersed in water form colloidal solutions. The sugars are further classified into two major divisions: (1) simple sugars or *monosaccharides* and (2) *oligosaccharides* (complex or compound sugars).

The simple sugars are called monosaccharides because they contain only one saccharide, or sugar, unit and cannot be hydrolyzed. Their chemical properties reveal that they are *aldehydic or ketonic derivatives of polyhydric alcohols* and usually contain a hydroxyl group adjacent to the carbonyl group. They are subclassified into dioses, trioses, tetroses, pentoses, and hexoses according to the number of oxygen (also, usually carbon) atoms present and, still further, into aldoses and ketoses depending on whether they contain an aldehyde or ketone group. The relation of the simple sugars to the polyhydric alcohols is exemplified below:

(1) Dihydric Alcohol

$$\begin{array}{l} CH_2OH \\ | \\ CH_2OH \end{array}$$

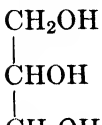
Ethylene glycol

Aldodiose

$$\begin{array}{l} CHO \\ | \\ CH_2OH \end{array}$$

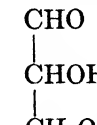
Glycollic aldehyde or glycolose

(2) Trihydric Alcohol



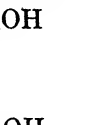
Glycerol

Aldotriose



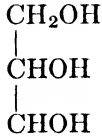
Glycerose or glyceric aldehyde

Ketotriose



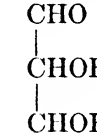
Dihydroxyacetone

(3) Tetrahydric Alcohol



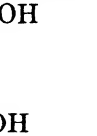
Erythritol

Aldotetrose



Threose and erythrose

Ketotetrose



Erythrulose

The oligosaccharides, complex or compound sugars, are condensation products of 2, 3, or 4 molecules of simple sugars and are subclassified into *disaccharides*, *trisaccharides*, and *tetrasaccharides* according to the number of molecules of simple sugar yielded upon hydrolysis.

The condensation of simple hexose sugars would yield the following complex sugars:

$$2\ C_6H_{12}O_6 - H_2O = C_{12}H_{22}O_{11}, \text{ a disaccharide}$$

$$3\ C_6H_{12}O_6 - 2H_2O = C_{18}H_{32}O_{16}, \text{ a trisaccharide}$$

$$4\ C_6H_{12}O_6 - 3H_2O = C_{24}H_{42}O_{21}, \text{ a tetrasaccharide}$$

The compound sugars are in reality "sugar-like" polysaccharides but the prefix *poly* is reserved for carbohydrates which upon hydrolysis yield a very large but indefinite number of simple sugars per molecule. Freudenberg has introduced the term *oligosaccharides* to represent the compound or complex sugars as a class distinct from the "non-sugar" polysaccharides. The disaccharides and other groups of oligosaccharides are subclassified into reducing and non-reducing sugars, depending on whether or not the carbonyl group is potentially functional.

The polysaccharides comprise a very heterogeneous group of compounds which yield a large and indefinite number of monosaccharide units on hydrolysis and are difficult to classify. One common system of classification is based primarily upon the class of monosaccharide yielded on hydrolysis; thus polysaccharides which yield a pentose sugar

on hydrolysis are called *pentosans*, $(C_5H_8O_4)_n$, and those yielding a hexose sugar are called *hexosans*, $(C_6H_{10}O_5)_n$. Each major class is subdivided according to the particular pentose or hexose sugar produced. Some polysaccharides, however, yield both pentose and hexose sugars on hydrolysis and are classed as mixed pentosans and hexosans; others yield sugar acids upon hydrolysis and frequently pentoses and hexoses as well.

CLASSIFICATION OF CARBOHYDRATES

I. Simple Sugars
  1. Diose, $C_2H_4O_2$: *glycolaldehyde*
  2. Trioses, $C_3H_6O_3$
     a. Aldotriose: D- and L-*glycerose* (glyceric aldehyde)
     b. Ketotriose: *dihydroxyacetone*
  3. Tetroses, $C_4H_8O_4$
     a. Aldotetroses (4 possible isomers): D- and L-*erythrose;* D- and L-*threose*
     b. Ketotetroses: *erythrulose*
     c. Hydroxymethyltetrose: *apiose* (β-hydroxymethyltetrose, from the glycoside apiin)
  4. Pentoses, $C_5H_{10}O_5$
     a. Aldopentoses (8 possible isomers): D- and L-*arabinose;* D- and L-*xylose;* D- and L-*ribose;* D- and L-*lyxose*
     b. Methylpentoses (one of the hydrogen atoms of the primary alcohol group in an aldopentose is replaced by a methyl group): L-*rhamnose* (component of many vegetable glycosides); L-*fucose* (sugar residue in fucosan found in seaweed of *Fucus* variety);[1] *rhodeose* (occurs in certain glycosides)
     c. Methoxymethylpentose: *digitalose* ($C_7H_{14}O_5$, in the glycoside digitalin)
     d. Methylthiopentose: ($C_6H_{12}O_4S$, from yeast)
     e. Ketopentoses (4 possible isomers): L-*ketoxylose*[2]
     f. 2-Desoxypentoses: D-2-*desoxyribose* ($C_5H_{10}O_4$, from thymonucleic acid); D-2-*desoxymethylpentose* ($C_6H_{12}O_4$; digitoxose, from digitalis glycoside)
  5. Hexoses, $C_6H_{12}O_6$
     a. Aldohexoses (16 possible isomers; all known, but only those starred (*) commonly occur in nature): D*- and L-*mannose;* D*- and L-*glucose;* D- and L-*idose;* D- and L-*gulose;* D*- and L*-*galactose;* D- and L-*talose;* D- and L-*allose;* D- and L-*altrose*
     b. Ketohexoses (8 possible isomers, all known): D*- and L-*fructose;* D- and L*-*sorbose;* D- and L-*tagatose;* D- and L-*psicose* (allulose)
     c. Aminohexoses: *chitosamine* or *glucosamine* (2-aminoglucose); *chondrosamine* (2-aminogalactose)
  6. Heptoses,[3] $C_7H_{14}O_7$
     a. Aldoheptoses (none naturally occurring, those known being synthetic): *glucoheptose, mannoheptose, galactoheptose*

[1] E. P. Clark, *J. Biol. Chem.*, **54**, 65 (1922).

[2] P. A. Levene and F. B. La Forge, *J. Biol. Chem.*, **18**, 319 (1914). See also, I. Greenwald, *J. Biol. Chem.*, **88**, 1 (1930); **89**, 501 (1930); **91**, 731 (1931).

[3] See C. S. Hudson, *Advances in Carbohydrate Chem.*, **1**, 1 (1945), for a complete list and discussion of the higher simple sugars.

b. Ketoheptoses (2 naturally occurring): *sedoheptose* (syn. sedoheptulose, D-altroheptulose);[4] D-*mannoketoheptose* (syn. D-mannoheptulose),[5] from the avocado; D-*glucoheptulose* (synthetic); L-*perseulose* (L-galaheptulose), by the action of *Bacterium xylinum* on natural perseitol from seeds of avocado

7. Octoses,[3] $C_8H_{16}O_8$
   a. Aldoöctoses (none naturally occurring): *glucoöctose*, *mannoöctose*, *galactoöctose*
8. Nonoses,[3] $C_9H_{18}O_9$
   a. Aldononoses (none naturally occurring): *gluconononose*, *mannononose*
9. Decoses,[3] $C_{10}H_{20}O_{10}$
   a. Aldodecoses (none naturally occurring): *glucodecose*

II. Oligosaccharides or Compound Sugars

1. Disaccharides
   a. Pentose-hexose saccharides, $C_{11}H_{20}O_{10}$ or $C_5H_9O_4$—O—$C_6H_{11}O_5$ (reducing—one carbonyl group potentially functional): *glucoapiose* (in the glycoside apiin); *vicianose*, 6-[β-L-arabinopyranosyl]-D-glucopyranose, in *Vicia angustifolia; primeverose*, 6-[β-D-xylopyranosyl]-D-glucopyranose, in *Primula officinalis*
   b. Methylpentose-hexose saccharides, CH_3—$C_5H_8O_4$—O—$C_6H_{11}O_5$: *glycorhamnoside* (in the glycoside convolvulin); *rutinose* (a glucorhamnose occurring in rutin)[6]
   c. Dihexoses, $C_{12}H_{22}O_{11}$ or $C_6H_{11}O_5$—O—$C_6H_{11}O_5$
   Type I. Reducing (one carbonyl group potentially functional)
   C-3 dihexoses (ring attached to carbon-3 of sugar serving as alcohol): *turanose*, 3-[α-D-glucopyranosyl]-β-D-fructopyranose
   C-4 dihexoses (ring attached to carbon-4 of sugar serving as alcohol): *maltose*, 4-[α-D-glucopyranosyl]-D-glucopyranose; *lactose*, 4-[α-D-galactopyranosyl]-D-glucopyranose; *cellobiose*, 4-[β-D-glucopyranosyl]-D-glucopyranose
   C-6 dihexoses (ring attached to carbon-6 of sugar serving as alcohol): *gentiobiose*, 6-[β-D-glucopyranosyl]-D-glucopyranose; *melibiose*, 6-[α-D-galactopyranosyl]-D-glucopyranose
   Type II. Non-reducing (carbonyl groups not functional): *sucrose*, α-D-glucopyranosyl-β-D-fructofuranoside; *trehalose*, α-D-glucopyranosyl-α-D-glucopyranoside

2. Trisaccharides
   a. Type I. Reducing (one carbonyl group potentially functional)
   Methylpentose-hexose saccharides: *rhamninose*, galactose-rhamnose-rhamnoside; *robinose*, galactose-rhamnose-rhamnoside
   Trihexose saccharide: *mannotriose*, D-glucose-D-galactose-galactopyranoside
   b. Type II. Non-reducing (no carbonyl groups functional)
   Trihexoses: *raffinose*, 6-[α-D-galactopyranosyl]-α-D-glucopyranosyl-β-D-fructofuranoside; *melezitose*, 3-[α-D-glucopyranosyl]-β-D-fructopyranosyl-

[4] F. B. La Forge and C. S. Hudson, *J. Biol. Chem.*, **30**, 61 (1917).

[5] F. B. La Forge, *J. Biol. Chem.*, **28**, 511 (1917).

[6] M. C. Charaux, *Bull. soc. chim. biol.*, **6**, 641 (1924); M. J. Rabate, *ibid.*, **12**, 974 (1930).

α-D-glucopyranoside; *gentianose*, 6-[β-D-glucopyranosyl]-α-D-glucopyranosyl-β-D-fructofuranoside

3. Tetrasaccharides
   a. Type II. Non-reducing: *stachyose*, 6-[D-galactopyranosyl]-D-galactopyranosyl-α-D-glucopyranosyl-β-D-fructofuranoside

III. Polysaccharides[7]

1. Polysaccharides composed of only one type of sugar residue
   a. Pentosans: *e.g., araban*
   b. Methyl pentosans: *e.g., fucosan*
   c. Hexosans: *e.g., starch, cellulose, glycogen, inulin, mannan, galactan, lichenin, levan, dextran, laminarin*
2. Polysaccharides composed of more than one type of sugar unit
   a. Pentosans: *e.g., araboxylan*
   b. Hexosans: *e.g., galactomannan*
3. Polysaccharides composed of one type of uronic acid unit—polyuronides
   a. Pectic acid
   b. Alginic acid
4. Polysaccharides composed of aldose (pentose or hexose) and uronic acid units
   a. Plant gums: *e.g., gum arabic, damson gum, gum tragacanth*
   b. Mucilages: *e.g., linseed mucilage*
   c. Pectins
   d. *Pneumococcus* polysaccharides
5. Polysaccharides containing hexose units esterified with an inorganic acid
   a. Certain seaweed polysaccharides: *e.g., agar*
6. Polysaccharides containing amino sugar units
   a. Amino-glucosans: *e.g., chitin, mucosin*
   b. Amino-galactosans: *e.g., chondroitin*

INTRODUCTION TO STEREOISOMERISM OF THE SUGARS

With the exception of the simplest aldose (glycollic aldehyde, $CH_2OH \cdot CHO$) and the simplest ketose (dihydroxyacetone, $CH_2OH \cdot CO \cdot CH_2OH$), the carbohydrates contain one or more asymmetric carbon atoms so that a number of optical isomers is possible. This greatly complicates the study of carbohydrates. The notable achievement of the great German chemist, Emil Fischer (1852–1919) in synthesizing and establishing the configuration of the various aldopentoses and hexoses predicted by stereochemical theory, which had been enunciated by van't Hoff and Le Bel a few years previously (1874), is one of the outstanding chapters in the history of organic chemistry.

Optical Activity and Specific Rotatory Power. Biot, in 1815, discovered the phenomenon of *optical activity* or *optical rotation*. When a beam of polarized light is passed through crystals of certain inorganic compounds (quartz, sodium chloride), certain organic liquids (tur-

[7] This classification does not take into consideration those complex biological substances consisting of a polysaccharide in combination with a protein.

pentine) or solutions of some organic compounds, such as the sugars, the plane of polarization is rotated. Such substances are said to be *optically active.* If the plane of polarization is rotated to the right or in the clockwise direction, the substance is said to be dextrorotatory, and the rotation is positive; conversely, substances which rotate the plane of polarization to the left, or counterclockwise, are said to be levorotatory and to exhibit a negative rotation. The rotatory power of organic liquids and solutions is measured by means of a *polariscope* or *polarimeter.* For a description of the principle and methods of polarimetry the reader should consult a textbook of physical chemistry.

The angle of rotation of the plane of polarization produced by a solution of an optically active compound in the polariscope was shown by Biot in 1853 to depend on several factors including the nature of the substance, the length of the column of the solution through which the plane polarized light passes, the concentration of the solution, the nature of the solvent, the wave length of the light, and the temperature. When all the variables except the nature of the substance are fixed, the rotation is specific for any given substance and is called the *specific rotation* or *specific rotatory power.* It is usually determined with sodium light at a temperature of 20°C., and the specific rotation, designated by the symbol $[\alpha]_D^{20}$, in a given solvent is defined as the angle of rotation through which the plane of polarized light, corresponding in wave length to that of the D line (5893 Å.) of the solar spectrum, is rotated in passing through, at 20°C., a 1-decimeter column of the substance having a concentration of 1 gram per milliliter. That is,

$$[\alpha]_D^{20} = \frac{\alpha v}{lw} \tag{166}$$

where α is the observed angle of rotation; l the length of the column in decimeters; w, the weight of the substance in grams; and v, the volume of the solution in milliliters.

The *molecular rotation* of a substance is obtained by multiplying its specific rotation by the molecular weight. To avoid large numbers, it is customary to use one one-hundredth of the actual values.

Molecular Asymmetry. Chemical compounds which have the property of rotating the plane of polarized light fall into two classes:

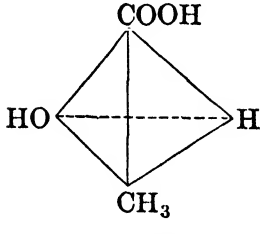
1. Crystalline substances, such as quartz and sodium chlorate, in which the optical activity is associated with crystal structure and disappears when they are fused or dissolved.

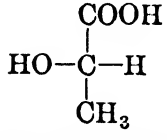
2. Organic and inorganic compounds which rotate the plane of polarized light whether in the solid, liquid, or gaseous state, or in solu-

tion. *The optical activity must therefore be due to some special structure of the molecule.*

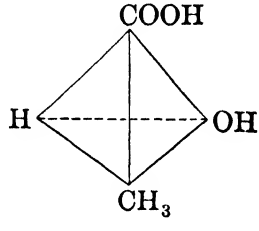
The optical activity of the great majority of carbon compounds, including the sugars, is of the latter class. Here the optical activity is associated with the presence of one or more *asymmetric* carbon atoms, *i.e.*, where the four carbon valences are satisfied with different atoms or groups. For every optically active compound an isomer exists which possesses the same general physical and chemical properties and differs only in the direction (*not* the extent) in which it rotates the plane of polarized light. Such pairs of isomers are called *enantiomorphs, antimers,* or *optical antipodes.*[8] Models of such isomeric pairs are mirror images of each other and cannot be superimposed.

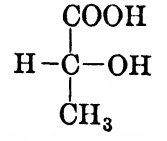
Lactic acid is a common example of a compound containing an asymmetric center, and the spatial and Fischer projection formulas of the two optical isomers are shown below.





Dextrorotatory lactic acid





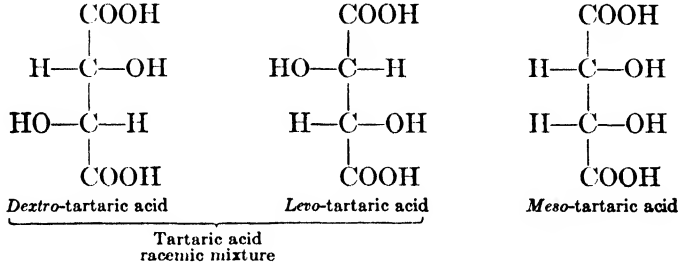
Levorotatory lactic acid

The dextrorotatory form, known as sarcolactic acid, is obtained from muscle tissue, and the levorotatory form is produced in the fermentation of lactose by *Bacterium laevalacticum* Migula (*Bacillus acidi levolactici*). When synthesized in the laboratory an equimolecular mixture of the *dextro* and *levo* forms, known as *racemic acid,* is obtained; this mixture is optically inactive because of the mutual or external compensation of the two constituents.

One of the above configurations has been labeled the dextrorotatory form, and the other the levorotatory form, but this is purely arbitrary as there is no present method of ascertaining which of the configurations actually represents the lactic acid which is dextrorotatory.

[8] An exception occurs in tartaric acid in which the crystalline forms of the enantiomorphs differ.

In compounds which contain two similar asymmetric carbon atoms, a fourth or *meso* form exists which is optically inactive owing to *internal compensation*. The isomers of tartaric acid are classical examples of molecules of this type.



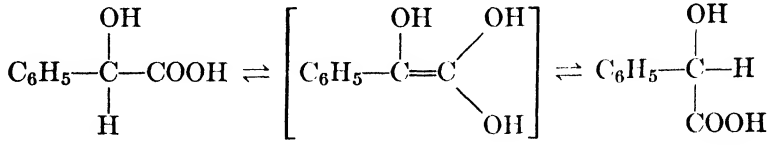
The two central carbon atoms are asymmetric and have the same groups attached to them. If the bottom half of either the dextro or levo forms is rotated through 180° in the plane of the paper, it will be seen that it has precisely the same configuration as the corresponding top half. Thus, if each asymmetric carbon atom in *dextro*-tartaric acid is represented by $A+$, we have the combination $A+A+$, and the compound is dextrorotatory; similarly in *levo*-tartaric acid we have $A-A-$, and the compound is levorotatory. In *meso*-tartaric acid, however, we have $A+A-$; the top half of the molecule is antipodal to the lower part, and consequently the rotations within the molecule neutralize each other. The compound is optically inactive owing to *internal compensation*, and it cannot be resolved into two optically active compounds.

The classical researches of Louis Pasteur (1822–1895) on the isomerism of the tartaric acids played an important part in the development of the theory of optical isomerism. Before his work only *dextro*-tartaric acid and a sodium ammonium salt of the racemic acid were known. He showed that the racemic salt was an equal mixture of a dextro and levo form; this was the first time that the resolution of an optically inactive compound into its component optically active parts had been achieved. He also discovered *meso*-tartaric acid.[9]

Racemization. Resolution of Racemic Mixtures. The term racemization is applied to the process which occurs when an optically active substance is one-half converted into its enantiomorph, thus producing an optically inactive mixture of equal quantities of the dextro

[9] The reader is referred to L. F. Fieser and M. Fieser, *Organic Chemistry*, D. C. Heath & Co., Boston, 1944, for a more detailed account of Pasteur's classical researches on the optical isomerism of the tartaric acids.

and levo isomers. For example, racemization of mandelic acid occurs when it is treated with alkali:



Alkalinity favors enolization, which involves the migration of the α-hydrogen to the oxygen atom and the formation of a double bond between the carbon atoms. This destroys the asymmetry of the α-carbon. As enolization is readily reversible, and in the reverse process either of the two bonds may open with equal ease, an equimolecular mixture of the dextro and levo forms results.

The selective racemization of carbon-2, adjacent to the aldehyde group, in an aldotetrose or higher simple sugar is termed *epimerization*. In these sugars two or more asymmetric centers are present. Since in epimerization the inversion is confined to the asymmetric center adjacent to the carbonyl group, it does not result in the production of enantiomorphs; the configuration of the other asymmetric carbons remains unchanged, the products differing only in the configuration of carbon-2. Aldose sugars related in this manner are called *epimers*, a term introduced by Votoček [10] in 1911.

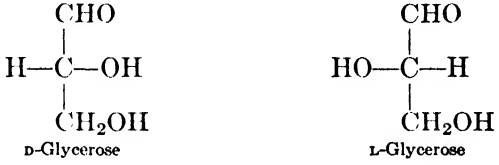
Racemic mixtures usually result when optically active substances are synthesized in the laboratory, and special methods must be employed to separate the optically active components, since enantiomorphs usually differ only in their optical activity. The methods described previously for resolving racemic amino acid mixtures (see p. 296) also apply to sugar enantiomorphs. The mechanical method of separation and the biochemical method are generally of little value; most useful is the chemical method in which the racemic mixture is combined with a second optically active substance. Thus a racemic acid *A* may be combined with an optically active base, such as an alkaloid *B*, of the levo form; two salts are formed, *levo-A-levo-B* and *dextro-A-levo-B*. These compounds are *diastereoisomers* and *not* enantiomorphs; they differ in optical activity and solubility and may be separated by fractional crystallization. Similarly basic enantiomorphs may be separated by the formation of diastereoisomeric salts with an optically active acid.

Optical Isomerism of the Aldose Sugars. Unlike tartaric acid, the terminal groups in the aldose sugars are not identical, and the formula is therefore not divisible into two equal halves. Each asym-

[10] E. Votoček, *Ber.*, **44**, 362 (1911).

metric carbon atom is different from the other since only two of the attached groups (H and OH) are the same. When two asymmetric centers, *A* and *B*, are present the following four combinations are possible: $A+B+$; $A+B-$; $A-B+$ and $A-B-$. All the isomers are optically active, and *internal compensation is impossible.* Similarly with three asymmetric centers eight isomers are possible. Thus, in compounds in which the terminal groups are different, the number of optically active isomers is 2^n, where n is the number of asymmetric centers.

Stereoisomerism in the aldose sugar series is conveniently illustrated by starting with the aldotriose sugar, glycerose, which contains one asymmetric center and accordingly may exist in two forms.



D-Glycerose L-Glycerose

By the successive addition of CHOH groups adjacent to the CHO group, the aldose sugar series is obtained.[11] Those derived from D-glycerose in this manner are shown in the chart on p. 528. As each new asymmetric center is added, an epimeric pair of sugars is obtained (D-erythrose and D-threose; D-arabinose and D-ribose; D-glucose and D-mannose). For each of the sugars listed there is an optical antipode or enantiomorph of equal but opposite specific rotatory power obtained in a similar manner from L-glycerose.

In the sugars, the carbon atoms are numbered serially beginning with the end of the chain nearest the carbonyl group; thus in aldoses the carbonyl carbon is carbon-1 and in ketoses it is carbon-2. As will be seen later, the higher aldoses are known to have a closed chain structure involving carbon-1 (the aldehydic carbon) and carbon-5 (usually). This introduces an additional asymmetric center (carbon-1) so that each of the aldopentoses and aldohexoses represented by the open chain formulas normally exists in two stereoisomeric forms, making a total of 16 aldopentoses and 32 aldohexoses.

Meaning of D and L in Carbohydrate Chemistry. Several of the sugars in this chart are levorotatory, yet they are designated as D sugars; accordingly, in carbohydrate chemistry, the prefix D or L *does not represent the direction of rotation* but rather the *configuration* of the asymmetric carbon atom farthest away from the reducing group and

[11] Increasing the chain length of an aldose sugar by introducing a CHOH group in this position is accomplished by the cyanohydrin reaction described on p. 560.

CONFIGURATION OF THE D-ALDOSES
D(+)-Glucose
D(+)-Mannose
D(+)-Allose
D(+)-Altrose
D(+)-Galactose
D(+)-Talose
D(−)-Gulose
D(−)-Idose
D(−)-Arabinose
D(−)-Ribose
D(−)-Lyxose
D(−)-Xylose
D(−)-Erythrose
D(−)-Threose
D(+)-Glyceric aldehyde

adjacent to the terminal CH_2OH group. When the OH group on this asymmetric carbon is represented on the right in the Fischer projection formula, the position which has been arbitrarily selected to represent the configuration of D-glycerose, the sugar is designated as a D sugar.

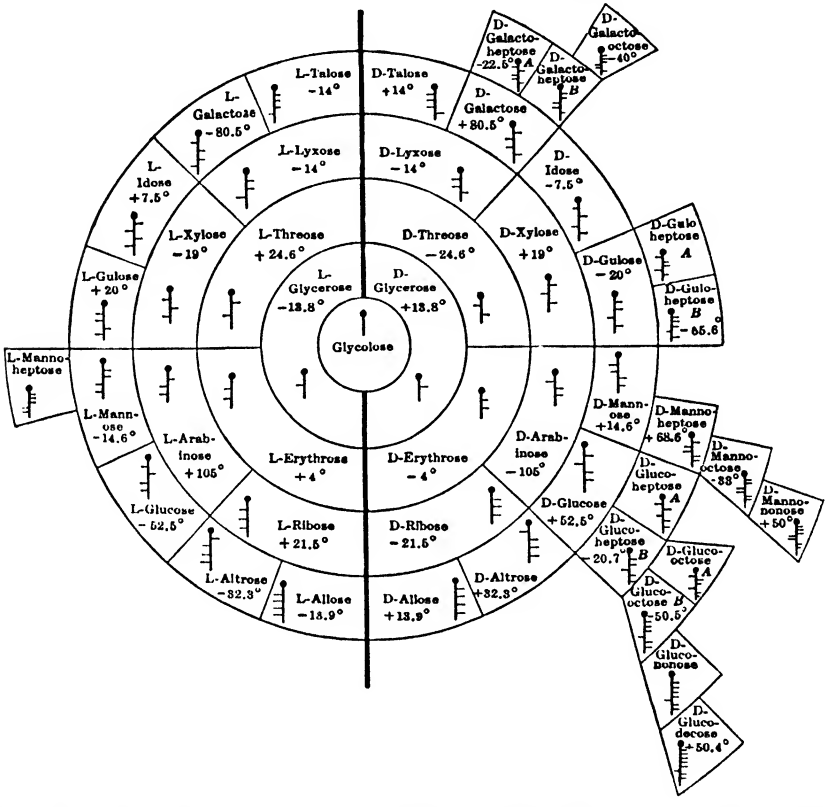


FIG. 97. The structural and configurational relationships of the aldose sugars, together with their specific rotatory powers. The aldehyde group is represented by a black dot, the carbon chain by a straight line; the lines perpendicular thereto represent the OH groups.

Conversely, all sugars derived from L-glycerose (in which the OH group on the penultimate carbon is represented on the left) are L sugars. This system was first proposed by Rosanoff,[12] in 1906. He emphasized that, in order to avoid ambiguities in designation, it is necessary to select glycerose, a sugar with only one asymmetric center as the reference sugar.[13]

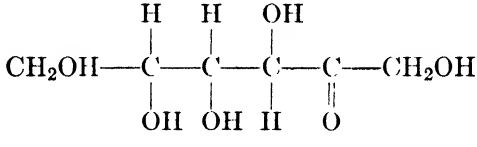
[12] M. A. Rosanoff, *J. Am. Chem. Soc.*, **28**, 114 (1906).

[13] Prior to the adoption of this system there was a great deal of confusion owing to the employment of the prefix *d* by Fischer to designate any sugar derived from *d*-

In the carbohydrate literature, the prefix *d* or *l* has been quite commonly used to designate the configuration of the penultimate carbon. However, it has also been used to indicate the direction of rotation. Thus to avoid confusion the letters D and L will be used here, a practice which has now been adopted by several scientific journals.

The structural and configurational relationships of the aldose sugars are represented diagrammatically in Fig. 97, which is based on a modification by Willaman and Morrow [14] of a diagram originally published by Rosanoff.[12] Glycolose is shown at the center; in each succeeding circle, the sugars which arise by the successive addition of primary alcohol groups between the first and second carbon atoms of the parent sugar are represented, the D-sugars being placed in the right semi-circle and the L-sugars in the left semi-circle.

Optical Isomerism of the Ketoses. In the ketose sugars, there is one less asymmetric center than in the aldoses of equal chain length; therefore the number of isomers is greatly reduced. As a class the ketoses are not so important biochemically, as the aldoses. D-Fructose,



which has a configuration similar to D-glucose, is the most common, but it is a structural unit of relatively few of the complex sugars and polysaccharides.

OXIDE RING STRUCTURE OF MONOSACCHARIDES

Thus far, the simple sugars have been considered as open-chain compounds but the reactivity of D-glucose, the phenomenon of mutarotation, and the existence of two methylglucosides were early recognized as incompatible with this view.

Reactivity. Although D-glucose exhibits many of the properties of common aldehydes its reactivity, when compared with hydroxyaldehydes such as glycollic aldehyde, is not so great as would be expected if it possessed the simple aldehyde formula. The cyanohydrin reaction proceeds

glucose. Later it became evident that, in synthesizing what Fischer termed *d*-gulose from *d*-glucose, the positions of the CHO and CH_2OH groups were interchanged. This was equivalent to inverting all the asymmetric centers and led to an error in his configurations for the D and L forms of threose, xylose, gulose, and idose.

[14] J. J. Willaman and C. A. Morrow, *J. Am. Chem. Soc.*, **45**, 1273 (1923).

with difficulty, and it fails to add sodium bisulfite or to give the Schiff test.[15]

Mutarotation. A freshly prepared solution of D-glucose exhibits a specific rotatory power of +110° which gradually decreases on standing to a constant value of +52.5°. This was discovered in 1846 by Dubrunfaut, who called the phenomenon birotation because the rotation had fallen by approximately one half. All the aldoses and D-fructose exhibit a change in rotation when freshly dissolved. Depending on the sugar, the rotation may increase or decrease, and this property is now called mutarotation, as suggested by Lowry.[16]

α- and β-Methylglucosides. Fischer,[17] in 1893, attempted to prepare the methylacetal of D-glucose by treating it with absolute methanol and hydrogen chloride. Instead of the usual reaction, $RCHO + 2CH_3OH \rightarrow RCH(OCH_3)_2 + H_2O$, he found that only one methyl group entered the molecule, *i.e.*, a hemiacetal instead of an acetal was obtained.

The following year, a second isomer was obtained by the same reaction.[18] The two compounds, designated as α- and β-methylglucosides by Fischer, are crystalline non-reducing substances which are hydrolyzed by acids and do not undergo mutarotation. They differ in crystalline form, melting point, solubility, specific rotatory power, and behavior toward the enzymes maltase and emulsin.[19,20]

	α-Methyl-D-glucoside	*β-Methyl-D-glucoside*
Melting point	165–166°C.	104°C.
$[\alpha]_D^{20}$	+159°	−34°
Action of maltase	Hydrolyzed	No action
Action of emulsin	No action	Hydrolyzed

When solid anhydrous glucose (chiefly α-D-glucose) is used to prepare the methylglucosides the equilibrium mixture contains 77 per cent of the α- and 23 per cent of the β-isomer. β-Methyl-D-glucoside may be isolated from the isomeric mixture of the two methanol condensation products after destruction of the α-form by treatment with maltase and fermentation of the liberated glucose with yeast.

[15] The Schiff test for an aldehyde involves the addition of the unknown to a solution of fuchsin which has been decolorized with sulfur dioxide. If an aldehyde is present, the red color is restored.

[16] T. M. Lowry, *J. Chem. Soc.*, **75**, 211 (1899).

[17] E. Fischer, *Ber.*, **26**, 2400 (1893).

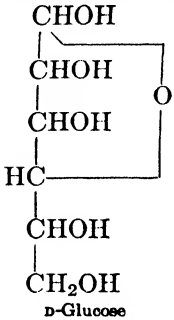
[18] W. A. van Ekenstein, *Rec. trav. chem.*, **13**, 183 (1894).

[19] E. Fischer, *Ber.*, **27**, 2985 (1894).

[20] E. F. Armstrong, *J. Chem. Soc.*, **83**, 1305 (1903).

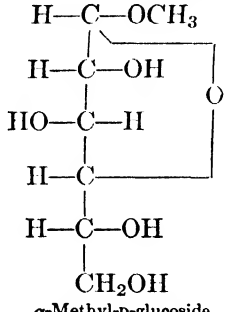
To explain these results, Fischer assumed a 1,4-oxide, γ-oxide, or butylene oxide ring structure for these compounds, an assumption which later proved to be incorrect. This structure had previously been proposed for glucose by Tollens [21] to explain mutarotation and its low reactivity.

Tollens' Formula

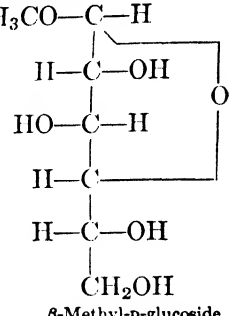


D-Glucose

Fischer's Formulas



α-Methyl-D-glucoside
$[\alpha]_D^{20} = 159°$



β-Methyl-D-glucoside
$[\alpha]_D^{20} = -34°$

The Tollens formula for D-glucose was derived by analogy with the γ-hydroxyacids which readily lose water to form internal anhydrides or lactones. With glucose no water is split out in the ring formation depicted, and the name *lactal* has been sometimes used to designate this tautomeric form. As D-glucose contains an aldehyde group and hydroxyl groups it might be expected to form an inner cyclic hemiacetal as depicted by Tollens. As a result of the ring formation carbon-1 is asymmetric and can exist in the D and L configurations. A spontaneous change of the one form into the other in solution to produce an equilibrium mixture of the two diastereoisomers would explain mutarotation.

Preparation of α- and β-D-Glucose. Tanret [22] provided experimental proof of this view by the isolation of two isomeric forms of glucose.[23] When crystallized from solution at ordinary temperatures, α-D-glucose, with an initial rotation when dissolved in water of +110° changing slowly to an equilibrium value of +52.5°, was obtained. The other, β-D-glucose, which was prepared by crystallization of the sugar from solution above 98°C., showed a rotation change in solution from an initial value of +19° to the equilibrium value of +52.5°.

The natural inference that the α- and β-D-methylglucoside are related to α- and β-D-glucose was simply and convincingly established in 1903

[21] B. Tollens, *Ber.*, **16**, 921 (1883).

[22] C. Tanret, *Bull. soc. chim.*, [3] **13**, 728 (1895).

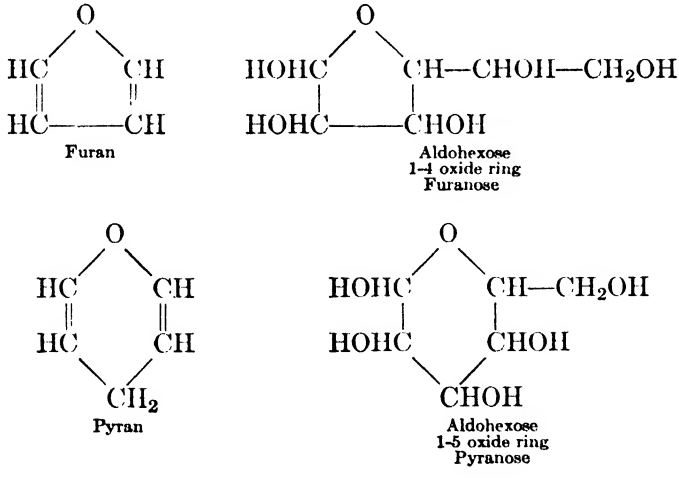
[23] Tanret thought he had also isolated a third isomer, γ-glucose, but this was proved to be crystals of an equilibrium mixture of α- and β-D-glucose.

by Armstrong.[20] He followed the changes in optical activity of the α- and β-methylglucoside as they underwent hydrolysis by maltase and emulsin, respectively. The glucose initially liberated from the α-methylglucoside was found to be of high rotation and that from the β-derivative of low rotation.

In depicting the configurations of α- and β-D-glucose on p. 532, the OH group was written on the right for the α-isomer. This is in accordance with the widely accepted rule proposed by Hudson [24] that in the D series the more dextrorotatory member of the α-β pair is designated α and the OH is assigned to the right in the Fischer projection formula. The reverse is followed in the L series; for example, the enantiomorph of the α-D-glucose is designated as α-L-glucose. Experimental evidence supporting this convention will be considered later.

Furanose and Pyranose Formulas. The relatively low reactivity of D-glucose, its ability to undergo mutarotation, and the existence of two methyl glucosides (and two pentaacetates) require the existence of a ring structure.

In the absence of chemical evidence to determine the point of ring closure, the 1-4 or butylene oxide ring was for many years considered the most plausible on the basis of the Baeyer strain theory. It was not



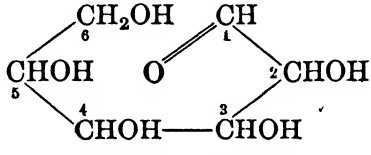
until 1926 that Haworth and his associates at the University of Birmingham, by means of methylation studies, secured chemical proof that D-glucose normally possesses a 6-membered ring, *i.e.*, a 1-5, δ-oxide, or amylene oxide structure. However, it and other sugars may also exist, under certain conditions, in the 5-membered ring form (1-4, γ-oxide

[24] C. S. Hudson, *J. Am. Chem. Soc.*, **31**, 66 (1909).

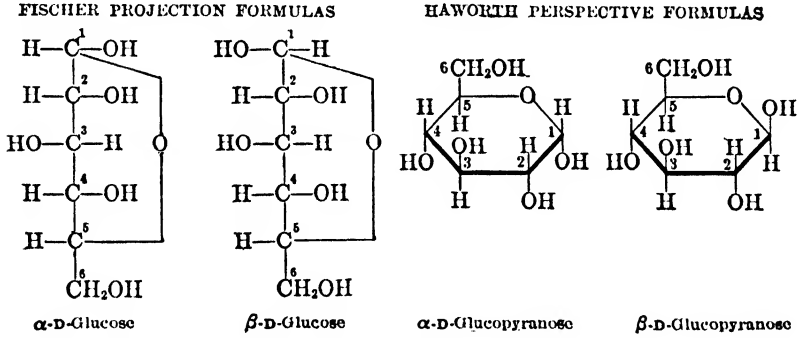
or butylene oxide ring). The proofs for these two ring structures will be given later (Chapter 23) but it is appropriate here to introduce the modern method used to represent and name them.

Haworth [25] pointed out the structural relationships of the 1-4 and 1-5 oxide formulas to furan and pyran, respectively, and suggested the names *furanose* and *pyranose* to designate them.

He emphasized that the open-chain projection formulas which represent the carbon atoms as lying in a straight line do not correctly depict the spatial relationships; because of the tetrahedral angle of 109° 28′ between adjoining carbon valences, the open-chain formula is more accurately represented as an open hexagon.



Such a formula brings the supposed aldehyde group into close proximity to the hydroxyl groups attached to carbon-4 and carbon-5, and it would therefore be anticipated that either of these carbon atoms might be involved in ring formation. Haworth proposed that *perspective formulas* be used to represent the ring forms of the sugars. The projection and perspective formulas for the α- and β-isomers of normal D-glucose (which, as previously stated, possess a 1-5 oxide ring structure) are:



In the Haworth perspective formulas, carbon atoms 1 to 5 and the ring oxygen are represented in a single plane perpendicular to the plane of the paper. The heavy lines at the bottom of the formulas represent the sides of the hexagon nearest the observer. The valences of the carbon atoms which are not involved in forming the ring stand above or

[25] W. N. Haworth, *The Constitution of Sugars*, Longmans, Green and Co., New York, 1929.

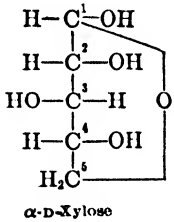
below it, and their projections lie on planes parallel to the plane of the paper. Thus, when the carbon atoms in the ring are numbered clockwise, the groups which are normally written on the right-hand side in the projection formula are depicted as projecting below the plane of the hexagon, and the corresponding groups on the left-hand side are projecting above. A correction has to be made in the relative positions of the groups on carbon-5, as represented by the Fischer projection formula, because ring formation requires the rotation of the carbon atom carrying the oxygen of the hydroxyl group which becomes involved in the ring through approximately the tetrahedral angle (109° 28′). This has the effect of bringing the hydrogen atom at carbon-5 on the opposite side of the carbon chain.

Unlike the Fischer projection formulas, which may be rotated at will in the plane of the paper, the rotation of the Haworth perspective formula through 180° in the plane of the paper results in the representation of the enantiomorph. However, the ring formula may be rotated about a central vertical axis in such a manner as to maintain the same groups above and below the plane of the ring.

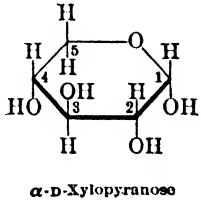
The heterocylic oxygen atom is not written at the top as it is convenient to have the aldehylic carbon at an angle in depicting the structures of the complex sugars.

The projection and perspective formulas for two aldopentoses and D-fructose follow. Since simple sugars have not been isolated in the

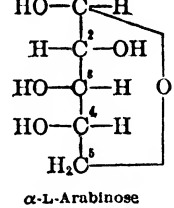
ALDOPENTOSES



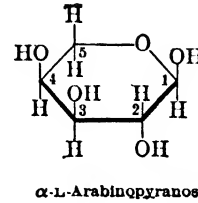
α-D-Xylose



α-D-Xylopyranose

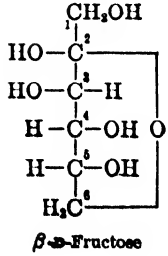


α-L-Arabinose

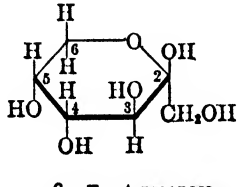


α-L-Arabinopyranose

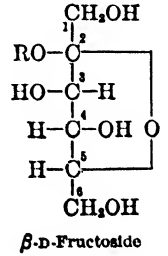
ALDOHEXOSES



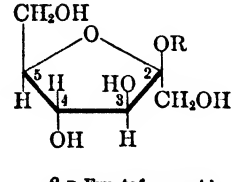
β-D-Fructose



β-D-Fructopyranose



β-D-Fructoside



β-D-Fructofuranoside

furanose form, this ring structure is shown for fructose with the active hydrogen of the potential reducing group substituted. The substituent R may be a glucopyranose unit as in sucrose, or a chain of other fructofuranose units as in the polysaccharide *inulin*.

α-, β-Isomerism. Mutarotation. As indicated previously, the optical rotation of the α- or β-isomer of a simple sugar in solution attains the same equilibrium value. The solubilities of the isomers differ, and, if the specific rotatory power of one form is known, that of the other can be calculated from the solubilities of the original form and that of the isomeric mixture at equilibrium, on the assumption that only two forms are present.[26] Application of this method to sugars available in both isomeric forms gives calculated results in good agreement with the actual values. Polarographic studies, however, strongly point to the existence of small amounts of other forms of sugars in aqueous solution.

The mutarotation of D-glucose follows the monomolecular law very closely, the velocity being greatly accelerated by acids and bases. The velocity constants are the same for the α- and β-forms, and hence mutarotation is a true reversible change.

In studying the effects of acids and bases on the velocity of mutarotation of D-glucose in water, Hudson [27] developed the following equation for the velocity constant:

$$K_{25^\circ} = 0.0096 + 0.258\,[H^+] + 9{,}750\,[OH^-] \qquad (167)$$

The catalytic effect of hydroxyl ions is therefore about 40,000 times greater than that of hydrogen ions. This is applied in the analysis of sugars by optical methods; the equilibrium is hastened by the addition of a few drops of ammonium hydroxide. From the above equation and the velocity of the mutarotation of D-glucose in pure water, Hudson [28] calculated the dissociation constant of water to be 1.0×10^{-14}, a result which is in excellent agreement with other methods.

When the *cis* and *trans* relations of the hydroxyl groups are considered, the *trans* form always predominates in the equilibrium mixture.[29] These calculations are based on the assumption that only α- and β-forms are present at equilibrium. Like D-glucose the mutarotations of the α- and β-forms of D-mannose, D-lyxose, and D-xylose follow the equation for a first-order reaction quite closely. On the other hand, the

[26] This method was developed by C. S. Hudson, *J. Am. Chem. Soc.*, **26**, 1065 (1904), and by T. M. Lowry, *J. Chem. Soc.*, **85**, 1551 (1904).

[27] C. S. Hudson, *J. Am. Chem. Soc.*, **29**, 1571 (1907).

[28] C. S. Hudson, *J. Am. Chem. Soc.*, **31**, 1136 (1909).

[29] W. N. Haworth and E. L. Hirst, *J. Chem. Soc.*, 1221 (1928).

mutarotations of D-galactose, D-talose, L-arabinose, D-ribose, and the heptoses of the galactose, talose, and idose series do not obey the monomolecular law, an indication that more than two isomers are present in the solutions at equilibrium.

Sugar	*trans Form*	*Proportion of trans Form at Equilibrium, per cent*
D-Glucose	β	66
D-Mannose	α	62
D-Galactose	β	69
D-Xylose	β	65
L-Arabinose	α	58
D-Lyxose	α	75

Evidence that the α-β-shift in aqueous solutions of D-glucose involves the opening of the ring to form an intermediate aldehydrol form has been obtained by Lowry [30] and Lowry and Faulkner.[31] Mutarotation requires a catalyst, and the sugar must receive a proton from a proton donor and yield a proton to a proton acceptor. Thus, it only takes place in an amphoteric solvent, such as water which is both a proton donor and acceptor. It was shown that mutarotation does not take place in pure dry ortho-cresol (a proton donor) or in pure dry pyridine (a proton acceptor), but in a pyridine-cresol mixture the velocity of mutarotation was many times greater than in water.

D-Fructose is the only known ketose which exhibits any appreciable mutarotation. The α-form has an initial rotatory power of −63.6° and the β-form of −132.5°, both giving an equilibrium value of −93.8° (15°C.). The equilibrium value decreases very markedly with increasing temperature, whereas that of D-glucose is not materially affected, an indication that the mechanisms involved are different. From studies of the catalytic effect of acids and bases, of the temperature coefficients for the reaction rates, of heats of reaction, and of other physical constants, Isbell [32] concludes that the mutarotation of D-fructose consists of a pyranose-furanose ring change. This view is supported by the fact that the mutarotation of D-fructose liberated from sucrose by enzymic hydrolysis occurs at the same rate as the mutarotation of a solution prepared from crystalline D-fructose, but in the opposite direction. Since the fructose residue in sucrose is known to exist in the furanose form and crystalline D-fructose is known to be a pyranose, the mutarotation of fructose liberated from sucrose must involve a partial con-

[30] T. M. Lowry, *J. Chem. Soc.*, **127**, 1371 (1925).
[31] T. M. Lowry and I. J. Faulkner, *J. Chem. Soc.*, **127**, 2883 (1925).
[32] H. S. Isbell and W. W. Pigman, *Bur. Standards J. Research*, **20**, 773 (1938).

version to the pyranose form; conversely, the mutarotation of a solution prepared from D-fructose must involve a pyranose-to-furanose change.

QUANTITATIVE RELATIONS BETWEEN ROTATORY POWER AND CONFIGURATION. RULES OF OPTICAL ROTATION

Extensive studies of the numerical values of the optical activities of sugars and their derivatives by Hudson and his associates have played a large part in the history of sugar chemistry.

Hudson's Rules of Isorotation. One phase of these investigations consisted in testing the principle of optical superposition formulated by van't Hoff which stated that *the optical rotation of a substance is the algebraic sum of the contributions of the several asymmetric centers.* In applying this rule to the sugars and their derivatives Hudson [33] postulated that their molecular rotation depends on two factors: (1) the rotation due to the potentially reducing group (carbon-1 in aldoses) which he termed A, and (2) the rotation due to the remainder of the molecule, designated B. For the D-series, the molecular rotation would be due to $A+B$ for the α-form and to $-A+B$ for the β-form. From a study of the optical activities of a large number of sugars and their derivatives, Hudson formulated his well-known rules of isorotation which may be stated as follows:

Rule 1. The rotation of carbon-1 is affected in only a minor degree by changes in the structure of the rest of the molecule. That is, the difference ($2A$) between the molecular rotation of the α- and β-isomers will be a constant, characteristic of the nature of the hydroxyl or substituent group on carbon-1.

Rule 2. Changes in the structure of carbon-1 affect in only a minor degree the rotation of the remainder of the molecule; *i.e.*, the sum ($2B$) of the molecular rotations of the α- and β-isomers of a sugar and its derivatives will be a constant characteristic of the particular sugar.

Although the van't Hoff law must be regarded merely as an approximation, there were many instances in which the rule held quite well. Mannose, lyxose, and rhamnose did not obey rule 2, nor the phenyl radical rule 1.

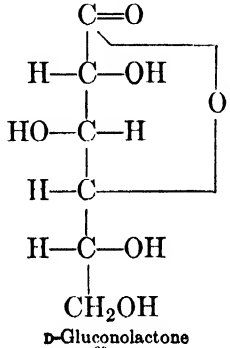
Hudson's Lactone Rule. In a study of the optical rotatory power of the lactones of twenty-four aldonic acids, Hudson [34] observed that the sign of rotation of the lactone was without exception determined by the

[33] C. S. Hudson, *J. Am. Chem. Soc.*, **31**, 66 (1909).

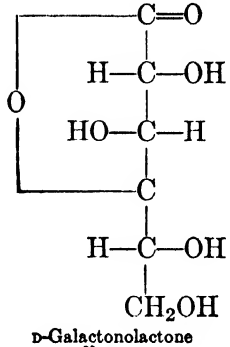
[34] C. S. Hudson, *J. Am. Chem. Soc.*, **32**, 338 (1910); **61**, 1525 (1939).

configuration of carbon-4 where lactonization took place. This led to the lactone rule:

When the oxygen ring in a γ-lactone is on the right side, as represented in the Fischer projection formula, dextrorotation is favored; if on the left side, levorotation is favored.



D-Gluconolactone
$[\alpha]_D^{20} = 68°$



D-Galactonolactone
$[\alpha]_D^{20} = -77°$

If a ring structure other than the γ-oxide formula was assigned, no such consistent relation was found. These observations led to the recognition that the sugar lactones are γ-lactones.

Levene and Hudson Hydrazide Rule. This rule expresses the relation between the rotation of the phenylhydrazides of the aldonic acids, $RCO \cdot NH \cdot NHC_6H_5$, and the configuration of carbon-2, a relation first noted by Levene.[35] *If the hydroxyl on carbon-2 is to the right, the hydrazide is dextrorotatory; if to the left it is levorotatory.* This rule also applies to the sugar amides,[36] the acetylated nitriles of the sugar acids,[37] and the aldo-benzimidazoles.[38]

These rules of optical rotation have been very useful in sugar chemistry. Evidence of the configuration of carbon-2 in aldonic acids can be obtained from the hydrazide rule. Upon degrading an aldose to a lower sugar or adding additional asymmetric centers through the cyanohydrin synthesis, the lactone rule may be applied to obtain the configuration of various asymmetric centers.

[35] P. A. Levene, *J. Biol. Chem.*, **23**, 145 (1915).
[36] C. S. Hudson, *J. Am. Chem. Soc.*, **40**, 813 (1918).
[37] V. Deulofeu, *Nature*, **131**, 548 (1933).
[38] N. K. Richtmeyer and C. S. Hudson, *J. Am. Chem. Soc.*, **64**, 1612 (1942).

CHAPTER 23

The Monosaccharides

CHEMICAL REACTIONS OF THE SIMPLE SUGARS

Since the simple sugars are polyhydroxy compounds containing either a potential aldehyde or ketone group, they naturally exhibit the usual reactions of alcohols and of aldehydes or ketones, although, as pointed out in the previous chapter, the pentoses and hexoses are not so reactive as would be anticipated if the carbonyl groups existed chiefly in the free condition. In addition, they undergo many reactions which are not given by simple alcohols or carbonyl compounds. In dealing with the chemical reactions of the simple sugars it is, in most cases, convenient to employ open-chain formulas.

Reactions as Alcohols. *Formation of Metallic Derivatives.* Sugars act as very weak acids. Michaelis and Rona [1] give the following dissociation constants at 18°: glucose, 6.6×10^{-13}; fructose, 9.0×10^{-13}; galactose, 5.2×10^{-13}; mannose, 10.9×10^{-13}; sucrose, 2.4×10^{-12}; maltose, 18.0×10^{-13}; lactose, 6.10×10^{-13}. That the sugars are *very* weak acids may be seen by comparison with the dissociation constants of some of the weak acids: acetic acid, 1.8×10^{-5}; butyric acid, 1.5×10^{-5}; lactic acid, 1.4×10^{-4}; boric acid (H^+ and $H_2BO_3^-$), 6.4×10^{-10}; and hydrocyanic acid, 7.2×10^{-10}. Ethyl alcohol forms sodium ethylate very easily; glycerol forms glycerolates with three sodium atoms. Sugars fall in the same class. Their importance lies in the fact that these metallic compounds decompose much more easily than the original sugar.

Ester Formation. ACYL DERIVATIVES. When the simple sugars are treated with acid anhydrides and acid chlorides, the hydroxyl groups are esterified according to the general reactions

$$ROH + (CH_3CO)_2O \rightarrow ROOCCH_3 + CH_3COOH$$

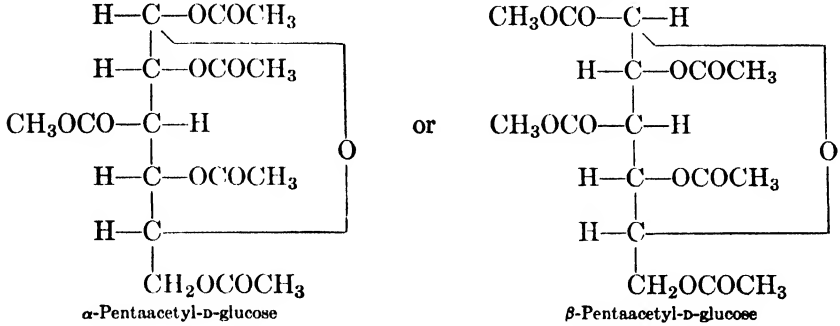
$$ROH + CH_3COCl \rightarrow ROOCCH_3 + HCl$$

The sugar esters, particularly the acetates and benzoates, have been widely used in studies of sugar structure as well as in the preparation of various carbohydrate derivatives. Esters of the higher fatty acids,

[1] L. Michaelis and P. Rona, *Biochem. Z.*, **49**, 232 (1913).

such as palmitic and stearic, have also been prepared, and they show characteristic properties.

Upon acetylation, ordinary D-glucose forms an α- or β-pentaacetyl derivative now known to have the following ring structure:

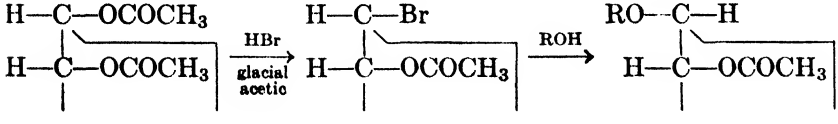


α-Pentaacetyl-D-glucose β-Pentaacetyl-D-glucose

The relative proportions of the α- and β-forms which are produced depend on the method of acetylation. Crystalline D-glucose exists in the α-form, and, if the acetylation is carried out rapidly by adding zinc chloride to boiling acetic anhydride to catalyze the reaction, the product is chiefly the α-derivative. The α-form also predominates when glucose is acetylated in pyridine solution at 0°C., as mutarotation does not occur in this solvent. If α-D-glucose is heated with acetic anhydride and sodium acetate the reaction proceeds slowly and mutarotation precedes acetylation, so that the β-pentaacetate predominates.

The sugar acetates are insoluble in water, do not mutarotate, and can be separated by fractional crystallization from solution in organic solvents. By saponification with alcoholic or aqueous alkali the acetyl groups may be removed. This reaction serves as a convenient method for determining the number of hydroxyl groups present in the molecule. The acyl derivatives have little value for determining ring structure, because of the tendency of the acetyl groups to migrate to unsubstituted hydroxyls in later chemical reactions.

The acetyl group on the reducing carbon is readily replaced by halogens (or the nitro group), and these compounds react readily with alcohols in the presence of a condensing agent such as silver oxide or silver carbonate to form glycosides. This is known as the Königs-Knorr reaction and it serves as a general method for preparing glycosides, such as the oligosaccharides. It is to be noted that a Walden inversion occurs.



Phosphoric acid esters are of great biochemical importance since phosphorylation has been found to be the primary step in the metabolism of D-glucose by microorganisms as well as by the higher animals. The nature of these compounds and their role in carbohydrate metabolism is discussed in Chapter 26. Pentose phosphates are obtained upon the degradation of nucleic acids (p. 406). The tannins are also carbohydrate esters.

TOSYL ESTERS. When a carbohydrate is treated with the chloride of *p*-toluenesulfonic acid in the presence of pyridine (or other organic bases), esters containing the *tosyl* group ($-SO_2 \cdot C_6H_4CH_3$) are formed.

$$RCH_2OH + ClSO_2 \cdot C_6H_4CH_3 \rightarrow RCH_2OSO_2 \cdot C_6H_4CH_3 + HCl$$

The primary alcohol group is more reactive than secondary alcohol groups, and this reagent esterifies the former the more readily. As the reaction is carried out in an alkaline medium, the tosyl group may be introduced into the acetone sugars without affecting the acetone residues (p. 544). The tosyl group which has esterified a primary alcohol group will react in acetone solution with sodium iodide to form an iodine derivative and sodium *p*-toluene sulfonate: $RCH_2OSO_2 \cdot C_6H_4CH_3 + NaI \rightarrow RCH_2I + C_6H_4CH_3SO_3Na$. Tosyl derivatives are used for purposes of characterization and also for the preparation of anhydro sugars and glycoseens.

Sugar Ethers. As alcohols, the sugars would be expected to form ethers, but it was not until 1903 that the first practical method for preparing sugar ethers was developed, by Purdie and Irvine,[2] who used methyl iodide and silver oxide. An alternative and more convenient and economical procedure employing dimethylsulfate and sodium hydroxide was introduced into sugar chemistry by Haworth.[3]

In the methylation of polysaccharides, such as starch, it has been found difficult to secure complete etherification of all the hydroxyl groups with these reagents. Haworth and his co-workers [4] found that the methylation of starch could be more readily accomplished if it was first converted into the acetyl derivative under mild conditions either by treatment with pyridine and acetic anhydride or with glacial acetic acid and acetic anhydride in the presence of chlorine and sulfur dioxide as catalysts. The acetylated starch is then simultaneously deacetylated and methylated by treatment in acetone solution with dimethylsulfate and sodium hydroxide. Complete methylation may also be accom-

[2] T. Purdie and J. C. Irvine, *J. Chem. Soc.*, **83**, 1021 (1903).

[3] W. N. Haworth, *J. Chem. Soc.*, **107**, 13 (1915).

[4] D. K. Baird, W. N. Haworth, and E. L. Hirst, *J. Chem. Soc.*, 1201 (1935); W. N. Haworth, E. L. Hirst, and J. I. Webb, *J. Chem. Soc.*, 2681 (1928).

plished by adding metallic sodium and methyl iodide to a suspension of moisture-free starch in liquid ammonia.[5] Another procedure is to secure a partial methylation with dimethylsulfate and dissolve the product in anisole; the etherification is then completed by adding sodium dissolved in liquid ammonia to the anisole solution and methylating with methyl iodide.[6]

Purdie's reagent [2] has strong oxidizing properties, and the sugars are therefore first converted to the methylglycoside by treatment with absolute methanol and hydrogen chloride. Methylsulfate will react with sugars to produce glycosides but, even when the methylsulfate-alkali etherification method is employed, Haworth found it advisable in methylating the complex sugars to cause the preliminary formation of a glycoside at low temperatures before employing the more drastic conditions which are required to bring about the methylation of the remaining hydroxyl groups. *It must be emphasized that the methyl group on the reducing carbon is held in an acetal linkage and not in an ether linkage.* This is shown by the fact that the glycosidic methyl group can be removed by heating with dilute mineral acids whereas those in the other positions are stable.

With normal D-glucose, 2,3,4,6-tetramethyl-methyl-D-glucoside is obtained, and, after purification by vacuum distillation, the glycosidic methyl group can be hydrolyzed off with aqueous hydrochloric acid and the resulting tetramethylglucose crystallized.

Methylation studies have played a very important part in establishing the ring structure of carbohydrates since the methylglycosides may be purified by vacuum distillation and many can be crystallized with ease. They have characteristic melting points and optical rotations, but their distinctive value lies in the fact that they do not undergo stereochemical changes such as wandering of constituents, changes in ring structure, and racemization. By completely methylating a sugar the ring is fixed, and by identifying the particular carbons to which the methoxy groups are attached one can determine the position of the oxide ring. These derivatives, however, have the disadvantage that the methyl groups cannot be removed from the molecule without causing decomposition of the sugar.

In determining the structure of the complex sugars, a number of sugar ethers (such as 2,3,4,6-tetramethyl-D-glucose; 2,3,4,6-tetramethyl-D-galactose; 1,3,4,5-tetramethyl-D-fructose; 2,3,4-trimethyl-D-glucose; and 2,3,6-trimethyl-D-glucose) had to be prepared and their properties investigated. These sugar ethers serve as reference compounds in

[5] K. Freudenberg and H. Boppel, *Ber.*, **71**, 2505 (1938); **73**, 609 (1940).

[6] K. Hess and Kan-Hou Lung, *Ber.*, **71**, 815 (1938).

identifying the products obtained upon the hydrolysis of the methylated oligosaccharides and polysaccharides.

TRITYL ETHERS. When polyhydric alcohols are condensed with triphenylchloromethane (trityl chloride) in the presence of pyridine only the primary alcohol groups readily form ethers.

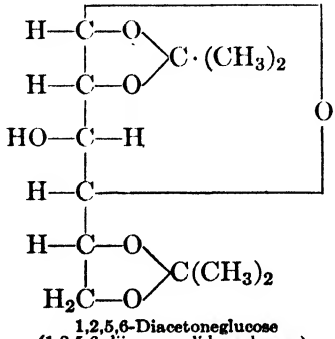
$$RCHOH\cdot CH_2OH + ClC\cdot(C_6H_5)_3 \xrightarrow{\text{pyridine}} RCHOH\cdot CH_2OC(C_6H_5)_3 + HCl$$

The sugars react in a similar manner. The trityl group is easily split off by acids, and hence, if the secondary alcohol groups have been previously blocked by more firmly bound substituents such as alkyl or acyl groups, this reaction provides a means of obtaining sugar derivatives which contain only a reactive primary alcohol group.

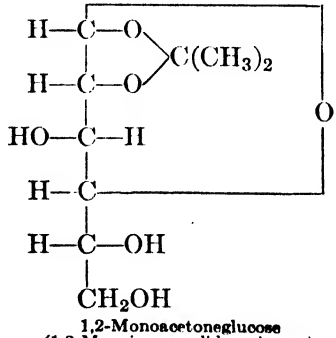
$$\text{6-Tritylglucose} \xrightarrow{(CH_3CO_2)O} \text{2,3,4-triacetyl-6-tritylacetylglucoside} \xrightarrow[CH_3COOH]{HBr} \text{2,3,4-triacetylglucose}$$

Acetone Derivatives. Sugars will condense with acetone in the presence of an acidic dehydrating agent such as sulfuric acid, zinc chloride, anhydrous copper sulfate, and phosphoric acid to form acetone derivatives. The acetone sugars, discovered by Emil Fischer [7] in 1895, are quite stable to alkalies but they are readily hydrolyzed by dilute acids. Many of them are crystalline, and they have been of great importance in studies of ring structure, particularly of the furanoses.

D-Glucose forms a diacetone derivative, diisopropylideneglucose, which contains only one free hydroxyl group (on carbon-3). On controlled hydrolysis, one isopropylidene residue is hydrolyzed off, yielding a crystalline monoacetoneglucose. It has been proved that these derivatives have the following structure:



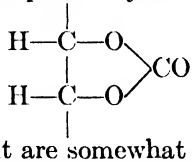
1,2,5,6-Diacetoneglucose
(1,2,5,6-diisopropylideneglucose)



1,2-Monoacetoneglucose
(1,2-Monoisopropylideneglucose)

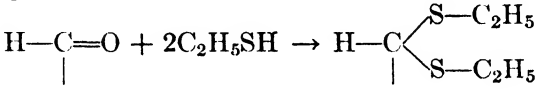
[7] E. Fischer, *Ber.*, **28**, 1165 (1895).

These glucose derivatives are non-reducing and have a furanose structure in which the ring is stabilized. Galactose forms a 1,2,3,4-diacetone sugar with a pyranose ring structure and a free hydroxyl group on carbon-6.

Sugar Carbonates. Through the use of carbonyl chloride, Haworth and Porter [8] prepared crystalline dicarbonates of the sugars containing the grouping  which are analogous to the diacetone derivatives but are somewhat more stable to dilute acids.

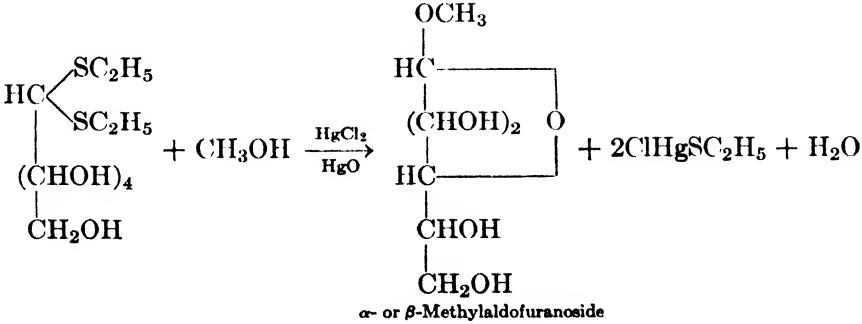
Properties Associated with the Carbonyl Group. *Glycoside Formation.* As discussed on p. 531, the simple sugars form glycosides (inner cyclic acetals) under acetal-forming conditions (treatment with absolute methanol containing hydrogen chloride). They do not reduce Fehling's solution or mutarotate and are characterized by their stability toward bases and their ease of hydrolysis with acids.

Sugar Mercaptals. When sugars are treated in strong hydrochloric acid with mercaptans, the mercaptan condenses with the carbonyl group to form mercaptals:



The condensation with the formation of the mercaptals is exceedingly rapid and takes place in a very few minutes even at 0°C. The mercaptals are insoluble in water and are easily crystallized from organic solvents.

The sugar mercaptals serve as starting materials for the synthesis of furanosides. In the presence of mercuric chloride and yellow mercuric oxide, they will react with an alcohol to produce either the α- or β-furanoside,[9] depending upon the temperature.

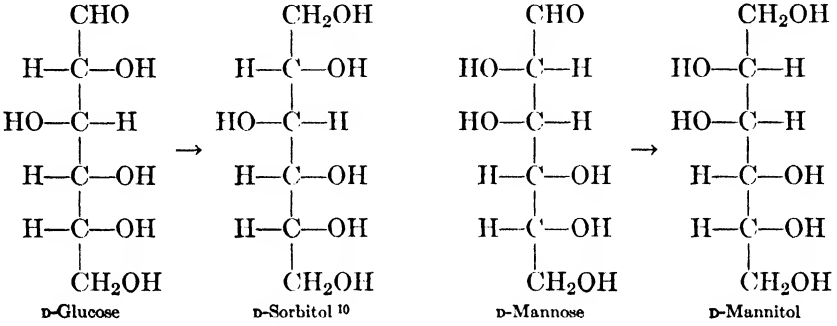


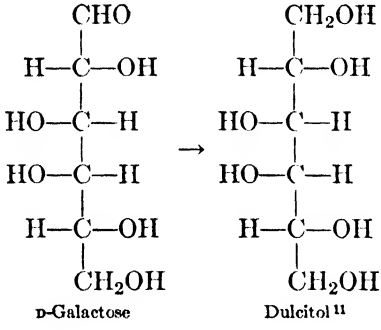
α- or β-Methylaldofuranoside

[8] W. N. Haworth and C. R. Porter, *J. Chem. Soc.*, 151 (1930).

[9] J. W. Green and E. Pacsu, *J. Am. Chem. Soc.*, **59**, 1205 (1937).

Reduction. The products formed upon the reduction of a simple sugar depend on the nature of the reducing agent. Reduction with sodium amalgam or platinum black in alkaline solution gives the corresponding polyhydric alcohol. From an aldose sugar only one alcohol is usually obtained.

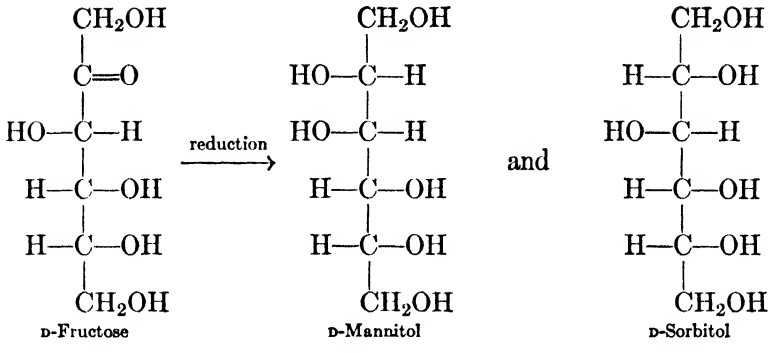




Upon reduction of a ketose, two epimeric alcohols are produced. Thus, D-fructose gives D-mannitol and D-sorbitol.

[10] Both D-glucose and L-gulose yield the same alcohol (sorbitol) upon reduction. Hudson, *Advances in Carbohydrate Chem.*, I, 14 (1945), notes that the reduction of D-glucose to yield natural sorbitol is older than its synthesis by the reduction of L-gulose. For that reason natural sorbitol has been named D-sorbitol to relate it to D-glucose. A similar problem in nomenclature occurs with saccharic acid, since the same isomer results from the oxidation of the above two hexoses. By analogy, the oxidation product of D-glucose is designated as D-saccharic acid—EDITORS.

[11] By convention, this compound might be designated D-dulcitol. However, if it is rotated through 180° in the plane of the paper, it is designated L-dulcitol. Accordingly the configuration symbol has no significance in this particular case and is omitted.



The configurations of D-mannitol and D-sorbitol are known from those of D-glucose and D-mannose. They differ only in the configuration of carbon-2, and hence the carbonyl group in D-fructose must be on this carbon and the configurations of the remaining asymmetric centers must be identical in these three sugars.

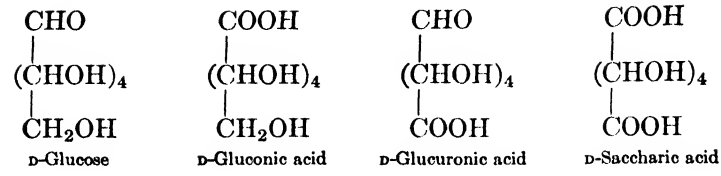
It should be noted that the reduction of the sugars with sodium amalgam occurs in alkaline solution, and hence other isomers may be obtained as a result of tautomeric changes (p. 566). Catalytic reduction with hydrogen and Raney nickel also yields the polyhydric alcohol corresponding to the sugar. This procedure is preferable to reduction with sodium amalgam in alkaline solution because enolization does not occur and the alcohol corresponding to an aldose sugar is obtained almost pure and in practically quantitative yields.

Electrolytic reduction of a simple sugar yields polyhydric alcohols, and D-glucose has been reduced commercially in this manner on a large scale. D-Sorbitol and D-mannitol are both obtained, and hence the conditions under which the reduction is effected result in epimerization. The epimers are separated by making use of the fact that D-sorbitol is the more soluble in cold ethanol.

Reduction of an aldose or ketose with hydriodic acid in the presence of phosphorus yields the corresponding straight-chain hydrocarbon as the ultimate product.

Oxidation. The monosaccharides are relatively strong reducing agents and yield several products which vary with the oxidizing agent used and the conditions under which the reaction is carried out.

OXIDATIONS WITH BROMINE, HYPOBROMITE, HYPOIODITE, AND NITRIC ACID. ALDOSES. When an aldose is oxidized in an acid medium, three different acids representing successive steps in oxidation may be obtained without splitting the molecule, as for example, with D-glucose:



The particular acid which is obtained depends on the nature of the oxidizing agent. Bromine water, hypobromite, and hypoiodite bring about selective oxidation of the carbonyl group to the aldonic acid.[12] If the carbonyl group is protected by glycoside formation, bromine water in the presence of pyridine and sodium hydroxide solution oxidizes the glycoside to the corresponding uronide. Uronic acids may also be prepared by the selective reduction of the monolactone of dicarboxylic sugar acids. Hot dilute nitric acid oxidizes aldoses to aldonic acids, but upon heating with strong nitric acid they are oxidized to dicarboxylic acids.

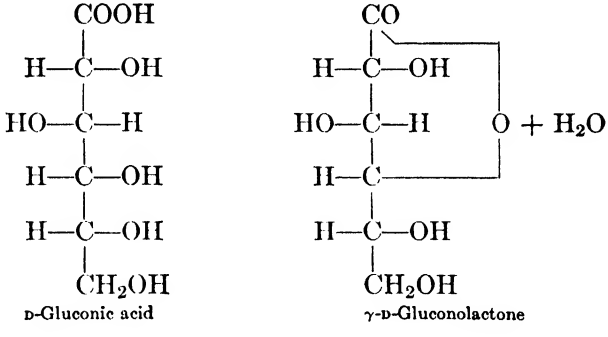
The aldonic acids are noteworthy for the ease with which they undergo epimerization or partial inversion of the asymmetry of carbon-2 upon heating with a weak base such as pyridine or quinoline, which prevents lactone formation, to produce an equilibrium mixture of the two isomers (epimers).[13] This represents a selective racemization of only one asymmetric center and hence does not result in an optically inactive mixture containing equal quantities of D- and L-forms. The equilibrium does not correspond to 50 per cent conversion, but varies from one compound to another because of the controlling influence of the other asymmetric centers on the opening of one of the double bonds in the intermediate enolic form (see p. 568). Since the epimeric aldonic acids are not enantiomorphs, their physical properties are dissimilar and they frequently may be separated by making use of differences in the solubility of their potassium, calcium, cadmium, or barium salts.

[12] C. S. Hudson and H. S. Isbell, *Bur. Standards J. Research*, **3**, 57 (1929), have described an efficient bromine-barium benzoate procedure for the oxidation of aldoses. The hypobromite oxidation of pure aldoses may be conveniently and efficiently carried out electrolytically by the method of H. S. Isbell and H. L. Frush, *Bur. Standards J. Research*, **6**, 1145 (1931). This procedure involves the continuous electrolysis of an aldose solution containing a small quantity of a bromide; the hypobromite ion formed is reduced by the sugar to hydrobromic acid which is again electrolyzed to the hypobromite ion. Hypoiodite oxidation may be carried out with barium hypoiodite, as described by W. F. Goebel, *J. Biol. Chem.*, **72**, 809 (1927), but more recently S. Moore and K. P. Link, *J. Biol. Chem.*, **133**, 293 (1940), have developed a procedure employing potassium iodite in methanol which brings about the efficient and rapid oxidation of the aldoses to potassium aldonates.

[13] The epimerization of aldonic acids was discovered by Emil Fischer, *Ber.*, **23**, 799 (1890).

This property of the aldonic acids has been of great value in the synthesis of the sugars since it provides a means of passing from glucose to mannose and, in a similar way, from galactose to talose, or from xylose to lyxose, since, after separation of the aldonic acids, their lactones can be readily reduced to the corresponding aldoses with sodium amalgam.

Upon heating, the aldonic acids, as well as the uronic and dicarboxylic sugar acids, spontaneously form internal esters to give stable 1-4 or γ-lactones.



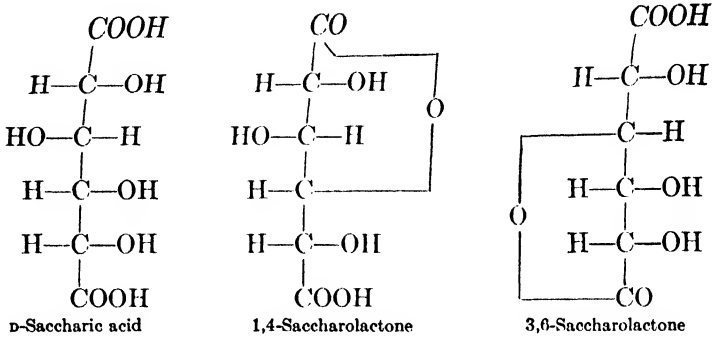
Convincing evidence that ring formation occurs at the γ-carbon atom was provided by Hudson's studies of the direction of optical rotation of twenty-four lactones in the sugar series which were discussed on p. 538; if a 1-4 oxide structure was uniformly assigned, a striking parallelism was exhibited between the position of the lactone ring and the sign of the rotation.

It has been shown, however, that when an aldonic acid is liberated from its salt with mineral acids it undergoes lactonization in two stages. First, an unstable lactone is rapidly formed which gradually disappears with the slow production of a stable γ-lactone.[14]

Conversely, Haworth and his associates [15] found that the rate of hydrolysis of methylated aldonolactones of the γ-sugar series in aqueous solution was much slower than that of methylated lactones from the "normal" sugars. These studies suggested that the normal sugars have a 1-5 oxide-ring structure which give rise to the rapid formation of an unstable δ-lactone which is slowly converted to a stable γ-lactone. This interpretation is supported by the fact that when aldoses are rapidly oxidized to aldonic acids by hypobromite, the immediate formation of a δ-lactone occurs. It is believed that the lactone is formed directly from the sugar without the intermediate formation of the aldonic acid.[16]

[14] P. A. Levene and H. S. Simms, *J. Biol. Chem.*, **65**, 31 (1925).
[15] W. Charlton, W. N. Haworth, and S. Peat, *J. Chem. Soc.*, **89** (1926).
[16] C. S. Hudson and H. S. Isbell, *Bur. Standards J. Research*, **8**, 327 (1932).

The dicarboxylic sugar acids can yield dilactones and might also be expected to form two isomeric γ-monolactones, as illustrated below for D-saccharic acid.[10]



D-Saccharic acid | 1,4-Saccharolactone | 3,6-Saccharolactone

Upon partial reduction, the 1,4-lactone would yield D-glucuronic acid, and the 3,6-lactone would yield L-guluronic acid. Therefore, when saccharolactone, which is a mixture of these two lactones, is reduced with sodium amalgam in acid solution there is formed a mixture of D-glucuronic acid and L-guluronic acid. The partial reduction of the dicarboxylic sugar acids in this manner provides a general means for the synthesis of the hexuronic acids.[17]

The dicarboxylic acids obtained by nitric acid oxidation of the various aldoses differ in their physical properties and are therefore useful in the identification of sugars. Thus, saccharic acid (from D-glucose) is soluble in water whereas mucic acid (from D-galactose) is very sparingly soluble. If a solution of saccharic acid is neutralized with potassium hydroxide and then acidified with acetic acid a sparingly soluble potassium salt forms which is characteristic for glucose. Nitric acid oxidation of the sugars frequently leads to decomposition, and oxalic acid is a very common product.

KETOSES. The ketoses are not so readily oxidized as the aldoses. They are not oxidized by bromine water or by hypoiodite, and this affords a method for differentiating between an aldose and a ketose. When a ketose undergoes oxidation (with, for example, hypobromite or mer-

[17] The synthesis of D-glucuronic acid by reduction of saccharolactone with sodium amalgam was first reported by E. Fischer and O. Piloty, *Ber.*, **24**, 521 (1891). K. P. Link and his associates have improved this partial reduction procedure and employed it for the synthesis of several hexuronic acids such as D,L-mannuronic, D,L-galacturonic, and D,L-alluronic acids, *J. Biol. Chem.*, **95**, 203 (1932); **100**, 407 (1933); **104**, 189, 737 (1934).

curic oxide) the molecule breaks at the carbonyl group, giving rise to two acids. D-Fructose, for example, yields glycolic acid ($CH_2OH \cdot COOH$) and trihydroxybutyric acid ($CH_2OH \cdot CHOH \cdot CHOH \cdot COOH$).

OXIDATION WITH CHROMIC ACID. With chromic acid and other vigorous oxidizing agents, the aldoses and ketoses may be completely oxidized to carbon dioxide and water.

OXIDATION WITH METALLIC IONS. In alkaline solutions containing cupric ions, such as Fehling's solution, the simple sugars are oxidized to the corresponding acids and to cleavage products, the cupric ion being reduced and precipitated as cuprous oxide. Similarly an ammoniacal solution of silver nitrate is reduced to metallic silver. The former reaction is used for the quantitative determination of reducing sugars. In alkaline solution, complex isomerizations and cleavages of sugars are prone to occur (see p. 566), and the nature of the products obtained upon the oxidation of reducing sugars with cupric ions depends on the alkali concentration, the temperature, and the time of the reaction. For this reason, the various reduction methods are calibrated with pure sugar solutions, and in analyzing samples for sugar content the specified methods must be strictly followed in order to secure reliable quantitative results.

OXIDATION WITH HYDROGEN PEROXIDE IN THE PRESENCE OF IRON CATALYSTS. With ferrous salts as catalysts, hydrogen peroxide oxidizes both aldoses and ketoses to osones. This is known as the *Fenton reaction.*

$$RCHOH \cdot CHO \xrightarrow[H_2O_2]{Fe^{++}} RCO \cdot CHO$$

$$RCO \cdot CH_2OH \xrightarrow[H_2O_2]{Fe^{++}} RCO \cdot CHO$$

Hydrogen peroxide in the presence of ferric iron, such as colloidal ferric hydroxide, converts aldonic acids to aldoses containing one less carbon atom than the acid. This reaction, known as the *Ruff degradation,* will be considered later along with other methods for descending the sugar series.

OXIDATION OF METHYLGLYCOSIDES WITH BARIUM BROMITE. Upon the oxidation of methylglycoside with barium bromite an unusual type of oxidation involving cleavage of the molecule was discovered by Jackson and Hudson.[18] Since the reducing group on carbon-1 was protected by glycoside formation, it was anticipated that the terminal CH_2OH would be oxidized to give a methylglycoside of the corresponding uronic

[18] E. L. Jackson and C. S. Hudson, *J. Am. Chem. Soc.,* **59,** 994 (1937).

acid; that is, a methyluronide. However, when α-methyl-D-mannopyranoside was oxidized with barium hypobromite, carbon-3 was eliminated as formic acid and both carbon-2 and carbon-4 were oxidized to carboxyl groups to give a dibasic acid which could be isolated as the strontium salt. This reaction is somewhat similar to that obtained in the oxidation of various aldohexosides with periodic acid or lead tetraacetate discussed in the succeeding section, except that with these reagents carbon-2 and carbon-4 are only oxidized to the aldehyde stage.

OXIDATIVE DEGRADATION WITH PERIODIC ACID OR LEAD TETRAACETATE. Lead tetraacetate and periodic acid, first employed in the splitting of glycols, were later applied to the study of carbohydrates and have provided an important new tool for determining both the ring structures and the α- and β-configurations of the glycosides. These reagents only cause oxidative cleavage between adjacent carbon atoms carrying hydroxyl groups. Application of the periodic acid cleavage to the methylglycosides revealed that the ring is broken only between the central carbon atoms to give a dialdehyde in which the original glycosidic and ring-forming carbons remain intact and united by an oxygen atom.[19]

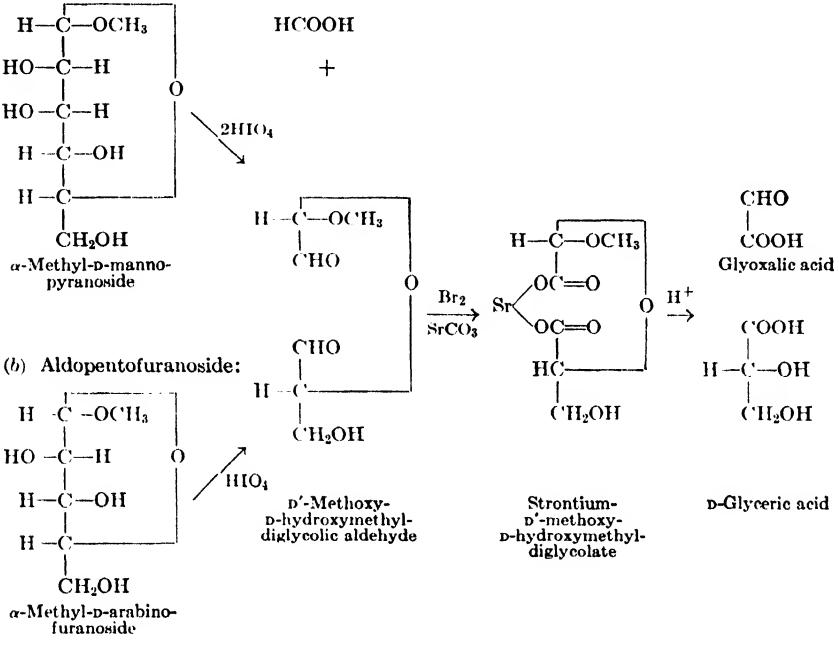
With an aldohexopyranoside, which has three adjacent carbon atoms carrying hydroxyl groups, the central one is eliminated as formic acid, and as the other two are oxidized to the aldehyde stage two moles of periodic acid are required. The two remaining asymmetric centers correspond to carbon-1 and carbon-5 of the original sugar, and their configurations represent the configurations of the glycosidic and ring-forming carbons in the parent glycoside. Hence, the methylaldohexopyranosides should give rise to four stereoisomeric dialdehydes, and this was found to be the case by Jackson and Hudson.[20] As the dialdehydes differ widely in optical rotation, periodic acid oxidation serves as a useful method for establishing the configurations of the reducing and ring-forming carbons of the parent glycoside. An aldopentofuranoside has only two carbon atoms carrying adjacent hydroxyls, and only one mole of periodic acid is consumed in the oxidation. Thus, the methylgly-

[19] This observation was first made by H. Hérrissey, P. Fleury, and M. Joly, *J. Pharm. Chem.*, **20**, 149 (1934), who employed periodic acid. Later W. S. McClenahan and R. C. Hockett, *J. Am. Chem. Soc.*, **60**, 2061 (1938), showed that lead tetraacetate can be used equally well for the cleavage of the glycosides. However, in oxidations employing lead tetraacetate in aqueous solution an additional mole of this reagent as compared with periodic acid is required where formic acid is split out because the formic acid produced is oxidized to carbon dioxide and water; J. M. Grosheintz, *J. Am. Chem. Soc.*, **61**, 3379 (1939).

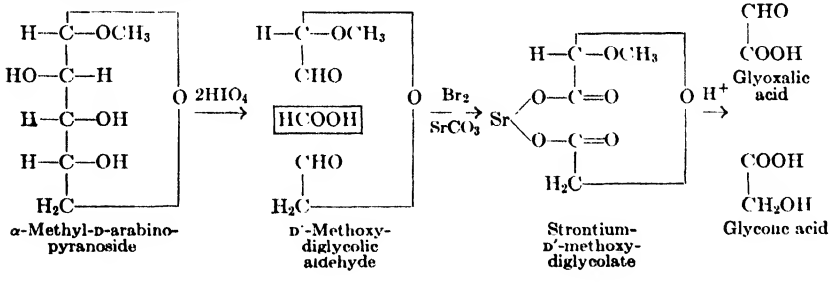
[20] E. L. Jackson and C. S. Hudson, *J. Am. Chem. Soc.*, **59**, 994 (1937); **61**, 959 (1939).

cosides of α-D-arabinofuranose and α-D-mannopyranose yield the same dialdehyde, but they differ in the quantity of periodic acid required for the oxidation.

(*a*) Aldohexopyranoside:



(*c*) Aldopentopyranoside:



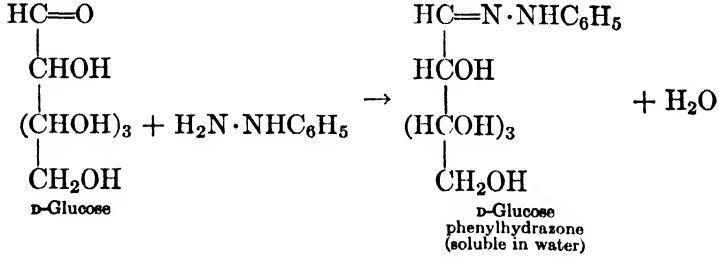
Jackson and Hudson subjected the dialdehydes to hypobromite oxidation in the presence of strontium carbonate, which reacts with any hydrobromic acid formed and thereby prevents the hydrolysis of the glycosidic group. The strontium salt of the dicarboxylic acid corresponding to the dialdehyde is formed. Upon hydrolysis, the methyl

group is lost, and the oxygen linkage splits between the carbon atoms involved in the oxide ring formation. From the nature of the two acids which are produced the carbon atom of the original sugar which was involved in the ring formation can be ascertained. Also, where one of the acids has an asymmetric carbon, the D- or L-configuration of the ring-forming carbon of the original sugar is revealed. Thus, aldohexopyranosides and aldopentofuranosides of the D-series yield D-glyceric acid, and those of the L-series yield L-glyceric acid. These reactions afford a very direct method of correlating the configuration of the sugars with glyceric acid.

Reactions of Sugars with Phenylhydrazine. Phenylhydrazones, Osazones. The reactions of the simple sugars (and of the complex sugars containing a functional carbonyl group) with phenylhydrazine are of great practical and historical interest. Emil Fischer's first paper in the field of carbohydrate chemistry, which he published in 1884, dealt with the application of phenylhydrazine to sugar chemistry. He had prepared the reagent ten years previously and had found that it condensed with simple aldehydes and ketones to form derivatives containing one phenylhydrazine residue. With the sugars, he obtained beautiful crystalline products which he called osazones (a combination of *-ose* and *hydrazone*) and this led him to conduct further researches in the carbohydrate field.

As a result of his studies, Fischer [21] concluded that the reaction involved the following three steps when an excess of phenylhydrazine is heated with an aldose in acetic acid solution.[22]

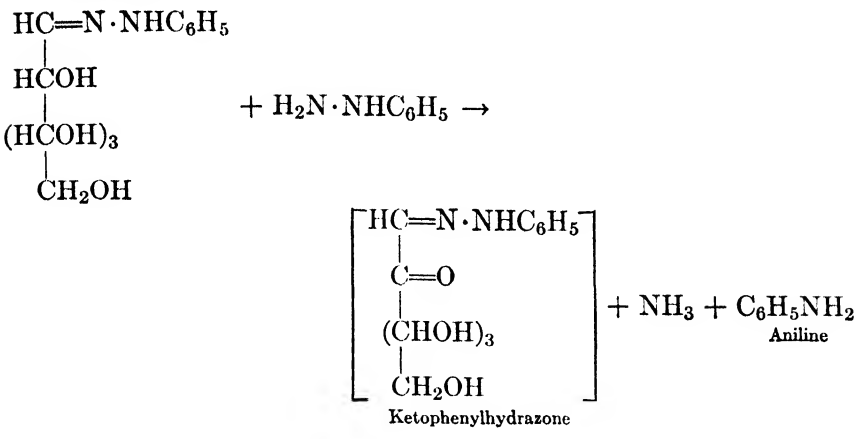
(*a*) Condensation



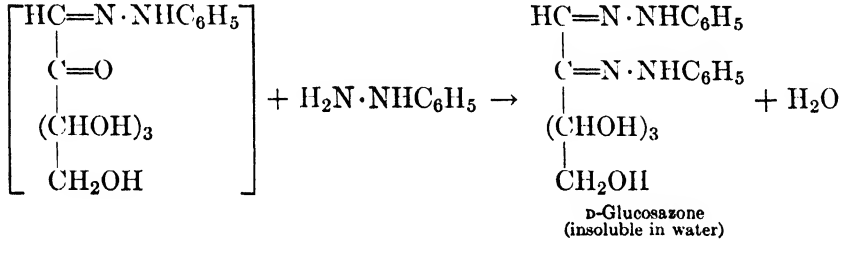
[21] E. Fischer, *Ber.*, **17**, 579 (1884); **20**, 821 (1887).

[22] Although written as straight-chain compounds the hydrazones and osazones may, like the sugars themselves, have a cyclic structure since many of them undergo mutarotation in aqueous alcohol or pyridine solutions. From spectroscopic data, however, L. L. Engel, *J. Am. Chem. Soc.*, **57**, 2419 (1935), concludes that the osazones are straight-chain compounds and suggests that their mutarotation is due to equilibria between them and their hydrolytic products.

(*b*) Oxidation (in excess of phenylhydrazine)



(*c*) Condensation

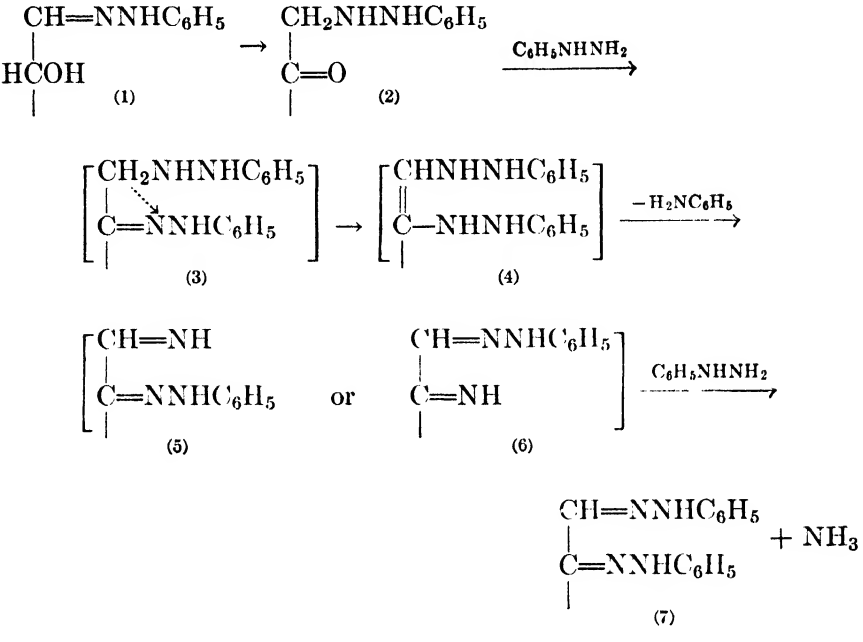


The first step in the reaction involves the condensation of phenylhydrazine with the oxygen of the carbonyl group to form a hydrazone. Next, the excess phenylhydrazine present is assumed to act as an oxidizing agent removing two hydrogens from carbon-2 to give a carbonyl group, the phenylhydrazine undergoing reductive fission to aniline and ammonia. In the final and third step, an additional molecule of phenylhydrazine condenses with the carbonyl group on carbon-2 to form the osazone.

Although accepted for many years, this explanation of osazone formation does not appear likely since the sugars are relatively weak reducing agents and phenylhydrazine does not function as an oxidizing agent in other reactions. In 1940, Weygand [23] suggested a more plausible explanation which involves an intramolecular oxidation-reduction by

[23] F. Weygand, *Ber.*, **73**, 1284 (1940).

means of an Amadori type of rearrangement (α-γ shift). One of the mechanisms he proposed is depicted below.



Weygand assumes that two hydrogens migrate from carbon-2 to saturate the carbon-nitrogen double bond on carbon-1 and form a ketonic derivative (2) which then reacts with another molecule of phenylhydrazine to give (3). This intermediate then isomerizes to (4) through an α-γ hydrogen shift; aniline then splits out through a reversal of an ordinary 1,4-addition reaction to give either of the imino compounds (5) or (6) which then reacts with phenylhydrazine to form the osazone (7) and ammonia. The cleavage of the intermediate (4) to form the imino derivatives (5) or (6) and aniline provides a means for the oxidation of carbon-2 without assuming direct oxidation by phenylhydrazine.

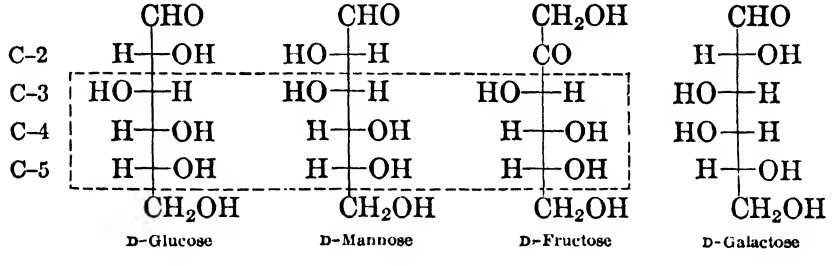
The phenylhydrazones can be decomposed to the original sugar by treatment with benzaldehyde; the phenylhydrazone of benzaldehyde is formed at the expense of the sugar phenylhydrazone.

Mannose phenylhydrazone differs from that of other sugars in being relatively insoluble in water; this property is an aid in the identification and determination of mannose.

The osazones are very sparingly soluble in water; because they have characteristic crystalline forms and specific rotations they have been of

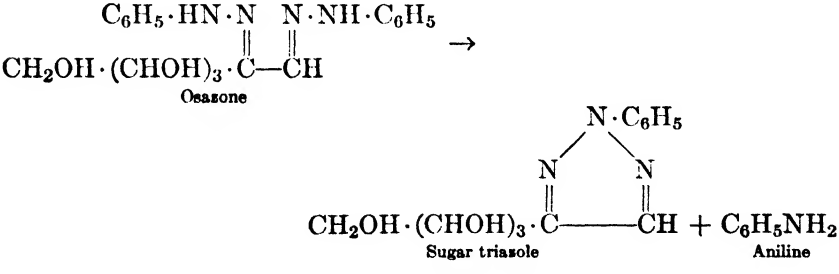
considerable importance in the identification of sugars and in the development of our knowledge of sugar structure. Their value is limited, however, by the proximity of their melting points, variations in melting point with rate of heating, and their tendency to decompose at or near the melting point, *i.e.*, their melting points are actually decomposition temperatures. The purity of the sugar and of the hydrazine reagent is also an important factor.

The asymmetry of carbon-2 in the aldoses is destroyed in osazone formation, and Fischer noted that D-glucose, D-mannose, and D-fructose yielded the same osazone. This proves that the configurations of carbon-3, carbon-4, and carbon-5 are identical in the three sugars. He noted that D-galactose, however, formed a different osazone, and a start was made in establishing the configurations of these sugars:



D-Fructose can be distinguished from D-glucose and D-mannose by employing an asymmetric disubstituted hydrazine, such as methylphenylhydrazine ($C_6H_5 \cdot N \cdot CH_3 \cdot NH_2$) in the reaction; fructose forms an osazone, but glucose and mannose only form the hydrazone.

Recently, Hann and Hudson [24] have found that, when osazones are boiled with aqueous copper sulfate solution, osatriazoles are formed:

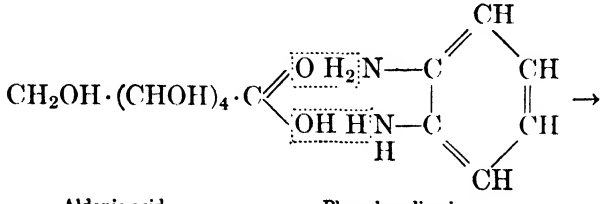


The sugar triazoles have sharp melting points and optical rotations

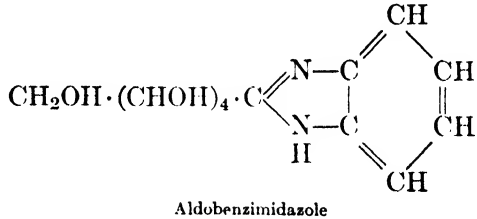
[24] R. M. Hann and C. S. Hudson, *J. Am. Chem. Soc.*, **66**, 735 (1944).

(with no mutarotation), and should prove of great value as a means of characterizing the osazones.

Aldobenzimidazole Sugar Derivatives. Moore and Link [25] have also stressed the limitations of the osazones for sugar identification and have pointed out that the aldose sugars may be more satisfactorily characterized as aldobenzimidazoles. These derivatives are readily prepared by condensation of the corresponding aldonic acids with *o*-phenylenediamine.



Aldonic acid *o*-Phenylenediamine



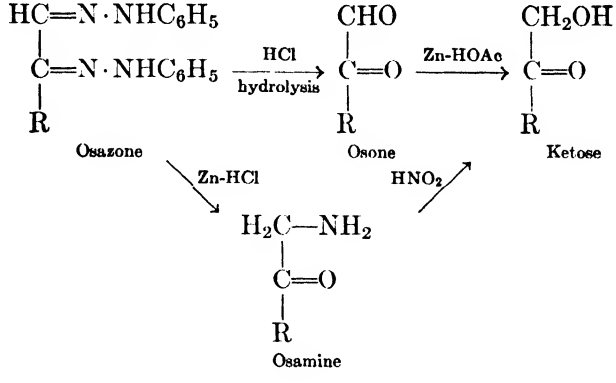
Aldobenzimidazole

These carbohydrate derivatives are stable white crystalline solids, sparingly soluble in water, with sharp melting points and relatively wide differences in specific rotatory power. They can readily be converted to crystalline acid salts, such as the hydrochlorides and picrates, to provide additional constants for identification. The aldonic acid salts required as intermediates in the formation of the benzimidazoles may be efficiently and rapidly prepared by oxidation of the aldose sugars with potassium hypoiodite in methanol. Moore and Link have detailed the procedures involved in the systematic identification of seven aldoses, namely, arabinose, lyxose, xylose, rhammose, glucose, mannose, and galactose.

Osone and Osamine Formation. Fischer found that by treating an osazone with fuming hydrochloric acid both phenylhydrazine residues were hydrolyzed off to yield a ketonic aldehyde, a colorless strongly reducing syrup called an *osone*, which could readily be reduced to a ketose upon treatment with zinc dust and acetic acid. If the osazone

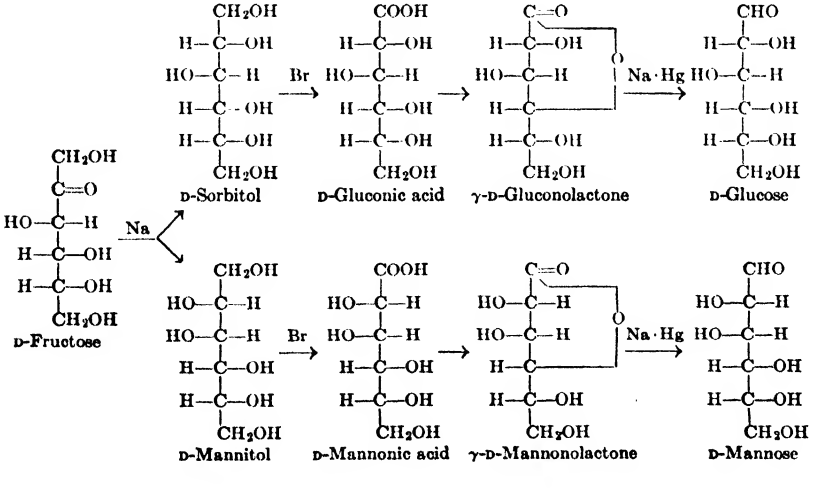
[25] S. Moore and K. P. Link, *J. Biol. Chem.*, **133**, 293 (1940).

itself was reduced by treatment with zinc dust and hydrochloric acid, an *osamine*, an amino derivative of a ketose, was produced which yields the corresponding ketose upon treatment with nitrous acid.

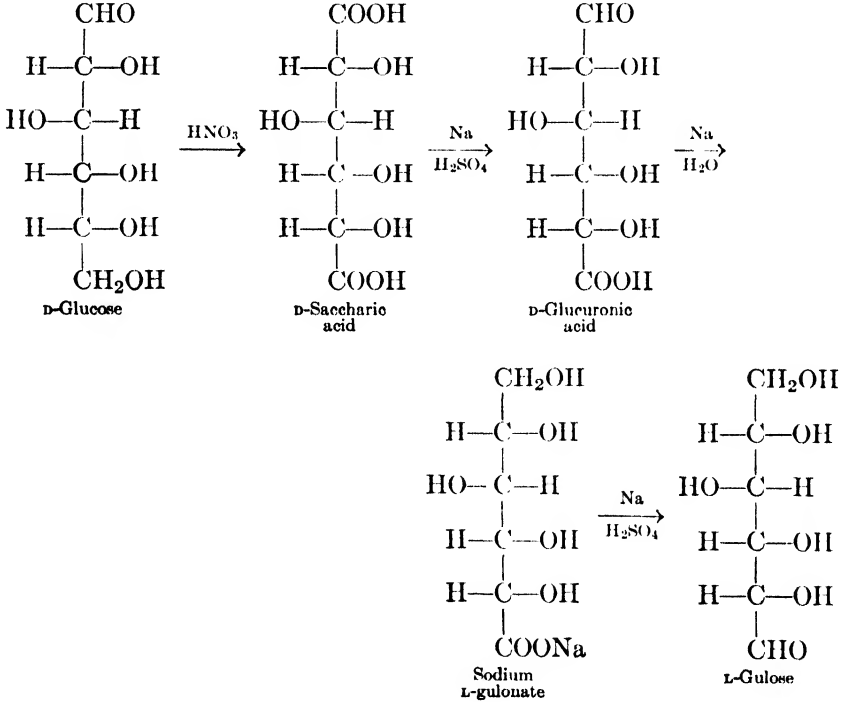


These reactions afford a means of preparing the corresponding ketose from an aldose, namely: aldose → osazone → osone (or osamine) → ketose. Thus D-glucose can be converted into D-fructose.

The reverse of this process, the transformation of a ketose to an aldose, can be brought about by reducing the ketose to the corresponding alcohol, oxidizing this to the aldonic acid, and then reducing the lactone of the acid. In the reduction of the ketose two epimeric alcohols are formed so that two epimeric aldoses result; thus, D-fructose yields D-glucose and D-mannose.



Transformation of an Aldose from D- to L-Series. *The transformation of a* D- *to an* L-*aldose may be brought about by a series of reactions which result in an exchange of the aldehyde and primary alcohol groups.* For example, D-glucose may be oxidized with nitric acid to the dicarboxylic acid and the lactone selectively reduced to D-glucuronic acid with sodium amalgam in dilute sulfuric acid. Upon further reduction with sodium amalgam in water, sodium L-gulonate is formed, the lactone of which yields L-gulose when reduced with sodium amalgam in dilute sulfuric acid. These reactions are summarized below, the lactone forms being omitted. This series of reactions was carried out by Fischer and



led to the confusion in the nomenclature of the D- and L-series of gulose, idose, lyxose, and threose mentioned in Chapter 22.

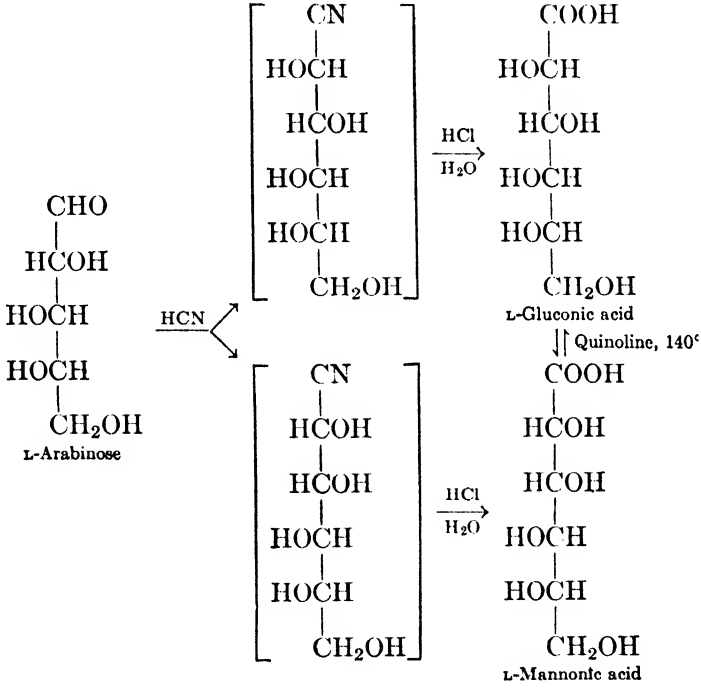
Cyanohydrin Synthesis of Kiliani. Ascent of the Aldose Series. Early

$$\underset{\text{Acetaldehyde}}{CH_3\cdot CHO} + HCN \rightarrow \underset{\text{Acetaldehyde cyanohydrin}}{CH_3\cdot CHOH\cdot CN} \xrightarrow[HCl]{H_2O} \underset{\text{D,L-Lactic acid}}{CH_3\cdot CHOH\cdot COOH}$$

$$\underset{\text{Acetone}}{(CH_3)_2\cdot CO} + HCN \rightarrow \underset{\text{Acetone cyanohydrin}}{(CH_3)_2\cdot COH\cdot CN} \xrightarrow[HCl]{H_2O} \underset{\alpha\text{-Hydroxyisobutyric acid}}{(CH_3)_2\cdot COH\cdot COOH}$$

in the history of organic chemistry, it was discovered that aldehydes and ketones reacted with hydrogen cyanide to form cyanohydrins or nitriles which, upon hydrolysis with strong hydrochloric acid, yielded α-hydroxy acids containing one more carbon atom than the original carbonyl compound.

The cyanohydrin reaction was first applied to the reducing sugars by Kiliani,[26] and his discovery that these polyhydroxy acids readily lose water to form stable lactones which could be crystallized was of great significance since it was the classical method used by Fischer for increasing the chain length in the aldose series. The reaction, as carried out by Fischer [27] with L-arabinose which he obtained by the hydrolysis of plant gums, is illustrated below.



Since the new asymmetric carbon is being added to an asymmetric molecule, its configuration is influenced by that of the other asymmetric centers, and the two aldonic acids are formed in unequal amounts. In

[26] H. Kiliani, *Ber.*, **18**, 3066 (1865); **19**, 221, 767, 772, 1128, 3029 (1886); **20**, 282 339 (1887); **21**, 915 (1888); **22**, 521 (1889).

[27] E. Fischer, *Ber.*, **23**, 2611 (1890).

the addition of hydrogen cyanide to arabinose, mannonic acid is produced in very much greater quantity than the epimeric gluconic acid; with lyxose, galactonic acid is produced in excess of talonic acid; with xylose, gulonic acid is favored over idonic acid; with ribose, allonic and altronic acid are produced in nearly equal amounts.

As mentioned previously, Fischer discovered that the lactones of the aldonic acids could be reduced with sodium amalgam to yield the corresponding aldose. Moreover, he found that the aldonic acids could be partially epimerized by treatment with pyridine or quinoline and that their phenylhydrazides crystallized readily, thereby aiding in the isolation and purification of these acids. By the addition of these reactions to the cyanohydrin procedure of Kiliani, the synthesis of the higher carbon aldoses was made possible. By these means, one could advance from formaldehyde, one carbon at a time, to the higher sugars.

As already indicated, Fischer synthesized L-glucose and L-mannose from L-arabinose, and from D-xylose he prepared two new synthetic hexoses, D-gulose and D-idose. He extended this work to the other naturally occurring aldoses, D-glucose, D-mannose, D-galactose, and D-rhamnose. D-Glucose and D-mannose were carried to aldononoses, and later Phillipe [28] carried the reaction with D-glucose to the decose stage.[29]

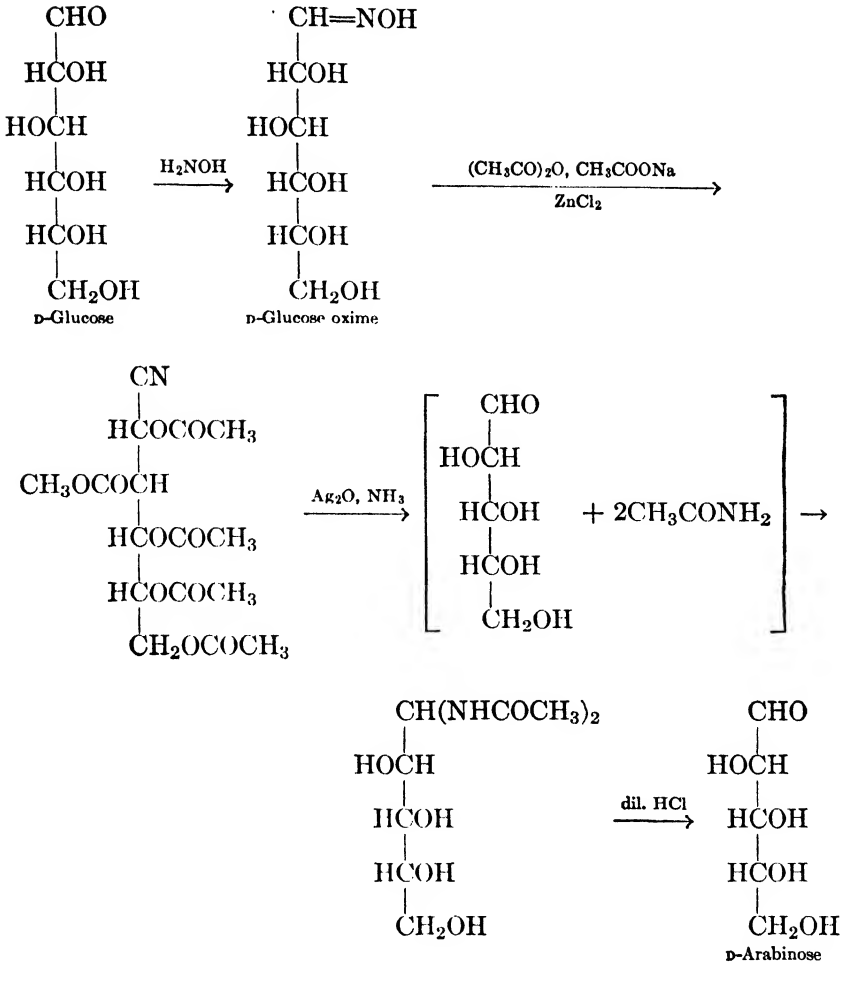
Degradation of an Aldose Sugar. Three general methods have been developed for converting an aldose sugar to the next lower member.

WOHL DEGRADATION. In the Wohl [30] procedure, the aldose is condensed with hydroxylamine to form the aldoxime, which upon treatment with acetic anhydride and anhydrous zinc chloride is converted into the acetyl derivative of the corresponding aldonic acid nitrile. Treatment of the acetylated nitrile with ammoniacal silver nitrate results in ammonolysis of the acetyl groups and the loss of hydrogen cyanide to give an aldose containing one less carbon atom. This is the reverse of the Kiliani synthesis. The aldose reacts with the acetamide produced to give a diacetamide derivative which may be readily hydrolyzed with dilute hydrochloric acid to give the free aldose.

[28] L. H. Phillipe, *Ann. Chem. Phys.*, [8] **26**, 289, (1912).

[29] An excellent discussion of the cyanohydrin synthesis and configurations of higher carbon sugars and alcohols will be found in a paper by C. S. Hudson, *Advances in Carbohydrate Chem.*, **1**, 1 (1945).

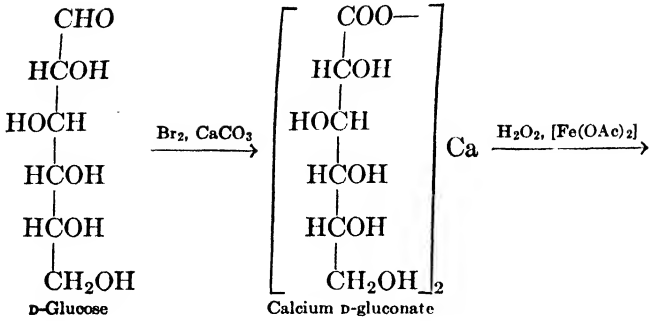
[30] A. Wohl, *Ber.*, **26**, 730 (1893).



Zemplén [31] modified the Wohl method by eliminating the nitrile group with sodium methoxide in chloroform, thereby obtaining much better yields.

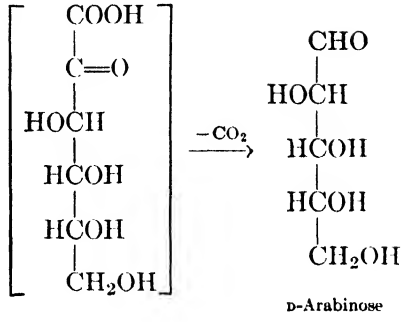
RUFF DEGRADATION. In the Ruff method [32] the aldose is converted to the calcium salt of the corresponding aldonic acid by oxidation with bromine in the presence of calcium carbonate. The calcium salt of the aldonic acid is then treated with hydrogen peroxide in the presence of basic ferric acetate. A 2-keto-aldonic acid, which loses carbon dioxide to yield the next lower aldose, is probably formed.

[31] G. Zemplén and D. Kiss, *Ber.*, **60B**, 165 (1927).
[32] O. Ruff, *Ber.*, **31**, 1573 (1898).



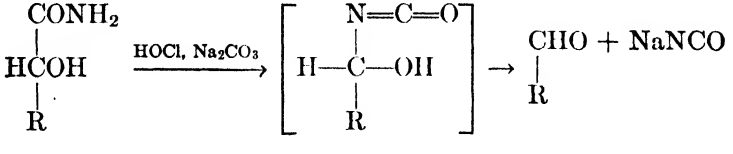
D-Glucose

Calcium D-gluconate



D-Arabinose

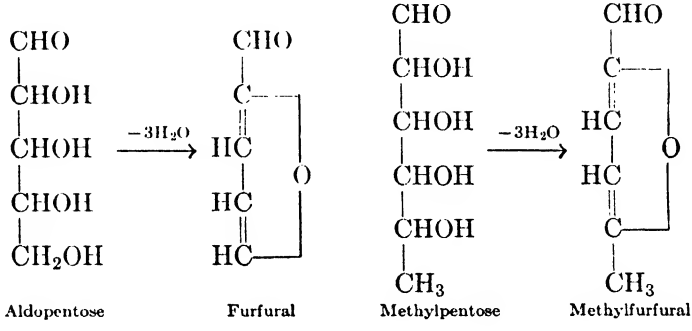
WEERMAN METHOD. In Weerman's method [33] the aldose is oxidized to aldonic acid, and its amide is prepared. Upon treatment of the aldonamide with sodium hypochlorite in the presence of sodium carbonate an isocyanate derivative is formed which breaks down to yield the next lower aldose and sodium isocyanate.

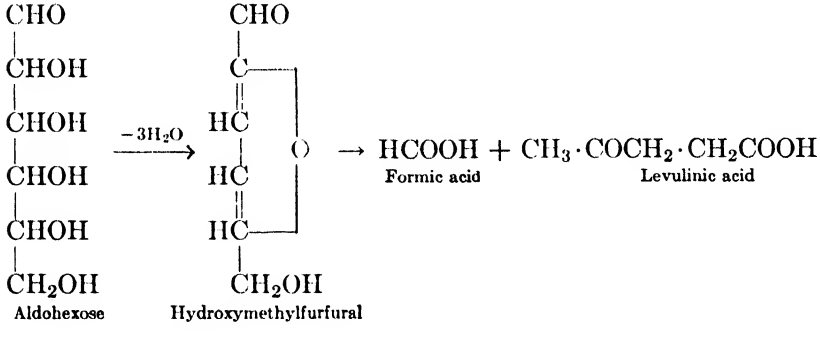


Action of Acids on Monosaccharides. The mechanism of the rearrangements and cleavage of the simple sugars upon treatment *in vitro* with acids and bases is of biological as well as of chemical interest since these reactions may provide clues to the nature of their reactions *in vivo*. The sugar molecule is most stable in neutral solution and in the absence of metallic salts; it is less stable in acid solution and least stable in alkalies.

[33] R. Weerman, *Rec. trav. chim.*, **37**, 16 (1918).

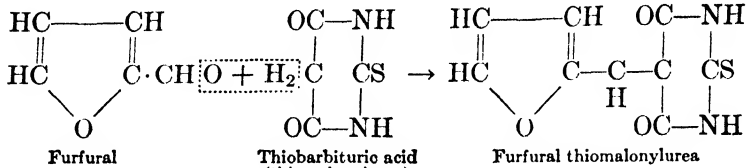
The effect of acids on the simple sugars depends on the acid concentration and the temperature. At ordinary temperatures the simple sugars are stable in weak non-oxidizing acids, whereas hot concentrated acids cause complex changes with the formation of humic substances. At intermediate concentrations, *e.g.*, 12 per cent hydrochloric acid, the elements of water (3 molecules per molecule of sugar) are extracted. In the case of a pentose, furfural is formed in quantitative yield; methyl pentoses give methylfurfural. By a similar elimination of 3 molecules of water from an aldohexose, hydroxymethylfurfural will be formed, but only small quantities are recovered, as this derivative decomposes in the hot acid solution to give formic and levulinic acids.





The stoichiometric formation of furfural and methylfurfural from pentoses and methylpentoses forms the basis for the quantitative estimation of pentoses and pentosans. The sample is distilled under controlled conditions with a solution of hydrochloric acid (12 per cent HCl by weight), and the furfural is determined by oxidation, precipitation by various reagents, or by colorimetric means. The usual method is to form the relatively insoluble furfural-phloroglucin derivative which is

isolated and weighed. This method suffers from the fact that the phloroglucin derivative is slightly soluble, its composition varies with the conditions of precipitation, and hydroxymethylfurfural (of hexose origin) is precipitated in small yield. These objections can be overcome by condensation of the furfural with thiobarbituric acid to form insoluble furfuralthiomalonylurea, a compound of definite and uniform composition.



Furfural and methylfurfural are precipitated quantitatively, and the presence of hexoses or hexosans in the sample under analysis does not interfere since hydroxymethylfurfural forms a soluble condensation product.[34]

Furfural and its derivatives have the property of forming colored complexes with phenolic substances such as α-naphthol, resorcinol, and orcinol, a property which is the basis of the Molisch, Seliwanoff, and Bial tests for carbohydrates, ketoses, and pentoses, respectively.[35] Several furfural compounds develop a brown color on storage, and they have recently been shown to be involved in the browning which occurs during and subsequent to the heat processing of many foods, as in the dehydration of eggs, milk, malt, fruit, and vegetables.

Action of Alkalies on the Simple Sugars. Sugars in dilute alkalies are much less stable than they are in acid solutions, and they undergo

[34] See A. W. Dox and G. P. Plaisance, *J. Am. Chem. Soc.*, **38**, 2156 (1916); also A. J. Bailey, *Ind. Eng. Chem., Anal. Ed.*, **8**, 389 (1936).

[35] *Molisch test for carbohydrates* (α-naphthol reaction): A few drops of 5 per cent alcoholic solution of α-naphthol is mixed with the solution under test, and concentrated sulfuric acid is added to the inclined test tube to form a lower layer. The formation of a reddish violet zone at the junction between the two layers indicates the presence of furfural or a derivative.

Seliwanoff test for ketoses (resorcinol-hydrochloric acid reaction): When a ketose sugar solution is boiled with Seliwanoff's reagent (0.05 g. resorcinol dissolved in dilute HCl—1:2), a red color with or without the separation of a brown-red precipitate is produced. If the boiling is prolonged aldoses may give the test.

Bial test for pentoses (orcinol-hydrochloric acid reaction): When a pentose solution is heated with Bial's reagent (orcinol in HCl, containing ferric chloride) a green color develops on cooling, and a flocculent green precipitate may form.

much more complex transformations. The transformations which occur vary with the kind and concentration of alkali, the temperature, and the presence or absence of oxygen or oxidizing agents. In order of increasing complexity, the reactions of the sugars in alkalies may be classified as follows:

(1) Polymerization (aldol condensation.)
(2) Rearrangements without cleavage.

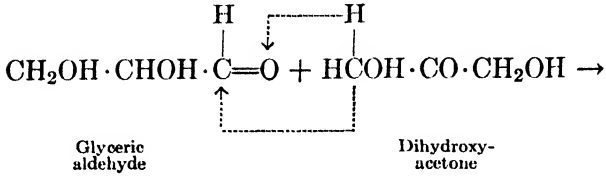
(*a*) Mutarotation and attainment of equilibrium between the α- and β-forms.
(*b*) Epimerization or Walden inversion of the asymmetric center adjacent to the carbonyl group (Lobry de Bruyn transformation).
(*c*) Progressive wandering of the reducing group along the carbon chain.
(*d*) Intramolecular oxidation and reduction. Saccharinic acid formation.

(3) Degradative cleavage with the production of various alcohols, aldehydes, and acids.

These various types of reactions, however, do not take place independently, and consequently the problem of elucidating the mechanism of these complex reactions is an extremely difficult one. Although this aim has not been fully achieved, the reaction products can be largely accounted for by enolization of the carbonyl group to form enediols. These enediols are assumed to be the reactive substances, and they undergo various transformations depending on the conditions present.

Effects of Low Hydroxyl-Ion Concentration. ALDOL CONDENSATION. Upon mild alkali treatment the aldehydes and the simple sugars undergo polymerization by aldol condensation, a reaction which played an important part in the early methods of synthesizing the carbohydrates. The first synthesis of a sugar was apparently accomplished in 1861 by Butlerow, who treated formaldehyde with saturated calcium hydroxide solution and secured a sweet syrup having the properties of a carbohydrate. In 1886, Loew improved the process and named the product "formose." This was subsequently shown to be a mixture of sugars formed by aldol condensation. Fischer later used glycerol as a starting point for the synthesis of sugars. Upon careful oxidation with nitric acid or with bromine water and sodium carbonate, glycerol yielded a mixture of glyceric aldehyde and dihydroxyacetone. Under the influence of baryta water these trioses underwent aldol condensation to produce a mixture of two ketohexoses, originally named α- and β-acrose.

Later it was shown that α-acrose was D,L-fructose and β-acrose was D,L-sorbose.



$CH_2OH \cdot (CHOH)_3 \cdot CO \cdot CH_2OH$
α- and β-Acrose

Fischer and Landsteiner [36] synthesized a tetrose (erythrose) by the aldol condensation of two molecules of glycolaldehyde in a 1 per cent solution of sodium hydroxide.

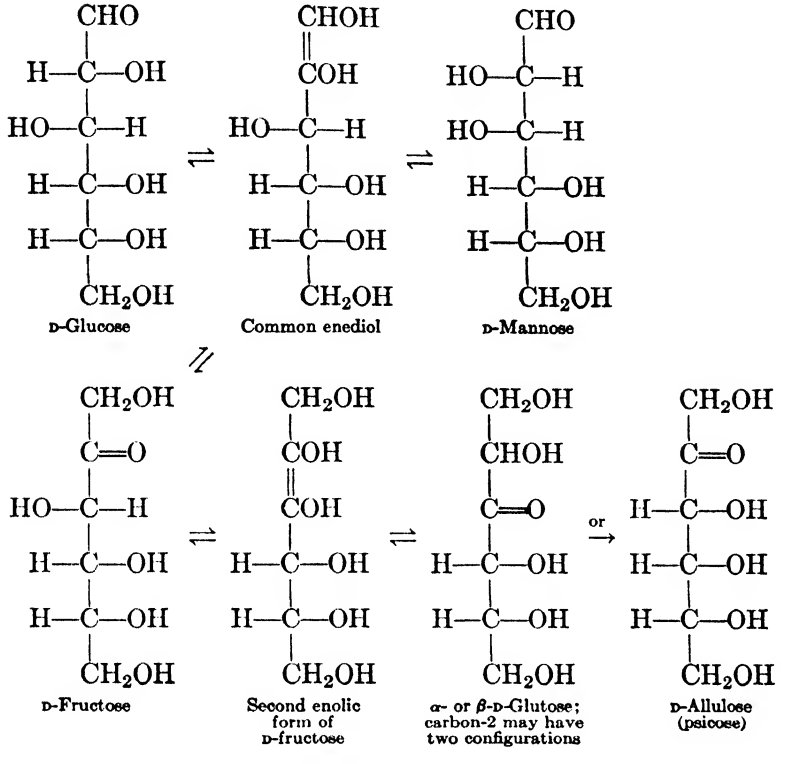
MUTAROTATION OR WALDEN INVERSION OF THE REDUCING CARBON. This transformation, which gives rise to α- and β-isomers, has been discussed on p. 536. In aqueous solution, it is attributed to the intermediate formation of the straight-chain aldehydrol form of the sugars.

LOBRY DE BRUYN TRANSFORMATION. In a series of papers beginning in 1895, Lobry de Bruyn and van Ekenstein [37] reported studies on the action of various alkalies on simple sugars, studies which were stimulated by Lobry de Bruyn's observation that the optical activity of dilute solutions of D-glucose containing a small amount of potassium hydroxide gradually decreases to almost zero, with an accompanying decrease in alkalinity. They showed that hexoses upon treatment with dilute alkalies undergo rearrangements which result in an equilibrium mixture of epimers and other isomeric forms. For example, starting with either D-glucose, D-mannose, or D-fructose, a mixture of these three sugars, together with D-allulose and α- and β-D-glutose, was obtained.

The interconversion of the three parent sugars is explained by the formation of a common enediol from each; by the subsequent shift of a hydrogen atom any one of the three sugars may be formed. D-Fructose, however, can give rise to a second enolic form which can revert to either D-fructose, D-allulose, or a 3-ketohexose, α- and β-D-glutose. In a similar manner, the D-galactose series yields an equilibrium mixture

[36] E. Fischer and K. Landsteiner, *Ber.*, **25**, 2549 (1892).

[37] C. A. Lobry de Bruyn, *Rec. trav. chim.*, **14**, 156 (1895); **16**, 257 (1897); C. A. Lobry de Bruyn and W. A. van Ekenstein, *ibid* **14**, 203 (1895); **16**, 274 (1897); **19**, 1 (1900).

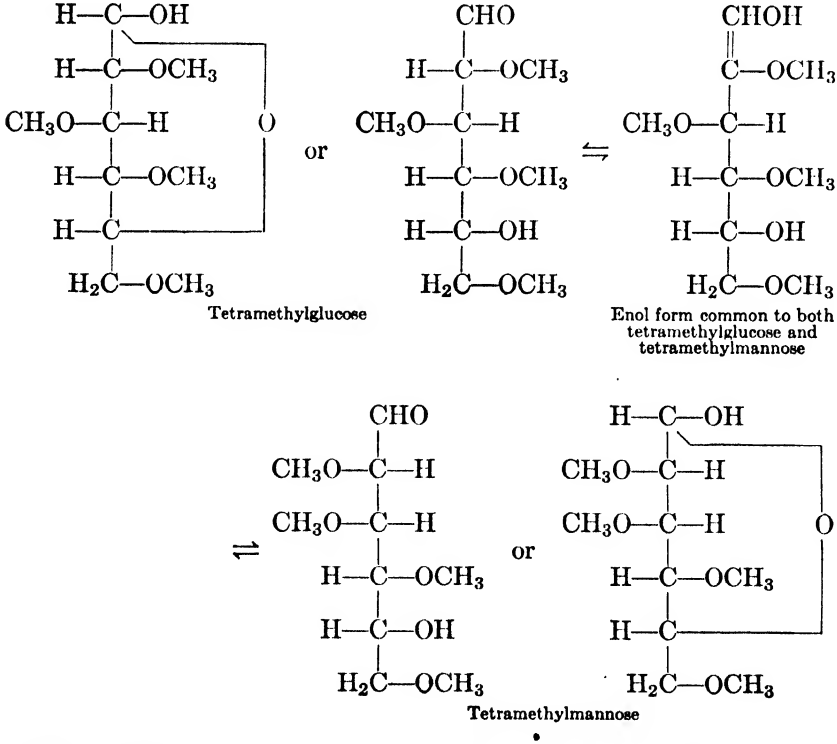


containing D-galactose, D-talose, D-tagatose, D-sorbose, and α- and β-D-galtose. In these rearrangements the sugars of the D-glucose series do not give rise to those of the D-galactose series or *vice versa*; the rearrangements thus appear to involve only carbons-1, -2, and -3, indicating that the sugar molecule is most reactive at the carbonyl end.

Although these transformations are represented as reversible, it should be emphasized that a true thermodynamic equilibrium does not result. since the initial sugar usually predominates in the reaction mixture, Moreover, there is considerable question as to whether glutose and galtose are definite compounds. Lobry de Bruyn and van Ekenstein were unable to crystallize these non-fermentable products and stated that they might probably be mixtures. In an extensive series of experiments on "glutose" formation from hexoses with lead hydroxide and phosphates, Spoehr and Strain [33] found wide variations in the yields, reducing power, and aldose content of the non-fermentable residues, and the "glutose" osazone was shown to be a mixture of phenylhydrazine

[33] H. A. Spoehr and H. H. Strain, *J. Biol. Chem.*, **85**, 365 (1929–30).

derivatives. Also Wolfrom and Lewis [39] found that upon treating D-glucose with lime water (*pH* 12.6) until polarimetric equilibrium was obtained the reaction product contained the following: D-glucose, 63.4 per cent; D-fructose, 30.9 per cent; D-mannose, 2.4 per cent; and non-sugar substances (probably saccharinic acids), 3.3 per cent. Lobry de Bruyn and, later, Nef assumed the enediol to be formed by the alternate addition and removal of the elements of water but the researches of Wolfrom and Lewis [40] with tetramethylglucopyranose and tetramethylmannopyranose support the simpler concept of enolization depicted here. When mild alkali treatment was employed either of these methylated sugars was converted at the same rate to an equilibrium mixture containing equal amounts of tetramethylglucose and tetramethylmannose, showing that a true equilibrium was established. As no ketoses were found, it was suggested that in these sugar ethers only the 1,2-enediol can form because further enolization is blocked by the methyl group.



[39] M. L. Wolfrom and W. L. Lewis, *J. Am. Chem. Soc.*, **50**, 837 (1928).

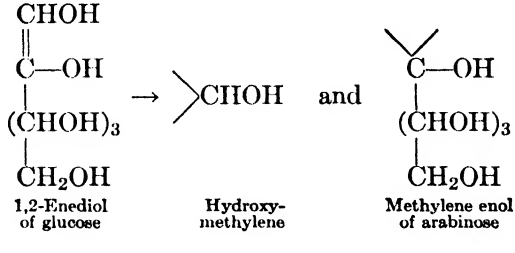
[40] M. L. Wolfrom and W. L. Lewis, *J. Am. Chem. Soc.*, **50**, 837 (1928); see also D J. Loder and W. L. Lewis, *J. Am. Chem. Soc.*, **54**, 1040 (1932).

These researches support the view that the Lobry de Bruyn transformation is due to the migration of a hydrogen atom to the carbonyl group and not to hydration followed by dehydration.

Because the sugars themselves undergo complex isomerizations and degradation in alkaline solution in addition to epimerization, the aldonic acids are employed when it is desired to bring about epimerization in the sugar series. By heating them in aqueous pyridine only the epimers are formed.[41] Saccharic acids, as well as the methyl derivatives of the γ- and δ-lactones, also undergo epimerization with this reagent.[42,43]

Effect of Strong Alkalies. As the hydroxyl-ion concentration is increased, the sugars undergo degradative changes; the nature of the products formed depends to a large extent on whether or not oxygen or an oxidizing agent is present. Early in the present century, Nef [44] carried out intensive studies for over ten years on the origin, isolation, and identification of the decomposition products which are formed from the sugars by the action of alkalies and alkaline oxidizing agents, such as the oxides of copper, mercury, and silver. He assumed that in the presence of strong alkali the hexoses form 1,2-, 2,3-, and 3,4-enediols by means of a progressive enolization along the carbon chain, and that these split at the double bond into a mixture of various carbohydrate-like fragments containing 1, 2, 3, 4, and 5 carbon atoms, the rupture at the double bond yielding radicals which contain divalent carbon.

Thus, the 1,2-enediol of glucose would give rise to hydroxymethylene and the methylene enol of arabinose.



1,2-Enediol of glucose — Hydroxymethylene — Methylene enol of arabinose

As divalent carbon should be an exceedingly reactive form, these radicals would absorb oxygen to yield formic acid from the hydroxymethylene, and D-arabonic acid (with some D-ribonic acid) from the methylene enol of arabinose.

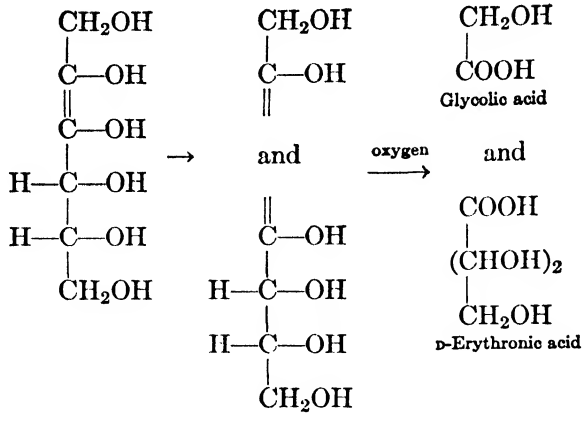
[41] E. Fischer, *Ber.*, **23**, 799 (1890).

[42] W. N. Haworth and C. W. Long, *J. Chem. Soc.*, 345 (1929).

[43] O. F. Hedenberg and L. H. Cretcher, *J. Am. Chem. Soc.*, **49**, 478 (1927).

[44] J. U. Nef, *Ann.*, **335**, 191 (1904); **357**, 214 (1907); **376**, 1 (1910); **403**, 204 (1914).

The 2,3-enediol would give the methylene enol of a diose and of D-erythrose,



which on exposure to oxygen would yield glycolic acid and erythronic acid (some L-threonic acid would be also formed). D-Erythronic acid is the chief optically active trihydroxybutyric acid found among the oxidation products of D-glucose when air or hydrogen peroxide is used as oxidizing agent in the presence of six equivalents of sodium hydroxide.

The 3,4-enediols would break up into two molecules of the methylene enol of glyceric aldehyde, $CH_2OH \cdot CHOH \cdot COH$, and oxidize to D- and L-glyceric acid.

Nef assumed that the initial fragments may: (1) undergo molecular rearrangement to form more stable compounds, (2) react with each other, one being oxidized and the other reduced, (3) combine or polymerize, and (4) undergo oxidation to form acids. On the basis of this theory he postulated that D-glucose in sodium hydroxide solution could yield a mixture containing no less than 116 compounds, and he actually succeeded in qualitatively identifying 93 of them in the form of either the pure compounds or known derivatives.

In 1935, Spengler and Pfannenstiehl [45] found that by using oxygen in place of air, the cleavage of an aldose sugar in 2 *N* potassium hydroxide solution takes place smoothly and an aldonic acid containing one less carbon atom than the original sugar is obtained in good yield. This technic in addition to those of Wohl, Ruff, and Weerman, thus serves as a useful method of degrading a sugar.

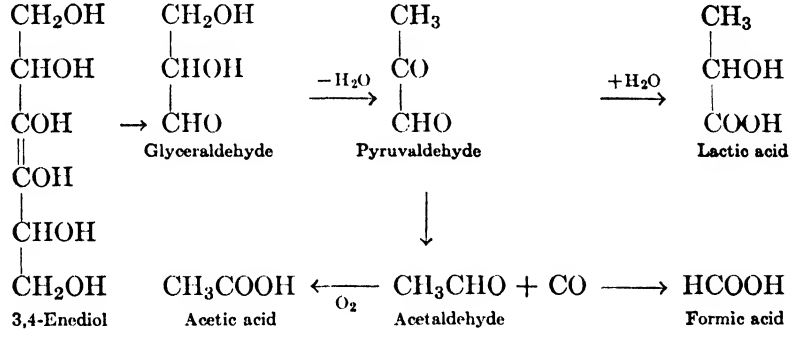
Several years later Evans and his co-workers [46] repreated and extended

[45] O. Spengler and A. Pfannenstiehl, German Patent 618,164 (1935).

[46] W. L. Evans, *et al.*, *J. Am. Chem. Soc.*, **47**, 3085, 3098, 3102 (1925); **48**, 2665, 2678, 2703 (1926); **50**, 486, 1496, 2267, 2543 (1928); **52**, 294, 3680, 4065 (1930); **53**,

the observations of Nef by carrying out extensive quantitative studies of the products formed under precisely controlled experimental conditions in regard to alkali concentration, oxidation potential (air, hydrogen peroxide, potassium permanganate, cupric hydroxide, etc.), temperature, and time. They accepted Nef's postulation of an equilibrium between 1,2-, 2,3-, and 3,4-enediols in alkaline solution and endeavored to correlate the products which were quantitatively isolated after the action of alkalies on glucose and galactose with these respective enediols.

Carbon dioxide obtained in the oxidation of hexoses was believed to arise principally from oxidative cleavage of the 1,2-enediol. An increase in the alkali concentration caused the production of larger amounts of oxalic acid due, it was believed, to the greater formation of the 2,3-enediol which cleaves to yield glycolic aldehyde followed by its oxidation to oxalic acid. They adduced evidence to show that pyruvaldehyde, lactic acid, acetic acid, and part of the formic acid arise from the fission of the 3,4-enediol. Thus:

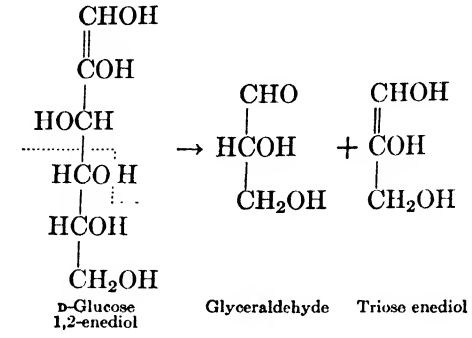


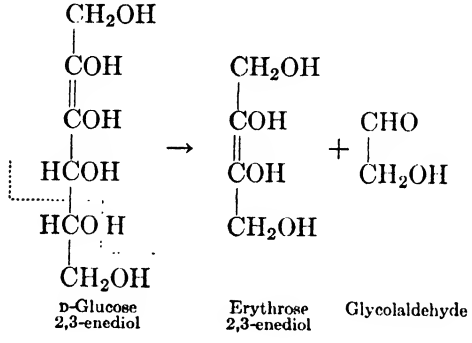
The amount of lactic acid produced was augmented by increasing the alkali concentration which, Evans and co-workers pointed out, would favor the formation of the 3,4-enediol as well as the transformation of pyruvaldehyde into lactic acid. Although the enediol theory with scission at the double bond satisfactorily explained the results obtained by Evans, *et al.*, it is now known that the bond energy of a carbon-to-carbon double bond is greater than that of singly linked carbons. Moreover, the work of Schmidt [47] shows that the presence of a double bond in a compound strengthens the next single bond and weakens the next following; and hence cleavage should occur between the α- and β-carbons.

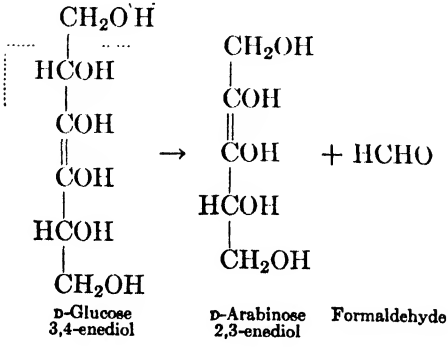
4384 (1931); **54**, 698 (1932); **55**, 4957 (1933); **57**, 200 (1935); *J. Org. Chem.*, **1**, 1 (1936); *J. Am. Chem. Soc.*, **58**, 797, 1661, 1890, 1950, 2388 (1936); and W. L. Evans, *Chem. Revs.*, **31**, 537 (1942).

[47] O. Schmidt, *Chem. Revs.*, **17**, 137 (1935).

According to this view, the scission of the three enolic forms of D-glucose should occur as follows:

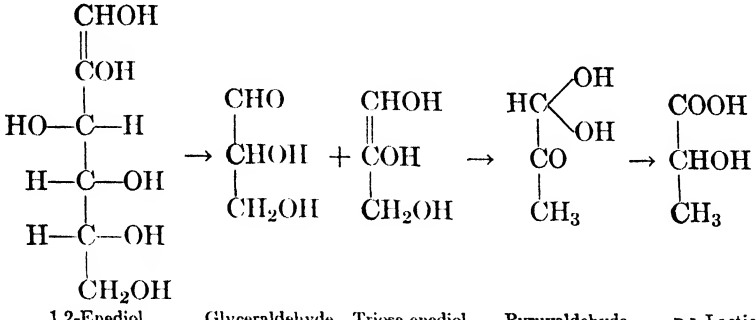






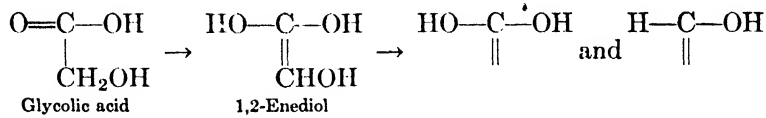
The enediols of glyceraldehyde, erythrose, and arabinose may conceivably rearrange to the normal form and undergo a different enolization and subsequent fragmentation; also, the 2,3-enediol of D-arabinose may undergo fragmentation into formaldehyde and the 2,3-enediol of erythrose before rearranging to the normal form.

The formation of lactic acid from the hexoses by the action of strong alkalies in the absence of oxidizing agents is explained by the Schmidt theory as follows:



1,2-Enediol of D-glucose, D-mannose, and D-fructose — Glyceraldehyde — Triose enediol — Pyruvaldehyde hydrate — D,L-Lactic acid

From a study of the quantities of carbonic, formic, acetic, lactic, and oxalic acids produced when aldonic acids were heated in barium hydroxide solution in the absence of air, Upson and his associates [48] concluded that the theory of Schmidt provided a more satisfactory explanation of the reaction mechanism than that advanced by Nef and Evans. In applying the Schmidt "double-bond rule" to the aldonic acids, they assumed that the carbon-to-oxygen double bond would influence the cleavage of the carbon chain in the same manner as a carbon-to-carbon double bond. As further evidence for the cleavage according to the Schmidt theory, they pointed out that, according to the Nef theory, glycolic acid would be expected to form 1,2-enediol and undergo cleavage to carbonic acid.



Glycolic acid — 1,2-Enediol

On the contrary, glycolic acid was found to be very stable toward alkalies. This is in accordance with the Schmidt theory, as this compound contains only two carbon atoms and hence would not undergo fragmentation.

Evans has pointed out that, whereas the Schmidt theory may be successfully applied in explaining the reactions which take place in the absence of air or oxidizing agents, the Nef theory of scission at the double

[48] W. D. Albert and F. W. Upson, *J. Am. Chem. Soc.*, **57**, 132 (1935); F. W. Upson, W. K. Noyce, and W. D. Albert, *J. Am. Chem. Soc.*, **61**, 779 (1939).

bond provides a more satisfactory explanation of the reaction products obtained in the presence of oxidizing agents. He suggested that the presence of oxidizing agents may involve a different process than that occurring in their absence. It is well known that oxidative cleavage normally occurs at a carbon-to-carbon double bond. This view is supported by Warshowsky,[49] from a quantitative study of the products obtained when D-glucose in potassium hydroxide solution was oxidized with molecular oxygen at 30°C. He obtained arabonic acid, formic acid, oxalic acid, and lactic acid in yields which accounted for about 96 per cent of the oxygen and the alkali consumed. Neither the Nef nor the Schmidt theory alone could account for the yields he obtained, although the chief products, arabonic acid and most of the formic acid, were considered to have resulted from cleavage according to the Nef theory. Although it would be possible to obtain arabonic acid by the cleavage of the 3,4-enediol between carbon-1 and carbon-2, in the manner which has been previously illustrated, ribonic, lyxonic, and xylonic acids should also be formed because of the loss of asymmetry of carbon-3 and carbon-4 in the formation of the enediol. However, arabonic acid was the only pentonic acid obtained.

Saccharinic Acid Formation. In addition to the reactions already described, the simple sugars when treated with strong alkalies undergo intramolecular oxidation and reduction. One part of the molecule is reduced and another part simultaneously oxidized to give rise to a series of so-called saccharinic acids which are isomeric with the original sugar. These rearrangements normally occur when a sugar is treated with strong alkali in the absence of oxygen. The saccharinic acids which arise from the hexoses are:

	Number of Possible Isomers
(*a*) Metasaccharinic acids $CH_2OHCHOHCHOHCH_2CHOHCOOH$	8
(*b*) Isosaccharinic acids $CH_2OHCHOHCH_2COH(COOH)(CH_2OH)$	4
(*c*) Saccharinic acids $CH_2OHCHOHCHOHCOH(COOH)(CH_3)$	8
(*d*) Parasaccharinic acids $CH_2OHCHOHCOH(COOH)(CH_2CH_2OH)$	4

[49] B. Warshowsky, Ph.D. thesis, University of Minnesota, Minneapolis, 1945.

Although the mechanism of saccharinic acid formation has been extensively studied, it is not completely understood.[50]

ESTABLISHMENT OF THE FORMULA FOR D-GLUCOSE

The structure of a simple sugar may be established by determining in sequence its empirical, molecular, structural, and configuration formulas by the correct interpretation of analytical data. As D-glucose is the central compound among the carbohydrates and has been very thoroughly investigated, it will be used as an example.

Proof of Molecular Structure. Elementary analysis of pure D-glucose (C = 40.0 per cent, H = 6.7 per cent, and O = 53.3 per cent) showed the empirical formula to be CH_2O, and from molecular weight determinations the molecular formula $C_6H_{12}O_6$ was established.

Glucose contains a carbonyl group since it reduces alkaline copper solutions, forms oximes with hydroxylamine, and hydrazones and osazones with hydrazines. On mild oxidation with bromine water it yields a monobasic acid (D-gluconic acid) containing six carbon atoms; hence the carbonyl group is on a terminal carbon atom.

When D-glucose is treated with acetic anhydride it forms an ester whose formula ($C_{16}H_{22}O_{11}$) shows it to be a pentaacetate.[51] Kiliani [52] established that glucose was a straight-chain pentahydroxy-aldehyde by preparing glucose cyanohydrin, hydrolyzing it to the corresponding acid which he then reduced with hydriodic acid in the presence of red phosphorus. He obtained *n*-heptylic acid, $CH_3 \cdot (CH_2)_5COOH$, which proves that the original six atoms formed a straight chain. Moreover, since the extra carbon atom introduced by the cyanohydrin reaction was found at the end of the chain, the carbonyl group is aldehydic. This evidence leads to the formula $CH_2OH \cdot (CHOH)_4 \cdot CHO$.

Establishment of the Open-Chain Configuration Formula of Monosaccharides. The Kiliani formula for glucose contains four asymmetric centers and, according to the Le Bel-van't Hoff theory, sixteen optical isomers are possible, of which four (D-glucose, D-mannose, and D- and L-galactose) occur in nature. During the period 1884–1894 Fischer succeeded in synthesizing and establishing the configuration of all the aldohexoses and lower sugars postulated by theory. The procedures

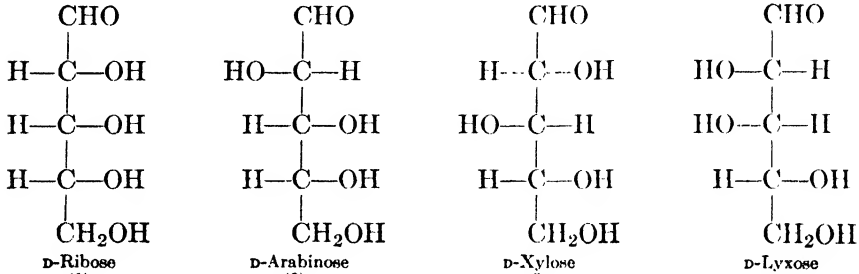
[50] The various theories which have been proposed to account for the formation of saccharinic acids are discussed by A. L. Raymond, in H. Gilman, *Organic Chemistry*, John Wiley & Sons, New York, 2nd ed. Vol. II, p. 1646, 1943.

[51] A. P. N. Franchimont, *Ber.*, **12**, 1940 (1879), prepared the first crystalline pentaacetate.

[52] H. Kiliani, *Ber.*, **19**, 767 (1886).

which he employed were osazone formation, reduction to alcohols, oxidation to acids, and the methods already described for building up (Kiliani synthesis) and degrading the sugar series (Wohl degradation). The use of these methods will be illustrated by showing how they were applied in establishing the configurations of the aldopentoses and aldohexoses.

Aldopentoses. There are eight possible aldopentoses corresponding to the D- and L-forms of arabinose, xylose, ribose, and lyxose. There are thus four alternative configurations for the D-forms:



D-Ribose (1) D-Arabinose (2) D-Xylose (3) D-Lyxose (4)

The L-forms are the mirror images of these. The facts and reasoning are as follows:

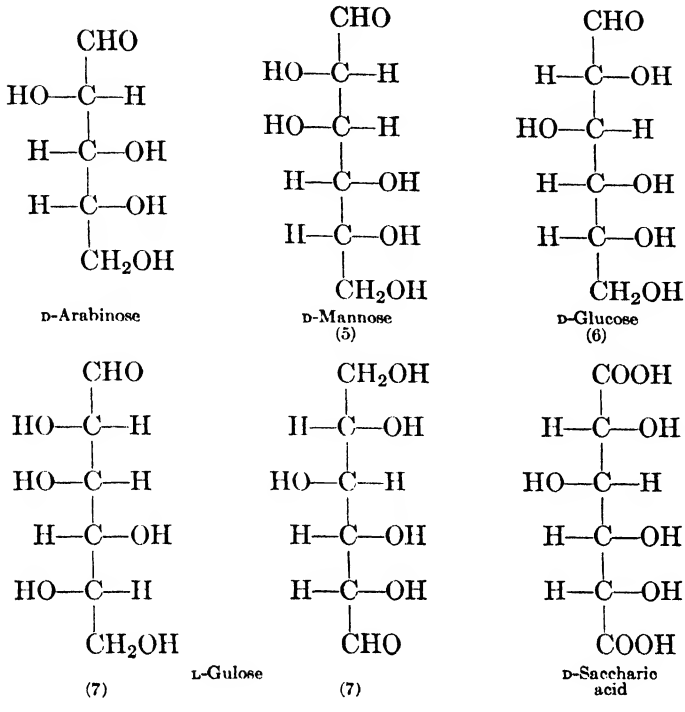
A. Arabinose and ribose form the same osazone, and hence their configuration must be identical except for carbon-2. Likewise, lyxose and xylose give identical osazones and are therefore also epimers. Arabinose and ribose must therefore be either (1) and (2) or (3) and (4).

B. On oxidation with nitric acid, arabinose and lyxose give *optically active* dibasic acids, whereas ribose and xylose give optically inactive acids. Hence arabinose and lyxose must be (2) and (4), and ribose and xylose (1) and (3).

C. Arabinose, on treatment with hydrogen cyanide followed by hydrolysis and oxidation, gives two optically active dicarboxylic acids, whereas the same treatment of lyxose gives one active and one inactive dicarboxylic acid. Therefore D-arabinose is (2), and D-lyxose is (4). Since D-arabinose is (2), D-ribose must be (1). Similarly, D-xylose must be (3); this conclusion is also confirmed by the fact that it yields an optically inactive dicarboxylic acid (trihydroxyglutaric acid) upon oxidation.

Aldohexoses. The configurations of D-glucose,[53] D-mannose, and L-gulose are derived from the following facts.

[53] See C. S. Hudson, *J. Chem. Ed.*, **18**, 353 (1941), for a discussion of Emil Fischer's discovery of the configuration of glucose.



A. When D-glucose is degraded, D-arabinose results; conversely, when a new asymmetric center is added to D-arabinose by the Kiliani procedure, two sugars, D-glucose and D-mannose, are obtained. Therefore D-glucose must be either (5) or (6).

B. Upon oxidation, D-glucose and L-gulose give the same dicarboxylic acid, D-saccharic acid. This means that the configuration of each asymmetric center in D-glucose must be the reverse of that in L-gulose and that the one can be derived from the other by interchanging the aldehyde and primary alcohol groups. If the CHO and CH_2OH groups are transposed in (5), the same sugar (D-mannose) results, since on rotation through 180° the two formulas can be superimposed. Therefore D-glucose [53] is (6), D-mannose is (5), and L-gulose is (7).

A similar line of reasoning establishes the configuration of the remaining aldohexoses.

Ketoses. The configurations of the ketoses are readily obtained from their relationship to the aldohexoses. Thus, D-fructose yields the same osazone as D-glucose and D-mannose (p. 557). Since osazone formation involves only carbon-1 and carbon-2, the configuration of carbon-3, carbon-4, and carbon-5 is the same in all three sugars. As D-fructose is a ketose sugar, the carbonyl group must be on carbon-2.

Establishment of the Ring Structure. *Pyranose Ring Structure.* Although a ring structure was early postulated for D-glucose to account for its low reactivity, its mutarotation, and the formation of the α- and β-methylglucosides, no methods were available for determining the size of the ring. As the aldonic acids formed γ-lactones, it was assumed for many years that the simple sugars also possessed a 1,4-oxide ring. The first chemical evidence that the normal or ordinary forms of the simple sugars had a 1,5-oxide ring, or pyranose structure, was obtained in 1926 by the application of the methylation and oxidation techniques developed by Haworth and his associates at the University of Birmingham. Later, other independent lines of evidence were obtained.

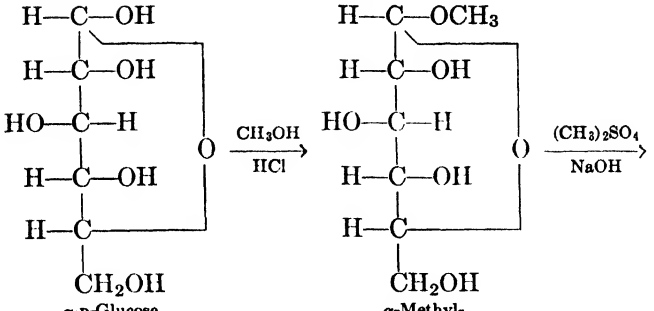
METHYLATION AND OXIDATION. It will be recalled that ordinary D-glucose forms α-methyl-D-glucoside when treated with hydrogen chloride in methanol and that, upon methylation with methyl iodide and silver oxide or dimethylsulfate and sodium hydroxide, tetramethyl-α-methyl-D-glucoside results. The glycosidic methyl group is readily hydrolyzed by dilute hydrochloric acid to yield tetramethyl-D-glucose in which all the hydroxyl groups have been converted to stable methyl ether linkages. In 1926, Hirst [54] oxidized tetramethyl-D-glucose with nitric acid and isolated and identified *i*-trimethoxy-glutaric acid and L-dimethyltartaric acid among the reaction products by converting them to their crystalline diamides. The formation of the former acid proves that the methyl ether derived from α-methyl-D-glucoside is 2,3,4,6-tetramethyl-D-glucose. The ring closure is therefore on carbon-5. As Armstrong had previously shown that α-methyl-D-glucoside yielded α-D-glucose upon enzymic hydrolysis, α-D-glucose is believed to possess the pyranose structure, on the assumption that no shift occurred in the ring during methylation. This assumption is now accepted. The successive steps in the reactions are shown on the facing page.

Similar studies with the ordinary crystalline forms of D-mannose, D-galactose, and D-fructose, as well as with the pentoses, showed that all existed in the pyranose form.

LACTONE STUDIES. Confirmatory evidence for the pyranose structure of the sugars has been obtained from the application of the Hudson lactone rule to the sugars themselves. In a critical comparison of the optical properties of several sugars, Haworth and his associates found that, in general, they obey the lactone rule if a six-atom ring structure is assumed.[55] Studies of the rates of lactonization of aldonic acids also

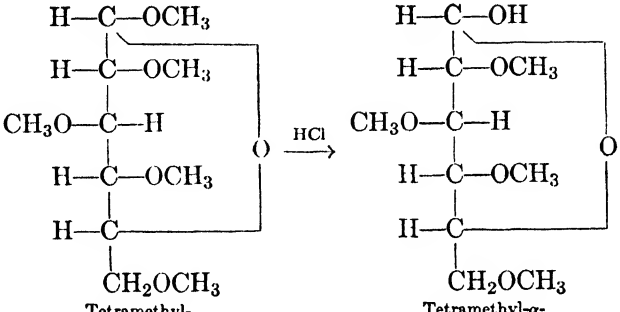
[54] E. L. Hirst, *J. Chem. Soc.*, 350 (1926).

[55] W. N. Haworth, *The Constitution of Sugars*, Longmans, Green and Co., New York, 1929.



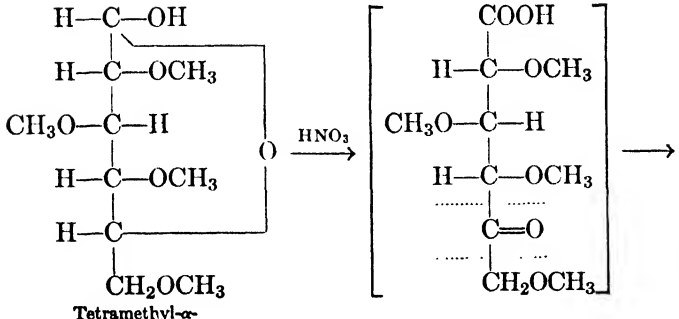
α-D-Glucose

α-Methyl-D-glucoside

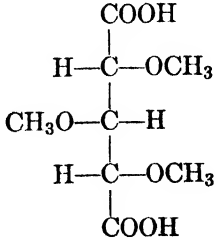


Tetramethyl-α-methyl-D-glucoside

Tetramethyl-α-D-glucose

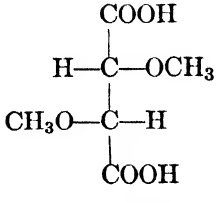


Tetramethyl-α-D-glucose



i-Trimethoxyxyloglutaric acid

and



L-Dimethyltartaric acid

support the pyranose structure in the normal sugars. In addition, it will be recalled that, when an aldose is rapidly oxidized to aldonic acids by hypobromite, δ-lactones are formed immediately without the intermediate production of free aldonic acids.[56] This indicates that the sugars possess a δ-oxide (pyranose) structure.

X-RAY STUDIES. Several x-ray investigations of D-glucose and other sugars in the crystalline form support the pyranose structure.[57] The five carbon atoms lie approximately in one plane, with the oxygen atom projecting slightly from the plane of the ring.

OXIDATIVE DEGRADATION WITH LEAD TETRAACETATE AND PERIODIC ACID. In a previous section (p. 552), it was pointed out that oxidative cleavage of the methylglycosides with lead tetraacetate or periodic acid affords a completely independent method of establishing the ring structure in these derivatives. For example, when the methylglycoside prepared from crystalline D-mannose is oxidized with these reagents, formic acid splits out, and oxidation of the resulting dialdehyde with bromine water in the presence of strontium carbonate, followed by hydrolysis of the glycosidic methyl group, yields glyoxalic acid and D-glyceric acid. This proves that the original mannoside contained a pyranose ring. When a methylglycoside prepared from the ordinary crystalline form of the pentose sugars is similarly treated, the end products, glyoxalic acid and glycolic acid, prove the presence of a pyranose ring in the original pentoside.

Existence of Simple Sugars in Forms Other than the Pyranose Ring Structure. Although it has been established that the ordinary crystalline forms of the pentoses and hexoses have the pyranose structure, both furanose and open-chain derivatives of the sugars have been prepared.

Several years after his original discovery of α- and β-methylglucoside, Emil Fischer obtained a third form, an impure syrupy mixture, which he called γ-methylglycoside, by reacting glucose with methanol in the presence of a considerable concentration of hydrogen chloride at a moderate temperature (10°–20°C.). This product was very easily hydrolyzed, and he regarded it as having a different ring structure from the α- and β-derivatives which he had previously discovered. In 1915, Irvine and his associates [58] found that the tetramethylglucose prepared from this so-called γ-glucoside differed in properties from the 2,3,4,6-tetramethylglucose prepared from the normal α- and β-methylglucosides. The furanose ring structure in this tetramethylglucose was proved by

[56] C. S. Hudson and H. S. Isbell, *Bur. Standards J. Research*, **8**, 327 (1932).

[57] E. G. Cox, T. H. Goodwin, and A. I. Wagstaff, *J. Chem. Soc.*, 978, 1495 (1935); H. Mark, *Chem. Revs.*, **26**, 169 (1940).

[58] J. C. Irvine, A. W. Fyfe, and T. P. Hogg, *J. Chem. Soc.*, **107**, 524 (1915).

Haworth and associates by methylation, conversion to the lactone, and its subsequent oxidation with nitric acid, as outlined below in connection with the establishment of the furanose ring structure in γ-methylglucofuranoside prepared from monoacetoneglucose.

Starting with monoacetoneglucose, Haworth and his associates [59] proved that glucofuranose derivatives exist. The isopropylidene residue, previously proved to be attached to carbon-1 and carbon-2, served to stabilize the furanose ring. The monoacetoneglucose was first treated with phosgene to prepare the 1,2-isopropylidene-5,6-carbonate derivative. The acetone residue is sensitive to acids and stable to alkalies, whereas the carbonate ester is stable to acids and sensitive to alkalies. Upon treatment of this mixed derivative with methanol and hydrogen chloride, the acetone residue was hydrolyzed off; the crystallizable α- and β-methylglucofuranoside-5,6-carbonate was formed, the furanose ring being stabilized by the carbonate group. Upon mild saponification with baryta water, the carbonate group was hydrolyzed off, giving the α- and β-methylglucofuranoside. The α-derivative was obtained in crystalline form, and upon methylation and hydrolysis tetramethylglucofuranose was obtained. Although this could not be obtained in a crystalline condition it formed a crystalline lactone upon oxidation with hypobromite, and the furanose structure of the lactone was established by the fact that it yielded L-dimethyltartaric acid upon oxidation with nitric acid.

The formation of γ-glucosides is common to the pentoses and hexoses. They are characterized by their ease of hydrolysis. The ring is very labile, and efforts to isolate the free sugars in the *furanose* form have failed because this form is unstable and quickly reverts to the pyranose structure.

In addition to the pyranose and furanose forms, derivatives of arabinose, glucose, galactose, and fructose have been prepared which have an open-chain structure.[60]

Configuration of the Glycosidic Carbon in α- and β-Isomers. Reference has previously been made to the system proposed by Hudson [61] in 1909 for designating the anomeric pairs of sugars and their derivatives. In the D-series the more dextrorotatory member of an α-,β-pair is des-

[59] W. N. Haworth, *et al.*, *J. Chem. Soc.*, 2436 (1927); 2796 (1929); 2254 (1932).

[60] P. A. Levene and G. M. Meyer, *J. Biol. Chem.*, **69**, 175 (1926); **74**, 665 (1927); M. L. Wolfrom, *J. Am. Chem. Soc.*, **52**, 2464 (1930); **57**, 2498 (1935); P. Brigl and H. Mühlschlegel, *Ber.*, **63**, 1551 (1930); F. Micheel, H. Ruhkopf, and F. Suckfull, *Ber.*, **68**, 1523 (1935); N. W. Pirie, *Biochem. J.*, **30**, 374 (1936); G. E. Felton and W. Freudenberg, *J. Am. Chem. Soc.*, **57**, 1637 (1935); E. Pacsu and F. V. Rich, *J. Am. Chem. Soc.*, **54**, 1697 (1932).

[61] C. S. Hudson, *J. Am. Chem. Soc.*, **31**, 66 (1909); **60**, 1537 (1938).

ignated as the α-form, and the hydroxyl group is assigned to the right in the Fischer projection formula; in the L-series, the more levorotatory anomer is named α-, and the hydroxyl group is assigned to the left. This proposal was not wholly empirical since it was based on his rules of isorotation.

A number of physical and chemical methods have been proposed for determining, experimentally, the configuration of the glycosidic carbon in α- and β-isomers in relation to a hydroxyl or other group attached to one of the other carbon atoms in the ring, the configurations of which are already known from Fischer's researches.[62]

Of the physical methods, the conductivity studies of Boeseken [63] with α- and β-D-glucose in boric acid solutions are frequently quoted as evidence in support of the Hudson convention. In α-D-glucose, this convention places the hydroxyl groups on carbon-1 and carbon-2 in the *cis* position. Boeseken observed that glycols with hydroxyl groups in the *cis* configuration formed a complex with boric acid with a resultant increase in electrical conductivity. In applying this observation to D-glucose he found that the conductivity of the α-form in boric acid solution decreased during mutarotation; with β-D-glucose the reverse was true. On this basis, the hydroxyl groups on carbon-1 and carbon-2 are *cis* to each other, and since the configuration of carbon-2 has previously been established as to the right, the hydroxyl group on carbon-1 must also be represented to the right in α-D-glucose. Later work, however, has shown that this method is not reliable in the case of mannose and rhamnose, and carbohydrate chemists today take the view that the results with D-glucose were merely fortuitous.

Another line of physical evidence rests on the observation that β-methyl-D-glucoside has a higher dissociation constant than the α-form, indicating a greater separation of the —OH and —OCH_3 groups on carbon-1 and carbon-2 in the former.

The first clear-cut evidence of a chemical nature was obtained from studies of the addition of methanol to the 1,2-anhydro sugars (glycosans). Thus, Bergmann and Schotte [64] found that a syrup presumed to be 1,2-anhydromannose readily reacted with methanol at room temperature to produce α-methylmannoside. Similarly, Brigl [65] and later Hickinbottom [66] found that methanol added to crystalline 1,2-anhydro-

[62] See F. Micheel, *Chemie der Zucker und Polysaccharide*, Akademische Verlagsgesellschaft m.b. H., Leipzig, 1939.

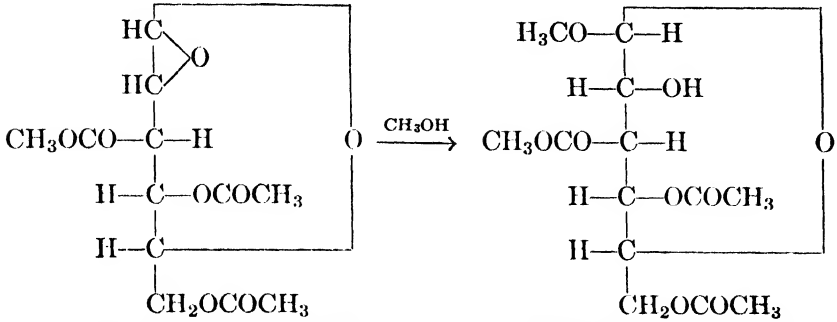
[63] J. Boeseken, *Ber.*, **46**, 2612 (1913).

[64] M. Bergmann and H. Schotte, *Ber.*, **54**, 1564 (1921).

[65] P. Brigl, *Z. physiol. Chem.*, **122**, 245 (1922).

[66] W. J. Hickinbottom, *J. Chem. Soc.*, 3140 (1928).

3,4,6-triacetylglucose to yield the triacetyl derivative of β-methyl-glucoside.



1,2-Anhydro-3,4,6-triacetylglucose β-Methyl-3,4,6-triacetylglucoside

Since it is a general behavior of ethylene oxide rings to yield compounds having *trans* configurations upon fission,[67] the methoxyl groups in α-methylmannoside and in β-methylglucoside must be *trans* to the respective hydroxyl groups on carbon-2. As Armstrong [68] in 1903 had correlated the α- and β-methylglucosides with α- and β-glucose by means of enzymatic hydrolysis with maltose and emulsin, the hydroxyl group on carbon-1 in β-methyl-D-glucose must be *trans* to that on carbon-2.

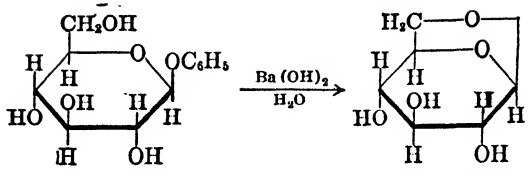
The conversion of β-D-glucose by destructive distillation into levoglucosan by loss of the elements of water between carbon-1 and carbon-6 to form a 1,6-anhydro derivative has been regarded as evidence in favor of the configuration assigned to carbon-1 since the glycosidic group can readily react with the hydroxyl group when they are on the same side of the pyranoid ring. But, it has recently been shown [69] that α-D-glucose can also form levoglucosan; therefore the production of this compound does not constitute proof of the configuration of carbon-1. However, the fact that β-phenyl-D-glucoside is readily and quantitatively converted into levoglucosan upon alkaline hydrolysis, whereas the α-derivative cannot be, is believed to be evidence that in β-D-glucose the hydroxyl groups on carbon-1 and carbon-6 bear a *cis* relation in the pyranose ring.[70]

[67] H. Ohle and C. A. Schultz, *Ber.*, **71**, 2302 (1938); S. Müller, M. Moricz, and G. Verner, *Ber.*, **72B**, 745 (1939).

[68] E. F. Armstrong, *J. Chem. Soc.*, **83**, 1305 (1903).

[69] R. M. Hann and C. S. Hudson, *J. Am. Chem. Soc.*, **64**, 2435 (1942).

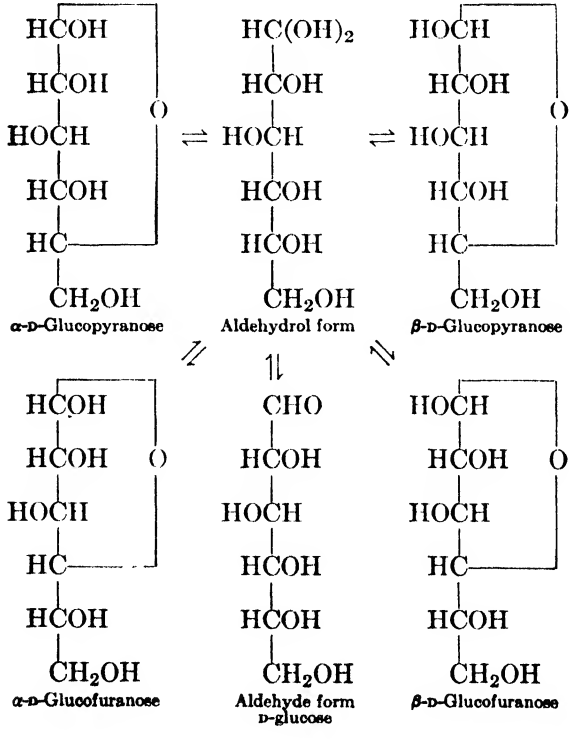
[70] E. M. Montgomery, N. K. Richtmeyer, and C. S. Hudson, *J. Am. Chem. Soc.*, **65**, 3 (1943).



β-Phenyl-D-glucoside — 1,6-Anhydroglucose (levoglucosan)

Application of this reaction to the phenylglucosides of xylose, mannose, and galactose gave results in accordance with the Hudson convention. Moreover, oxidation studies with periodic acid (see p. 552) have shown that the α-methylglycosides of glucose, mannose, galactose, gulose, and altrose have identical configurations for their glycosidic carbon atoms.

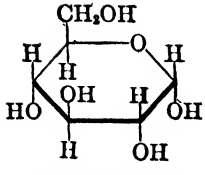
Simple Sugars in Solution. The properties of D-glucose (and other simple sugars) in solution indicate that several forms may be present in equilibrium, including the open-chain, enol, and aldehydrol forms, as well as the α- and β-furanose and pyranose forms.

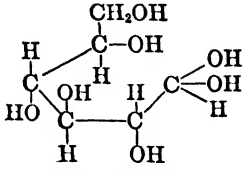


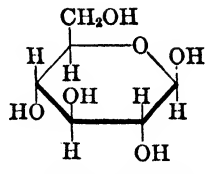
α-D-Glucopyranose ⇌ Aldehydrol form ⇌ β-D-Glucopyranose

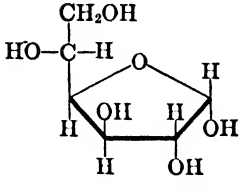
α-D-Glucofuranose — Aldehyde form D-glucose — β-D-Glucofuranose

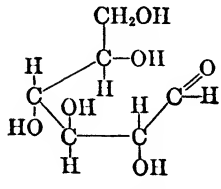
or

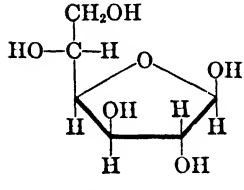

α-D-Glucopyranose


Aldehydrol form


β-D-Glucopyranose


α-D-Glucofuranose


Aldehyde form
D-glucose


β-D-Glucofuranose

The quantities of the open-chain, enol, and aldehydrol forms present are augmented when the alkalinity of the solution is increased. That the aldehyde form must be present is shown by the fact that D-glucose gives many of the aldehyde reactions; as this form of the sugar is removed by reacting with carbonyl reagents, more will be produced to restore the equilibrium. The concentration of the aldehyde form which is present in equilibrium, however, must be very low because D-glucose and other pentoses and hexoses fail to add sodium bisulfite or give the Schiff test. The presence of the carbonyl group in a compound is indicated by an absorption band at 2,800 Å. Solutions of aldoses do not give this characteristic band, and hence the concentration of the aldehyde form must be below the limit of spectroscopic investigation (0.3 per cent of the total sugar). Solutions of levulose and of sorbose, however, show faint absorption at 2,800 Å., which is indicative that low concentrations of the open-chain forms are present.[71] However, aldohexoses may be reduced at the mercury cathode, and this is attributed to the presence of the aldehyde form; the calculated percentage of the open-chain form ranges from 0.022 per cent for D-glucose to 1.10 per cent for D-allose. These polarographic studies revealed that the percentage of the open-chain form increased with increasing *p*H and increasing concentration.[72]

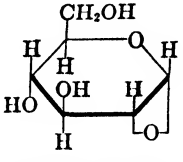
The marked influence of temperature on the optical activity of D-fructose solutions indicates that it is highly tautomeric. Crystalline derivatives corresponding to the keto, pyranose, and furanose forms have been prepared.

[71] W. Bednarcyk and L. Marchlewski, *Biochem Z.*, **300**, 42 (1938).

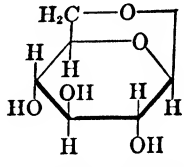
[72] S. M. Cantor, P. P. Quintin, and Q. P. Peniston, *J. Am. Chem. Soc.*, **62**, 2113 (1940).

Glycosans, Glycoseens, Glycals, and Desoxy Sugars. By the removal of hydrogen and oxygen or of oxygen alone, four important series of sugar derivatives are obtained.

Anhydro Sugars or Glycosans.[73] In the anhydro sugars or glycosans, the elements of water have been eliminated from a simple sugar with the formation of an internal ether. The ring may form between different carbon atoms, but in D-glucose the two best known are 1,2-anhydroglucose (α-glucosan) and 1,6-anhydroglucose (β-glucosan, or levoglucosan).



1,2-Anhydroglucose-α-glucosan



1,6-Anhydroglucose-β-glucosan (levoglucosan)

Pictet and his associates [74] prepared α-glucosan by heating α-D-glucose under reduced pressure, whereas the destructive distillation of starch and other polysaccharides composed of glucose residues yielded levoglucosan.[74]

The 1,2-anhydro sugars add methanol to form a methylglycoside in which the configuration of the methoxy group on carbon-1 is *trans* to the hydroxyl group on carbon-2. This reaction and the formation of levoglucosan by the alkaline hydrolysis of β-phenyl-D-glucoside serve to establish the configuration of the glycosidic carbon as discussed on p. 585.

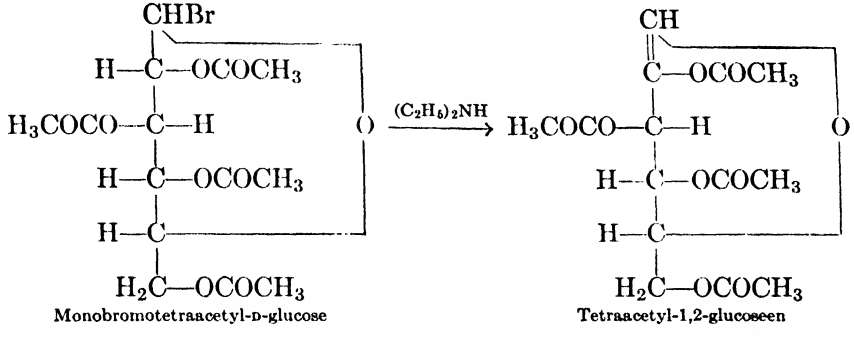
The 1,6-anhydro sugars are readily hydrolyzed by acids to yield the original sugar. This property makes them particularly useful in synthesizing various sugar derivatives which have substituent groups on carbon-2, carbon-3, or carbon-4. For example, 1,6-anhydro-D-mannopyranose forms a 2,3-monoacetone compound which has only one hydroxyl group (on carbon-4); by condensation with an acetobromo sugar, a disaccharide derivative is produced with a linkage between carbon-4 of the mannose residue and the glycosidic carbon of the other sugar; both the acetone residue and the 1-6 ring are readily hydrolyzed by aqueous acid.

[73] A detailed discussion of the chemistry of anhydro sugars by S. Peat will be found in *Advances in Carbohydrate Chem.*, **2**, 37 (1946).

[74] A. Pictet and P. Castan, *Helv. Chim. Acta*, **3**, 645 (1920); *Compt. rend.*, **171**, 243 (1920); A. Pictet and J. Sarasin, *Helv. Chim. Acta*, **1**, 78 (1918); A. Pictet and M. Cramer, *Helv. Chim. Acta*, **3**, 640 (1920).

Glycoseens. Like the glycosans or anhydro sugars, the glycoseens may be considered to be formed from the simple sugars by elimination of the elements of water, but in these compounds a double bond is introduced, the best known being the 1,2- and 5,6-glucoseens.

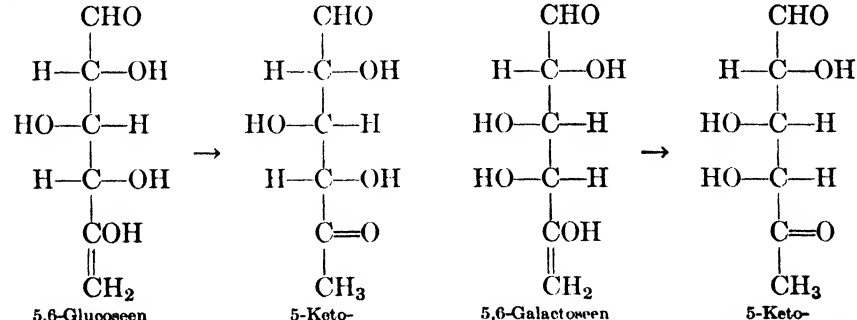
The 1,2-glycoseens are only known as acetyl derivatives and may be formed by treating the acetobromo sugar (p. 541) with diethylamine whereby hydrogen bromide is eliminated between carbon-1 and carbon-2. The reactions with D-glucose are



Monobromotetraacetyl-D-glucose — Tetraacetyl-1,2-glucoseen

The 5,6-glycoseens may be prepared by reacting the 6-bromo or 6-iodo derivative with silver fluoride in pyridine solution, or by treating a 6-tosyl diacetone sugar with alcoholic ammonia.[75]

Two biologically important 5,6-glycoseens are 5,6-glucoseen and 5,6-galactoseen, since they ketonize to form 5-keto-6-desoxy sugars:

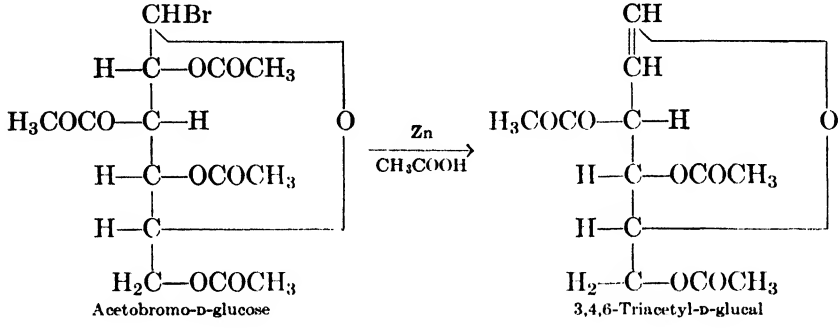


5,6-Glucoseen — 5-Keto-6-desoxyglucose (isorhamnose) — 5,6-Galactoseen — 5-Keto-6-desoxygalactose (fuconose)

Glycals. The glycals differ from the parent aldose sugar in having a double bond between carbon-1 and carbon-2 through the loss of two hydrogen and two oxygen atoms. They are prepared by reducing the

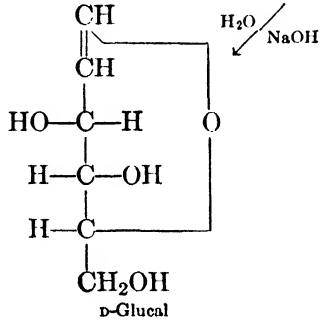
[75] See H. Gilman, *Organic Chemistry*, John Wiley & Sons, New York, Vol. II, 2nd ed., p. 1625, 1943.

corresponding acetobromo sugar with zinc dust in acetic acid, and then removing the remaining acetyl groups by saponification.



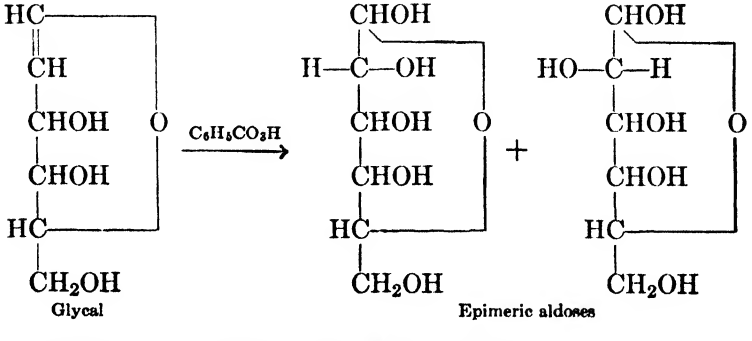
Acetobromo-D-glucose

3,4,6-Triacetyl-D-glucal



D-Glucal

Since the asymmetry of carbon-2 is destroyed, epimeric sugars yield the same glycal. Bergmann and Schotte [76] introduced a useful method of preparing epimeric aldoses by the action of perbenzoic acid on a glycal.



Glycal

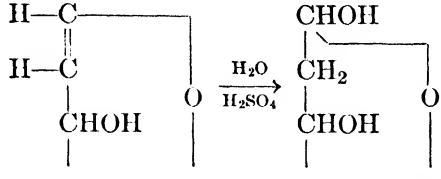
Epimeric aldoses

[76] M. Bergmann and H. Schotte, *Ber.*, **54B**, 440 (1921).

The proportions of the two epimers which are formed vary widely depending on the glycal and the conditions of the oxidation.

Desoxy Sugars. The desoxy sugars, as the name implies, are formed from the parent sugar by the removal of an oxygen atom from either a secondary or a primary alcohol group to produce a $—CH_2$ or a $—CH_3$ group, respectively. They are of biological interest because 2-desoxy and 6-desoxy sugars occur in nature.

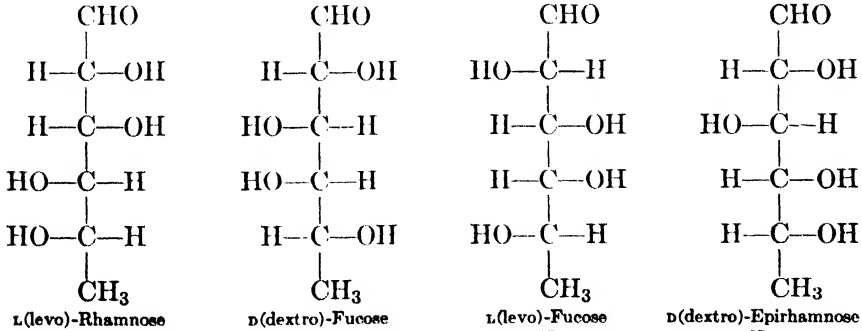
The 2-desoxy sugars can be readily prepared by the direct addition of water to a glycal in the presence of sulfuric acid as a catalyst.



They are more reactive than the parent sugar and give the Schiff test for aldehydes. Unlike the pentoses or methylpentoses, which yield furfural or methylfurfural upon treatment with hydrochloric acid, the naturally occurring 2-desoxy-D-ribose (or 2-desoxy-D-arabinose) yields levulinic acid. This was responsible for the long-held belief that levulinic acid was a constituent of thymus nucleic acid.

The removal of an oxygen atom from the terminal $—CH_2OH$ group of a hexose gives rise to 6-desoxy sugars, several of which occur in nature.[77]

As these hexose reduction products may be regarded as derived from a pentose by the replacement of a hydrogen of the terminal $—CH_2OH$ group with a $—CH_3$ group, they are normally called methylpentoses, but in order to bring out their relation to the hexoses they are also named by adding the suffix *methylose* to the name of the parent hexose.



[77] They may be synthesized by replacing the terminal tosyl group in a monoacetone tosyl derivative with iodine and reducing the iodo derivative.

Four methylpentoses which occur in plants as glycosides or as polymers are L-rhamnose, D-epirhamnose, D- and L-fucose. L-Fucose occurs in gum tragacanth and L-rhamnose in gum arabic.

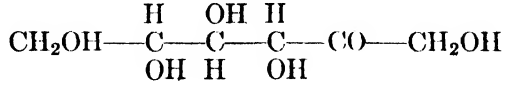
Fischer Synthesis of Monosaccharides. As previously noted, a sugar syrup was first synthesized by Butlerow in 1861 through the action of saturated calcium hydroxide solution on paraformaldehyde. Several years later Loew, using formaldehyde, repeated the synthesis and named the syrup *formose*. Later, Fischer oxidized glycerol with bromine water in the presence of sodium carbonate and obtained a mixture of ketohexose sugars, called α- and β-acrose, as a result of aldol condensation of the glyceric aldehyde and dihydroxyacetone produced by oxidation of the glycerol (p. 567). These sugars were so named because he had previously obtained them by treating acrolein dibromide with barium hydroxide.

$$2CH_2Br\cdot CH_2Br\cdot CHO + 2Ba(OH)_2 \rightarrow \underset{\alpha\text{-and }\beta\text{-Acrose}}{C_6H_{12}O_6} + 2BaBr_2$$

In 1890, Fischer isolated α-acrose as the osazone, from which he prepared the osone. Upon reduction of the latter he obtained *α-acritol*, which proved to be identical with D,L-mannitol, thus establishing that α-acrose was either D,L-mannose, or D,L-fructose. The former possibility was eliminated because α-acrose did not form an insoluble phenylhydrazone, a property which is characteristic of mannose. From the synthetically prepared *α-acrose*, Fischer prepared the D- and L-forms of fructose, mannose, and glucose as well as L-gulose and L-idose, according to the scheme shown on the opposite page.[78]

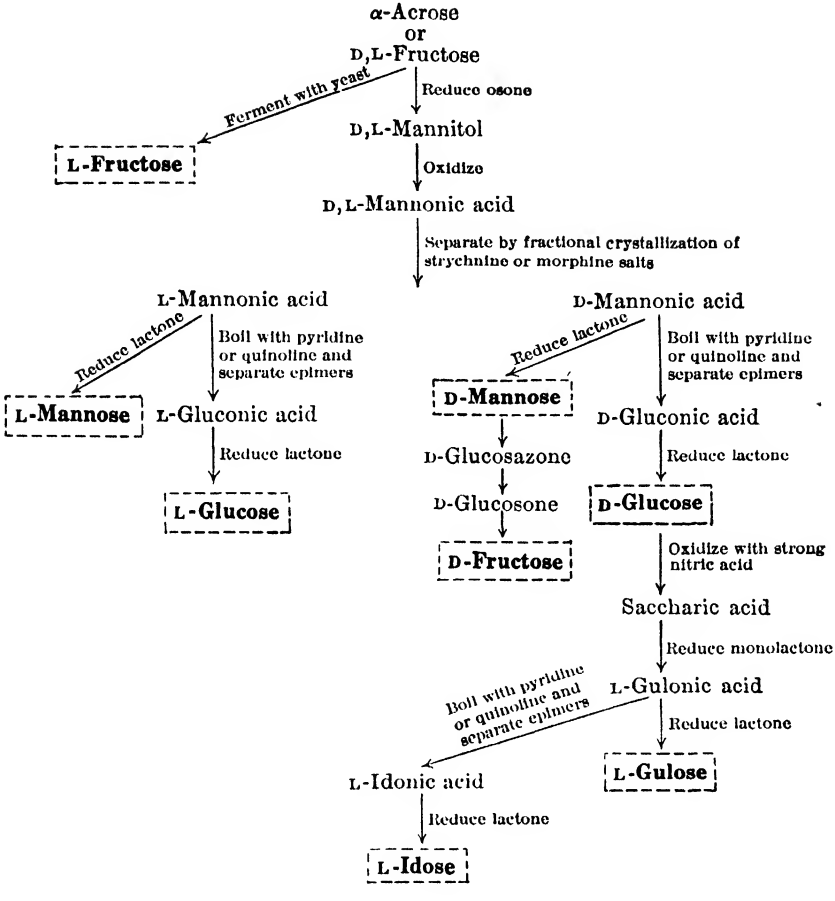
In 1924, β-acrose was shown to be principally D,L-sorbose, although a pentose, 2-araboketose, was also present.[79]

The formula for D-sorbose is

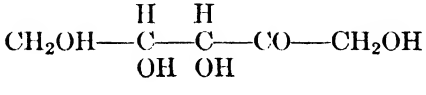


[78] These reactions have been detailed in discussing the chemical reactions of the sugars. It will be recalled that Fischer designated the gulose and idose he prepared in this manner as D forms because they were derived from saccharic acid prepared from D-glucose. However, as described on p. 560, the reactions involved in the preparation of gulose from D-glucose result in a transposition of the aldehyde and primary alcohol groups of the original glucose, thus causing the inversion from the D- to the L-series. Similarly, in the preparation of idose, the aldehyde group replaces the primary alcohol group on carbon-6 of the original glucose.

[79] W. Küster and F. Schoder, *Z. physiol. Chem.*, **141**, 110 (1924).



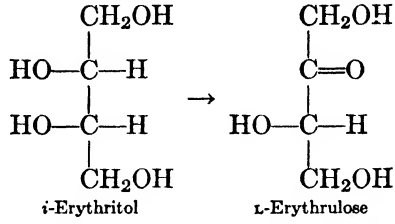
and that for 2-araboketose is



OCCURRENCE OF THE MONOSACCHARIDES IN NATURE

Trioses and Tetroses. The triose sugars, D-glyceraldehyde, and dihydroxyacetone are intermediate products of carbohydrate metabolism. The ketotriose L-erythrulose is produced by the bacterial oxi-

dation of *i*-erythritol with *Acetobacter xylinum* [80] and *Acetobacter suboxydans*.[81]



i-Erythritol L-Erythrulose

The remarkable property of bacteria of the genus *Acetobacter* in oxidizing a secondary alcohol group of certain sugar alcohols was first discovered by Bertrand [82] who noted that *Acetobacter xylinum* oxidized carbon-5 of D-sorbitol to L-sorbose. He found that the organism attacked a secondary alcohol group on a penultimate carbon only when the hydroxyl group on the adjacent carbon was in the *cis* configuration. From Bertrand's rule, it might be expected that *i*-erythritol would yield both D- and L-erythrulose but only the L-form is produced.

Pentoses. Small amounts of aldopentoses have been reported free in vegetables, and they are sometimes secreted in urine in an abnormal metabolic condition known as pentosuria. These carbohydrates, however, occur principally in the combined form and are widely distributed in plants as various types of polysaccharides and in a number of glycosides; a pentose unit is also present in plant and animal nucleic acids.

The principal aldopentoses which are obtained by the hydrolysis of plant products are D-xylose and L-arabinose; D-arabinose, D-ribose, and certain desoxyaldopentoses are also found.

D-Xylose (together with L-arabinose) is a hydrolytic product of xylan (wood gum) which is found in the alkali-soluble hemicellulose fraction of wood, in straw and the outer coating of seeds (wheat bran, oat hulls, cottonseed hulls). L-Arabinose is obtained along with other carbohydrate products upon the hydrolysis of vegetable gums, such as gum arabic and cherry tree gums, and the pectins; in animals it occurs along with D-arabinose in some cases of pentosuria. D-Arabinose is of much less wide occurrence than L-arabinose. It is obtained upon the hydrolysis of the glycosides barbaloin and isobarbaloin of aloes and of the polysaccharide of tubercle bacilli.

[80] G. Bertrand, *Compt. rend.*, **130**, 1330 (1900); *Bull. soc. chem.*, [3] **23**, 681 (1900); *Ann. chim. phys.*, [8] **3**, 181 (1904); H. Müller, C. Montigel, and T. Reichstein, *Helv. Chim. Acta*, **20**, 1468 (1937).

[81] R. L. Whistler and L. A. Underkofler, *J. Am. Chem. Soc.*, **60**, 2507 (1938).

[82] G. Bertrand, *Compt. rend.*, **130**, 1330 (1900).

Either D-ribose or 2-desoxy-D-ribose is obtained on the hydrolysis of nucleic acids, and a D-ribose unit is present in certain of the nucleotides which serve as hydrogen carriers in carbohydrate metabolism.

Methylpentoses (6-desoxyhexoses, or methyloses) are found as constituents of plant polysaccharides and natural glycosides; L-rhamnose (6-desoxy-L-mannose) is a component of many glycosides particularly in combination with flavonol derivatives, and is also found among the saponins. D-Epirhamnose (6-desoxy-D-glucose) and D-fucose (6-desoxy-D-galactose) are constituents of the glycoside convolvulin. L-Fucose (6-desoxy-L-galactose) occurs as the polymeric anhydride fucosan, a polysaccharide found in the cell walls of marine algae. Several 2,6- or bisdesoxyhexoses, such as cymarose (probably 2,6-desoxy-3-methyl-D-allose or altrose) and digitoxose (2,6-desoxy-D-allose or altrose) occur as components of natural glycosides. L-Xyloketose has been found in certain cases of pentosuria and may possibly be derived from D-xylose by reduction to xylitol and the bacterial oxidation of carbon-4 in a manner analogous to the formation of L-sorbose from D-sorbitol by the sorbose bacterium.

Hexoses. Of the sixteen possible aldohexoses only D-glucose, D-mannose, and D- and L-galactose are found in nature; of these only D-glucose is commonly found in the free state. D-Glucose, dextrose, or grape sugar is found free in plant saps and fruits, where it is usually associated with D-fructose, and in the circulating medium of animals. D-Glucose is a constituent sugar of most oligosaccharides, of the polysaccharides starch, glycogen, and cellulose, and of many natural glycosides. A derivative of D-glucose, 2-acetyl-glucosamine, is a component of mucoitin sulfuric acid (the prosthetic group of certain of the glycoproteins). The polysaccharide chitin, which is the structural element of the exoskeletons of insects and the *Crustaceae*, is composed of acetyl-glucosamine units.

D-Mannose occurs as mannans and is commonly prepared from vegetable ivory, the seed of the tagua palm, *Phytelephas marcrocarpa*.[83] In the animal kingdom mannose occurs in the trisaccharide aminoglucose-dimannose, which is the repeating unit in the carbohydrate constituent of egg albumin, serum albumin, and serum globulin.

D-Galactose occurs as one of the sugar residues in several oligosaccharides (lactose, melibiose, raffinose, and stachyose) and is widely distributed in polysaccharides, being especially abundant in algae and lichens, in the pectins, and in the wood of the Western larch. Many of the polysaccharides contain galacturonic acid residues in addition to the

[83] C. S. Hudson and H. L. Sawyer, *J. Am. Chem. Soc.*, **39**, 470 (1917); H. S. Isbell, *Bur. Standards J. Research*, **26**, 35 (1941).

galactose units. Snails and their eggs are said to contain galactogen, an analogue of glycogen. Natural glycosides in which galactose is the sugar component are relatively rare, but several of the saponins and a few of the flavone derivatives and anthocyanins are galactosides. Chondrosamine (2-acetyl-aminogalactose) is one of the groups present in chondroitin sulfuric acid, the prosthetic group of certain of the glycoproteins. Galactose is also a component of the galactolipids or cerebrosides, kerasin, phrenosin, and nervone, which are present in brain tissue.

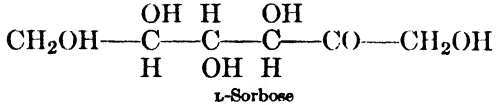
D-Fructose, levulose, or fruit sugar occurs free in plant saps and juices along with D-glucose. It is a component of the oligosaccharides, sucrose, raffinose, melezitose, and gentianose. Its polysaccharide is inulin, the reserve carbohydrate in the tubers of the dahlia, chicory, girasole (Jerusalem artichoke). Methods have been developed for the preparation of D-fructose from inulin.[84] Ordinary crystalline fructose consists chiefly of the β-form. It is the sweetest of all the sugars, is very hygroscopic, and is difficult to crystallize. According to Biester and Wood,[85] the relative sweetness of the sugars on the basis of sucrose as 100 is as shown in Table 55.

TABLE 55. RELATIVE SWEETNESS OF THE VARIOUS SUGARS

(Data of Biester, Wood, and Wahlin)

Sucrose	100.0	Maltose	32.5
Fructose	173.3	Rhamnose	32.5
Invert sugar	130.0	Galactose	32.1
Glucose	74.3	Raffinose	22.6
Xylose	40.0	Lactose	16.0

L-Sorbose is found in stored mountain ash berries where it probably originates from the oxidation of carbon-5 of D-sorbitol by the action of certain bacteria of the genus *Acetobacter*.[86]



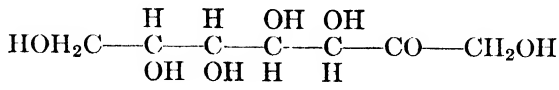
L-Sorbose

[84] E. L. Jackson, *Bur. Standards Scientific Papers*, No. 519, p. 587 (1926).

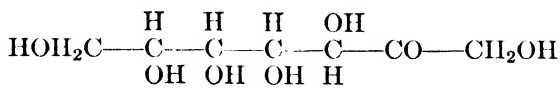
[85] A. Biester, M. W. Wood, and C. S. Wahlin, *Am. J. Physiol.*, **73**, 387 (1935). For a detailed review of research on the sweetness of sugars see A. T. Cameron, *The taste sense and the relative sweetness of sugars and other sweet substances*, Sugar Research Foundation Scientific Report Series No. 9 (Dec., 1947).

[86] E. I. Fulmer, J. W. Dunning, J. F. Guymon, and L. A. Underkofler, *J. Am. Chem. Soc.*, **58**, 1012 (1936), found that *Acetobacter suboxydans* is preferable to *Acetobacter xylinum* for the oxidation of D-sorbitol. With the former organism, conditions have been established whereby yields of 80 to 85 per cent of L-sorbose may be obtained.

Two naturally occurring ketoheptoses [87] have been discovered: D-mannoketoheptulose,



which exists in the free state in the fruit of the avocado tree (*Persea gratissima* Gaert),[88] and sedoheptulose (D-altroheptulose),



which is found free in the common stone crop (*Sedum spectabile* Bor).[89]

Hudson [90] has proposed a systematic nomenclature for the heptoses (and higher carbon synthetic sugars) by combining the names of two hexoses to which they are related. For example, in sedoheptulose, carbon atoms 3, 4, 5, and 6 have the configurations of D-altrose, whereas carbon atoms 3, 4, and 5 have the configuration of D-fructose, and the systematic name is D-altro-D-*fructoheptose*, the hexose syllable which refers to the carbon atoms nearer the carbonyl group being italicized.

Two additional ketoheptoses, volemulose and perseulose, are obtained by the action of *Acetobacter xylinum* on the naturally occurring heptitols D-volemitol and D-perseitol.

Occurrence of the Sugar Alcohols.[91] The higher sugar alcohols are widely distributed in plants. They are water-soluble, they crystallize readily, and they are not fermented by yeast but are attacked by a variety of microorganisms.

Tetritols. The tetrahydric alcohol *i*-erythritol is found in lichens, algae, and fungi. Some species are reported to utilize it as a carbohydrate, but yeast, most species of bacteria, and the higher animals are incapable of metabolizing it. The sorbose bacterium oxidizes it to erythrulose. Erythritol is very sweet, more than twice as sweet as sucrose.

Pentitols. Four pentitols, two of which are inactive and two active, are possible. The two *meso* forms are D-ribitol (or adonitol), formed by the reduction of D-ribose, and D-xylitol, formed by the reduction of D-xylose. The two active forms, D- and L-arabitol, are obtained by the reduction of either D- and L-arabinose or L- and D-lyxose, respectively.

[87] For a discussion of the heptoses and higher sugars see paper by C. S. Hudson, *Advances in Carbohydrate Chem.*, **1**, 1 (1945).

[88] F. B. La Forge and C. S. Hudson, *J. Biol. Chem.*, **28**, 511 (1917).

[89] F. B. La Forge and C. S. Hudson, *J. Biol. Chem.*, **30**, 61 (1917).

[90] C. S. Hudson, *J. Am. Chem. Soc.*, **60**, 1537 (1938).

[91] For a discussion of the occurrence and metabolism of the sugar alcohols see the paper by C. J. Carr and J. C. Krantz, Jr., *Advances in Carbohydrate Chem.*, **1**, 175 (1945).

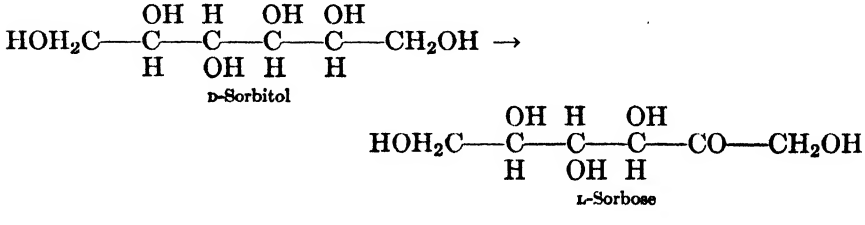
The ribitol residue occurs in riboflavin, and D-arabitol has been found in the fungus *Ustilago virens* which infects rice in Japan, and in the lichen *Lobario pulmonario*. L-Rhamnitol, $CH_3(CHOH)_4 \cdot CH_2OH$, is a homolog of this series formed by the reduction of L-rhamnose.

Hexitols. Of the ten possible hexitols only D-mannitol, dulcitol, and D-sorbitol, the reduction products of D-mannose, D-galactose, and D-glucose, respectively, are of natural occurrence. Of these, D-mannitol is the most abundant. It is found in the sago, pineapple, olive, date palm, and persimmon, the first three containing 5 to 10 per cent; vegetables, such as asparagus, carrot, and sweet potato, contain approximately 1.0 per cent D-mannitol, and lesser amounts are found in celery, turnip, onion, green pea, cauliflower, and the string bean. The mannas, the sweet exudates of various plants, are the best natural sources, that of the plane tree containing 80 to 90 per cent. It is a normal constituent of silage where it is formed by the bacterial reduction of D-fructose (from sucrose). Many varieties of fungi contain from 10 to 20 per cent of D-mannitol. It also occurs in concentrations of 25 per cent or more in many species of seaweeds. D-Mannitol is only sparingly soluble in cold water and is virtually insoluble in all organic solvents except those which are water-soluble, such as acetic acid, methanol, and pyridine. It is not hygroscopic, is neutral in reaction, and is equal in sweetness to sucrose. It is absorbed from the alimentary tract of animals but is only partially converted to glycogen in the liver; most of it is excreted unchanged in the urine.

Dulcitol (optically inactive), formed by the reduction of D-galactose, is not widely distributed in the higher plants but is found in red algae. It is a precursor of glycogen in the white rat but is less completely utilized than D-mannitol or D-sorbitol.[91]

D-Sorbitol is found in many fruits of the *Rosaceae* as well as in vegetables. In edible plants, the richest sources are the service, rowan, or sorb berries and the mountain ash berry. They contain 5 to 10 per cent of D-sorbitol, and the European varieties are used as food. Sorbitol has been extensively studied as a substitute for carbohydrates in the diet of diabetics with rather conflicting results.[91]

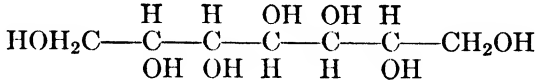
As noted previously, D-sorbitol is oxidized at carbon-5 by *Acetobacter xylinum* and *Acetobacter suboxydans*, thereby producing L-sorbose.



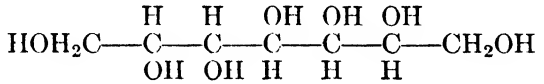
This reaction is utilized in the synthesis of ascorbic acid (see p. 910). Sorbitol has a sweetness of 54 compared with sucrose as 100 and is readily soluble in water, forming solutions which are characterized by high viscosity, low hygroscopicity, non-volatility, and great chemical stability; as a result it is particularly useful as a conditioner, plasticizer, or humectant.

D-Sorbitol and D-mannitol are now synthesized commercially by the reduction of D-glucose, and, as they are inexpensive, they and their derivatives are finding extensive applications as solvents and as emulsifying and dispersing agents for foods and vitamins.

Heptitols. D-Perseitol,



occurs along with D-mannoheptulose in the fruit of the avocado tree (*Persea gratissima* Gaert). As no fermentable sugars are present in this fruit the carbohydrate metabolism of this plant is unique. D-Volemitol,



was first discovered in a rather rare species of mushroom, *Lactarius volemus* Fr., and later it was found in the roots of *Primula officinalis* Jacq., *elatior* Jacq., and *grandiflora* Lmk.

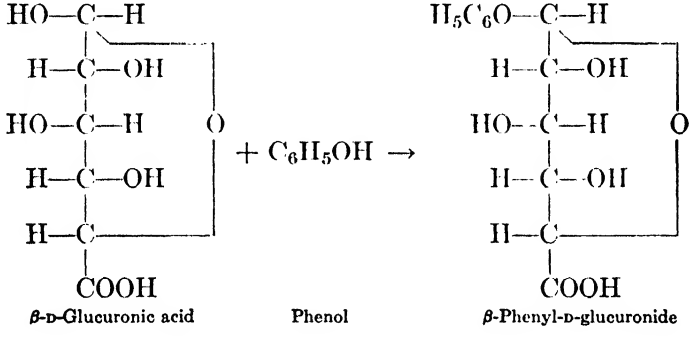
Occurrence of the Sugar Acids. *Ascorbic Acid.* Ascorbic acid or vitamin C is the lactone of a 2-keto-aldonic acid, namely the 2,3-enediol of L-gulofuranolactone; it is considered in Chapter 36.

Uronic Acids. Of the various acids which are obtainable by oxidizing the terminal alcohol groups of aldoses, the hexuronic acids are of the greatest biochemical importance. In these acids, the primary alcohol group on carbon-6 of the parent sugar is oxidized to a carboxyl group. They have the pyranose ring structure, and they can exist in α- and β-forms. Several molecules can condense to form polymeric anhydrides called uronides, which may possess either an α- or β-linkage. The hexuronic acids are reducing substances and are unstable toward hot mineral acids, in which they decompose to give carbon dioxide and furfural.

In the higher plants D-glucuronic, D-galacturonic, and D-mannuronic acid are widely distributed as constituent units of certain types of polysaccharides such as the vegetable gums, vegetable mucilages, and the pectins. D-Glucuronic and D-galacturonic acid also occur in certain of the "type-specific" polysaccharides of bacteria.

In the animal kingdom, D-glucuronic acid is a constituent of mucoitin sulfuric acid, and galacturonic acid is a component of chondroitin sulfuric acid, the prosthetic groups of glycoproteins.

D-Glucuronic acid together with glycine and sulfuric acid and, to a lesser extent, cysteine, glutamine, acetic acid, and ornithine, are used by mammals as a means of detoxicating poisonous substances which have gained access to the body.[92] Two types of linkage occur in the conjugation of the toxic substance with D-glucuronic acid, the glucosidic type and the ester type. Phenol, trichloromethanol, chloral, and camphor are in part detoxicated by the formation of β-glucuronides.



β-D-Glucuronic acid Phenol β-Phenyl-D-glucuronide

This detoxication mechanism serves as an excellent biological method for obtaining D-glucuronic acid (which is difficult to make in the laboratory).

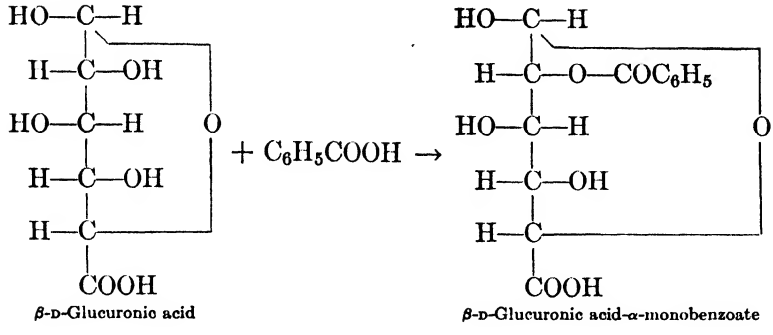
Chloral, $CCl_3 \cdot CHO$, must first be reduced to trichloroethanol, $CCl_3 \cdot CH_2OH$, and camphor, which contains a ketone group, to the corresponding secondary alcohol, borneol, before condensation.

When sodium benzoate is ingested by mammals, part of the benzoic acid is eliminated in combination with glycine as hippuric acid (benzoylglycine, $C_6H_5CO \cdot NH \cdot CH_2COOH$) and part as an ester of glucuronic acid.

The proportions of benzoic acid which are detoxicated by conjugation with glycine and with glucuronic acid vary in different mammals.[92] The compound formed with benzoic acid was originally named benzoylglucuronic acid, but Quick [93] has shown that an ester rather than a uronide is formed since the aldehyde group remains free, and it is therefore more correctly named β-D-glucuronic acid-α-monobenzoate. The

[92] For a review of detoxication mechanisms see A. M. Ambrose and C. P. Sherwin, *Ann. Rev. Biochem.*, **2**, 377 (1933); also B. Harrow and C. P. Sherwin, *ibid.*, **4**, 263 (1935).

[93] A. J. Quick, *J. Biol. Chem.*, **69**, 549 (1926); **70**, 59, 397 (1926); **98**, 537 (1932).



β-D-Glucuronic acid — β-D-Glucuronic acid-α-monobenzoate

main source of D-glucuronic acid is the carbohydrate store in the body, and the completely diabetic dog can synthesize glucuronic acid from the glucose which would otherwise be excreted.

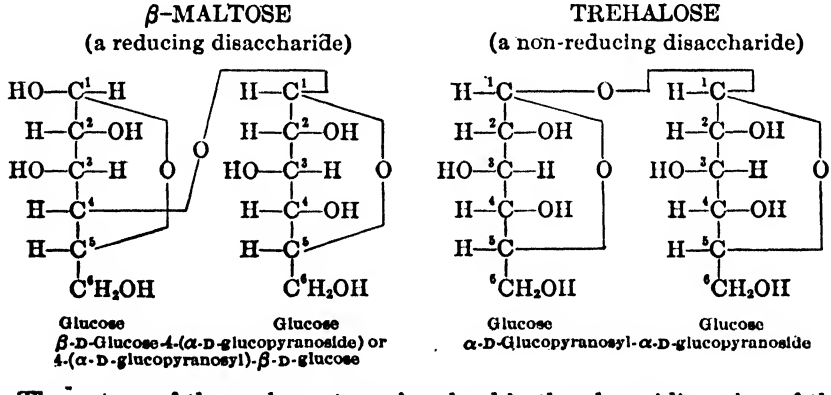
It is of interest to note that free glucuronic acid is metabolized with difficulty by animals and cannot be utilized for conjugation. This would seem to indicate that the toxic substance first condenses with glucose to form a β-glucoside, and then the secondary alcohol group on carbon-6 of the glucopyranose ring is oxidized to a carboxyl group. To test this hypothesis, Pryde and Williams [94] fed β-phenyl- and β-bornyl-glucosides to dogs and found that they were not excreted as the corresponding glucuronides; on the contrary, hydrolysis took place, and the phenol was excreted as phenylsulfate. Several of the natural plant glycosides yield toxic substances such as benzaldehyde and hydrocyanic acid upon hydrolysis, and one of their functions in plants may be to protect them against poisonous substances.

[94] J. Pryde and R. T. Williams, *Biochem. J.*, **28**, 131 (1934).

CHAPTER 24

The Oligosaccharides—Complex or Compound Sugars

The oligosaccharides may be considered to be derived by the mutual union of two, three, or four monosaccharide residues by loss of the elements of one, two, or three molecules of water, respectively, to give disaccharides, trisaccharides, and tetrasaccharides. The names di-, tri-, and tetrasaccharide indicate the number of single sugar residues. They are glycosidic condensation products of the simple sugars, the second sugar unit serving as the "aglycone group," *i.e.*, the glycosidic hydroxyl of one of the constituent sugars is substituted in the same manner as is glucose in the α- and β-methylglucosides. If the union occurs in such a way that the reducing group of one of the sugars is left intact, the complex sugar which is formed will mutarotate, exhibit reducing properties, form an osazone, and give the other carbonyl reactions of the sugars. If, on the other hand, the linkage between the sugars involves the glycosidic hydroxyl groups of all the component sugars, the complex sugar is non-reducing, will not mutarotate, form an osazone, or give the other reactions characteristic of the carbonyl group. The disaccharides, maltose and trehalose, both of which contain two D-glucopyranose residues, are examples of these two types:



The nature of the carbon atoms involved in the glycosidic union of the sugar units serves as a basis for subclassifying the compound sugars. Thus the disaccharides are divided into two groups: *non-reducing*, in

which the glycosidic union is through the carbonyl carbon of each component, as in trehalose, and *reducing*, in which the glycosidic union is from the carbonyl carbon of the one sugar to a non-carbonyl carbon of the other sugar, as in maltose. The reducing disaccharides are subclassified according to the number of the carbon atom of the sugar residue which serves as the alcohol. With the exception of D-fructose, the monosaccharide residues in the naturally occurring compound sugars have the pyranose ring structure, and hence the linkage in the reducing oligosaccharides may involve carbon-1, carbon-2, carbon-4, or carbon-6 of the hexose sugar serving as the alcohol. When D-fructose serves as the glycosidic component it occurs in the furanose form (except in turanose). The more common naturally occurring compound sugars are condensation products of hexoses, and glucose is one of the constituent sugars.

Systematic Nomenclature. In the systematic nomenclature of the compound sugars, different methods have been used by various authors. The two which are now most commonly employed will be exemplified in naming the reducing disaccharides. In the one method, the hydrogen attached to the oxygen on carbon-1 of the cyclic form of an aldose is considered as being removed in the condensation of the two sugar molecules to give rise to a glycosidic radical. The name of the sugar component serving as the alcohol and containing the unsubstituted potential carbonyl group ends in "ose," and the carbon atom involved in the linkage is designated by the number. If known, the configuration of the hydroxyl group on the reducing carbon, and the stereochemical nature of the glycosidic linkage are designated by the letters α and β in the usual manner. Thus β-maltose would be named β-D-glucose-4-(α-D-glucopyranoside). This is sometimes written simply as 4-α-glucosidoglucose. Since the reducing sugar residues in the reducing oligosaccharides are present in the pyranose form, it is not necessary to indicate their ring structures in the systematic names.

In the other common method of nomenclature, both the oxygen and hydrogen of the glycosidic carbon are considered as being removed in the condensation, thereby giving rise to a *glycosyl* radical. The structural name of β-maltose according to this system is therefore 4-(α-D-glucopyranosyl)-α-D-glucose. In the first system the suffix "oside" indicates the glycosidic component, whereas in the second system this component is designated by the suffix "osyl." In the non-reducing disaccharides the two component sugars are jointly glycosidic, and the suffix "osyl" is used for one and "ide" for the other. Thus, trehalose is named α-D-glucopyranosyl-α-D-glucopyranoside.

The classification and systematic names of the naturally occurring oligosaccharides are shown in Chapter 22 (p. 521).

Determination of Structure. The determination of the structure of the oligosaccharides is rather difficult, and only the constitutions of the common naturally occurring ones have been established. For example, the determination of the structure of a reducing disaccharide involves (1) the nature of the two constituent sugars, (2) the component which serves as the alcohol and glycosidic portion, respectively, (3) the carbon atom of the sugar serving as the alcohol portion which is involved in the glycosidic linkage, (4) the stereochemical nature of the glycosidic linkage (α or β), (5) the position of the internal oxygen rings in the constituent sugars, and (6) the configuration (α or β) of the carbonylic hydroxyl group in the sugar serving as the alcohol.

The problems enumerated under (2), (3), (4), and (6) have been solved by the methylation and oxidation technics of Haworth and his associates. All free hydroxyl groups are methylated, and the methylated sugar is hydrolyzed into its component monosaccharide derivatives by acid hydrolysis. From a study of these methylated derivatives, which can be separated and identified by comparison with methylation reference compounds of known structure, it is possible to locate the carbon atoms involved in the glycosidic linkage and in the oxide rings. The stereochemical configuration of the glycosidic component is ascertained by determining whether the sugar is hydrolyzed by an α- or β-glucosidase and by a study of the optical relations involved. The configuration of the carbonyl hydroxyl is ascertained in the same manner as for the simple sugars (p. 583).

The steps involved in establishing the structure of an oligosaccharide will be elucidated by consideration of maltose.

1. Maltose yields 2 molecules of D-glucose on hydrolysis and hence is a glucose glucoside.

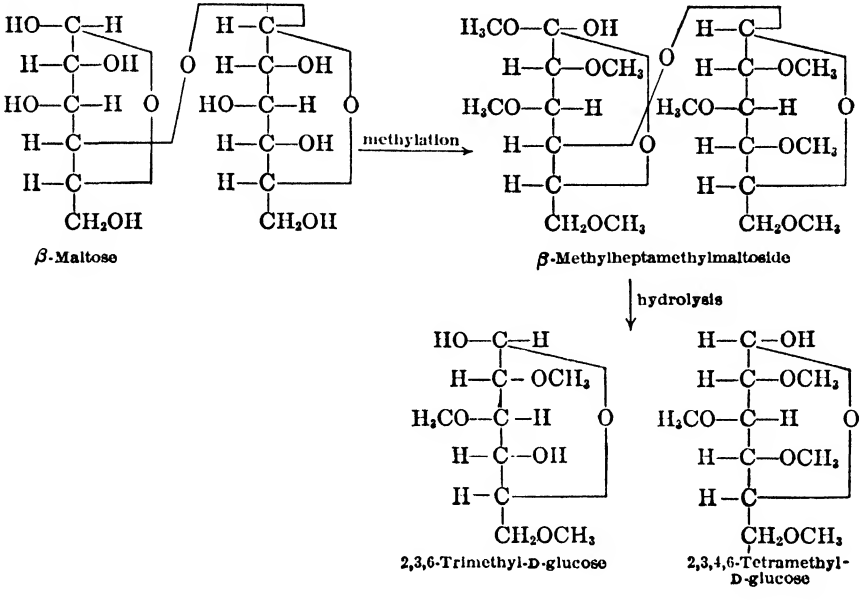
2. Maltose is a reducing sugar and adds one molecule of hydrogen cyanide; hence, one of the carbonyl groups must be functional.

3. It is hydrolyzed by maltase, an enzyme which hydrolyzes α-glucosides; also α-D-glucose is initially formed on hydrolysis. Maltose must, therefore, be an α-glucoside.

4. Upon complete methylation[1] and hydrolysis, an octamethyl derivative, methyl-heptamethyl-maltoside, is formed which on acid

[1] In methylating an oligosaccharide, the Fischer method of preliminary glycoside formation cannot be applied because the methanol solution of hydrogen chloride would cause hydrolysis of the glycoside linkage. Moreover, the Purdie methylation reagents (methyl iodide and silver oxide) cannot be used because of their oxidizing action on the reducing group. Haworth and Leitch, *J. Chem. Soc.*, **115**, 809 (1919), overcame these difficulties by preliminary formation of the methylglycoside with methyl sulfate at low temperatures before employing the more stringent conditions

hydrolysis yields 2,3,6-trimethyl-D-glucose and 2,3,4,6-tetramethyl-D-glucose.



As the glucosidic portion of the molecule yields 2,3,4,6-tetramethyl-D-glucose [2] this establishes that it has the pyranose structure. The glucose unit serving as the alcohol portion of the molecule yields 2,3,6-trimethyl-D-glucose,[3] but the isolation of this trimethyl ether does not complete the determination of the structure since either carbon-4 or carbon-5 may be involved in the glucoside linkage and carbon-5 or carbon-4, respectively, in the ring structure. It is not safe to assume a similar ring structure in the free and in the combined form; this became evident when it was found that fructose exists in the furanose form in oligosaccharides [4] and is spontaneously transformed into the pyranose form upon liberation of the free sugar.

5. The problem of which carbon is involved in the glucoside linkage and which in the oxide ring was solved by Haworth and Peat [5] in 1926. By oxidation of maltose with bromine water, the aldose portion of the

required to methylate the remaining hydroxyl groups with the methyl sulfate-sodium hydroxide reagent. A final treatment with methyl sulfate and silver oxide is then generally used to insure that methylation is complete.

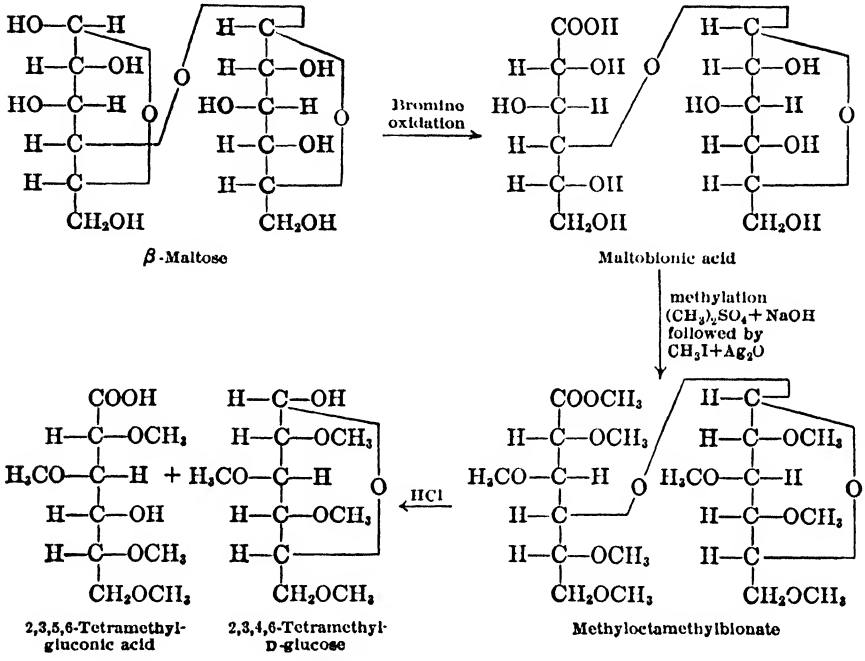
[3] A. W. Purdie and J. C. Irvine, *J. Chem. Soc.*, **87**, 1022 (1905).

[3] J. C. Irvine and I. M. A. Black, *J. Chem. Soc.*, 862 (1926); C. J. A. Cooper, W. N. Haworth, and S. Peat, *J. Chem. Soc.*, 876 (1926).

[4] Turanose is an exception.

[5] W. N. Haworth and S. Peat, *J. Chem. Soc.*, 3094 (1926).

molecule is oxidized to a monocarboxylic acid, maltobionic acid, thereby opening the ring. Methylation of the calcium salt of this acid leads to the methyl ester of octamethylbionic acid, and hydrolysis gives both 2,3,4,6-tetramethyl-D-glucose and 2,3,5,6-tetramethyl-D-gluconic acid. The latter was isolated and identified as the crystalline phenylhydrazide derivative.



The position of the free hydroxyl group on carbon-4 was also established by the formation of a γ-lactone, as revealed by its rate of hydrolysis [6] and by its oxidative degradation to L-(dextro)-dimethyl tartaric acid (see p. 580). Carbon atom 4 is, therefore, the one which is involved in the glycoside or biose linkage.

6. When the stable crystalline form of maltose, $C_{12}H_{22}O_{11} \cdot H_2O$, is dissolved in water, its initial specific rotation is +112.6° (or +118° for anhydrous maltose), and the rotation increases to a constant value of +129.5° (or +136° for the anhydrous form).[7] Crystalline maltose, therefore, exists in the β-form and, in accordance with the Hudson convention, the hydroxyl group is assigned to the left in the Fischer projection formula (p. 533).

[6] W. Charlton, W. N. Haworth, and S. Peat, *J. Chem. Soc.*, 89 (1926).

[7] C. S. Hudson and E. Yanovsky, *J. Am. Chem. Soc.*, **39**, 1035 (1917).

These reactions establish that the stable crystalline form of maltose is 4-(α-D-glucopyranosyl)-β-D-glucose. The sugar may also be named β-D-glucose-4-(α-D-glucopyranoside).

Similar methods have been employed in determining the structure of other common disaccharides. In some compounds, such as sucrose, it has been necessary to identify the hydrolytic products of the methylated disaccharides by degradative oxidation in a manner similar to that employed in establishing the structures of glucopyranose and glucofuranose; the methoxydicarboxylic acids which result are identified as the amides or other derivatives, and the nature of these acids establishes the biose linkage and the ring structures.

Zemplén [8] determined the position of the biose linkage of reducing oligosaccharides in a fundamentally different manner from that involved in methylation, *i.e.*, by stepwise degradation by means of a modification of Wohl's degradation method. The carbon chain of the sugar residue containing the reducing group is shortened by the successive removal of carbon atoms until a disaccharide results which will form a hydrazone, but not an osazone. Thus, maltose yields first arabinose-glucoside and secondly erythrose-glucoside. The latter will not form an osazone showing that carbon-2 in the erythrose part, which corresponds to carbon-4 in the original glucose unit, is blocked. Hence, the biose linkage in maltose is carbon-4.

Studies of the rates of lactonization of substituted acids and lactones arising from the disaccharides are also useful. In the lactonization of sugar acids which have free hydroxyl groups on carbon-4 and carbon-5, two simultaneous reactions occur: (1) the very rapid formation of an unstable δ-, or 1,5-lactone, and (2) the slow formation of a stable γ-, or 1,4-lactone. In the reverse hydrolytic process the δ-lactones are rapidly and extensively hydrolyzed whereas the γ-lactones undergo only slight hydrolysis. In the substituted acids arising from the disaccharides, either one or the other type of lactone formation may be precluded or both may take place unhindered, depending on the position of the substituents. Studies of the rate and extent of lactone formation and hydrolysis, therefore, provide a clue to their structure.

Difficulties arise in establishing the stereochemical nature of the glycosidic linkages in non-reducing disaccharides, especially where one of the monosaccharides is joined with the α- and the other with the β-linkage; in such a case the use of enzymes fails to indicate the stereochemical configuration of the component sugar units.

The structure of the trisaccharides has been studied by the methylation technics already discussed. As will be shown later, the nature of the

[8] G. Zemplén, *Ber.*, **59**, 1254 (1926).

partial hydrolytic products resulting from the use of specific enzymes and dilute acids also aids in establishing their structures.

General Properties. The oligosaccharides resemble the simple sugars in being crystalline, soluble in water and aqueous ethanol, and possessing a sweet taste. The majority of them crystallize as hydrates, sucrose being an exception.

Since they are glycosidic condensation products, they are hydrolyzed by mineral acids (and more slowly by organic acids) into the component simple sugars. The hydrogen-ion activity is the important factor, and the rate of hydrolysis of sucrose by acids, which can readily be followed polarimetrically, was one of the early methods employed to measure the relative strength of acids. The acid hydrolysis of sucrose is a classic example of a monomolecular reaction; in equal times the same fraction of the unchanged sucrose present at the commencement of the period is hydrolyzed. The ease of hydrolysis of different oligosaccharides varies widely. Sucrose is rapidly hydrolyzed at 20°C. by normal sulfuric acid, but lactose needs prolonged heating at 80°C. to accomplish the same degree of hydrolysis. Armstrong states that the relative rates of hydrolysis are: sucrose 1,240, maltose 1.27, lactose 1.00. Each of the naturally occurring disaccharides is hydrolyzed by its own specific enzyme, *e.g.*, sucrase, maltase, lactase.

DISACCHARIDES

Reducing Disaccharides. Although reducing disaccharides are known which are condensation products of pentoses or methylpentoses with a hexose (see p. 521), the common ones are dihexoses ($C_{12}H_{22}O_{11}$). Glucose is one of the constituent sugars present in all of them, and, with the exception of lactose, all are partial hydrolytic products of more complex carbohydrates. The reducing dihexoses are conveniently classified according to the carbon atom (carbon-3, carbon-4, or carbon-6) of the reducing component which is involved in the biose linkage.

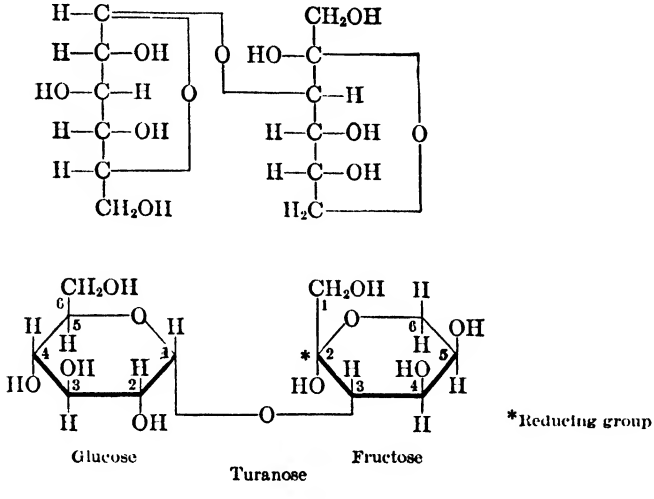
C-3 Dihexoses. In the C-3 dihexoses the biose ring is attached to carbon-3 of the sugar serving as the alcohol.

Turanose[9] is 3-(α-D-glucopyranosyl)-β-D-fructopyranose, or β-D-fructose-3-(α-D-glucopyranoside). It was discovered by Alekhine in 1889 as a partial hydrolysis product of the trisaccharide melezitose, but the preparation of the sugar in crystalline form was not reported until 1929.[10] It is hydrolyzable by acids with difficulty. Until recently the

[9] For a detailed discussion of the occurrence, structure, and properties of turanose and melezitose, see C. S. Hudson, *Advances in Carbohydrate Chem.*, **2**, 1 (1946)

[10] C. S. Hudson and E. Pacsu, *Science*, **69**, 278 (1929).

linkage in turanose was assigned to carbon-6 of the fructose residue, but several independent lines of investigation led to the above structure.[11-14]



This sugar is unique in that fructose is the reducing sugar component.

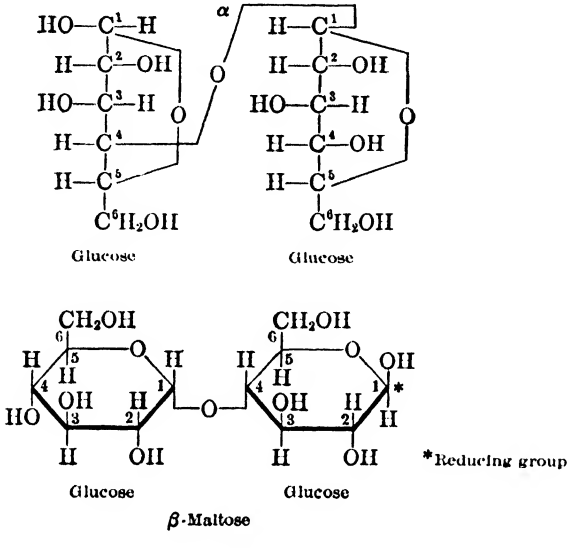
Other C-3 disaccharides have been obtained by the partial hydrolysis of polysaccharides. The first established occurrence of a 1,3-biose link in a disaccharide prepared from a natural source was 3-(D-galactopyranosyl)-L-arabinose which was obtained by the acid degradation of gum arabic;[15] 3-(D-galactopyranosyl)-D-galactose was later found as one of the hydrolytic products.[16]

C-4 Dihexoses. These dihexoses have the biose ring attached to carbon-4 of the sugar serving as the alcohol.

Maltose, 4-(α-D-glucopyranosyl)-D-glucose or D-glucose-4-(α-D-glucopyranoside), was first isolated by DeSaussure in 1819 from the products of hydrolysis of starch. Aside from traces of glucose, it is the end product of the action of the amylases on starch, dextrin, and glycogen. It is, therefore, an important constituent of germinating cereals and malt and is formed by salivary and pancreatic amylase from the above polysaccharides during digestion. Maltose is an intermediate in the acid hydrolysis of starch and is hydrolyzed by acids and by the enzyme maltase to two molecules of D-glucose.

[11] H. S. Isbell and W. W. Pigman, *J. Research Natl. Bur. Standards*, **20**, 787 (1938).
[12] W. Pacsu, E. J. Wilson, Jr., and L. Graf, *J. Am. Chem. Soc.*, **61**, 2675 (1939).
[13] E. Pacsu, *J. Am. Chem. Soc.*, **62**, 2568 (1940).
[14] H. S. Isbell, *J. Research Natl. Bur. Standards*, **26**, 35 (1941).
[15] F. Smith, *J. Chem. Soc.*, 744 (1939).
[16] J. Jackson and F. Smith, *J. Chem. Soc.*, 79 (1940).

In the stable crystalline state it exists as β-maltose monohydrate; its water of crystallization is incompletely lost at 100°C. under atmospheric pressure. As it contains one potential aldehyde group, it is a reducing sugar, it undergoes mutarotation, and it forms a characteristic osazone.



The specific rotation $[\alpha]_D^{20}$ in water of the equilibrium mixture of the α- and the β-forms is +136° (anhydrous maltose); the calculated rotations for the α-form are +168°, and for the β-form +118°.

In attempting to synthesize maltose by treating glucose with concentrated hydrochloric acid at 10° to 15°C., Emil Fischer [17] isolated an unfermentable dextrin-like substance of low reducing power which he named "isomaltose." This name has also been applied to an unfermentable product found in "hydrol," the mother liquor from the crystallization of dextrose in the acid hydrolysis of starch. The isomaltose is believed to result from a synthesis brought about by a reversal or "reversion" of the hydrolytic process. The evidence of several workers [18-23] indicates that, although the unfermentable product obtained from "hydrol" or by the action of hydrochloric acid on glucose contains several

[17] E. Fischer, *Ber.*, **28**, 3024 (1895).

[18] H. Berlin, *J. Am. Chem. Soc.*, **48**, 1107, 2627 (1926).

[19] G. Zemplén and Z. Bruckner, *Ber.*, **64**, 1852 (1931).

[20] A. Georg, *Thèse de Doctorat*, Genèvre, 1926.

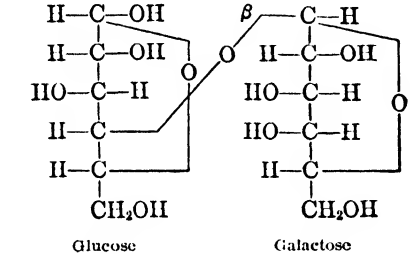
[21] G. H. Coleman, M. A. Buchanan, and P. T. Paul, *J. Am. Chem. Soc.*, **57**, 1119 (1935).

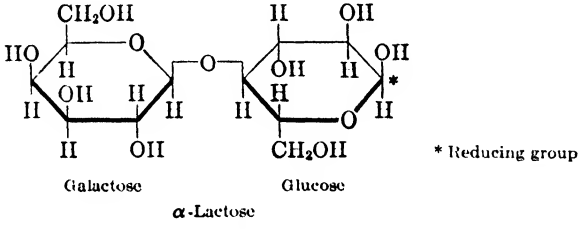
[22] K. Myrbäck, *Svensk Kem. Tid.*, **53**, 264 (1941).

[23] K. Ahlborg and K. Myrbäck, *Biochem. Z.*, **308**, 187 (1941).

carbohydrates (including gentiobiose), the principal constituent is 6-(α-D-glucopyranosyl)-D-glucose. This sugar differs from gentiobiose only in the stereochemical nature of the glucosidic linkage and would be more appropriately called isogentiobiose than isomaltose.

Lactose or milk sugar,[24] 4-(β-D-galactopyranosyl)-D-glucose or D-glucose-4-(β-D-galactopyranoside), was first isolated by Bartolettus in 1633. It occurs in the milk of all mammals in concentrations ranging





from about 2.0 to 8.5 per cent. Cow's milk normally contains about 4.8 per cent and human milk 6.0 per cent. This sugar is formed in the mammary gland from D-glucose supplied by the blood. There is no definite proof that any other organ of the body can transform glucose into galactose. It, therefore, appears that nature has provided for an adequate supply of galactose (in the form of lactose) for the synthesis of the galactosides of the brain and nerve tissues of the rapidly growing young mammal.

The ordinary lactose of commerce is α-lactose monohydrate which melts at 201.6°C. and has a specific rotation of +89.4° (for the anhydrous sugar). When dissolved in water it undergoes mutarotation; the equilibrium solution contains 37.8 per cent of the α- and 62.2 per cent of the β-form to give a constant specific rotation of +55.5°. If lactose is crystallized from solution above 93.5°C., anhydrous β-lactose, with a specific

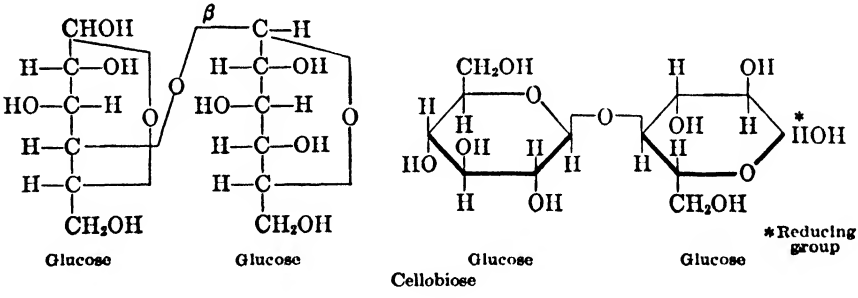
[24] For proofs of the structure of lactose see W. N. Haworth and G. C. Leitch, *J. Chem. Soc.*, **113**, 118 (1918); W. N. Haworth and C. W. Long, *J. Chem. Soc.*, 544 (1927). An excellent review of the properties, commercial preparation, and utilization of lactose is given by E. O. Whittier, *J. Dairy Sci.*, **27**, 505 (1944).

rotation of +35.0° and a melting point of 252.2°C., is produced. When an equilibrated solution is very rapidly dried by means of a drum or spray drier, the resulting solid is an amorphous glass containing the α- and β-forms in approximately the equilibrium ratio present in solution. By drying more slowly above 93.5°C., the ratio of the β- to the α-form is increased, and under appropriate conditions practically pure β-lactose can be obtained.

α-Lactose monohydrate is less soluble in water and less sweet than the anhydrous β-form. The solubility of lactose (equilibrium mixture of α- and β-forms) at 0°C. is only 12 g. per 100 g. of water as compared with 179 g. per 100 g. of water for sucrose. Its sweetness, in comparison with sucrose as 100, is 16.

Lactose forms a characteristic osazone and is hydrolyzed by acids and the enzyme lactase, which is a β-galactosidase, to glucose and galactose. It is not fermented to ethyl alcohol by common yeast, *Saccharomyces cerevisiae*, but it is readily converted to lactic acid by many microorganisms.

Cellobiose,[25] 4-(β-D-glucopyranosyl)-D-glucose or D-glucose-4-(β-D-glucopyranoside), is obtained by the partial hydrolysis of cellulose by acids



or by the enzyme cellulase, and is related to maltose in the same way as β-methylglucoside is related to α-methylglucoside. It is most readily obtained from cellulose in the form of derivatives, such as by acetolysis with acetic anhydride and concentrated sulfuric acid to give the octa-acetate[26,27] or by treatment with acetylbromide and glacial acetic acid to give acetobromocellobiose.[28]

Cellobiose is quite soluble in water and crystallizes readily. Its specific rotation is +34.6°. For many years cellobiose was the only crystalline

[25] For proofs of the structure of cellobiose see W. N. Haworth, C. W. Long, and J. H. G. Plant, *J. Chem. Soc.*, 2809 (1927).

[26] Z. H. Skraup and J. König, *Ber.*, **34**, 1115 (1901).

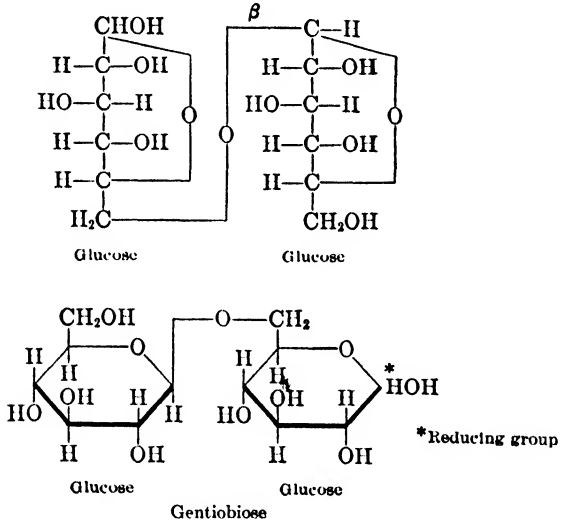
[27] F. C. Peterson and C. C. Spencer, *J. Am. Chem. Soc.*, **49**, 2822 (1927).

[28] P. Karrer and R. Widmer, *Helv. Chim. Acta*, **4**, 700 (1921).

product besides D-glucose which was isolated in degrading cellulose, and it has played a prominent role in studies on the structure of that polysaccharide. It is hydrolyzed by boiling with acids or by emulsin, a β-glucosidase, to give two molecules of D-glucose. Yeast does not ferment it, although a cellobiase is widely distributed in plants and animals. The sugar is utilized by *Aerobacter aerogenes* but is not attacked by *Escherichia coli*, and it is, therefore, useful for differentiating between these two types of organisms.

C-6 Dihexoses. These dihexoses have the biose ring attached to carbon-6 of the sugar serving as the alcohol.

Gentiobiose,[29] 6-(β-D-glucopyranosyl)-D-glucose or D-glucose-6-(β-D-glucopyranoside), is obtained by treating gentianose (see p. 620) with



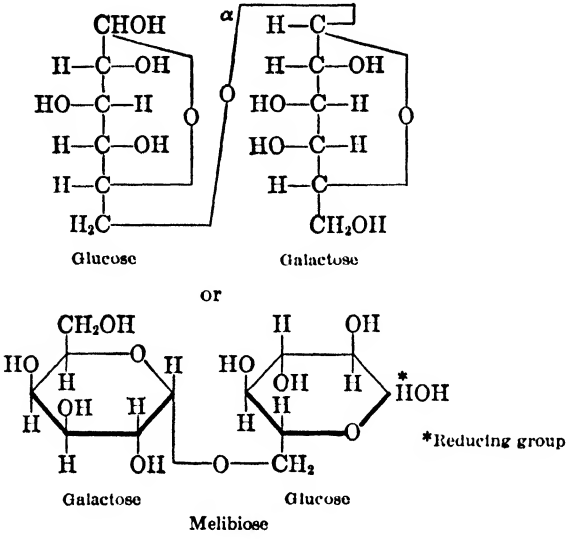
weak acids,[30] but it is more readily prepared by the synthetic action of emulsin on a concentrated solution of glucose.[31] It is the sugar component of the glycoside amygdalin, found in bitter almonds, and of the saffron pigment α-crocin. Small quantities of gentiobiose are obtained upon the hydrolysis of starch and by the action of concentrated hydrochloric acid on D-glucose. It melts at 190–193°C., has a specific rotation of +9.8°, and yields two molecules of D-glucose upon hydrolysis. It is not fermented by yeast.

[29] For proofs of the structure of gentiobiose see W. N. Haworth and B. Wylam, *J. Chem. Soc.*, **123**, 3120 (1923); B. Helferich, *Z. angew. Chem.*, **41**, 871 (1928).

[30] Em. Bourguelot and H. Hérissey, *Compt. rend.*, **132**, 571 (1901).

[31] Em. Bourguelot, H. Hérissey, and J. Coirre, *Compt. rend.*, **157**, 732 (1913).

Melibiose,[32] 6-(α-D-galactopyranosyl)-D-glucose or D-glucose-6-(α-D-galactopyranoside), is obtained along with D-fructose when the trisaccharide raffinose is selectively hydrolyzed with invertase.[33] Bottom



fermentation yeasts, but not top fermentation yeasts, contain an enzyme which hydrolyzes the sugar to D-glucose and D-galactose and are, therefore, capable of fermenting it.

Non-Reducing Dihexoses. The two important non-reducing dihexoses are sucrose and trehalose. The former occurs only in plants that contain chlorophyll and hence are capable of photosynthesis, whereas the latter takes its place in the yeasts and fungi. In these disaccharides the two sugar units are linked through their glycosidic hydroxyl groups; therefore they do not contain a functional carbonyl group and do not reduce Fehling's solution, form osazones, or undergo mutarotation in solution.

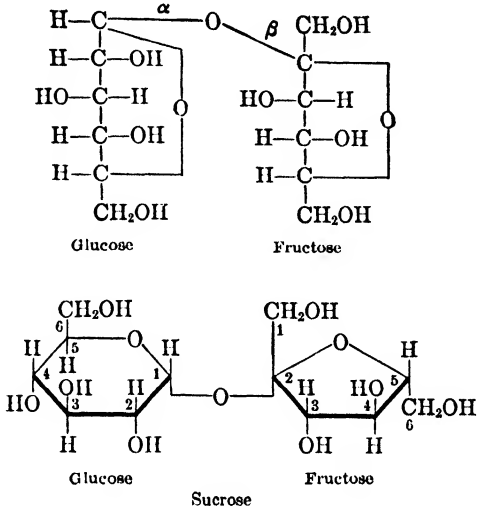
Sucrose is α-D-glucopyranosyl-β-D-fructofuranoside. This structure for sucrose was established by Haworth and his associates after a long series of investigations. The methylation of this sugar tends to stop at the heptamethyl stage, but in 1915 Haworth [34] succeeded in forming the completely methylated octamethylsucrose; upon mild hydrolysis and

[32] For proof of the structure of melibiose see W. N. Haworth, J. V. Loach, and C. W. Long, *J. Chem. Soc.*, 3146 (1927).

[33] M. Berthelot, *Ann. chim. phys.*, [3] **46**, 66 (1856).

[34] W. N. Haworth, *J. Chem. Soc.*, **107**, 12 (1915).

separation of the products, 2,3,4,6-tetramethylglucose and a tetramethylfructose were obtained.[35] The latter was dextrorotatory, whereas tetramethylfructopyranose from D-fructose [36] was strongly levorotatory.



This showed that the fructose residue in sucrose had a different ring structure from that of the free sugar, but it was not until several years later that the furanose structure of the fructose residue was established by the methylation and oxidation studies of Haworth and his associates.[37, 38] The structure they assigned has recently been confirmed [39] by successive oxidation with periodic acid and bromine water (in the presence of strontium carbonate) followed by acid hydrolysis and identification of the resulting acids, as described in Chapter 23 (p. 552). When sucrose is hydrolyzed, the D-fructofuranose rapidly undergoes a ring shift to the pyranose form. Sucrose comprised the first example of a γ-fructose residue in a natural product. Subsequently, it was found that in oligosaccharides (with the exception of turanose) and polysaccharides, it is present as a furanose ring structure. Although the stereochemical nature of the two glycosidic linkages is not definitely known, the evidence from enzymic and mutarotation studies points to

[35] W. N. Haworth and J. Law, *J. Chem. Soc.*, **109**, 1314 (1916).
[36] T. Purdie and D. McL. Paul, *J. Chem. Soc.*, **91**, 289 (1907).
[37] W. N. Haworth and E. L. Hirst, *J. Chem. Soc.*, 1858 (1926).
[38] J. Avery, W. N. Haworth, and E. L. Hirst, *J. Chem. Soc.*, 2308 (1927).
[39] P. Fleury and J. Courtois, *Compt. rend.*, **214**, 366 (1942).

an α-linkage for the glucose residue and a β-linkage for the fructose component.[40,41]

Many unsuccessful attempts have been made to achieve the chemical synthesis of sucrose. Pictet and Vogel [42] claimed that they had synthesized sucrose by coupling tetraacetylfructofuranose with tetraacetylglucopyranose in the presence of a dehydrating agent, but other investigators [43-45] have been unable to confirm their findings. One difficulty in effecting this synthesis chemically is that there are four possible stereochemical configurations for the joint glycosidic linkage of D-glucopyranose and D-fructopyranose. In 1944, however, Hassid, *et al.*,[46] accomplished the synthesis enzymatically by the action of a sucrose phosphorylase preparation obtained from the bacterium *Pseudomonas saccarophilia* on a mixture of the potassium salt of α-D-glucose-1-phosphate and D-fructose at *p*H 6.85. The enzyme preparation was freed from invertase, and the equilibrium reaction is considered to occur as a result of a dephosphorylytic condensation of the two monosaccharides:

$$\alpha\text{-D-glucose-1-phosphate} + \text{D-fructose} \xrightarrow[p\text{H } 6.85]{} \text{sucrose} + \text{phosphate}$$

In this enzymatic synthesis, the glucose-1-phosphate is supplied in the α-form, suggesting that the glucose residue in sucrose exists in the same form; α-glucose-1-phosphate also serves as a substrate for the enzymatic synthesis of starch, in which the glucose residues are joined by α-linkages. As fructose is supplied in the pyranose form, the ring change to the furanose form must be effected by the enzyme preparation. The mechanism of the ring change does not involve phosphorylation of carbon-6 since fructofuranose-6-phosphate cannot be substituted for fructopyranose in the synthesis.[47] This phosphorylase is highly specific with respect to the phosphorylated component of the substrate since no synthetic action is observed if glucose-1-phosphate is replaced by phosphoric esters of other aldoses or ketoses, including maltose-1-phosphate.[46-49] It is not so highly specific, however, in regard to the ketose

[40] C. S. Hudson, *J. Am. Chem. Soc.*, **31**, 655 (1909).

[41] E. F. Armstrong and K. F. Armstrong, *The Carbohydrates*, p. 181, Longmans, Green and Co., London, 1934.

[42] A. Pictet and H. Vogel, *Helv. Chim. Acta*, **11**, 436 (1928); *Ber.*, **62**, 1418 (1929).

[43] G. Zemplén and A. Gerecs, *Ber.*, **62**, 984 (1929).

[44] J. C. Irvine, J. W. H. Oldham, and A. F. Skinner, *J. Am. Chem. Soc.*, **51**, 1279 (1929).

[45] J. C. Irvine and J. W. H. Oldham, *J. Am. Chem. Soc.*, **51**, 3609 (1929).

[46] W. Z. Hassid, M. Doudoroff, and H. A. Barker, *J. Am. Chem. Soc.*, **66**, 1416 (1944).

[47] M. Doudoroff, *J. Biol. Chem.*, **151**, 351 (1943).

[48] M. Doudoroff, N. Kaplan, and W. Z. Hassid, *J. Biol. Chem.*, **148**, 67 (1943).

[49] H. A. Barker, W. Z. Hassid, and M. Doudoroff, *Science*, **100**, 51 (1944).

component since two analogs of sucrose have been prepared by the action of the bacterial phosphorylase on mixtures of glucose-1-phosphate with L-sorbose and D-ketoxylose, respectively.[50] The non-reducing disaccharide produced with L-sorbose has been shown to be α-D-glucopyranosyl-α-L-sorbofuranoside.[51]

Sucrose is very widely distributed in green plants, where it is frequently associated with D-glucose or D-fructose or both. It appears to function either as a temporary reserve (principally in the leaves) or as one of the forms in which carbohydrate is translocated. The important industrial sources are the sugar cane, sorghum, sugar beet, sugar maple, and the sugar palm, but the very pure crystalline sucrose of commerce is prepared from the sugar cane and the sugar beet. The sugar cane, a tropical plant native to India which contains about 18 per cent sucrose, was the earliest commercial source. As a result of a search for a plant source which could be grown in temperate climates under field cultivation, sucrose was discovered in the beet by the German professor Marggraf in 1747, and the first factory was established by Achard, one of his pupils, in 1802. The early development of the beet sugar industry was greatly stimulated as a result of subsidies granted by Napoleon, a step taken to counteract the British blockade of France which cut off imports of cane sugar during the Napoleonic wars. At that time the sugar beet contained only about 8 per cent sucrose, but as a result of plant breeding and improved cultural practices, strains of sugar beets have been developed which contain an average of 16 to 20 per cent sucrose, depending on soil and climatic conditions. For several years prior to World War II, the annual world production of sucrose varied from 20 to 23 million tons, of which 32 to 35 per cent was from sugar beets. During the present century, the per capita consumption of sucrose in the United States has increased from about 68 pounds to over 100 pounds per capita; the average American today obtains approximately one-seventh of his total calories in this form.

Unlike most sugars, sucrose crystallizes readily from water, and the crystals exist in the anhydrous form. It is very soluble in water; at room temperature it is soluble in about one-half its weight of water, and at 50°C. in one-fourth its weight of water. It melts at 160°C., and, when heated at or just above its melting point for some time, it soldifies on cooling to form a light-brown glassy mass, called barley sugar. When heated to 210°C., it undergoes partial decomposition to form *caramel*, an

[50] M. Doudoroff, W. Z. Hassid, and H. A. Barker, *Science*, **100**, 315 (1944).

[51] W. Z. Hassid, M. Doudoroff, H. A. Barker, and W. H. Doro, *J. Am. Chem. Soc.*, **67**, 1394 (1945).

amorphous light-brown substance which is widely used as a flavoring and coloring material.

Sucrose, like other sugars, forms compounds with bases which are called sucrates or saccharates; thus, with calcium hydroxide, mono-, di-, and tricalcium sucrates are formed. The tricalcium derivative, $C_{12}H_{22}O_{11} \cdot 3Ca(OH)_2$, is very insoluble in water, and its formation at temperatures of 10° to 15°C. provides a means of recovering sucrose from molasses; the sugar is liberated upon treatment of the saccharate with carbon dioxide. The soluble calcium saccharates are formed upon treatment of sucrose with calcium hydroxide at room temperature, and the alkaline solution, known as viscogen, causes cream to "whip" much more readily when added in small amounts.

Sucrose has a specific rotation in water of +66.5°. It is readily hydrolyzed by acids and by the specific enzyme *sucrase* (invertase) to yield equal amounts of the pyranose forms of D-glucose and D-fructose. As the levorotation due to D-fructose exceeds the dextrorotation due to D-glucose, the hydrolysis results in a change in sign. The process is, therefore, called *inversion*, and the equimolecular mixture of the two sugars is known as *invert sugar*.

$$\text{Sucrose} + \text{water} \rightarrow \text{D-glucose} + \text{D-fructose}$$

$$[\alpha]_D^{20} = +66.5° \qquad [\alpha]_D^{20} = +52.2° \quad [\alpha]_D^{20} = -93°$$

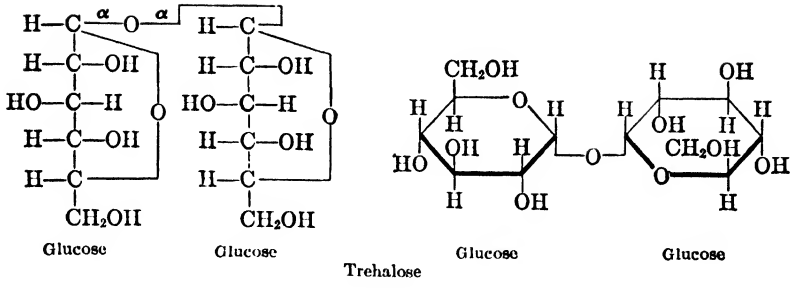
$$\text{Invert sugar } [\alpha]_D^{20} = -20.4°$$

On the basis of sucrose as 100, the sweetness of D-glucose is 74.3, and that of D-fructose is 173.3; thus the hydrolysis of sucrose results in an increase in relative sweetness from 100 to about 123 on a weight basis. During World War II, large quantities of sucrose were converted to invert sugar syrup as a means of securing maximum sweetness from the limited quantities of sucrose which were available to the food industries.

Invertase is widely distributed in the plant and animal kingdom; as ordinary yeasts contain the enzyme they are capable of fermenting sucrose.

Trehalose,[52] α-D-glucopyranosyl-α-D-glucopyranoside, is widely distributed in yeasts, molds, and fungi. As these lower plants do not contain chlorophyll, they do not form starch but store trehalose as a reserve food supply. It has a sweet taste and a high specific rotation (+197°). In contrast to sucrose, it is difficult to hydrolyze with acids. It is hy-

[52] For proofs of the structure of trehalose see H. H. Schlubach and K. Maurer, *Ber.*, **58**, 1178 (1925); H. Bredereck, *Ber.*, **63**, 959 (1930); C. S. Hudson, *J. Am. Chem. Soc.*, **38**, 1566 (1916); E. L. Jackson and C. S. Hudson, *J. Am. Chem. Soc.*, **61**, 1530 (1939).



drolyzed by the specific enzyme trehalase which occurs in certain species of yeasts, molds, and fungi (*e.g.*, *Aspergillus niger*).

TRISACCHARIDES

The trisaccharides, listed in Chapter 22 (p. 521), include both reducing and non-reducing types. Some occur in plants as the sugar component of certain plant glycosides from which they can be liberated by mild hydrolysis; others occur in the free state. The reducing types are found as constituents of plant glycosides; of these *rhamnose* and *robinose* are methylpentose-hexose saccharides, and *mannotriose* is a trihexose. The best-known trisaccharides are the *non-reducing trihexoses*, raffinose, melezitose, and gentianose, having the molecular formula $C_{18}H_{32}O_{16}$. They may be considered as two disaccharides combined in one molecule with the central sugar residue in common.

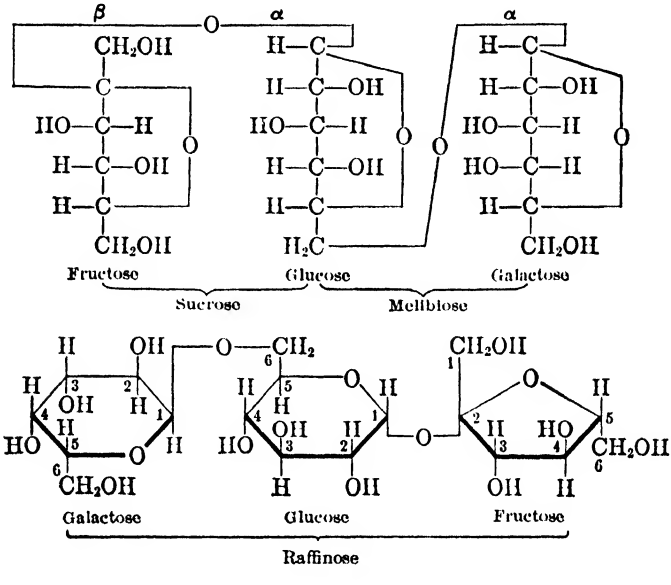
Raffinose (melitose, melitriose) is 6-(α-D-galactopyranosyl)-α-D-glucopyranosyl-β-D-fructofuranoside. It is the best-known trisaccharide. It occurs in small quantities in the sugar beet and, as it has the high specific rotation of +104.5°, a correction has to be made in the polarimetric determination of sucrose in its presence. Raffinose occurs in appreciable quantities in the cottonseed, soybean, and other oil seeds and in the so-called Australian manna. Cottonseed meal contains about 8 per cent raffinose and is the material from which this sugar is commonly prepared. It crystallizes as the pentahydrate.

In structure it comprises melibiose and sucrose with the central glucose residue in common.

Raffinose can be hydrolyzed by enzymes in two ways, and its behavior serves as a striking example of the specificity of enzyme action. Invertase,[53] a β-D-fructofuranosidase, hydrolyzes the sucrose part of the

[53] There are two enzymes, α-D-glucosidase and β-D-fructofuranosidase, which can hydrolyze sucrose. Some yeast extracts contain both enzymes, but in the purification of invertase the α-glucosidase is usually lost. See R. Kuhn, *Z. physiol. Chem.*, **29**, 57 (1923); also R. Kuhn and G. E. von Gundherr, *Ber.*, **59**, 1655 (1936).

molecule to give melibiose and D-fructose. (The sucrose residue is also hydrolyzed to these products by weak acids or a low hydrogen-ion activity and thus shows the same ease of hydrolysis that characterizes sucrose.) Emulsin, which contains an α-D-galactosidase (as well as a β-glucosidase), on the other hand, hydrolyzes the melibiose residue to



yield galactose and sucrose. Bottom yeasts contain both enzymes, and, hence, like strong mineral acids, they can hydrolyze raffinose completely. Raffinose is valueless as a food until it is hydrolyzed, and in man its partial utilization is apparently due to invertase action in the small intestine and bacterial action in the large intestine.[54]

The behavior of raffinose toward enzymes coupled with methylation and oxidation studies [55] has led to the structure previously depicted.

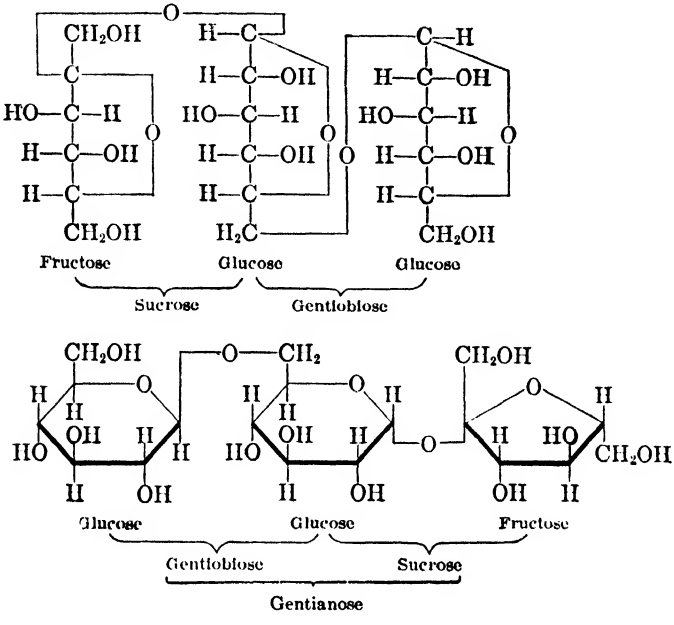
Gentianose,[56] 6-(β-D-glucopyranosyl)-α-D-glucopyranosyl-β-D-fructofuranoside, occurs in the Gentian root, *Gentiana lutea*. It may be regarded as composed of sucrose and gentiobiose with one glucose residue in common. Whereas raffinose is an α-galactosido-sucrose, gentianose is a β-glucosido-sucrose. Hydrolysis with emulsin, which contains a β-D-

[54] S. Kuriyama and L. B. Mendel, *J. Biol. Chem.*, **31**, 125 (1917).

[55] W. N. Haworth, E. L. Hirst, and D. A. Ruell, *J. Chem. Soc.*, **123**, 3125 (1923); W. Charlton, W. N. Haworth, and W. J. Hickinbottom, *J. Chem. Soc.*, 1527 (1927).

[56] Gentianose was discovered by A. Meyer, *Z. physiol. Chem.*, **6**, 135 (1882).

glucosidase, yields sucrose and D-glucose, whereas hydrolysis with invertase or dilute acids gives gentiobiose and D-fructose.



Melezitose,[57] 3-(α-D-glucopyranosyl)-?-D-fructofuranosyl-α-D-glucopyranoside, occurs in the exudations of plants belonging to such widely different genera as *Pinus*, *Larix*, *Pseudotsuga*, *Alhagi*, *Tilia*, *Populus*, *Laburnum*, and *Bambusa* (*e.g.*, larch, Douglas fir, scrub pine in America, and the Camel's thorn, which grows in the deserts of the Orient). These exudations are favored by dry conditions (perhaps, also, by the attack of sucking insects), and they harden into a "manna." When floral nectar is scarce, these exudations or "honeydew" are collected by bees, and, consequently, melezitose is very frequently found in "honeydew" honey. This sugar crystallizes very readily and in this form cannot be utilized by bees during the winter when they are unable to leave the hive to secure water; as a result the colonies die. Severe droughts in the Middle Atlantic States in 1917, 1928, 1943, and 1944 resulted in much "honeydew" honey which has been the principal source of melezitose used by scientists for studies of this sugar and turanose.

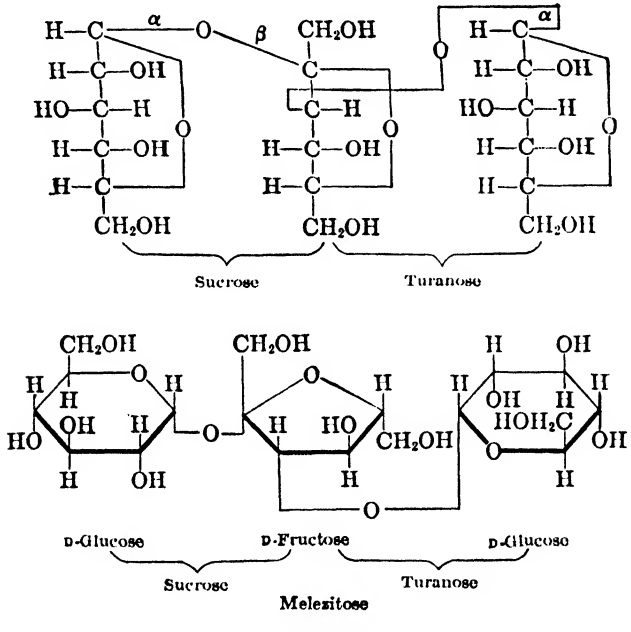
Apparently the honeybee can use melezitose to some extent, as both the adults and larvae live longer when fed solutions of pure melezitose

[57] A detailed discussion of the occurrence, properties, and structure of melezitose is given by C. S. Hudson, *Advances in Carbohydrate Chem.*, **2**, 1 (1946).

than when only water is available.[58] The efficiency of utilization of carbohydrates by the honeybee is sucrose > fructose > maltose > melezitose > glucose > trehalose > dextrin > galactose > lactose (the adults cannot utilize dextrin, galactose, or lactose).[59]

Melezitose exists both as a dihydrate and as a monohydrate, the latter being the stable form. The specific rotation $[\alpha]_D^{20}$ of the monohydrate is +88.5°, which corresponds to +91.7° for the anhydrous sugar.

The structure of melezitose has now been definitely established with the exception of the stereochemical nature of the fructofuranoside linkage.[57] Partial hydrolysis yields turanose and D-glucose, and, by analogy with raffinose and gentianose, it might be predicted that melezitose has the structure of turanose and sucrose with fructose common to both. The probable formula is



However, neither sucrose nor any other non-reducing disaccharide has ever been obtained upon the hydrolysis of melezitose with either acids or enzymes.

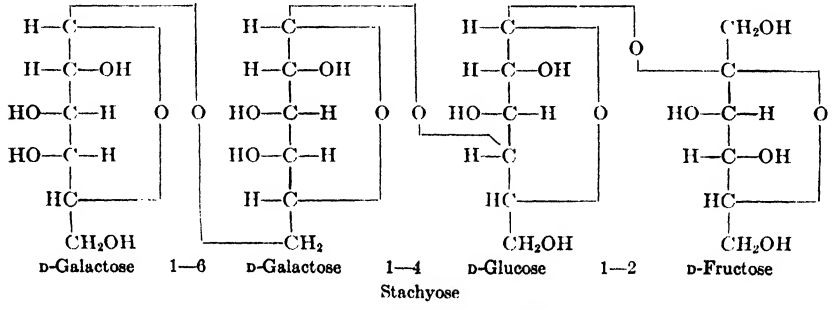
TETRASACCHARIDES

The only known tetrasaccharide is *stachyose*, a non-reducing tetrahexose ($C_{24}H_{42}O_{21}$) which was isolated in crystalline condition from the

[58] E. F. Phillips, *Z. Untersuch. Lebensm.*, **64**, 383 (1932).
[59] L. M. Bertholf, *J. Agr. Research*, **35**, 429 (1927).

tubers of the Japanese artichoke, *Stachys tuberifera.*[60] It is also found in the seeds of certain *Leguminoseae*, such as the soybean.

On complete hydrolysis it yields one molecule of D-fructose, one of D-glucose, and two of D-galactose. On mild hydrolysis the molecule of D-fructose is split off, yielding a trisaccharide. Methylation studies[61] indicate that it is a raffinosegalactopyranoside, but the stereochemical nature of the glycosidic linkages has not been fully established.



Stachyose

[60] A. von Planta and E. Schulze, *Ber.*, **23**, 1692 (1890).

[61] M. Onuki, *Sci. Papers Inst. Phys. Chem. Research, Tokyo*, **18**, 357 (1932); **20**, 201 (1933); also *J. Agr. Chem. Soc. Japan*, **9**, 90 (1933).

CHAPTER 25

The Polysaccharides

The polysaccharides comprise a large and abundant group of complex, naturally occurring carbohydrates of high molecular weight which serve either as reserve nutrients (such as starches, glycogen, and inulin) or as skeletal materials (such as cellulose and chitin) from which relatively rigid mechanical structures are built. Some polysaccharides, such as certain mannans and galactans, however, serve both functions.

The polysaccharides vary widely in physical and chemical behavior; some, such as glycogen and inulin, readily disperse in warm water to form colloidal solutions, whereas others, like cellulose, are extremely insoluble and have little or no biochemical reactivity. Upon complete hydrolysis with mineral acids some polysaccharides, such as starch and cellulose, yield a single monosaccharide, whereas others, like arabic acid, yield two or more monosaccharides as well as uronic acids; some, like the pectins, yield non-carbohydrate substances in addition to monosaccharides and uronic acids. The individual monosaccharide units are linked together in glycosidic fashion to form long chains which may possess either a linear structure, as in cellulose, or a branched structure, as in glycogen.

The hexose sugars, D-glucose, D-galactose, D-mannose, and their uronic acids occur in polysaccharides in the pyranose form, and D-fructose is found in the furanose form. In polysaccharides containing pentose residues, D-xylose occurs in the pyranose form, and L-arabinose exists as furanose units. Until relatively recently it was thought that cellulose, starch, and glycogen, the most extensively studied polysaccharides, were composed of D-glucopyranose units joined entirely by 1,4-linkages. In the last few years, however, it has been demonstrated that starch and glycogen contain some 1,6- as well as 1,4-linkages. Structural studies of plant gums and mucilages have demonstrated that 1,2-, 1,3-, 1,5-, and 1,6-linkages occur more commonly in these polysaccharides than the 1,4-type.

The complexity and diverse nature of the polysaccharides render them difficult to classify. The common system is based on the nature of the monosaccharides obtained upon complete hydrolysis (see p. 522). This method of classification is rather unsatisfactory because it places

carbohydrates of such diverse properties and functions as starch and cellulose in the same group. A more convenient classification based on their functions, relative ease of hydrolysis, and the nature of their hydrolytic products is:

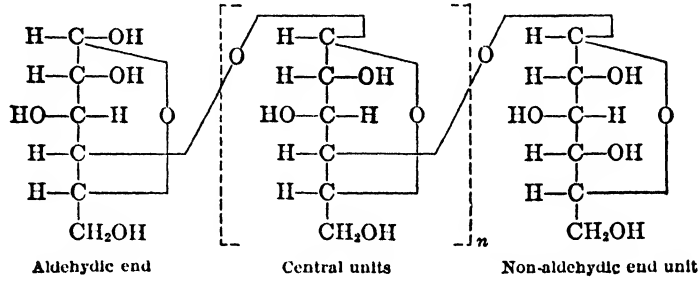
Group I. Nutrient polysaccharides
Starch, glycogen, inulin (and other polyfructosans), galactogen, lichenin
Group II. Pectins, plant gums and mucilages, hemicelluloses
Group III. Skeletal polysaccharides
Cellulose, chitin

Investigation of the structure of polysaccharides is exceedingly difficult because of their complexity, their lower reactivity as compared with the sugars, and the difficulty of isolating and purifying them without change. As a result, the structure of many of the polysaccharides has not been established, and even in the case of those about which a great deal is known, there still remains much to be done before their structures can be regarded as completely established. To establish the structure of a polysaccharide, it is necessary to determine the nature of the component units, the carbon atoms which are involved in the glycosidic linkage, the stereochemical nature of the glycosidic linkage, the average length of the chain and whether it is linear or branched. Complete hydrolysis by mineral acids and analysis of the products will reveal the nature of the constituent units. Partial hydrolysis by chemical or enzymatic methods may yield disaccharides, the structures of which aid in establishing the atoms involved in the biose linkage and the stereochemical nature of the linkage. Thus, on partial hydrolysis cellulose yields cellobiose, a β-glucoside in which the two glucose units are joined by a 1,4-linkage. Assuming that the cellobiose was not produced in the reaction products by synthesis from glucose, its presence in the hydrolytic products indicates that the cellulose molecule consists of "repeating units" of cellobiose. Complete methylation of a polysaccharide, whereby all free hydroxyl groups are converted into methoxyl groups, followed by hydrolysis will also reveal which carbon atoms are involved in the linkage by discovering where free hydroxyl groups are located in the cleavage fragments. The complete methylation of some polysaccharides is difficult to accomplish, and special techniques must be employed.

Haworth and his associates [1] introduced a method, known as the *end group assay,* for estimating the chain length of a polysaccharide. The

[1] W. N. Haworth and H. Machemer, *J. Chem. Soc.*, 2270 (1932).

principle of this method may be exemplified with a straight chain of D-glucopyranose units joined by α-1,4-linkages as in maltose:



Upon methylation and hydrolysis, the central glucose residues and the aldehydic end of the molecule will yield 2,3,6-trimethylglucose, and the non-aldehydic end unit will yield 2,3,4,6-tetramethylglucose. The relative proportions of these two derivatives will thus indicate the number of glucose units composing the linear chain. In applying this technic to starch (see p. 634) one obtains results corresponding to a chain length of 24 to 30 glucose units, or a molecular weight of about 5,000. However, determinations of the molecular weights of various starches and their methylated or acetylated derivatives by physical methods gives values ranging between 200,000 and 500,000. This discrepancy led the Birmingham school [2,3] to postulate that the chain length of starch revealed by the end group assay method represents the "chemical molecule" in which the glucose units are united by primary valences. These chemical molecules are presumably associated by means of secondary valences to form large aggregates or micelles of colloidal dimensions. However, as will be described later in the discussion of starch, subsequent methylation studies revealed that ordinary starches contain a fraction which is branched. The existence of numerous non-reducing terminal glucose residues at the ends of short branches would greatly increase the yield of tetramethylglucose and lead to a great underestimate of the chain length. The results of recent studies support the earlier view of Emil Fischer that the monosaccharide units in polysaccharides are linked only by primary valences, and the "association theory" has now been largely abandoned.

The extent to which polysaccharides are oxidized by periodic acid is useful in structural studies.[4] It will be recalled that this reagent selec-

[2] D. K. Baird, W. N. Haworth, and E. L. Hirst, *J. Chem. Soc.*, 1201 (1935).

[3] W. N. Haworth, *Chemistry & Industry*, **54**, 859 (1935).

[4] V. C. Barry, T. Dillon, and W. McGettrick, *J. Chem. Soc.*, 183 (1942); V. C. Barry, *ibid.*, 578 (1942).

tively oxidizes the hydroxyls on adjacent carbon atoms to —CHO groups. If the linkage between the units of the polysaccharides is 1,4-, the necessary pair of adjacent —CHO groups is provided by carbon-2 and carbon-3, and cleavage occurs. However, if the linkage is 1,3-, no glycol is available for reaction except in the terminal non-aldehydic group, and in this case the number of carboxyl groups formed by subsequent oxidation with bromine would appear to be a measure of the chain length of the polysaccharide.

The extent to which polysaccharides (and oligosaccharides) are converted into acetobromomonosaccharide derivatives upon treatment with a solution of hydrogen bromide in glacial acetic acid and acetyl bromide may be used to distinguish between 1,4- and 1,6-glycosidic linkages; the acetates of 1,4-glycosides are completely converted to the acetyl glycosyl halides, whereas those of 1,6-glycosides are not hydrolyzed.[5]

Recent studies have shown that the hydrolytic products of methylated polysaccharides may be quantitatively separated by chromatographic adsorption techniques which appear to provide a simpler method for small amounts of material than fractional distillation.[6-11]

THE STARCHES[12]

Starch constitutes the reserve carbohydrate of the plant, functioning in much the same capacity as fat in the animal organism. Next to cellulose and related hemicelluloses, it is the most abundant and widely distributed plant substance. In the great majority of cases, it occurs as minute birefringent granules in the seed, tuber, root, or pith of the plant. Occasionally, starch-like material is deposited in the plant in amorphous non-granular form, as in certain lichens and mosses and also in transitory state in green leaf tissue. Until more is known of these amorphous poly-

[5] Allene Jeanes and G. E. Hilbert, papers presented before the Division of Sugar Chemistry and Technology, American Chemical Society, New York, September, 1944.

[6] J. K. N. Jones, *J. Chem. Soc.*, 333 (1944).

[7] D. J. Bell, *J. Chem. Soc.*, 473 (1944).

[8] E. J. Norberg, I. Auerbach, and R. M. Hixon, *J. Am. Chem. Soc.*, **67**, 342 (1945).

[9] G. H. Coleman, D. E. Rees, R. L. Sundberg, and C. M. McCloskey, *J. Am. Chem. Soc.*, **67**, 381 (1945).

[10] W. H. McNeely, W. W. Binkley, and M. L. Wolfrom, *J. Am. Chem. Soc.*, **67**, 527 (1945).

[11] B. W. Lew, M. L. Wolfrom, and R. M. Goepp, Jr., *J. Am. Chem. Soc.*, **67**, 1865 (1945); **68**, 1449 (1946).

[12] Contributed by Thomas John Schoch, Chemist, Corn Products Refining Company, Argo, Illinois.

saccharides, it is perhaps advisable to exclude them from the starch category, on the basis that they represent primitive or incomplete phases of starch production. As a matter of definition, starch may therefore be described as the natural glucose polysaccharide occurring as microscopic birefringent granules and giving blue or red colorations with iodine.

Physical Structure of Starch. The granule may be considered as the botanical unit of starch, seemingly built up by deposition of starch substance about a central nucleus or hilum. Each variety of starch has its own characteristic size, shape, and markings,[13] by which the skilled microscopist can usually identify the source of the starch (Fig. 98). Thus potato starch consists of relatively large egg-shaped granules, 15 to 100 μ in size, with pronounced "oyster-shell" striations about an eccentrically placed hilum. These striations may result from periodicity in growth, since it has been reported that starch grown under constant light conditions shows no such markings.[14] Under the polarizing microscope, the potato starch granule exhibits a brilliant birefringence cross through the hilum. The granules of tapioca starch show a diversity of size and shape, some rounded, some cup-shaped or truncated, 5 to 35 μ in diameter with an eccentric hilum. Wheat starch is likewise mixed, containing large lenticular granules 15 to 35 μ in diameter and small rounded granules 2 to 8 μ in diameter. The hilum is centrally located, and striations are faint. Corn starch contains both rounded and angular or polygonal granules, from the floury interior and the horny exterior of the endosperm, respectively. The polygonal shape appears to result from internal pressures developed during drying of the corn, since the granules from immature "milk-stage" corn are all rounded. Certain less common starches possess more complex granule structure; *e.g.*, wrinkled-seeded pea starch has a deeply fissured compound granule, seemingly comprised of wedge-shaped segments.

The optical anisotropy of the starch granule indicates some sort of orderly orientation of the starch substance. Under pressure, the granules tend to fracture into radial segments through the hilum, and this behavior has led to the concept of the granule as a spherocrystal, composed of radially arranged needles or "trichites" of the starch substance. However, from x-ray diffraction patterns it is evident that the orientation within the granule extends down to the molecular level. Two major

[13] For a very complete description of the various starches with numerous photomicrographs, see E. T. Reichert, *The Differentiation and Specificity of Starches in Relation to Genera, Species, etc.*, Carnegie Institution of Washington, Pub. 173, Parts 1 and 2 (1913).

[14] H. L. van de Sande Bakhuyzen, *Proc. Soc. Exptl. Biol. Med.*, **23**, 302 (1926).

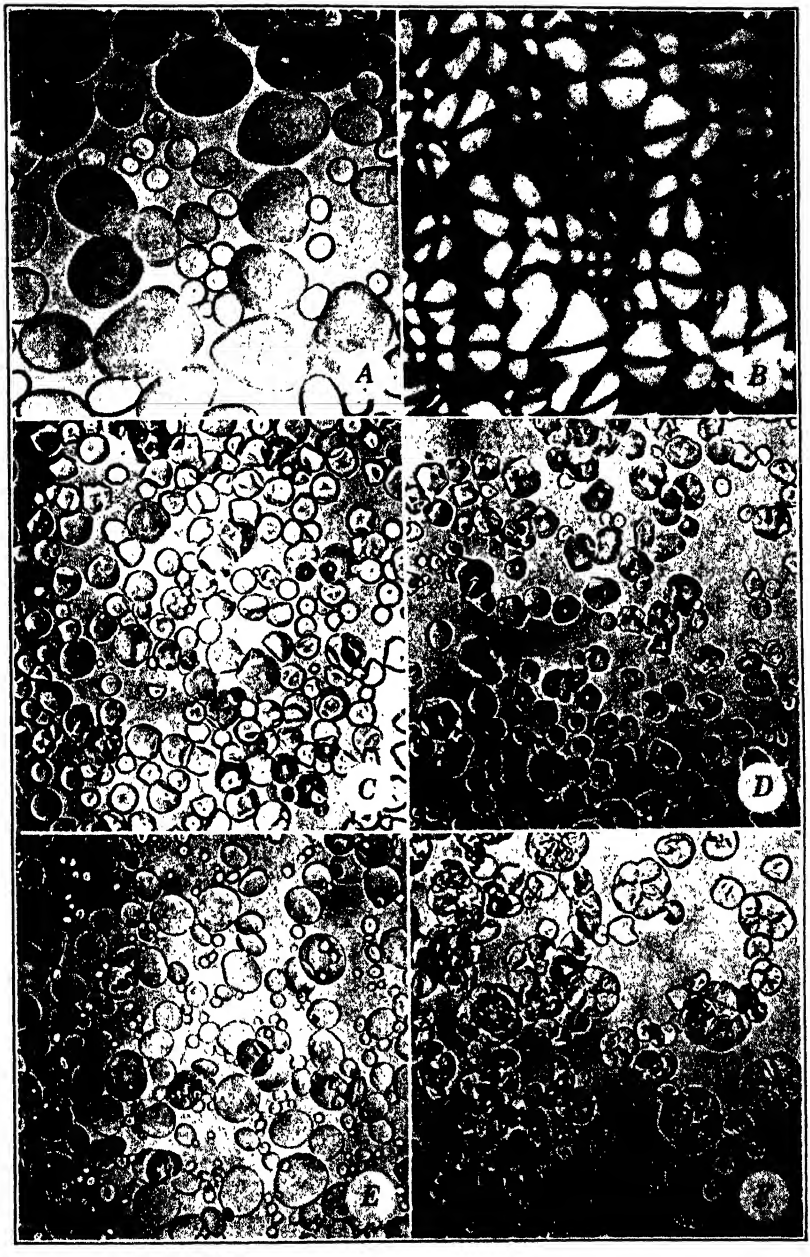


FIG. 98. Photomicrographs of (*A*) potato starch, (*B*) potato starch under polarized light, (*C*) tapioca starch, (*D*) corn starch, (*E*) wheat starch, (*F*) wrinkled-seeded pea starch. (Courtesy of Northern Regional Research Laboratory, Peoria, Ill.)

types of x-ray spectra are obtained with the granular starches, a so-called A-pattern with the cereal starches and a B-pattern with the tuber starches.[15] Although it has not yet been possible to translate these diffraction data into terms of molecular spacings, they necessarily represent an orderly molecular lattice.

In their granular form, the various starches are completely insoluble in cold water. When an aqueous starch suspension is heated to a certain point, the granules suddenly commence to swell, with simultaneous loss of optical birefringence. The temperature at which this transition occurs is usually termed the gelatinization point of the starch or more properly the gelatinization range, since the larger granules commence to swell at a lower temperature than the smaller ones. The various starches have different gelatinization ranges, the root and tuber starches generally swelling at lower temperature than the cereal starches. Thus potato starch loses its birefringence in the range of 59° to 67.5°C. and corn starch at 62° to 70°C., although these temperatures will be influenced by such factors as the source and pretreatment of the starch, the rate of heating, presence of electrolyte, and *p*H of the medium.

As the starch suspension is progressively heated above the gelatinization temperature, the granules continue to swell, and the mixture becomes viscous and more translucent, eventually attaining the thick consistency of a boiled starch paste. This transition from granular to pasted starch may be followed by a variety of methods: (1) the loss of birefringence and progressive swelling may be observed microscopically by use of an electrically heated stage,[16] (2) the increased translucency of the pasted starch may be determined photoelectrically,[17] or (3) the rise in viscosity may be charted as a function of temperature by use of a recording viscometer.[18] The so-called viscosity of a starch paste does not become apparent until the granules are swollen to such an extent that they touch and jostle one another (Fig. 99). Even though the granules of a boiled starch paste are swollen to such a degree that their outlines are no longer discernible under the microscope, the thick consistency of such a paste is primarily due to the presence of these swollen granule masses. Vigorous mechanical agitation or prolonged boiling breaks down these structural effects to give starch solutions of greatly reduced

[15] J. R. Katz and Th. B. van Itallie, *Z. physik. Chem.*, **A150**, 90 (1930).

[16] C. K. Francis and O. C. Smith, *Ind. Eng. Chem.*, **8**, 509 (1916); A. W. Dox and G. W. Roark, Jr., *J. Am. Chem. Soc.*, **39**, 742 (1917).

[17] D. H. Cook and J. H. Axtmayer, *Ind. Eng. Chem., Anal. Ed.*, **9**, 226 (1937); W. L. Morgan, *ibid.*, **12**, 313 (1940).

[18] T. J. Schoch, *Cereal Chem.*, **18**, 121 (1941); C. A. Anker and W. F. Geddes, *ibid.*, **21**, 335 (1944); C. C. Kesler and W. G. Bechtel, *Analytical Chem.*, **19**, 16 (1947).

viscosity. Autoclaving provides the most effective means for dissolving the last remnants of granule structure, although extreme care must be exercised to avoid attendant hydrolytic degradation of the starch substance.

The starch granule gelatinizes at room temperature in dilute caustic alkali or in concentrated solutions of the lyotropic salts, particularly

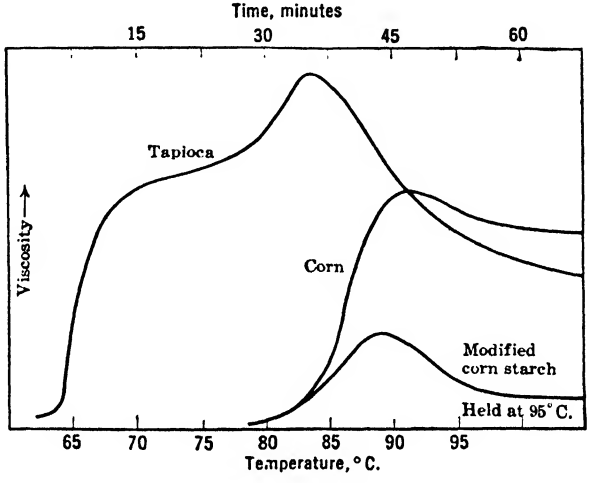


FIG. 99. Viscosity changes during the pasting of various starches, as measured with the Brabender Amylograph. The starch suspension (35 grams in 450 ml. of water) was heated at a constant rate to 95°C., then held at this temperature. The granules of corn starch lose their birefringence in the range of 65° to 70°C., but the viscosity does not commence to rise until 80°C., when the granules are swollen approximately four times their original diameter. The double "hump" is characteristic of most tapioca starches. The modified corn starch is a commercial acid-treated product, employed as a textile size.

ammonium thiocyanate or sodium salicylate. Here again, the pasty or gel-like consistency of such systems probably reflects structural viscosity effects persisting from the original granule organization. In contrast, sulfate markedly impedes gelatinization, and a starch suspension in molar magnesium sulfate may be heated to 100°C. without loss of birefringence of the granules.

When starch pastes or autoclaved solutions are allowed to stand, the starch substance tends to become progressively less soluble. This reversion to insoluble state, or "retrogradation," is reflected in a variety of changes; the starch paste or solution becomes more opaque and viscous, often setting to a firm gel. At dilute concentrations, the starch substance separates as an insoluble precipitate which cannot be redissolved by heating, even at autoclave temperatures. Retrogradation

is accelerated by low temperatures. Thus, if a starch paste is frozen and thawed, the starch is completely transformed to a pulpy sponge from which the water may be removed by squeezing or pressing. Retrograded starch, whether of tuber or cereal origin, gives a B-type x-ray pattern, and the process of insolubilization therefore seems to represent some sort of orderly crystallization of the starch substance.

All starches contain trace amounts of non-carbohydrate substances. Potato starch contains 0.06 to 0.07 per cent phosphorus,[19] apparently in chemical combination with the carbohydrate since glucose-6-phosphate has been isolated from its hydrolysis products.[20] Fatty acid occurs in the common cereal starches to the extent of 0.5 to 0.7 per cent. Formerly, this was thought to be esterified with the starch, since the associated lipids could not be removed by extraction with such solvents as petroleum ether or carbon tetrachloride.[21] More recently, it has been found that these fatty acids can be completely removed by extraction with such hydrophilic solvents as methanol;[22] conversely, fat-free starch can be impregnated with fatty acid by use of a methanol medium, and this added fat cannot be removed by petroleum ether extraction. The affinity of starch for fatty acid is therefore explained as an adsorption effect.[23] Specific differences in the pasting characteristics of various starches can in part be explained by the influence of these associated lipids. Thus potato starch gives clear slimy pastes, whereas corn starch pastes are more opaque and of a "short" consistency. The addition of an equivalent amount of fatty acid to potato starch gives an opaque paste without any slimy characteristics. Similarly, defatted corn starch possesses much greater clarity. Removal of fatty acid lowers the gelatinization temperature by some 3° to 4°C. Certain starches likewise contain traces of silica and nitrogenous material. Although it seems unlikely that these substances are chemically combined with the starch, no means have yet been found to effect their removal.

Chemical Structure of Starch. Under ideal conditions, acid hydrolysis of starch gives a theoretical yield of glucose which must, therefore, constitute the sole building block of the starch molecule. The enzyme β-amylase degrades starch to give substantial yields of maltose, indicating the probability of α-1,4-linkages between a majority of the

[19] M. Samec, in *A Comprehensive Survey of Starch Chemistry*, edited by R. P. Walton, pp. 51–61, Chemical Catalog Co., New York, 1928.

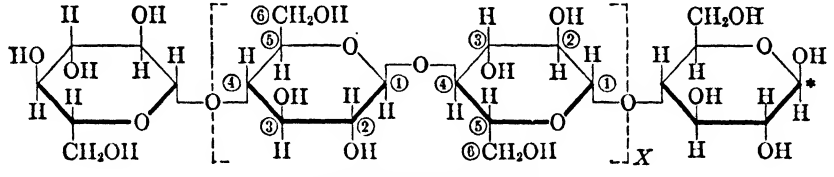
[20] T. Posternak, *Helv. Chim. Acta*, **18**, 1351 (1935).

[21] T. C. Taylor and H. A. Iddles, *Ind. Eng. Chem.*, **18**, 713 (1926).

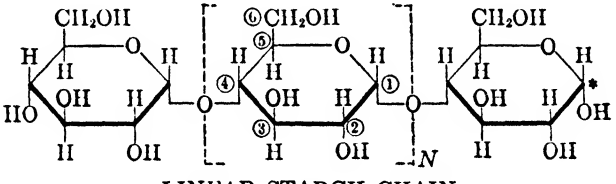
[22] J. W. Evans and D. R. Briggs, *Cereal Chem.*, **18**, 443, 465 (1941); J. W. Evans, *ibid.*, **18**, 468 (1941); T. J. Schoch, *J. Am. Chem. Soc.*, **60**, 2824 (1938); **64**, 2954 (1942).

[23] L. Lehrman, *J. Am. Chem. Soc.*, **64**, 2144 (1942).

glucose units. However, enzymes frequently cause rearrangement or synthesis. The presence of the α-1,4-glucosidic bond in starch has therefore been confirmed by chemical means, since acetolysis of starch with acetyl bromide yields approximately 20 per cent of acetobromomaltose.[24] By repeated treatments with dimethyl sulfate and alkali, starch is converted into the trimethyl derivative, and subsequent hydrolysis gives 2,3,6-trimethylglucose as the main product. It therefore appears that the main unit of starch is the stable glucopyranose ring, with free hydroxyls on carbons-2, -3, and -6 (Fig. 100). Such a structure would differ from



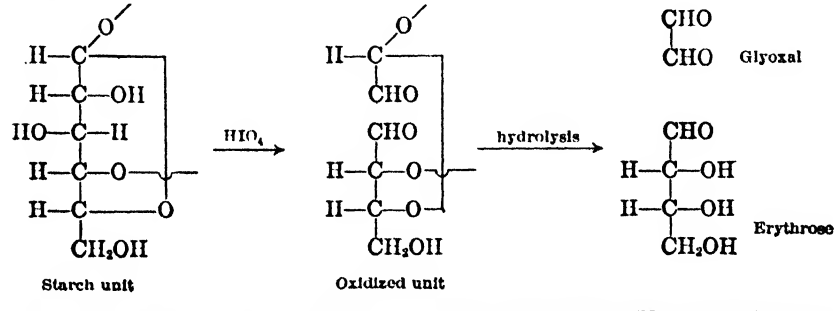
CELLULOSE CHAIN



LINEAR STARCH CHAIN

FIG. 100. Comparative structures of cellulose and linear starch chain. Conventional numbering of carbons indicated by circled numerals. Aldehydic terminus indicated by asterisk.

cellulose because the 1,4-glucosidic linkages are in α-configuration rather than β. Further evidence of this basic unit is provided by periodic acid oxidation of starch [25] followed by hydrolysis to give glyoxal and erythrose:



[24] P. Karrer and C. Nageli, *Helv. Chim. Acta*, **4**, 263 (1921); K. Freudenberg and K. Soff, *Ber.*, **B69**, 1252 (1936).

[25] E. L. Jackson and C. S. Hudson, *J. Am. Chem. Soc.*, **59**, 2049 (1937).

Periodic acid oxidizes the glycol grouping with consequent rupture of the carbon-to-carbon bond. If the basic glucose unit contained hydroxyl groups on three adjacent carbons, the center carbon would be split out as formic acid, contrary to the observed reaction.

However, methylation studies have likewise provided definite evidence that starch does not consist entirely of simple linear chains as depicted in Fig. 100. Although 2,3,6-trimethylglucose is the primary product, some 4 to 5 per cent of 2,3,4,6-tetramethylglucopyranose is likewise formed.[26] This is necessarily derived from the glucose unit at the non-aldehydic terminus of the chain, since this glucose has a free hydroxyl group on carbon-4. For almost a decade, this end group assay for tetramethylglucose was accepted as evidence that the starch molecule consisted of a linear chain of some 25 to 30 α-glucopyranose units. However, other evidence was not in accord with this viewpoint. Such a chain would possess an opposite aldehydic terminus which should exhibit a substantial reducing power, yet the reducing value of starch toward alkaline copper solutions [27] or toward ferricyanide is vanishingly small, corresponding to a molecular magnitude of several thousand glucose units. In addition, physical measurements of osmotic pressure and viscosity are not compatible with a molecular size as small as 25 to 30 glucose units. Also, in addition to tetramethylglucose, an approximately equivalent amount of the 2,3-dimethylglucose can be isolated from the hydrolysate of fully methylated starch. This indicates the possibility that a small proportion of the glucose units are involved in some sort of anomalous linkage which blocks the hydroxyl group on carbon-6. Finally, the action of β-amylase on starch is not in accord with a linear chain. This enzyme presumably attacks a linear starch chain from the non-aldehydic terminus, effecting the scission of two glucose units at a time in the form of maltose. Hence the yield of the latter should approach 100 per cent, but actually only some 60 per cent of maltose is so derived from starch, the balance of the carbohydrate remaining as a high molecular weight polysaccharide which resists further action by the enzyme (the so-called limit dextrin). Here again, there appears to be an anomalous linkage blocking amylase action. These conflicting data have finally been resolved by the proposal of a branched structure [28] for starch (Fig. 101), a very large tree-like molecule whose individual linear

[26] E. L. Hirst, M. M. T. Plant, and M. D. Wilkinson, *J. Chem. Soc.*, 2375 (1932).

[27] W. A. Richardson, R. S. Higginbotham, and F. D. Farrow, *J. Textile Inst.*, **27**, T131 (1936).

[28] H. Staudinger and E. Husemann, *Ann.*, **527**, 195 (1937); *Ber.*, **B71**, 1057 (1938).

branches are 25 to 30 glucose units in length. The branching is presumably effected through an α-1,6-linkage (isogentiobiose), though this point has not been experimentally confirmed.

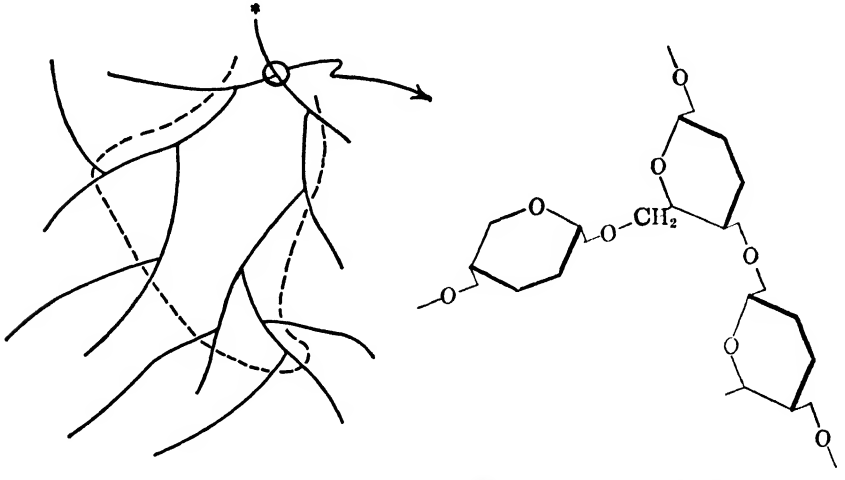


FIG. 101. Branched fraction of starch, showing the presumed α-1,6-linkage. The molecule may contain 50 to 70 branches, each 25 to 30 glucose units in length. Dotted line indicates the boundary of β-amylase action. Aldehydic terminus designated by asterisk.

Starch Fractionation. It has long been thought that starch contains two or more individual constituents, and a great amount of study has been devoted to their separation. Six general methods of fractionation have been developed, all of which accomplish at least a partial separation. However, there are three principal criticisms of most of these methods: (1) the low degree of separation achieved and consequent impurity of the resulting products, (2) the use of rigorous chemical treatment resulting in degradation of the starch substance, and (3) attendant physical changes (*viz.*, retrogradation) altering the character of the products. In addition, much confusion has arisen in the nomenclature of the starch fractions, with a multiplicity of ill-defined terms and frequent transposition in their usage. Recent evidence [29] has almost certainly established the presence of two components in the normal starches, a long linear type of α-glucopyranose polymer (Fig. 100) and a highly branched molecular structure (Fig. 101). To clarify the present discussion, these will be identified merely as the linear and branched components, though this treatment may be somewhat oversimplified.

[29] K. H. Meyer, in *Advances in Colloid Science*, Vol. 1, edited by E. O. Kraemer, pp. 143–182, Interscience Publishers, New York, 1942.

The following tabulation summarizes the conflicts in nomenclature of the starch fractions.

Investigator	*Method*	*Linear Component*	*Branched Component*
Baldwin [30]	Leaching	β-Amylose	α-Amylose
Samec [19]	Electrophoresis (potato starch)	Amyloamylose Amylose	Erythroamylose Amylopectin
Taylor [21] and Alsberg [31]	Ultrafiltration or electrophoresis (cereal starches)	α-Amylose Amylopectin	β-Amylose Amylose
Meyer [29]	Leaching	Amylose	Amylopectin
Pacsu [32]	Selective adsorption	β-Amylose	α-Amylose

To avoid use of any of the above terms, Schoch [33] has designated the linear and branched components obtained by selective precipitation methods as the A-fraction and the B-fraction, respectively.

The various techniques of fractionation may be generalized as follows.

Selective Retrogradation. When a dilute starch solution is allowed to stand for an extended period of time, the linear component gradually separates as an insoluble floc, the branched component remaining in solution.[34]

Selective Enzymolysis. If a starch paste or autoclaved solution is treated with malt amylase or other liquefying enzyme, the linear component tends to flocculate, the branched component undergoing considerable degradation.[35] Several reactions are probably involved. In the first place, retrogradation of the linear material is accelerated by enzymatic degradation of the branched component and consequent lowering of the protective colloid action afforded by the latter. Secondly, where fatty acid is present as in the cereal starches, it is preferentially bound or adsorbed by the linear component, contributing to the insolubility of the latter. Amylase action is greatly retarded by such insolubilization; consequently, the linear component is at least partially protected against enzymatic degradation.

[30] M. E. Baldwin, *J. Am. Chem. Soc.*, **52**, 2907 (1930).
[31] C. L. Alsberg, *Plant Physiol.*, **13**, 295 (1938).
[32] E. Pacsu, *J. Am. Chem. Soc.*, **63**, 1168 (1941).
[33] T. J. Schoch, *Advances in Carbohydrate Chem.*, **1**, 247 (1945).
[34] L. Maquenne, *Compt. rend.*, **137**, 88, 797 (1903).
[35] A. R. Ling and D. R. Nanji, *J. Chem. Soc.*, **123**, 2666 (1923); S. B. Schryver and E. M. Thomas, *Biochem. J.*, **17**, 497 (1923).

Electrophoresis. The organic phosphate of the root and tuber starches is associated almost exclusively with the branched component, imparting electrical polarity. Hence, when solutions of these starches are placed in an electric field, the branched component migrates toward the anode, leaving the linear component in solution.[19] Opposite behavior is observed with the cereal starches, which contain an appreciable amount of free fatty acid but little or no esterified phosphate. In such cases, the linear component adsorbs the fatty acid, thereby acquiring electrical polarity and migrating toward the anode.[21] The branched component remains in solution.

Selective Adsorption. When a cold, dilute starch solution is passed through a column packed with cotton, the linear component is preferentially adsorbed, the branched component remaining in solution. Subsequently, the linear component may be stripped from the cotton with hot water.[32]

Fractional Leaching. When a 1 to 3 per cent starch suspension is heated to a temperature just high enough to swell all the granules, a portion of the linear material dissolves and diffuses out of the swollen granule into the aqueous substrate.[29, 30] This may be separated by filtration or centrifugation and appears to represent the shorter and more mobile linear chains. It may be purified either by allowing it to retrograde from solution or by selective precipitation with such polar agents as butyl alcohol. The undissolved residue of swollen granules still contains at least half of the total linear material, rendered insoluble either by retrogradation or (in the case of the cereal starches) by adsorption of lipid material.

Selective Precipitation with Polar Agents. This recently developed method most closely approaches the ideal fractionation, since the fractions are isolated in a high state of purity, with minimum or negligible chemical degradation and without interference from retrogradation. It depends upon the fact that the linear component forms insoluble crystalline complexes with various polar organic materials, including alcohols and alicyclic hydroxy compounds, higher fatty acids, nitro-compounds, esters, ethers, and mercaptans—seemingly, any substance which contains a group capable of hydrogen bonding attached to a hydrophobic residue. Preferred fractionating agents include *n*-butyl alcohol,[36] commercial amyl alcohol,[33] nitropropane,[37] and a mixture of thymol and cyclohexanol.[38] The starch is first dissolved in water at 1 to 3 per cent concentration, either by prolonged boiling to disintegrate the granule

[36] T. J. Schoch, *J. Am. Chem. Soc.*, **64**, 2957 (1942).
[37] R. L. Whistler and G. E. Hilbert, *J. Am. Chem. Soc.*, **67**, 1161(1945).
[38] W. N. Haworth, S. Peat, and P. E. Sagrott, *Nature*, **157**, 19 (1946).

structure or by autoclaving within carefully controlled *p*H limits (5.9–6.3). The hot solution is then treated with the fractionating agent and allowed to cool. The linear component separates as microscopic needles, needle clusters (Fig. 102), or beautifully formed six-petalled rosettes. The product is most readily isolated by high-speed centrifuging (*e.g.*, by use of a laboratory supercentrifuge), the branched component re-

FIG. 102. Photomicrograph of the linear component of potato starch, as precipitated by complex formation with amyl alcohol. Stained with iodine. Magnification, 500 diameters.

maining in the centrifugate, whence it may be recovered by alcohol precipitation. The linear fraction may be purified by recrystallization, dissolving it in boiling water and cooling in the presence of the precipitating agent. According to this mode of fractionation, the percentage yields of linear component from various common starches are as follows: corn, 28 to 29; sago, 27; potato, 18 to 22; tapioca, 17.

An important aid in the fractionation of starch has been the development of accurate iodometric methods for estimating the linear component. The familiar blue starch-iodine coloration is due to an adsorption complex between the linear component and iodine. The branched component has no such affinity for iodine, giving only weak red or violet colorations. Thus the content of linear material in any starch sample

can be estimated by spectrophotometric methods,[39] the percentage light transmission being measured in the 610 to 710 mμ (red) band. The method is standardized on purified linear material obtained by fractional leaching or by selective precipitation. Alternatively, the starch sample may be titrated potentiometrically with standard iodine solution, giving a characteristic S-curve.[40] The iodine-binding capacity of the starch can be calculated from the "break" in the curve and related to the iodine affinity of the purified linear component. Results are in close accord with actual yields by selective precipitation methods. Thus defatted corn starch and its purified linear component bind 5.3 per cent and 19.0 per cent of iodine, respectively. The estimated content of linear component in corn starch would therefore be 28 per cent. Since fatty acids combine with the linear component and prevent adsorption of iodine, starch samples must be thoroughly defatted before analysis.

Structure and Behavior of the Starch Fractions. The evidence which has led to the concept of linear and branched structures for the starch fractions may be summarized as follows.[29] Methylation analysis of the linear component (obtained by fractional leaching) gives a yield of tetramethylglucose equivalent to a chain length of some 300 glucose units. Although the opposite aldehydic terminus of the chain cannot be accurately evaluated, its reducing value against ammoniacal silver oxide is of a similar magnitude. In contrast, while the branched component gives a much higher yield of tetramethylglucose corresponding to 25 to 30 glucose units, its reducing value toward alkaline copper, ferricyanide, or ammoniacal silver is much lower than that of the linear component. Relative content of terminal aldehyde can likewise be evaluated by digesting the starch material in hot alkali and determining the amount of acidic substances produced. Since alkaline degradation is initiated by enediol rearrangement of terminal aldehyde groups followed by progressive scission of acidic fragments, the so-called alkali lability provides a relative index of molecular weight.[41] By this method, the branched component of corn starch appears to have a molecular weight approximately three times that of the linear fraction.[36]

Confirmatory evidence for branched structure is provided by the action of β-amylase enzyme on the individual fractions.[29] The linear component gives a yield of maltose approaching 100 per cent, indicating a linear chain with no irregularities to obstruct enzymatic degradation. The branched component yields only 50 per cent maltose, leaving an un-

[39] R. M. McCready and W. Z. Hassid, *J. Am. Chem. Soc.*, **65**, 1154 (1943); R. W. Kerr and O. R. Trubell, *Paper Trade J.*, [15] **117**, 25 (1943).

[40] F. L. Bates, D. French, and R. E. Rundle, *J. Am. Chem. Soc.*, **65**, 142 (1943).

[41] T. J. Schoch and C. C. Jensen, *Ind. Eng. Chem., Anal. Ed.*, **12**, 531 (1940).

attacked "limit dextrin" which represents that residue of the molecule protected behind points of branching (Fig. 101). When this limit dextrin is treated with α-glucosidase to split off these interfering branch points, the product again becomes susceptible to β-amylase action. These observations indicate a highly random type of branching, possibly some 50 to 70 branches with an average length of 25 to 30 glucose units each, giving a tree-like molecule whose gross shape is roughly globular.

Physical evidence lends additional qualitative support to these concepts of molecular shape. The acetate of the linear component can be fabricated into filaments of high tensile strength,[42] a characteristic which is associated only with linear types of polymers. No such filaments can be obtained from the acetate of the branched component. Similar results have been obtained in flow polarization studies by placing solutions of the fractions (stained with iodine) in the gap between two concentric cylinders, one of which is rotated and the other static. Under such a gradient of shear, linear molecules tend to orient in the direction of flow, much like logs in a rapidly moving stream; this system then behaves as an optical grating and polarizes light. Only the linear starch component shows evidence of molecular alignment along the flow lines.[43]

Physical Behavior of the Starch Fractions. The development of the physical chemistry of high polymers has played an important role in elucidating the colloidal behavior of starch and its fractions. Long linear molecules containing a large proportion of hydrophilic groups (*e.g.*, hydroxyl or amino) develop intermolecular forces of attraction of the nature of hydrogen bonding or van der Waals forces. As a consequence, such molecules tend to associate in side-by-side fashion, giving a structure which is essentially crystalline. Thus, x-ray spectrography of cellulose and chitin reveals an orderly parallel-wise association of extended linear chains. Although the attractive force exerted by a single hydroxyl group is relatively small, the aggregate of such forces over a long linear chain may even exceed the energy of a primary valence linkage. Efforts to dissociate this intermolecular structure by physical means may therefore result in rupture of primary bonds.

The linear component of starch exhibits these association effects to a pronounced degree. Although starch can be isolated in dissolved form by fractional leaching, such solutions are highly unstable and retrograde spontaneously as an insoluble floc. This product gives a B-type x-ray pattern indicative of associated linear chains. It cannot be redissolved by heating, even at autoclave temperatures. Solution can only be ef-

[42] R. L. Whistler and G. E. Hilbert, *Ind. Eng. Chem.*, **36**, 796 (1944).
[43] R. E. Rundle and R. R. Baldwin, *J. Am. Chem. Soc.*, **65**, 554 (1943).

fected by reagents such as alkali hydroxides, which form stable complexes with the linear material and hence can pry apart the aggregated chains. Similarly, the linear fraction obtained by selective precipitation with amyl alcohol can be dissolved in hot water to give relatively clear fluid solutions, even at concentrations as high as 5 to 10 per cent. If such solutions are cooled to room temperature, they set up to rigid opaque gels which cannot be liquefied by heat. Such a gel may be considered as a "brush-pile" of associated linear chains.

However, the linear fractions from the various starches show different degrees of retrogradation, indicating inherent differences in molecular structure which have not yet been fully elucidated. Thus the linear components from corn and wheat starches undergo rapid retrogradation and gelation, whereas the corresponding fractions from tapioca and potato starches are much more stable. Viscosity measurements likewise indicate substantial variations in molecular shape. In a homologous series of linear polymers, the chain length is related to the intrinsic viscosity, the latter being derived by graphing the function (specific viscosity divided by concentration) against concentration and extrapolating to zero concentration. The intrinsic viscosities of the various linear fractions are in the order potato > tapioca > corn, and the relative chain lengths are therefore assumed to be in the same order.[44] Similarly, evaluation of terminal aldehyde groups by alkali lability measurement indicates that the linear component of potato starch is substantially larger than that of corn starch. In addition to these variations between the linear components of the various starches, there is undoubtedly a wide spectrum of molecular sizes in any individual linear fraction.

The linear fraction probably assumes an extended or stretched form in the granule or in retrograded state. In the presence of iodine, it may perhaps form a helical spiral, with the iodine located within the helix. This concept was first proposed to explain the blue starch-iodine color.[45] Since the hydroxyl groups of the starch would be located on the exterior surface of the helix, the interior would be essentially hydrocarbon, and the iodine in such an environment might therefore assume a color approaching that of its solutions in lipotropic medium. This interpretation has been questioned on the grounds that blue iodine complexes are obtained with a wide variety of organic and inorganic substances which cannot possibly possess helical structure. However, recent evidence from x-ray studies seems to substantiate the existence of a helix. The iodine complex of the linear component gives a V-type x-ray spectrum

[44] J. F. Foster and R. M. Hixon, *J. Am. Chem. Soc.*, **66**, 557 (1944).

[45] C. S. Hanes, *New Phytologist*, **36**, 101, 189 (1937); K. Freudenberg, E. Schaaf G. Dumpert, and T. Ploetz, *Naturwissenschaften*, **27**, 850 (1939).

whose unit cell dimensions correspond to closely packed helices with 6 glucose residues per "turn" of the spiral.[46] The complexes of the linear fraction with amyl alcohol or fatty acid yield the same V-pattern, and it is therefore presumed that these selective precipitants likewise assume an interior position within the helix.[47]

The branched component yields viscous opalescent solutions which are relatively stable in a physical sense. In fact, this fraction is largely responsible for the protective colloid action of starch. However, the exterior branches of the molecule are 25 to 30 glucose units in length, and even this relatively minor degree of linearity is sufficient to produce some aggregation effects. Thus, the granular organization of starch has been attributed to the branched component, through the formation of a radial network of interlacing branches.[29] Tangential association forces are thereby set up, imparting coherence and insolubility to the granule. The gelatinization phenomenon represents the gradual relaxation of these forces. Although solutions of the branched fraction are stable at room temperature for extended periods of time, insolubilization occurs if such solutions are frozen and thawed. Under certain conditions, the branched component may even be induced to crystallize, as indicated by the formation of a B-type x-ray pattern. The staling of bread has been ascribed to a similar reaction involving gradual association of the branched component throughout the bread crumb.[48] Unlike retrogradation of the linear component, these insolubilization reactions of the branched component are readily reversible by moderate heating.

It is of interest to compare the behavior of glycogen, which is considered to have a highly branched structure differing from the branched component of starch principally because its branches are much shorter (only 9 to 11 glucose units in length). Glycogen exhibits no such insolubilization on freezing. A counterpart of glycogen is present in dissolved state in milk-stage sweet corn,[49] possibly a forerunner of the branched component. This so-called corn glycogen is incapable of forming granules.

Waxy and Linear Starches. For almost a century, botanists have recognized the existence of certain peculiar types of starch which stain red with iodine [50] rather than the traditional blue. These starches were

[46] R. E. Rundle and D. French, *J. Am. Chem. Soc.*, **65**, 558, 1707 (1943).

[47] F. F. Mikus, R. M. Hixon, and R. E. Rundle, *J. Am. Chem. Soc.*, **68**, 1115 (1946).

[48] T. J. Schoch and D. French, *Cereal Chem.*, **24**, 231 (1947).

[49] D. L. Morris and C. T. Morris, *J. Biol. Chem.*, **130**, 535 (1939); J. B. Sumner and G. F. Somers, *Arch. Biochem.*, **4**, 7 (1944); W. Z. Hassid and R. M. McCready, *J. Am. Chem. Soc.*, **63**, 1632 (1941).

[50] A. Meyer, *Ber. deut. botan. Ges.*, **4**, 337 (1886).

reported from such different sources as maple cotyledons (*Acer pseudoplatanus*), certain orchids (*Goodyera*), and the seeds of "Job's tears." The so-called waxy or glutinous cereals are a much more important source, comprising particular varieties of maize, sorghum, rice, barley, and millet. The genetic waxy characteristic which gives rise to the red-staining cereal starches is recessive. The waxy cereals do not appear to be capable of synthesizing a linear fraction, and these starches therefore consist entirely of branched material, seemingly indistinguishable from the branched component of the normal blue-staining starches. No attempt has yet been made to explain the specific biochemical functions of the individual fractions in the plant.

Waxy rice has long been cultivated in China for certain types of food products. Recently, considerable quantities of waxy maize and sorghum starches have been produced in the United States as a replacement for Javanese tapioca starches when imports of the latter were cut off during World War II. In general, the properties of the waxy starches reflect the absence of the linear component. They give clearer pastes with little tendency to gel or retrograde.[51] Their principal utility is for certain food products and for those uses of starch where retrogradation effects must be avoided.

At the other end of the scale, it has been found that the starch from common wrinkled-seeded garden peas contains at least 65 to 75 per cent of linear component.[52] When heated in water, the granules swell only to a limited degree; retrogradation of the linear material then occurs, and further swelling is prevented. As a consequence, pea starch does not give a viscous paste.

Biosynthesis of Starch. Synthesis of starch substance in the plant is probably effected through the action of phosphorylase enzyme on glucose-1-phosphate. This reaction has been carried out *in vitro* with phosphorylase from potato juice to give a so-called synthetic starch which stains blue with iodine and exhibits pronounced retrogradation tendencies.[53] A small amount of polysaccharide (preferably glycogen or the branched component of starch) is necessary to activate the synthesis,[54] possibly by providing non-aldehydic terminal groups to build

[51] H. H. Schopmeyer, G. E. Felton, and C. L. Ford, *Ind. Eng. Chem.*, **35**, 1168 (1943); M. M. MacMasters and G. E. Hilbert, *ibid.*, **36**, 958 (1944).

[52] J. P. Nielsen and P. C. Gleason, *Ind. Eng. Chem., Anal. Ed.*, **17**, 131 (1945); G. E. Hilbert and M. M. MacMasters, *J. Biol. Chem.*, **162**, 229 (1946).

[53] C. S. Hanes, *Proc. Roy. Soc.*, **B128**, 421 (1940); **B129**, 174 (1940); G. T. Cori, C. F. Cori, and G. Schmidt, *J. Biol. Chem.*, **129**, 629 (1939); R. S. Bear and C. F. Cori, *J. Biol. Chem.*, **140**, 111 (1941).

[54] G. T. Cori, M. A. Swanson, and C. F. Cori, *Federation Proc.*, **4**, 234 (1945); P. H. Hidy and H. G. Day, *J. Biol. Chem.*, **152**, 477 (1944).

upon. In the presence of potato phosphorylase, the aldehydic carbon of glucose-1-phosphate is coupled with carbon-4 on the non-aldehydic terminus of the existing glucose chain, the phosphate being eliminated. Successive repetitions of this condensation may build up a linear chain of several hundred glucopyranose units. It is not known whether the activating agent remains as part of the synthesized linear chain.

In brain and liver tissue, there is a second enzyme system which is of itself incapable of synthesis, but which can induce points of branching, presumably in α-1,6-position. Thus, when both enzymes act conjointly, a linear chain is built up to a certain point, branching is then induced, and linear synthesis continues on both branches. This reaction has likewise been conducted *in vitro* to give a branched polysaccharide similar to glycogen or the branched starch fraction, red-staining toward iodine and stable in solution.[55] Although synthesis of the individual starch components can be explained by these mechanisms, no satisfactory reasons have been offered for the production of both linear and branched fractions in the same starch granule, or for the wide range in composition of the various starches. Since glycogen is presumably formed in the animal organism by similar means, it is of interest to note that Claude Bernard reported the formation of an abnormal blue-staining "glycogen" in paralyzed muscle.[56]

Technology of Starch.[57] Commercial production of potato and tapioca starches is effected merely by grinding the root or tuber with water to free the starch from its cellular matrix, screening the magma to remove cellulose, then passing this water suspension over long inclined troughs or "tables." The granular starch deposits on the tables and the lighter protein and fine fiber are floated off. Processing of corn is somewhat more elaborate, since the shelled corn must be steeped in very dilute sulfur dioxide solution to soften the protein and swell the kernel. The germ is then removed and the endosperm ground, screened through bolting cloth, and sedimented on tables. The principal uses for starch are the production of glucose syrups and crystalline glucose, for various food purposes, as an emulsifier and protective colloid, as an adhesive and binding agent, and for the sizing and finishing of textiles and paper.

A considerable amount of industrial starch is modified chemically to increase its water solubility, decrease its paste viscosity, or minimize

[55] E. J. Bourne and S. Peat, *J. Chem. Soc.*, 877 (1945); E. J. Bourne, A. Macey, and S. Peat, *ibid.*, 882 (1945).

[56] C. Bernard, "Leçons sur le diabête," p. 553 (1877), quoted from E. Pflüger, *Das Glycogen*, 2nd ed., p. 21, 1905.

[57] For a complete survey of the technology of starch, see R. W. Kerr, *Chemistry and Industry of Starch*, Academic Press, New York, 1944.

retrogradation tendencies of the linear component. The principal modes of modification follow.

Acid Hydrolysis. Granular starch may be treated with dilute acid at a temperature below the gelatinization range. This effects a slight and probably random hydrolysis of the starch substance, and the resulting product gelatinizes in hot water to give pastes of low viscosity (Fig. 99). These "thin-boiling" starches are used primarily for sizing purposes.

Enzyme Hydrolysis. The starch may be gelatinized in hot water, converted to the desired fluidity by hydrolysis with an amylolytic enzyme (from malt, from pancreatic preparations, or from various bacterial and fungus sources), then used directly as a sizing agent.

Dextrinization. Dry starch (either neutral or slightly acidified) is heated at 150° to 200°C. to give "dextrins" which are highly soluble in cold water and which are predominantly red-staining toward iodine. A slight initial hydrolysis probably occurs in this process, followed by recombination of the fragments to give a branched type of molecule. The dextrins are chiefly used as adhesives.

Oxidation. Granular starch is treated with alkaline hypochlorite solution to give products of low paste viscosity and reduced retrogradation tendencies. These oxidized starches have low iodine affinity, possibly because the linear chains have been warped by oxidation.

Etherification. Introduction of a small proportion of alkyl or hydroxyalkyl groups (*e.g.*, less than one substituent per glucose unit) yields water-soluble derivatives with low or negligible iodine affinity. These products show no tendency to gel or retrograde; as in the corresponding cellulose derivatives, the presence of substituent groups prevents the association of linear chains. High production costs have prevented commercial utilization of these products.

GLYCOGEN

Glycogen is the reserve polysaccharide of the animal kingdom and is found in the muscles and especially in the liver of many higher and lower animals. For this reason it is frequently called "animal starch." This designation is not particularly appropriate, however, because it also occurs in certain fungi (especially yeast), in certain algae, and in bacteria; also, a polysaccharide which is similar in properties and structure to glycogen has been isolated from a higher plant, *Zea mays.*[58, 59]

Unlike starch, glycogen is not differentiated morphologically, and it occurs in the organs largely in an insoluble form combined with

[58] D. L. Morris and C. T. Morris, *J. Biol. Chem.*, **130**, 535 (1939).

[59] W. Z. Hassid and R. M. McCready, *J. Am. Chem. Soc.*, **63**, 1632 (1941).

proteins from which it is liberated by hydrolysis of the proteins with alkali.

Glycogen is very resistant to the action of alkalies and is normally prepared by boiling liver (rich in glycogen) or mussels in 30 per cent potassium hydroxide solution, to destroy the proteins and other organic substances, and precipitating the glycogen with ethyl alcohol. Glycogens of widely varying molecular weight can be obtained by fractional precipitation from aqueous solutions with alcohol. It is a white amorphous powder which dissolves readily in cold water to form an opalescent, colloidal solution which gives a red-brown color with iodine, is precipitated by alcohol, does not reduce Fehling's solution. Its specific rotatory power $[\alpha]_D^{20}$ is about $+200°$.

In its chemical behavior, glycogen is very similar to starch. It is quantitatively hydrolyzed to D-glucose by acids and to maltose by the combined action of α- and β-amylase. With purified β-amylase, hydrolysis stops when 47 per cent of the original material is degraded to maltose; the remaining 53 per cent consists of a high molecular weight residual dextrin. Like starch it gives crystalline dextrins with *Bacillus macerans*. In the presence of inorganic phosphates, a phosphorylase obtainable from yeast, muscle, and liver catalyzes the following equilibrium reaction:

$$\text{Glycogen} + \text{inorganic phosphate} \rightleftharpoons \text{glucose-1-phosphoric acid}$$

For the synthetic reaction a trace of glycogen is required as an activator.[60]

Studies of the structure of glycogen reveal that, unlike ordinary starches, it consists entirely of branched molecules which vary considerably in size, depending largely on the source. The glucose residues are joined by α-1,4-linkages and branching occurs through α-1,6-linkages. Thus, Haworth and his associates[61] obtained 2,3,6-trimethylglucose, 2,3-dimethylglucose, and 7 to 9 per cent of 2,3,4,6-tetramethylglucose upon the hydrolysis of methylated glycogens from different sources. If it is assumed that methylation has been complete, the isolation of a dimethylglucose means that glycogen has a branched chain structure. From methylation studies of the residual dextrin which resists β-amylase action, Meyer and Fuld[62] obtained 18 per cent tetramethylglucose,

[60] G. T. Cori and C. F. Cori, *J. Biol. Chem.*, **131**, 397 (1939); **158**, 321, 341 (1945); G. T. Cori, M. A. Swanson, and C. F. Cori, *Federation Proc.*, **4**, 231 (1945).

[61] W. N. Haworth, E. L. Hirst, and F. A. Isherwood, *J. Chem. Soc.*, 577 (1937); W. N. Haworth, E. L. Hirst, and F. Smith, *ibid.*, 1914 (1939).

[62] K. H. Meyer and M. Fuld, *Helv. Chim. Acta*, **24**, 375 (1941).

which is equivalent to one end group for every 5.5 glucose residues. As the original methylated glycogen yielded 9 per cent of tetramethylglucose (equivalent to one end group for every 11 glucose units) and glycogen loses 47 per cent of its weight when hydrolyzed with β-amylase, this implies that the enzyme removes about 5 glucose units per end group. Since all the end groups are still present in the limit dextrin, Meyer [63] has concluded that the outer branches of the glycogen molecule which are hydrolyzed by the enzyme consist of 6 or 7 glucose residues connected with α-1,4-linkages; of these 5.5 are hydrolyzed by β-amylase, leaving 1 or 2 at the branching point; the latter comprise the

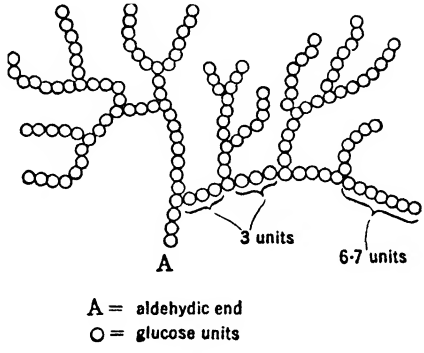


FIG. 103. Diagrammatic representation of the ramified structure of glycogen. (After Meyer.)

terminal groups of the residual dextrin. Only short chains, averaging 3 glucose residues, occur between the branches which stem from the 6 positions. This concept is illustrated in Fig. 103 and is in conformity with the fact that x-ray analysis reveals glycogen to be amorphous and that its solutions do not retrograde; the compact branched chain molecules cannot readily orient to form crystalline aggregates.

OTHER HEXOSANS

Inulin and Other Fructosans. The reserve polysaccharides of many plants yield D-fructose as the main product of acid hydrolysis. Inulin was the first to be discovered,[64] but more recently several other fructosans have been obtained from different plants. A bacterial fructosan, levan, is synthesized from sucrose (and raffinose) by a number of

[63] K. H. Meyer, *Advances in Enzymol.*, **3**, 109 (1943).

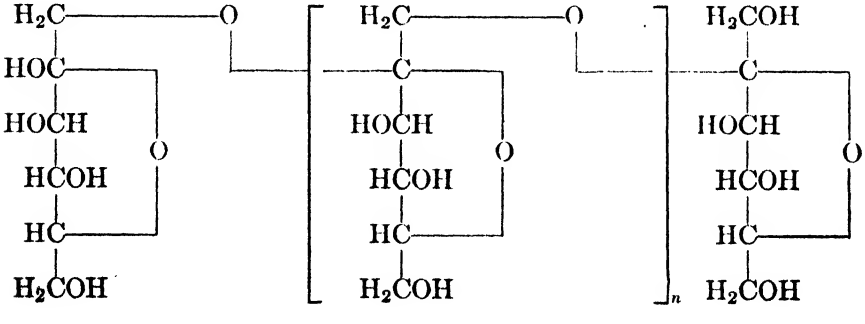
[64] Inulin was discovered in 1804, V. Rose, *Neves Allgem. J. Chem.* (*Gehlens*), **3**, 217, 1804, and was named inulin by T. Thomson in 1818.

different bacteria. In these polysaccharides the fructose residue is invariably present in the furanose form, but they differ in molecular size as well as in the nature of the glycosidic linkage.[65]

Inulin. Inulin replaces starch as the reserve polysaccharide in some plants, especially those of the genus *Compositae* where it is found in the roots and tubers of the dahlia, Jerusalem artichoke (girasole), and dandelion; in some other plant species starch and inulin occur together.

Inulin is a white powder which is only slightly soluble in cold water but readily dissolves in hot water to form a colloidal solution which does not gelatinize on heating; after heating in aqueous solution, glycerol, or ethylene glycol, it separates, upon precipitation with ethanol, in a form soluble in cold water. It does not give a color reaction with iodine, has very low reducing power, and is levorotatory, $[\alpha]_D^{20} = -40°$ (in aqueous solution). Molecular weight determinations give values of three to five thousand. It is readily hydrolyzed by acids to D-fructose but is not hydrolyzed by the amylases. It is hydrolyzed to D-fructose by the enzyme *inulase*, which is absent from the higher animals; inulin, therefore, is of little or no nutritional value to humans or other vertebrates.[66]

On hydrolysis of methylated inulin the chief product is 3,4,6-trimethylfructose with some 1,3,4,6-tetramethylfructose, thus revealing that the fructose residues are in the furanose form and adjacent units are linked through carbon-1 and carbon-2 (the glycosidic hydroxyl).[67] Like sucrose, the ease of hydrolysis of inulin by acids is associated with the fructofuranoside linkages.

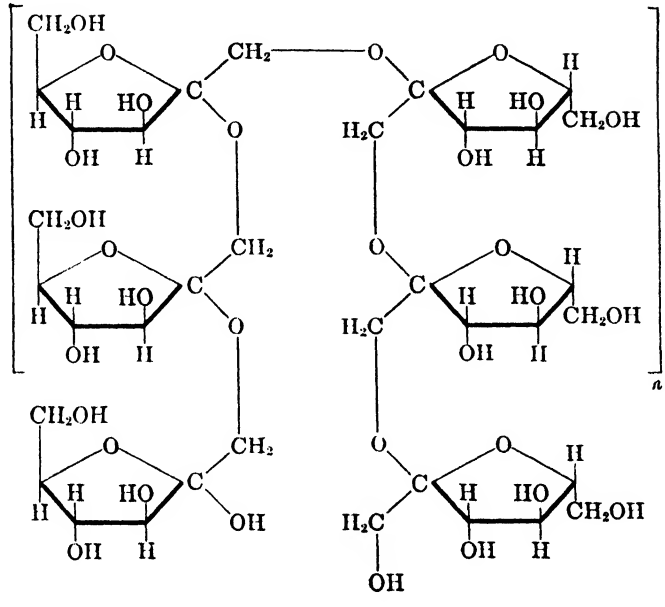


[65] The polyfructosans are discussed in detail by Emma J. McDonald, *Advances in Carbohydrate Chem.*, **2**, 253 (1946). See also the discussion by T. H. Evans and H. Hibbert, beginning on p. 204 of that volume.

[66] J. L. Bollman, *Proc. Staff Meetings, Mayo Clinic*, **8**, 67 (1933).

[67] W. N. Haworth and A. Learnes, *J. Chem. Soc.*, 619 (1928); W. N. Haworth, E L. Hirst, and E. G. V. Percival, *J. Chem. Soc.*, 2384 (1932).

This structure implies chains of the form:

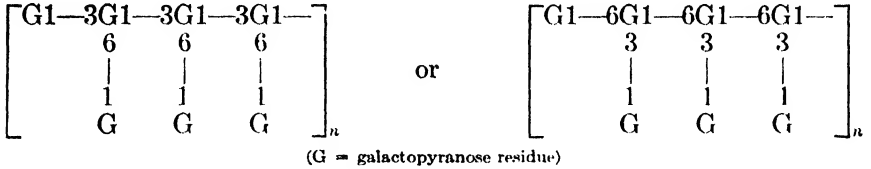


Other Fructosans. A number of fructosans differing from inulin have been prepared from the higher plants. These include asparagosin (from asparagus roots), sinistrin (from the sea onion, *Scilla maritima*, *Urginea maritima*), graminin (from rye), phlein (from tubers of timothy), poan (from roughstalk bluegrass, *Poa trivialis*), triticin (from quackgrass, couch grass, *Agropyron repens*), and irisin (from the roots of *Iris pseudacorus*). Those found in the grasses are present in the greatest concentration at the time of heading; during the subsequent rapid growth the amounts decrease, indicating that they serve as a transitory reserve.[68] Conditions which favor sugar storage appear to result in the accumulation of fructosan. The structure of these fructosans has not been established with certainty since many of the partially methylated hydrolytic products are not known in crystalline form. According to Schlubach and Sink,[69] asparagosin, sinistrin, and graminin, like inulin, yield chiefly 3,4,6-trimethyl-D-fructofuranose after methylation and hydrolysis and hence possess a 1,2-glycosidic linkage; on the other hand, phlein, poan (poain), and secalin give rise principally to 1,3,4-trimethyl-D-fructofuranose and hence contain a 2,6-linkage like the bacterial levans discussed in a subsequent section.

[68] H. K. Archbald, *New Phytologist*, **39**, 185 (1940).

[69] H. H. Schlubach and O. K. Sink, *Ann.*, **544**, 111 (1940).

Galactogen. Galactogen occurs along with glycogen in the albumin glands and eggs of the vineyard snail, *Helix pomatia*, and is made up entirely of galactose residues, some of which are L-galactose. Hydrolysis of the methylated polysaccharide yields approximately equal quantities of 2,3,4,6-tetramethyl-D-galactopyranose and 2,4-dimethyl-D-galactopyranose.[70,71] Since no trimethylgalactose has been found, the repeating unit must consist of a highly branched chain containing both 1,3- and 1,6-linkages.[71] All but the end galactose units must carry another galactose unit as a branch, the latter having 4 free hydroxyl groups. Either of the following structures is possible.[72]



(G = galactopyranose residue)

Optical studies indicate that there is a 6:1 ratio of D- to L-galactose, the L-form probably being one of the four terminal residues in the repeating unit depicted above. This is the only known occurrence of L-galactose in an animal product. As galactogen possesses colloidal properties and is non-reducing, it must be a large molecule made up of a number of the repeating units, each probably consisting of 7 galactose residues.[71] The stereochemical nature of the linkages is unknown, but as galactogen is levorotatory ($[\alpha]_D^{20}$ = approx. $-20°$) it is quite likely that a number of the glycosidic bonds are of the β-type.

Lichenin. This cellulose-like polysaccharide occurs as a cell-wall constituent in lichens, such as Iceland moss (*Cetraria islandica*), barley lichen (*Usnea barbata*), *Evernia vulpina*, and in smaller quantities in higher plants. Unlike cellulose, lichenin is soluble in hot water to form a colloidal solution; also it is more readily hydrolyzed than cellulose. Like cellulose it yields cellobiose upon acetolysis with acetic anhydride and sulfuric acid. Appropriate enzymes, free of cellobiase, will convert it quantitatively to cellobiose, and upon complete hydrolysis it yields D-glucose.[73-76] Upon methylation and hydrolysis, it yields 2,3,6-trimethyl-

[70] H. H. Schlubach and W. Loop, *Ann.*, **532**, 228 (1937).
[71] E. Baldwin and D. J. Bell, *J. Chem. Soc.*, 1461 (1938).
[72] D. J. Bell and E. Baldwin, *Nature*, **146**, 559 (1940); *J. Chem. Soc.*, 125 (1941).
[73] P. Karrer and B. Joos, *Biochem. Z.*, **136**, 537 (1923).
[74] H. Pringsheim and W. Kusenack, *Z. physiol. Chem.*, **137**, 265 (1924).
[75] W. Grassmann and H. Rubenbauer, *Munch. med. Wochschr.*, **78**, 1818 (1931).
[76] H. Pringsheim and W. Kusenach, *Z. physiol. Chem.*, **137**, 265 (1924).

glucose and only a small quantity of tetramethylglucose,[77,78] indicating that the chain is unbranched as in cellulose. The x-ray diagram differs from that of cellulose,[79] and the optical rotations of the acetate and methyl ether do not agree with those of the corresponding cellulose derivatives. The structure of the lichenin molecule has not been definitely established, but it has been suggested that it may contain both β-1,4- and β-1,6-linkages; also that there may be an alternating recurrence of 1,1- and 4,4-linkages in the molecule.[77]

Isolichenin, a starch-like polysaccharide, occurs along with lichenin and differs from it in being soluble in cold water, in giving a blue color with iodine, and in yielding maltose on enzymatic hydrolysis.[80]

PECTIC SUBSTANCES

The pectic substances are complex polysaccharides found in the cell walls of all young plant tissues and are characteristic constituents of fruits and fleshy-roots. They comprise mixtures of polysaccharides, but the chief component is pectic acid (a polygalacturonide) which is associated with an araban and frequently a galactan. They differ from those hemicelluloses which contain uronic acid residues in being much higher in uronic acid content. They are considered in detail in Chapter 28.

VEGETABLE GUMS AND MUCILAGES

The classification of the vegetable gums and mucilages in a separate group from the pectins and hemicelluloses is an indefinite and arbitrary one and is based on their solubility in water. Likewise there is not a sharp distinction between the vegetable gums and mucilages. However, the gums occur as exudates on the bark of trees either spontaneously or as a result of mechanical injury, and the mucilages are obtained by the aqueous or alkaline extraction of certain seeds. With the exception of gum tragacanth, the true gums yield a clear solution with water, whereas the vegetable mucilages swell and are only partly soluble, giving viscous liquids.

The plant gums are neutral salts of complex polysaccharide acids which contain methylpentose, pentose, hexose, and uronic acid residues joined glycosidically by linkages involving various carbon atoms within the one molecule to form branched chain structures. The component

[77] K. Hess and L. W. Lauridsen, *Ber.*, **73B**, 115 (1940).
[78] P. Karrer and K. Nishida, *Helv. Chim. Acta*, **7**, 363 (1929).
[79] R. O. Herzog, *Z. physiol. Chem.*, **152**, 119 (1926).
[80] H. Pringsheim, *Ber.*, **57**, 1594 (1924).

sugars include L-rhamnose, L-fucose, D-xylose, L-arabinose, D-mannose, and D-galactose; all except L-arabinose are present in gums in the pyranose form. In all the true gums so far examined, except gum tragacanth, the uronic acid component is D-glucuronic acid, and it exists in the pyranose form. D-Galacturnoic acid residues are present in gum tragacanth.

Although our knowledge of the plant mucilages is fragmentary, we know that they are salts of complex polysaccharide acids and bear a relationship to both the plant gums and the pectins. Like the gums they appear to possess complex branched chain structures composed of more than one sugar. Many also contain a galacturonic acid residue, which is the building unit of pectic acid.

Plant Gums. *Gum Arabic* (gum acacia) is collected as an exudate from the stems of the acacia tree, which grows in desert regions of Africa and Arabia. It is freely soluble in water and has very great "thickening-power"; it is of particular value in the confectionery industry because of its ability to prevent crystallization.

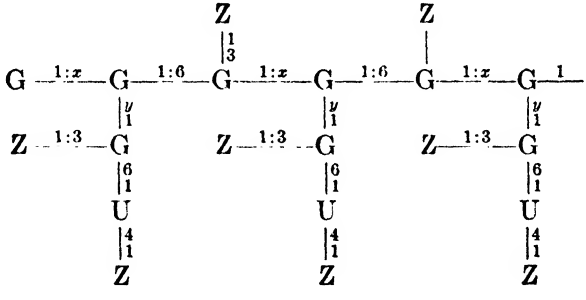
Gum arabic is the salt of a complex organic acid with calcium, magnesium, and potassium. The free acid is liberated by adding a slight excess of a mineral acid to a cold water solution of the gum and is obtained as an amorphous powder by precipitation with ethanol. Complete hydrolysis yields D-galactose, L-arabinose, L-rhamnose, and D-glucuronic acid. As a result of the careful and exhaustive studies of Smith and Jackson, [81] the structure of arabic acid can now be formulated in considerable detail. When the polysaccharide was degraded by partial hydrolysis, L-arabinose, L-rhamnose, and 3-galactopyranosyl-L-arabinose were removed, leaving a resistant residue of degraded arabic acid composed of D-galactose and D-glucuronic acid units in the proportions of three to one. Upon methylation and hydrolysis of the degraded arabic acid, 2,3,4,6-tetramethyl-D-galactose, 2,3,4-trimethyl-D-galactose, 2,4-dimethyl-D-galactose, and 2,3,4-tetramethyl-D-glucuronic acid were obtained in the relative molecular proportions of 1, 5, 3, and 3, respectively. This indicated that the nucleus of the repeating unit consisted of 9 galactopyranose residues and 3 glucopyruronic acid residues joined by both 1,3- and 1,6-linkages to give a branched chain having four terminal residues and only one reducing group. Three of the terminal residues must be uronic acid groups and the other galactose, since they were isolated as completely methylated derivatives. An aldobionic acid, 6-(β-D-glucopyranosyl)-D-galactose, was obtained as one of the products of partial hydrolysis; this yielded 2,3,4-trimethyl-

[81] F. Smith, *J. Chem. Soc.*, 744, 1724 (1939); J. Jackson and F. Smith, *ibid.*, 74, 79 (1940); F. Smith, *ibid.*, 1035 (1940).

D-galactose and 2,3,4-trimethyl-D-glucuronic acid upon methylation and hydrolysis. Isolation of this aldobionic acid indicates that some of the side chains consist of a terminal glucuronic acid group linked through at least one galactose residue with the main chain.

By direct methylation and hydrolysis of arabic acid itself, it was established that the sugar residues which are readily hydrolyzed off are joined to the relatively resistant nucleus in the form of L-arabinofuranose, L-rhamnopyranose, and 3-(galactopyranosyl)-L-arabinofuranose. In the hydrolytic products of undegraded methylated arabic acid practically all the glucuronic acid was found as the 2,3-dimethyl derivative, indicating that in arabic acid itself a sugar residue is attached to carbon-4 of each glucopyruronic acid unit. Moreover, as no trimethylgalactose was found in this hydrolysate and 2,4-dimethylgalactose was the only dimethyl derivative isolated, the galactose present in the side chain must have a residue attached to carbon-3.

In summarizing these studies, the following diagrammatic representation of the structure of the repeating unit of arabic acid has been advanced.[82]



G = D-galactopyranose
U = D-glucopyruronic acid
x = 3 or 6 and y = 6 or 3, respectively
Z = the L-rhamnopyranose, L-arabinofuranose, and 3-galactosido-L-arabinose which are liberated upon autohydrolysis, the individual disposition of which has not been determined

Similar studies of *damson gum*, which is exuded on the bark of damson trees, show that this gum also has a complex branched chain structure.[83] Like gum arabic it contains L-arabinose, D-galactose, and D-glucuronic acid units; it does not contain L-rhamnose, but instead a D-mannose derivative was found to be a constituent. A small quantity of D-xylose was also obtained upon hydrolysis.

[82] S. Peat, *Ann. Repts. on Progress Chem., Chem. Soc. London*, p. 159 (1941).
[83] E. L. Hirst and J. K. N. Jones, *J. Chem. Soc.*, 1174 (1938); 1482 (1939).

Cherry gum is similar to damson gum although it contains a larger percentage of pentose residues.[84] Upon hydrolysis it yields L-arabinofuranose, D-galactose, D-mannose, D-glucuronic acid in the molar proportions of 6, 2, 1, and 1, respectively, together with a small amount of D-xylose.

Gum tragacanth, obtained as an exudate upon incision of Asiatic shrubs of the genus *Astragalus*, [85] also contains a high proportion of L-arabinofuranose residues.

Plant Mucilages. Mucilages of a polyuronide nature are present in appreciable quantities in the seeds of mustard, flax, and quince.

The mucilage from the bark of the slippery elm, *Ulmus fulva*, yields an aldobionic acid on partial hydrolysis and has been identified as 2-(D-galacturonide)-L-rhamnose, the uronic acid having a pyranose ring.[86] This aldobionic acid has also been isolated from flaxseed mucilage;[87] flaxseed mucilage is of particular interest because it also contains L-galactose residues.[88]

The mucilage from Indian wheat, *Plantago fastigatia*, yields L-arabinose, D-galacturonic acid, and D-xylose.[89] A number of other mucilages have been studied, and their general composition is similar to those already described.[90]

The reserve carbohydrates of leguminous seeds such as the locust or carob bean, honey locust bean, and guar seed are called gums although they behave as mucilages by swelling in cold water. Because of this property, they are finding important applications in the food, textile, and canning industries. Unlike other vegetable gums and mucilages they do not appear to contain uronic acid residues but are galactomannans (or mannogalactans), since they are composed principally of D-mannose and D-galactose units. The ratios of mannose to galactose in these polysaccharides from carob bean, honey locust bean, and guar are 3.4:1,[91] 4.4:1,[92] and 1.8:1,[93] respectively. Conflicting results have been

[84] J. K. N. Jones, *J. Chem. Soc.*, 558 (1939).

[85] S. P. James and F. Smith, *J. Chem. Soc.*, 739, 746, 749 (1945).

[86] R. E. Gill, E. L. Hirst, and J. K. N. Jones, *J. Chem. Soc.*, 1469 (1939).

[87] R. S. Tipson, C. C. Christman, and P. A. Levene, *J. Biol. Chem.*, **128**, 609 (1939).

[88] E. Anderson, *J. Biol. Chem.*, **100**, 249 (1933).

[89] E. Anderson, L. A. Gillette, and M. G. Sceley, *J. Biol. Chem.*, **140**, 569 (1941).

[90] For reviews of the plant mucilages see S. Peat, *Ann. Repts. on Progress Chem., Chem. Soc. London*, p. 163 (1941); and W. Z. Hassid, *Ann. Rev. Biochem.*, **23**, 84 (1944). For an extensive discussion of the earlier literature on the mucilages see A. G. Norman, *The Biochemistry of Cellulose: The Polyuronides, Lignin, etc.*, Clarendon Press, Oxford, 1937.

[91] B. W. Lew and R. A. Gortner, *Arch. Biochem.*, **1**, 325 (1942–43).

[92] B. W. Lew, PhD. thesis, University of Minnesota, October, 1941.

[93] E. Heyne and R. L. Whistler, *J. Am. Chem. Soc.*, **70**, 2249 (1948).

reported on the nature of the glycosidic linkages, but results obtained with periodic acid oxidation indicate that 1,4-linkages are present in these gums.[93]

The tubers of *Orchidaceae* and certain *Liliaceae* also yield mucilages which do not appear to contain uronic acid residues.

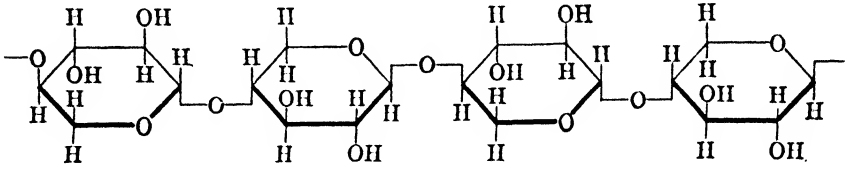
HEMICELLULOSES

The hemicelluloses comprise an ill-defined group of polysaccharides which accompany cellulose (and in the mature tissues, lignin) in the cell walls of plants. They are insoluble in water but may be extracted with dilute alkalies and are hydrolyzable by dilute acids to their constituent sugars or uronic acids. As thus defined, the hemicelluloses include two types of polysaccharides with different functions and from different locations in the cell wall: (*a*) short-chain polysaccharides that form part of the cellulose fabric itself and are oriented in the micellar structure; and (*b*) amorphous polysaccharides, which, together with lignin, are deposited in the sponge-like communicating interstices of the cell wall proper. The hemicelluloses associated with the cellulose and considered part of the cellulose aggregate are frequently called *cellulosans;* chemically they are relatively simple pentosans or hexosans. On the other hand, the incrusting polysaccharides, which may in part be linked to lignin, invariably contain uronic acid residues and are, therefore, often referred to as *polyuronide hemicelluloses.* The distinction between these two classes of hemicelluloses, however, is not sharp since it is based primarily on the partition which is effected by the procedure employed by Cross and Bevan [94] for the isolation of cellulose. Alternate chlorination and extraction with hot aqueous sodium sulfite to remove the lignin as a soluble sulfonic acid also removes the polyuronide hemicelluloses, leaving the true cellulose and cellulosans. This separation is, however, influenced by pretreatment and variations in the procedure so that it must be regarded as empirical. The chemistry of the hemicelluloses is largely undeveloped, but it is not unlikely that these substances will prove to be closely related to the polyuronides, pentosans, and hexosans, of which something is already known. The products of hydrolysis include D-glucose, D-glucuronic acid, D-xylose, D-galactose, D-galacturonic acid, L-arabinose, and D-mannose.

Cellulosans. *Pentosans.* The most common cellulosan is *xylan,* which can be obtained from lignified cell membranes, such as wood, straw, and the hulls of seeds, by extraction with 5 per cent alkali and precipitation with ethanol. By boiling with hot dilute acids, it is readily hydrolyzed to

[94] C. F. Cross and E. J. Bevan, *J. Chem. Soc.,* **38,** 667 (1880).

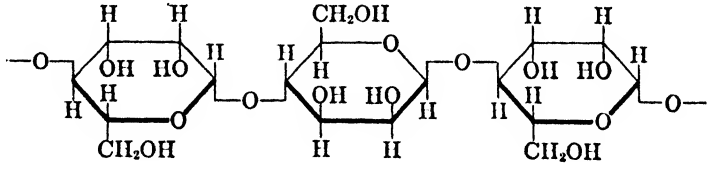
D-xylose. Beechwood, for example, yields about 16 per cent D-xylose and wheat straw about 25 per cent. Methylation and hydrolysis of xylan, in conjunction with optical rotation studies, have demonstrated that it consists of xylopyranose units joined in β-glycosidic linkages through carbon-4.[95] Xylan, therefore, has the same geometric pattern as cellulose (p. 659) and pectic acid (p. 729).



In addition to D-xylose, the xylan from Esparto grass apparently contains an end unit of L-arabinofuranose.[96]

Hexosans. In the so-called β-cellulose fraction of the polysaccharides from deciduous and coniferous woods, a *glucosan* (a 1,4-β-glucoside) is present which is soluble in 17.5 per cent sodium hydroxide solution. This alkali-soluble cellulosan consists of shorter chains than those present in the insoluble α-cellulose.

Mannans are prominent in the hemicellulose fraction of coniferous woods and also as cell-wall constituents of hard seeds, such as the endosperm (so-called vegetable ivory) of the tague tree, *Phytelepas macrocarpa*, a South American palm. Like cellulose, they are insoluble in water, of high mechanical strength, and structurally similar to cellulose since the D-mannose residues in the chain are joined by β-1,4-linkages.[97,98]



Actually, mannan occurs in the ivory nut in two forms, *A* and *B*. Mannan *A* is soluble in cold 5 per cent sodium hydroxide solution and consists of chains of 70 to 86 mannose units; mannan *B* is soluble in 10 per cent sodium hydroxide solution and is of much greater chain length.[98]

[95] H. A. Hampton, W. N. Haworth, and E. L. Hirst, *J. Chem. Soc.*, 1739 (1929).

[96] R. A. S. Bywater, W. N. Haworth, E. L. Hirst, and S. Peat, *J. Chem. Soc.*, 1983 (1937).

[97] K. H. Meyer and H. Mark, *Aufbau.*, 168 (1930).

[98] F. Klages, *Ann.*, **509**, 159 (1934); **512**, 185 (1934); F. Klages and R. Maurenbrecher, *Ann.*, **535**, 175 (1938).

Galactans have frequently been reported in the hemicellulose fraction of various plants.[99] The Western larch, *Larix occidentalis*, contains so-called ε-galactan, which is improperly classified as a hemicellulose because it is soluble in water. Larchwood may contain as much as 18 per cent of this carbohydrate.[100] It is a galactoaraban yielding approximately 6 units of D-galactose to 1 unit of L-arabinose on hydrolysis.[101,102] A galactoaraban has also been found in the seeds of the peanut.[99]

Polyuronide Hemicelluloses. The hemicellulose fraction of lignified plant tissues contains amorphous polysaccharides which yield a uronic acid and one or more simple sugars on hydrolysis. For example, the hemicelluloses from the sapwood of the white pine yield a monomethyluronic acid, D-xylose, and D-mannose,[103] whereas those from the cottonwood do not contain D-mannose.[104] The nature of the hexuronic acid or the linkages between the constituent units of these hemicelluloses have not been determined. However, galacturonic acid has been obtained from the sulfite liquors of beechwood,[105] and galacturonic and glucuronic acids have been shown to be present in the hemicelluloses of beechwood.[106] Since the hemicelluloses cannot be extracted from plant materials by water but become partly water-soluble after extraction by alkali and reprecipitation with acid, it has been suggested that these hemicelluloses are present in the wood in the form of esters. The carboxyl group of the uronic acid may be joined through an ester linkage with a cell-wall constituent such as lignin.

CELLULOSE

Cellulose is the most widely distributed skeletal polysaccharide and the most abundant and chemically resistant of all substances elaborated by living cells. As the name suggests, it is the main constituent of the cell walls of plants where it is found as a mixture of homologous polymers in association with other polysaccharides of related structure and with non-carbohydrate substances, of which the most important is lignin. The dry matter of wood contains from 40 to 50 per cent cellulose, 10 to 30 per cent hemicelluloses and other polysaccharides, and 20 to

[99] H. W. Buston, *Biochem. J.*, **29**, 196 (1935).
[100] A. W. Schorger and D. F. Smith, *Ind. Eng. Chem.*, **8**, 494 (1916).
[101] L. E. Wise and F. C. Peterson, *Ind. Eng. Chem.*, **22**, 362 (1930).
[102] E. L. Hirst, J. K. N. Jones, and W. G. Campbell, *Nature*, **147**, 25 (1941).
[103] E. Anderson, J. Kesselman, and E. C. Bennet, *J. Biol. Chem.*, **140**, 563 (1941).
[104] E. Anderson, R. B. Kaster, and M. G. Seeley, *J. Biol. Chem.*, **144**, 767 (1942).
[105] E. Hägglund, F. W. Lkingstedt, T. Rosenqvist, and H. Urban, *Z. physiol. Chem.*, **177**, 248 (1928).
[106] M. H. O'Dwyer, *Biochem. J.*, **17**, 501 (1923); **20**, 656 (1926).

30 per cent lignin. The seed hairs of the cotton plant are the purest commercial source of cellulose (85 to 97 per cent). The flax plant yields linen, the most important bast fiber, which contains 80 to 90 per cent of cellulose.

Commercial and scientific methods for preparing cellulose involve the removal of the substances associated with it and differ according to the composition of the raw material; for most commercial uses, it is usually sufficient to remove the greater part of the non-cellulosic substances. Wood has a high lignin content which can only be removed by rather drastic chemical treatment. Three processes are in common use in the preparation of chemical pulp: the wood in the form of small chips is cooked with a solution of sodium bisulfite and sulfurous acid (sulfite process), or with sodium hydroxide solution (soda process), or with a solution of sodium hydroxide containing sodium sulfate (sulfate process) which gives rise to the formation of sodium sulfide. The resulting chemical pulp still contains some lignin and considerable amounts of non-cellulosic polysaccharides. The remaining lignin may be removed by chlorine bleaching, after which the hemicelluloses and other polysaccharides may be largely removed by alkaline treatment.

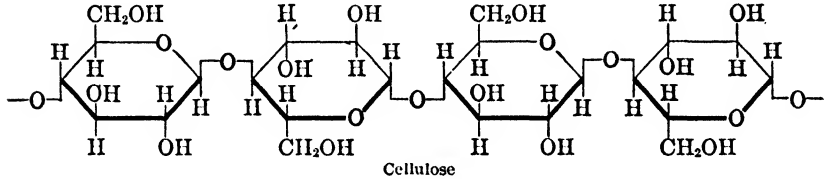
The cellulosans (principally xylan) may be extracted with 5 per cent sodium hydroxide solution, but a more concentrated solution is required to remove the less readily soluble hemicelluloses. Part of the cellulose of wood is soluble in 17.5 per cent sodium hydroxide and is called β-cellulose; the insoluble fraction is called α-cellulose and represents a much higher average degree of polymerization than the β-fraction. The cotton fiber, which contains α-cellulose associated with low percentages of non-cellulosic substances, yields relatively pure cellulose after extraction of fats and waxes with organic solvents, followed by treatment with dilute alkalies under pressure (to remove non-cellulosic polysaccharides), washing, and finally bleaching with hypochlorite; the product is termed standard cellulose. The cellulose of bast fibers, such as flax straw, is freed from pectins and other associated substances by fermentation processes (retting) and appropriate mechanical treatment.

Because of its industrial importance, cellulose has been extensively investigated, and the voluminous literature has recently been brought together in a comprehensive monograph edited by Ott.[107]

Chemical Structure. Complete hydrolysis of the purest form of cellulose by concentrated mineral acids at low temperatures yield D-glucose as the sole constituent sugar and in nearly quantitative yield. Acetolysis (simultaneous hydrolysis and acetylation) by a mixture of

[107] E. Ott, *Cellulose and Cellulose Derivatives*, Vol. V, *High Polymers*, Interscience Publishers, New York, 1943.

acetic anhydride and sulfuric acid yields the octaacetate of cellobiose and higher oligosaccharides (cellotriose, cellotetrose, and cellopentose) as intermediate products of hydrolysis; in this reaction, the oligosaccharides cannot arise from glucose. Methylation and optical rotation studies of cellobiose, in conjunction with the fact that it is hydrolyzed by emulsin but not by maltase, show that cellobiose is composed of D-glucopyranose residues joined by β-1,4-glucosidic linkages (*cf.* p. 612). Haworth and his associates have succeeded in completely methylating cellulose by treating cellulose acetate dissolved in acetone with methyl sulfate and alkali; hydrolysis followed by fractionation yielded mainly 2,3,6-trimethyl-D-glucopyranose and 2,3,4,6-tetramethyl-D-glucopyran-



Cellulose

ose, the relative proportions of which were equivalent to one end group in a chain of 100 to 200 glucose units.[108] Cellulose is therefore formulated as a linear chain of glucose residues mutually joined by β-1,4-linkages. Confirmation of this structure has been obtained by x-ray studies.

Cellulose, even when carefully prepared, apparently contains a small number of carboxyl groups which may be formed by the oxidation of the potential reducing group at the end of the chain or of primary alcohol groups at carbon-6. The possibility that the carboxyl groups may be associated with small quantities of polyuronides present as impurities cannot be excluded.

As already mentioned, the end group assay method indicates that the chain length of cellulose lies between 100 and 200 glucose units, which corresponds to molecular weights between 20,000 and 40,000. The chemical treatment to which the cellulose is subjected may bring about some degradation, and small quantities of smaller molecules with free end groups would greatly influence the result. Moreover, the isolation of the tetramethyl derivative is only approximately quantitative. Subsequent work [109] has shown that the yield of 2,3,4,6-tetramethylglucose (end group) depends on whether the methylation is carried out in air or in an atmosphere of nitrogen. When methylated under *nitrogen* no

[108] W. N. Haworth and H. Machemer, *J. Chem. Soc.*, 2270 (1932).

[109] K. Hess and F. Neumann, *Ber.*, **70**, 782 (1937); W. N. Haworth, E. L. Hirst, L. N. Owen, S. Peat, and F. J. Averill, *J. Chem. Soc.*, 1885 (1939); W. N. Haworth, R. E. Montonna, and S. Peat, *J. Chem. Soc.*, 1899 (1939).

end groups are found among the hydrolytic products, but measurements of particle size by physical methods show that repeated methylations cause a gradual degradation of the molecule to a minimum value of approximately 200 glucose units. When cellulose is methylated in *air*, end groups are found upon hydrolysis, and repeated methylations bring about a rapid and extensive degradation of the molecule as shown both by increases in the yield of end groups and by physical methods of measuring particle size. These results support the view that the end groups found in experiments where cellulose is methylated in air are due to oxidative degradation in the presence of alkali.

The molecular weight of cellulose dissolved in an ammoniacal solution of cupric hydroxide (Schweitzer's reagent) and those based on the molecular weights of cellulose esters and other derivatives determined by osmotic pressure, viscosity, and sedimentation methods have given values ranging from 70,000 to 450,000. These must be regarded as minimal values since the possibility of degradation cannot be excluded. Cellulose, like other high polymers, is a polymolecular material, as shown by the fact that, when cellulose or cellulose derivatives are fractionally precipitated from solution by a non-solvent, the less soluble fractions have higher average molecular weights than the more soluble ones. Cellulose from various sources and from different preparations may show great differences in mean chain length as well as in the degree of homogeneity. In general, the more homogeneous celluloses are the most suitable for industrial purposes.

Fiber Structure. It had long been known that cellulose fibers were anisotropic, but Nageli [110] was the first to propose that they were built up of submicroscopic, anisotropic crystalline particles which he called micelles. X-ray analysis has established that the cellulose fiber is essentially a crystalline body with submicroscopic disorganized (amorphous) regions. Early x-ray investigators were unable to reconcile their data with chemical evidence concerning cellulose structure. Independent glucose units were regarded as the fundamental building unit, but in 1926 Sponsler [111] put forward the modern concept that the glucose residues form chains which are arranged parallel to the fiber axis, *i.e.*, the three-dimensional model of a unit cell deduced from x-ray data is only part of the cellulose molecule. Sponsler and Dore [112] attempted to construct such a model which would be in complete accord with the

[110] C. Nageli, *Die Starkelkorner*, F. Schulthub, Zurich, 1858; C. Nageli and S. Schwendener, *Das Mikroskop*, W. Engelemann, Leipzig, 2 Aufl., 1877.

[111] O. L. Sponsler, *J. Gen. Physiol.*, **9**, 677 (1926).

[112] O. L. Sponsler and W. H. Dore, in *Colloid Symposium Monograph*, Vol. IV, p. 174, Chemical Catalog Co., New York, 1926.

x-ray data, physical properties, and chemical structure, and this was fully achieved a few years later by Meyer and his co-workers,[113] who postulated a monoclinic unit cell. The diagram of the basic cell they proposed for native cellulose is reproduced in Fig. 104.

Each corner and center of the unit cell is occupied by a cellobiose unit which is connected with similar units in adjacent cells to build up the cellulose chain running parallel to the *b* axis. As the cellobiose units at each corner are shared by 4 neighboring unit cells, only one is counted for each, which, together with the one in the center, gives a total of 2 cellobiose (or 4 anhydroglucose) units per cell. The planes of the glucose units lie in the *ab* plane. The unit cell is thus composed of two sets of independent chains which have a diagonal screw axis. Theoretically, the two chains could run either in the same or in opposite directions. The evidence obtained with mercerized cellulose favors the supposition that equal numbers of chains are oriented in opposite directions in these fibers; since it is inconceivable that a complete reversal of one set of chains takes place in this process, it is supposed that the same arrangement exists in native cellulose. Mark [114] has pointed out that three different kinds of forces must be operative in holding the lattice structure together in the three different directions in space. The strongest force is along the *b* axis, where the glucose units are held together by primary valence linkages (β-1,4 glucosidic bonds). Along the *c* axis, the nearest distance between atomic centers is approximately 3.1 Å., which corresponds very closely to the distance to be expected if relatively weak van der Waals forces were operative in this direction. The oxygen atoms of the glucose rings along the *a* axis (*ab* plane), however, are only 2.5 Å. apart; they are closer than oxygen atoms of different molecules in other organic crystals where only van der Waals forces are acting. This suggests that stronger forces are operative along the *ab* plane, and it is assumed that hydrogen bonding occurs between the two oxygen atoms.

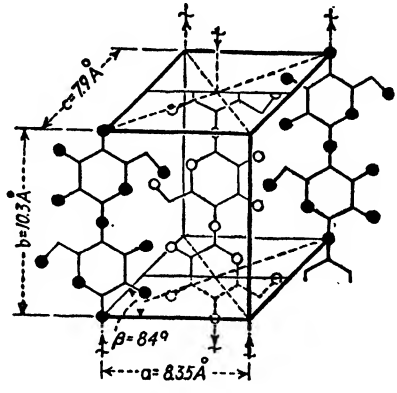


FIG. 104. Unit cell of native cellulose. (After Meyer and co-workers.)

[113] K. H. Meyer and H. Mark, *Ber.*, **61**, 593 (1928); *Z. physik. Chem.*, **2B**, 115 (1929); K. H. Meyer and L. Misch, *Ber.*, **70B**, 266 (1937); *Helv. Chim. Acta*, **20**, 232 (1937).

[114] H. Mark, *Chem. Revs.*, **26**, 169 (1940).

This bond is rather easily broken, and the assumption of its existence explains the ease with which swelling of cellulose occurs in this direction.

As previously indicated, the unit cell is repeated in all directions to build up submicroscopic crystal areas, or crystallites, and the cellulose fiber itself is an aggregate of the crystallites separated by amorphous or intercrystalline areas. Several theories, which have been advanced concerning the nature of the crystallites and their arrangement in the cellulose fiber, have been discussed by Mark.[115]

The *micellar theory*,[116] first proposed by Nageli, regards the submicroscopic crystallites as discrete particles or micelles which are much longer than they are thick. The individual micelles in native cellulose are arranged with their long axis approximately parallel with, and turned spirally around, the fiber axis. Seifriz [117] proposed that the micelles in the fiber have a "brick-wall" arrangement.

The *continuous structure theory* takes cognizance of the fact that sedimentation, viscosity, and other measurements indicate that the cellulose molecule is several times longer than that estimated for a micelle, and that it is unnecessary to assume chains which are broken lengthwise into discrete individual crystals in order to explain the x-ray data. In this theory it is assumed that in some places the long continuous chains are sufficiently close together and regular in arrangement to form a crystal lattice; in others, the chains have an irregular arrangement, and these are considered "amorphous" or intercrystalline areas.

As a compromise between the opposing micellar and continuous structure theories, the *secondary structure theory* put forward by Zwicky [118] postulates that the unit cell lattice has another much larger lattice consisting of crevices superimposed upon it. This theory was primarily designed to explain such physical properties of cellulose fibers as tensile strength, elasticity, swelling, and shrinkage which depend on a discontinuous crystal structure.

The *micellar network* or *fringe micellar theory* represents a more recent compromise between an extreme micellar and extreme chain structure. In this theory the crystalline regions are separated by an area of partially parallel but disorganized chains. It is assumed that long crystalline particles exist with an irregular outer form and that these crystalline

[115] H. Mark, *J. Phys. Chem.*, **44**, 764 (1940).

[116] The names applied to these theories are those used by W. A. Sisson, pp. 215–222 in E. Ott, *Cellulose and Cellulose Derivatives*, Interscience Publishers, New York, 1943.

[117] W. Seifriz, *Am. Naturalist*, **63**, 423 (1929).

[118] F. Zwicky, *Proc. Natl. Acad. Sci. U. S.*, **15**, 253, 816 (1929); *Phys. Revs.*, **38**, 1772 (1931); **40**, 63 (1932); **43**, 765 (1933).

particles are delineated by irregularly arranged primary valence chains which pass through several micelles and tie them together. This results in a network structure containing chains that are partly arranged in a crystalline manner and partly in an amorphous fashion.

In the *cellulose particle theory* put forward by Farr and her associates,[119] it is postulated that the fundamental building units of the cellulose fiber are tiny microscopic crystalline ellipsoidal particles which she observed in the living cytoplasm of young cotton fibers. She believes that these crystalline particles are responsible for the crystalline x-ray pattern of cellulose fibers. During growth they unite end to end to form fibrils. In the fiber, they are imbedded in a pectin-like matrix which she believes is largely responsible for the viscosity of cuprammonium solutions of cellulose, since her observations indicate that the crystallite particles are merely dispersed and not actually dissolved. Sisson [116] points out that this theory provides an explanation of the formation of cellulose and the morphological structure of native fibers, but it is not in accord with the chemical properties of cellulose and is of little value in explaining the properties of regenerated fibers. Evidence in support of the view that very small fibrils are the structural unit of native cellulose has been furnished by the electron microscope.[120]

The major axis of the unit cell in natural fibers is oriented paraliel with and frequently at some special angle to the fiber axis, the extent of orientation varying along the fiber to give crystalline and amorphous areas. When cellulose is dissolved either in cuprammonium solution, or in the form of derivatives, the orientation is largely destroyed, and, if the cellulose is regenerated without orientation, the threads or sheets which result have a much lower tensile strength than when it is oriented by applying tension during the regeneration process. As the orientation increases, the tensile strength, swelling power, luster, and shrinkage increase, while the elongation and depth of dyeing decrease. Of the natural fibers, linen has the highest degree of orientation, and its tensile strength is comparable with that of the best steel. Woods in which the orientation is parallel to the fiber axis split more easily, have a higher tensile strength, and undergo less longitudinal shrinkage than woods in which the orientation is of a more random nature.

In many reactions which cellulose undergoes (such as the sorption of water, the uptake of metallic ions by the carboxyl groups, partial hydrolysis with concentrated hydrochloric acid, and oxidation with chromic

[119] W. K. Farr and S. H. Eckerson, *Contrib. Boyce Thompson Inst.*, **6**, 189 (1934); W. K. Farr, *J. Phys. Chem.*, **41**, 987 (1937); **42**, 1113 (1938); *J. Applied Phys.*, **8**, 228 (1937); *Contrib. Boyce Thompson Inst.*, **10**, 71 (1938); **12**, 181 (1941).

[120] R. B. Barnes and C. J. Burton, *Ind. Eng. Chem.*, **35**, 120 (1943).

acid) the crystal lattice is not destroyed. This indicates that those reactions take place largely at the surface of the crystallites without changing their internal structure, and they are therefore known as *micellar surface* or *intermicellar* reactions. In other reactions, such as those involved in the formation of soluble derivatives, the x-ray diagram is altered, and they are known as *intramicellar* reactions.

Reactions of Cellulose. As cellulose is a long-chain linear compound with all but one of the potential aldehyde groups of the glucose residues involved in the glycoside linkages, it lacks the reducing power of the sugars (which contain potential carbonyl groups), and its reactions are due principally to the hydroxyl groups on carbon atoms -2, -3, and -6 of the residues. As an aliphatic alcohol, it yields addition compounds with alkalies and certain complex salts, forms alcoholates, esters, and ethers, and can be oxidized stepwise, usually with degradation, to form oxidation products containing carbonyl or carboxyl groups.

Cellulose is relatively inert toward mild chemical reagents. Strong alkalies, such as 18 per cent sodium hydroxide solution, cause pronounced swelling to form "hydrocellulose" and "mercerize" the fiber. Very little swelling occurs in the longitudinal direction, but the swelling is very pronounced in the lateral direction. Dilute sulfuric acid forms a hydrocellulose, whereas stronger sulfuric acid converts the fibers into vegetable parchment. Dilute nitric acid forms an oxycellulose. Complete nitration and acetylation results in the introduction of 3 nitro or acetyl groups per glucose residue. The nitrates are of great importance as explosives, and the nitrates, acetates, and ethers as plastics and synthetic fibers. Upon destructive distillation, cellulose yields acetone, acetic acid, and methanol.

Enzymes capable of hydrolyzing cellulose are absent from man and other animals but many bacteria, some protozoa, the "shipworm" (*Teredo*), and the snail can decompose cellulose. The ability of ruminant animals to obtain energy from cellulose and related polysaccharides is due to the action of microorganisms in the alimentary tract. These microorganisms convert the cellulose to short-chain fatty acids (mainly formic to butyric), carbon dioxide, hydrogen, and methane, and the host utilizes the fatty acids for the formation of tissue carbohydrates and fats.

CHITIN

Chitin, a nitrogen-containing polysaccharide, occurs in certain fungi and is common among invertebrate animals. It is found in certain coelenterates, mollusks, and annelids and is very prevalent in the *Arthropoda*, particularly insects and crustaceans. In the *Arthropoda*, it is

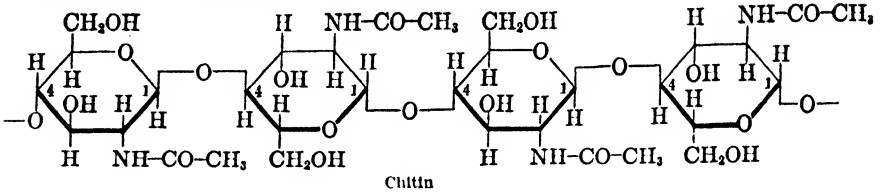
present in the exoskeleton of the body and appendages, it is a prominent constituent of the lenses of the eyes, it lines the alimentary canal (either as part of the epithelial lining or as a membranous tube surrounding the food mass, known as the peritrophic membrane), and it is present in the respiratory (tracheal) and excretory ducts. In the larger *Crustacea* (crabs, lobsters, etc.), the chitin matrix of the exoskeleton is impregnated with as much as 75 per cent of calcium carbonate. In the cuticle of insects, protein is always associated with chitin and is capable of forming a hard tough cuticle in the absence of chitin, *e.g.*, insect eggs.[121]

Chitin is colorless, insoluble in water, alcohol, ether, dilute acids, and alkalies, but in the form of a thin membrane it is freely permeable to water, electrolytes, and emulsified oils unless it is covered or impregnated with impermeable waxes. It will dissolve, with some degradation, in strong mineral acids; powerful acid hydrolysis yields glucosamine and acetic acid.[122] When heated with strong alkalies, such as saturated potassium hydroxide solution, under pressure it undergoes more or less deacetylation to form acetic acid and derivatives called *chitosans*. Chitosan has weak basic properties and forms water-soluble salts with acetic acid and mineral acids except sulfuric acid. Upon treatment with acetic anhydride, acetic acid, and zinc chloride, it is acetylated and undergoes extensive hydrolysis.

Chitin has proved to be very difficult to study chemically. It was uncertain whether it was a derivative of glucose or of mannose until the experiments of Karrer and Mayer [123] and of Haworth, Lake, and Peat [124] provided strong evidence for the former. The biological analogy between cellulose and chitin and the similarity in their x-ray diffraction patterns led Meyer and Mark [125] to conclude that the units of acetyl glucosamine are united by β-1,4-linkages to form long oriented chains in the same manner as the glucose units in cellulose. X-ray studies [126,127] of fibrous structures reveal that the unit cell contains 8 acetylglucosamine units; alternate parallel chains run in opposite directions, and the rings follow one another in a diagonal screw sequence, as shown in Fig. 105. The structure of chitin proposed on the basis of x-ray studies was soon confirmed by the isolation of *chitobiose*, the repeating unit (which differs from cellobiose only by the replacement of the —OH

[121] G. Fraenkel and K. M. Rudall, *Proc. Royal Soc. London*, **B129**, 1 (1940).
[122] A. G. Richards, *Entomol. News*, **55**, 18 (1944); *Science*, **105**, 170 (1947).
[123] P. Karrer and J. Mayer, *Helv. Chim. Acta*, **20**, 407 (1937).
[124] W. N. Haworth, W. H. G. Lake, and S. Peat, *J. Chem. Soc.*, 271 (1939).
[125] K. H. Meyer and H. Mark, *Ber.*, **61**, 1936 (1928).
[126] K. H. Meyer and G. W. Pankow, *Helv. Chim. Acta*, **18**, 589 (1935).
[127] G. L. Clark and A. F. Smith, *J. Phys. Chem.*, **40**, 863 (1936).

group on carbon-2 by an acetylamino group [128] and by the presence of β-glycosidic linkages.[129]



Chitin

Chitin is resistant to attack by most known enzymes and to the usual processes of decay; positive chitin tests have been obtained on fossilized insect remains.[130] However, it is attacked by *Bacillus chitinovorus*

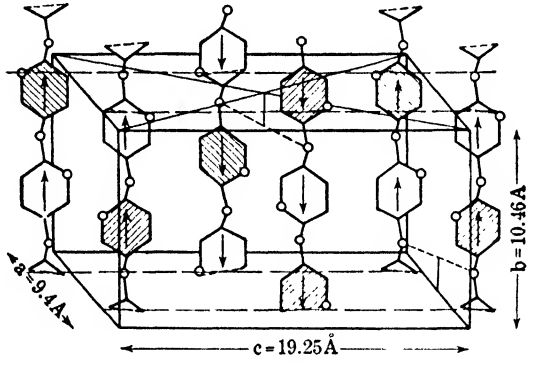


FIG. 105. The arrangement of the glucosamine units in the unit cell of chitin as shown by x-ray data. (After Clark.)

Benecke, the bacterium which is apparently responsible for its destruction in the soil. Certain snails (*Helix* sp.), a slug (*Arion ater*), a fungus-feeding beetle (*Platydema tricuspis*), and the larva of a parasitic wasp (*Pseudagenia carbonaria*) are supplied with a chitinase capable of hydrolyzing chitin.[131]

POLYSACCHARIDES FROM SEAWEEDS

The polysaccharides from seaweed comprise three groups: (*a*) the water-soluble sulfuric acid esters (agar, carrageenin, and fucoidin), (*b*) the water-soluble reserve carbohydrates, such as laminarin, which con-

[128] M. Bergmann, L. Zervas, and E. Silberweit, *Ber.*, **64**, 2436 (1931).

[129] L. Zechmeister, W. Grassmann, G. Tóth, and R. Bender, *Ber.*, **65**, 1706 (1932); L. Zechmeister and G. Tóth, *Ber.*, **66**, 522 (1933).

[130] E. Abderhalden and K. Heyns, *Biochem. Z.*, **259**, 320 (1933).

[131] C. M. Yonge, *Science Progress*, **32**, 638 (1938).

sist exclusively of glucose units and are analogous in their functions to starch, and (*c*) the alkali-soluble polyuronides, represented by algin, which are analogous to pectin.

Agar, carrageenin, and algin are frequently classed with the vegetable gums and vegetable mucilages since they form colloidal systems with similar properties. However, like the pectins, they are chiefly cell-wall constituents, and the first two differ chemically from the gums and mucilages in consisting of only 1 or 2 monosaccharide units. Tseng [132] has recently proposed the term *phycocolloid* to embrace the polysaccharides from brown and red seaweeds which are strongly hydrophilic.

The unique colloidal properties of agar, algin, and carrageenin have greatly stimulated their production and use in recent years. They are odorless and tasteless and are not liquefied by organisms which digest gelatin. Agar is still the preferred material for preparing bacteriological culture media, but it is also finding increasing applications in the food and pharmaceutical industries. The animal body does not possess enzymes capable of digesting agar; consequently, it cannot be utilized as a food and is sometimes prescribed to furnish bulk. Algin largely replaced gelatin as an ice cream stabilizer during World War II and has proved very satisfactory. Carrageenin is now the principal agent used for suspending cocoa particles in chocolate milk preparations.

Water-Soluble Sulfuric Acid Esters. *Agar* is obtained from certain species of agar-bearing red seaweeds (agarophytes), such as *Gelidium cartilageneum* and *Gracilaria confervoides* in America, and *Gelidium amansii* in Japan. The terms agar and agar-agar have been used to designate materials of varying purity and properties, and Tseng [132] has proposed that agar be defined as "the dried amorphous, gelatin-like, non-nitrogenous extract from *Gelidium* and other agarophytes, being the sulfuric acid ester of a linear galactan, insoluble in cold water, but soluble in hot water, a one per cent neutral solution of which sets at 35° to 50°C. to a firm gel, melting at 80° to 100°C."

As it occurs in nature, agar is the calcium salt of the agar acid and may be designated as calcium agarinate. The free agarinic acid will not gelate and is so strongly acid that it undergoes autohydrolysis on heating. A 1 per cent solution has a *p*H of 2.0.[133]

Structural studies [134-137] reveal that agar consists of a long chain of

[132] C. K. Tseng, *Science*, **101**, 597 (1945).

[133] W. F. Hoffman and R. A. Gortner, *J. Biol. Chem.*, **55**, 371 (1925).

[134] E. G. V. Percival and J. C. Somerville, *J. Chem. Soc.*, 1615 (1937).

[135] E. G. V. Percival and T. H. Soutar, *J. Chem. Soc.*, 1475 (1940); R. B. Duff and E. G. V. Percival, *J. Chem. Soc.*, 830 (1941).

[136] W. G. M. Jones and S. Peat, *J. Chem. Soc.*, 225 (1942).

[137] V. C. Barry and T. Dillon, *Chem. and Ind.*, 167 (1944).

D-galactopyranose residues connected by 1,3-glycosidic linkages, probably of the β-type. The chain is terminated at the reducing end by 1 L-galactopyranose residue which is esterified with sulfuric acid at carbon-6 and attached to the rest of the chain through carbon-4. There are at least 140 galactose units per non-reducing end group and probably up to 53 galactose units per $-SO_3H$ group.

Carrageenin, or Irish moss mucilage, is the name of a polysaccharide obtained from Carrageen or Irish moss, *Chondrus crispus*, which grows along the Atlantic coast.[138] It is presumed to occur as a mixture of sodium, potassium, and calcium salts. Carrageenin resembles agar in its properties and structure. However, it has a higher ash content (about 20 per cent) and requires concentrations of 3 per cent or more to form firm gels which "melt" at lower temperatures than agar gels. Like agar it is a sulfuric acid ester containing D-galactopyranose units joined by 1,3-glycosidic linkages, which appear to be of the α-type. Unlike agar, the sulfuric ester group is attached to carbon-4.[139,140]

Fucoidin is obtained from *Laminaria* and *Fucus* and is apparently a sulfuric acid ester of a polysaccharide which on hydrolysis gives rise to pentoses and methyl pentoses, particularly fucose.

Water-Soluble Reserve Carbohydrates. *Laminarin* occurs as a reserve carbohydrate in the kelp *Laminaria cloustoni* and consists exclusively of glucose residues which are combined by 1,3-glycosidic linkages; as it is levorotatory the linkages probably have the β-configuration.[141] By hydrolysis of laminarin with oxalic acid or an enzyme present in snail juice, a new disaccharide named laminaribiose has been prepared; it is probably glucose-3-β-glucoside.[142] An insoluble polyglucose with β-1,3-glycosidic linkages has also been obtained from yeast, *Saccharomyces cerevisiae*.[143] Laminarin and this yeast polyglucose are of special interest since they are the only polysaccharides known in which glucose units are joined by 1,3-glycosidic linkages. However, this linkage is the common one in the polysaccharides of plant gums which contain galactose residues.

Alkali-Soluble Polyuronides from Seaweed. *Algin* is a polysaccharide which is obtained by alkaline extraction of various species of kelps, such as horsetail kelp (*Saminaria digitata*), broadleaf kelp (*Lam-*

[138] *Chondrus* and *Gigartina stellata* (*G. mamillosa*) are commonly harvested together, and the name carrageenin applies to the polysaccharides from both.

[139] P. Haas and T. G. Hill, *Ann. Applied Biol.*, **7**, 352 (1921).

[140] J. Buchanan, E. E. Percival, and E. G. V. Percival, *J. Chem. Soc.*, 51 (1943).

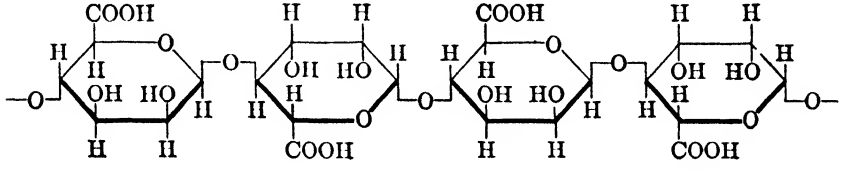
[141] V. C. Barry, *Sci. Proc. Roy. Dublin Soc.*, **20**, 59 (1939).

[142] V. C. Barry, *Sci. Proc. Roy. Dublin Soc.*, **22**, 423 (1941).

[143] W. Z. Hassid, M. A. Joslyn, and R. M. McCready, *J. Am. Chem. Soc.*, **63**, 295 (1941).

inaria saccharina), and giant kelp (*Macrocystis pyrifera*). As used commercially, the term applies to sodium alginate and, less frequently, to ammonium alginate. In nature, algin probably occurs as a mixture of the free acid and its sodium, potassium, calcium, and magnesium salts.

Alginic acid, $(C_6H_8O_6)_n$, is composed entirely of D-mannuronic acid residues, which yield 2,3-dimethyl-D-mannuronide on methylation and hydrolysis; hence the ring and bridge oxygens involve carbon-4 and -5, and the carboxyl groups are all free to react.[144,145] Although a definite decision cannot be made between pyranose and furanose ring structure, the former is more likely because of the extreme resistance of alginic acid to hydrolysis and its large negative rotation. The molecule may, therefore, be represented as containing β-1,4-linkages.



Alginic acid

It is noteworthy that alginic acid, as well as pectic acid, exhibits the geometric pattern of cellulose rather than of starch.

POLYSACCHARIDES OF MICROORGANISMS

Bacterial Polysaccharides.[146] Several bacteria elaborate polysaccharides, the formation of which may be endocellular, exocellular, or capsular. Certain of these polysaccharides are responsible for the immunizing powers of many bacteria, the structure of the polysaccharide determining the specific immunological response to the organisms. Among the bacterial polysaccharides are cellulose, dextrans, levans, and the specific polysaccharides of pathogenic bacteria.

Bacterial Cellulose. Cellulose is secreted in the form of tough, extremely hygroscopic membranes by cultures of the genus *Acetobacter* and other organisms when grown in nutrient solutions containing a suitable carbohydrate substrate (*e.g.*, D-fructose, D-glucose, D-galactose, and D-mannose) or polyalcohols (*e.g.*, D-mannitol, glycerol). Physical and

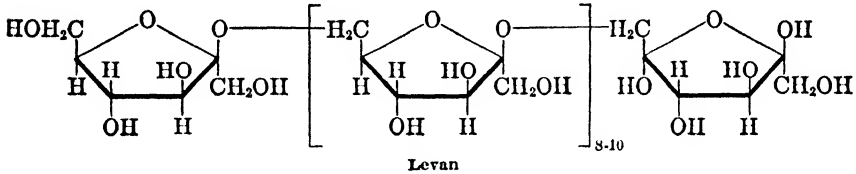
[144] V. C. Barry and T. Dillon, *Sci. Proc. Roy. Dublin Soc.*, **21**, 285 (1936).

[145] E. L. Hirst, J. K. N. Jones, and W. O. Jones, *J. Chem. Soc.*, 1880 (1939).

[146] A comprehensive review of the polysaccharides of microorganisms by T. H. Evans and H. Hibbert will be found in *Advances in Carbohydrate Chem.*, **2**, 203 (1946).

chemical studies [147-149] have established that bacterial cellulose is identical with native plant cellulose, and bacteria were used as a source of cellulose in Germany during World War I. More recently, products similar to parchment and mercerized cotton have been prepared from it.

Bacterial Levans. Levans are produced by the action of microorganisms, such as *Bacillus mesentericus, B. vulgatus, B. subtilis, B. megatherium, Phytomonas pruni, P. prunicola, Streptococcus salivarius,* and Gram-negative milk and soil actinomycetes, on sucrose and raffinose solutions.[150,151] Tests with several of these microorganisms reveal that other sugars, including D-fructose, will not serve as suitable substances for levan formation. Although the molecular size of the levans produced by different microorganisms varies as indicated by sedimentation, ultramicroscopic, and other studies, the fructofuranose residues are invariably joined through carbon-2 and carbon-6.[152-154]



Levan

Because of the ease with which levan formation takes place, it may prove to be a valuable test for the identification of bacterial species.[155]

Recently, the enzymic synthesis of levan (and also dextran, to be discussed later) has been accomplished. Hestrin and his co-workers [156] have produced levan from sucrose and raffinose by means of bacteria-free enzyme preparations of levansucrase from *B. subtilis, B. polymyxa,* and *Aerobacter levanicum.* The amount of sucrose consumed can be entirely accounted for as levan and reducing sugars of which fructose comprised a part; no appreciable interconversion of aldose and ketose occurred. With raffinose as the substrate, the products were levan,

[147] J. Eggert and F. Luft, *Z. physik. Chem.*, **B7**, 468 (1930).
[148] H. Hibbert and J. Barsha, *Can. J. Research*, **5**, 580 (1931).
[149] E. Franz and E. Schiebold, *J. makromol. Chem.*, **1**, 4 (1943).
[150] W. L. Owen, *J. Bact.*, **8**, 421 (1923).
[151] F. C. Harrison, H. L. A. Tarr, and H. Hibbert, *Can. J. Research*, **3**, 449 (1930).
[152] H. Hibbert and F. Brauns, *Can. J. Research*, **4**, 596 (1931).
[153] S. W. Challinor, W. N. Haworth, and E. L. Hirst, *J. Chem. Soc.*, 676 (1934).
[154] R. R. Lyne, S. Peat, and M. Stacey, *J. Chem. Soc.*, 237 (1940).
[155] E. A. Cooper, *J. Soc. Chem. Ind. London*, **58**, 229 (1939); E. A. Cooper and J. F. Preston, *Biochem. J.*, **29**, 2267 (1935).
[156] S. Hestrin, S. Avineri-Shapiro, and M. Aschner, *Biochem. J.*, **37**, 450 (1943); S. Hestrin and S. Avineri-Shapiro, *ibid.*, **38**, 2 (1944); S. Avineri-Shapiro and S. Hestrin, *ibid.*, **39**, 167 (1945).

melibiose, and D-fructose. The following equations represent the net results of the transformations:

$$(n + m)\text{Sucrose} + (n)H_2O \rightarrow (n + m)\text{D-glucose} + (n)\text{D-fructose} + \text{levan}$$

$$(n + m)\text{Raffinose} + (n)H_2O \rightarrow (n + m)\text{melibiose} + (n)\text{D-fructose} + \text{levan}$$

where the levan represents the polymer from m D-fructofuranose residues of the substrate. About 1 mole of substrate was hydrolyzed for each mole converted to levan and aldose, *i.e.*, the ratio of m/n was approximately 1.0. Since the phosphate ion is not required for the synthesis of levan (or dextran) it comprises a mechanism other than phosphorylation for the synthesis of polysaccharides in nature.

Dextrans. The dextrans are polysaccharide slimes which are synthesized by certain strains of chain-forming cocci, such as *Leuconostoc mesenteroides* (*Betacoccus arabinosaceus*) and *L. dextranicum*, which ferment sucrose.[157,158] At times these slimes have been a serious problem in the wine and beet sugar industries. Their presence in sugar sirups greatly increases the viscosity, retarding filtration and crystallization. Sulfur dioxide has been used to inhibit bacterial growth, and dextran may be removed from sugar solutions with lime, but traces of dextrans are said to exist in most sucrose preparations, even in the C.P. reagent grades.

The dextrans are strongly dextrorotatory. Structural studies indicate that the dextrans elaborated by various strains of *L. mesenteroides* and of *L. dextranicum* differ in chemical structure; some are linear and others are branched, although the D-glucopyranose residues are joined principally by 1,6-linkages.[158-162]

The dextrans have antigenic properties. For example, precipitin reactions have been obtained with anti-*Leuconostoc* sera prepared with the homologous organisms; the antigenic activity of the dextran was not dependent on the presence of traces of nitrogenous impurities, and it was therefore concluded that the dextran was a hapten.[163]

Hehre and Sugg [164] have described the enzymic synthesis of a dextran from sucrose by cell-free extracts of a culture of *L. mesenteroides* which

[157] H. L. A. Tarr and H. Hibbert, *Can. J. Research*, **5**, 414 (1931).

[158] F. L. Fowler, I. K. Buckland, F. Brauns, and H. Hibbert, *Can. J. Research*, **B15**, 486 (1937).

[159] W. Z. Hassid and H. A. Barker, *J. Biol. Chem.*, **134**, 163 (1940).

[160] I. Levi, W. L. Hawkins, and H. Hibbert, *J. Am. Chem. Soc.*, **64**, 1959 (1942).

[161] E. C. Fairhead, M. J. Hunter, and H. Hibbert, *Can. J. Research*, **B16**, 151 (1938).

[162] S. Peat, E. Schlüchterer, and M. Stacey, *J. Chem. Soc.*, 581 (1939).

[163] T. H. Evans, W. L. Hawkins, and H. Hibbert, *J. Exptl. Med.*, **74**, 511 (1941).

[164] E. J. Hehre and J. Y. Sugg, *J. Exptl. Med.*, **75**, 339 (1942).

is serologically similar to that formed by the living organism. The reaction proceeds without ordinary inversion according to the overall equation

$$(m)\text{Sucrose} \longrightarrow \text{dextran (D-glucose} - H_2O)_m + (m)\text{D-fructose}$$

Although *L. mesenteroides* can phosphorolyze sucrose to produce D-glucose-1-phosphate in the presence of inorganic phosphate,[165] Hehre [166] has shown that this phosphorylated sugar is not an intermediate in the formation of dextran from sucrose.

Polysaccharides of Pathogenic Bacteria. Serologists had ascribed the specific serological properties of bacterial cultures entirely to proteins and complex lipids until the researches of Heidelberger, Avery, Goebel, and co-workers [167] showed that the capsular carbohydrates of *Pneumococcus* were responsible for immunological type specificity and that they could function both as haptens and as antigens.[168] In investigating the polysaccharide of type III *Pneumococcus*, they isolated a colloidal, strongly acidic carbohydrate with an apparent micellar weight lying between 1,000 and 5,600, which was composed largely of an aldobionic acid ($C_{11}H_{19}O_{10}\cdot COOH$) consisting of D-glucose and D-glucuronic acid, probably linked through carbon-6 of the glucose residue to carbon-1 of the glucuronic acid. After this initial discovery, many workers became active in this field, and specific polysaccharides are now known to be elaborated by many types of bacteria. These bacterial polysaccharides occur naturally in combination with protein or polypeptide residues, and the amount and nature of the protein constituent appears to determine whether the polysaccharide functions only as a hapten or also has antigenic properties.

The polysaccharides of *Pneumococcus* types I–XXXII are all optically active, and the products of acid hydrolysis include D-glucose, hexuronic acids or aldobionic acids, and amino sugars; some contain acetyl groups which are apparently present as acetylated amino groups.[169] Those which contain nitrogen lose their specificity when the nitrogen is removed by treatment with nitrous acid. Type I polysaccharide, for example, yields 28 per cent of D-galacturonic acid and an amino sugar derivative containing acetyl groups, whereas that of type II yields 70

[165] B. O. Kagan, S. N. Lyatker, and E. M. Tsvasman, *Biokhimiya*, **7**, 93 (1942); *Chem. Abs.*, **37**, 4760 (1943).

[166] E. J. Hehre, *Proc. Soc. Exptl. Biol. Med.*, **54**, 240 (1943).

[167] M. Heidelberger and O. T. Avery, *J. Exptl. Med.*, **38**, 73 (1923); **40**, 301 (1924); M. Heidelberger, W. F. Goebel, and O. T. Avery, *ibid.*, **42**, 727 (1925).

[168] O. T. Avery and W. F. Goebel, *J. Exptl. Med.*, **58**, 731 (1933).

[169] R. Brown, *J. Immunol.*, **37**, 445 (1939).

per cent D-glucose.[170] Type III polysaccharide yields cellobiuronic acid upon hydrolysis [171] in which the repeating cellobiuronic acid units are linked through carbon-3 of the D-glucuronic acid residue, most probably by β-linkages.[172] This polysaccharide thus contains alternate 1,3- and 1,4-β-linkages:

$$\overset{4}{\text{—G—}}\overset{1,3}{\text{—U—}}\overset{1,4}{\text{—G—}}\overset{1,3}{\text{—U—}}\overset{1}{\text{—}}$$

where G and U represent D-glucose and D-glucuronic acid, respectively.

The polysaccharides of other pathogenic bacteria are usually not capsular, and their chemical structures are unknown. They usually have more than one simple sugar component, among which L-rhamnose, D-arabinose, D-glucose, D-mannose, D-galactose, and acetylamino sugars have been isolated from the hydrolytic products of different polysaccharides. For example, the polysaccharide of the tubercle bacillus, *Mycobacterium tuberculosis*, strain H-37, yields D-mannose and D-arabinose; [173] that of *Corynebacterium diptheriae* yields D-galactose, pentoses, and amino sugars; [174] that of *Clostridium perfringens* yields principally D-glucose and D-mannose; [175] and those of *Eberthella typhosa* and *Salmonella typhimurium* yield D-glucose, D-mannose, and D-galactose.[176] A water-soluble polysaccharide with immunological properties produced by the fungus *Coccidioides immitis*, the causal agent of coccidioidomycosis in man and certain animals, has been found to contain D-galacturonic acid, D-glucose, and an unidentified sugar in an approximate ratio of 1:6:3, respectively.[177]

Polysaccharides of Molds and Yeasts. Several polysaccharides elaborated by molds and yeasts have been studied. Thus, different species of *Penicillium* produce the polysaccharides *varianose*, *mannocarolose*, *galactocarolose* and *luteose*, respectively, from D-glucose. Varianose consists of 6 to 8 β-D-galactopyranose residues joined by 1,4-linkages with a D-glucopyranose residue at the non-reducing end and either L-altrose or D-idose at the other end.[178] Mannocarolose is a mannan,[179] galactocarolose consists of D-galactofuranose residues which

[170] M. Heidelberger, W. F. Goebel, and O. T. Avery, *J. Exptl. Med.*, **42**, 701 (1925).
[171] R. D. Hotchkiss and W. F. Goebel, *J. Biol. Chem.*, **121**, 195 (1937).
[172] R. E. Reeves and W. F. Goebel, *J. Biol. Chem.*, **139**, 511 (1941).
[173] M. Heidelberger and A. E. O. Menzel, *J. Biol. Chem.*, **118**, 79 (1937).
[174] E. Chargaff, *Rept. Proc. 3rd Intern. Congr. Microbiol.*, 223 (1939).
[175] M. H. Svec and E. McCoy, *J. Bact.*, **48**, 31 (1944).
[176] G. G. Freeman and T. H. Anderson, *Biochem. J.*, **35**, 564 (1941); G. G. Freeman, *ibid.*, **36**, 340 (1942); **37**, 601 (1943).
[177] W. Z. Hassid, E. E. Baker, and R. M. McCready, *J. Biol. Chem.*, **149**, 303 (1943).
[178] W. N. Haworth, H. Raistrick, and M. Stacey, *Biochem. J.*, **29**, 2668 (1935).
[179] W. N. Haworth, H. Raistrick, and M. Stacey, *Biochem. J.*, **29**, 612 (1935).

are apparently linked through carbon-1 and carbon-5,[180] and luteose is composed of β-D-glucose residues joined by 1,6-linkages.[181]

Yeast produces a mannan, sometimes referred to as yeast gum, which consists entirely of D-mannopyranose residues which are linked in quite a different manner from that in the mannan of the ivory nut. Upon methylation and hydrolysis, Haworth and his co-workers [182] obtained approximately equimolecular amounts of tetramethyl-β-mannose, 3,4-dimethyl-D-mannose, and three trimethyl-D-mannoses (about equal amounts of the 3,4,6- and 2,4,6-derivatives and less than 10 per cent of 2,3,4-trimethyl-D-mannose). The small amount of 2,3,4-trimethyl-mannose is believed to represent the non-reducing group of the central chain. The mannan thus consists of D-mannopyranose residues joined by 1,2-, 1,3- and 1,6-linkages to form a branched chain.

[180] W. N. Haworth, H. Raistrick, and M. Stacey, *Biochem. J.*, **31**, 640 (1937).

[181] C. G. Anderson, W. N. Haworth, H. Raistrick, and M. Stacey, *Biochem. J.*, **33**, 272 (1939).

[182] W. N. Haworth, E. L. Hirst, and F. A. Isherwood, *J. Chem. Soc.*, 784 (1937); W. N. Haworth, R. L. Heath, and S. Peat, *ibid.*, 833 (1941).

CHAPTER 26

Metabolism of Carbohydrates [1]

The study of carbohydrate metabolism has been an intensive and fruitful field of biochemical research, especially during the past two decades. The developments in this field have been intimately associated with concurrent research in enzymology, biological oxidations and reductions, and related fields. Some of these topics are necessarily presented elsewhere in the book, but an attempt is made in this chapter to present an integrated picture of various aspects of carbohydrate metabolism.

An outstanding feature of carbohydrate metabolism which has become increasingly evident is the *similarity of many of the intermediate reactions in animals, plants, and microorganisms.* Thus much of the knowledge of carbohydrate metabolism has broad application. However, since differences in the metabolic pattern are in some instances well established, excessive generalization is to be avoided.

The primary function of carbohydrates in living organisms is to serve as a source of energy for biological functions. Carbohydrates may also have a special role as components of some compounds, such as the ribose and desoxyribose in the nucleic acids (p. 406) and the carbohydrate components of cerebrosides (p. 815), various glycosides (p. 712), and some proteins (p. 402). In addition, carbohydrates may be used to form structural components such as cellulose, chitin, pectin, and hyaluronic acid. Ingested or assimilated carbohydrates may be either metabolized directly or converted to disaccharides or polysaccharides to be metabolized later. That which is metabolized for energy may be converted to partial breakdown products or oxidized completely to carbon dioxide and water, depending principally on the organism and the oxygen supply. The utilization of carbohydrate occurs through the participation of a series of chemical reactions carried out by interaction with specific tissue enzymes and coenzymes. Products derived from carbohydrates enter the general metabolic pool, and the organism may use them for

[1] This chapter contributed by Dr. Paul D. Boyer, Associate Professor of Agricultural Biochemistry, University of Minnesota.

the formation of a variety of substances, especially lipids and amino acids. Energy may be derived from carbohydrate with or without oxygen utilization. The breakdown of glucose without oxygen utilization, commonly called *anaerobic breakdown, glycolysis,* or *fermentation* of glucose, liberates only a small part of the energy of the glucose molecule. The complete metabolism of glucose in the presence of oxygen, termed *aerobic breakdown,* liberates the bulk of the energy of the glucose molecule. Anaerobic and aerobic metabolism of glucose do not involve separate pathways. Under anaerobic conditions, pyruvate produced by glycolysis of glucose or glycogen is reduced to lactate. Under aerobic conditions the pyruvate is oxidized to CO_2 and water. The glycolytic process occurs concurrently with the aerobic phase, but usually at a much diminished rate.

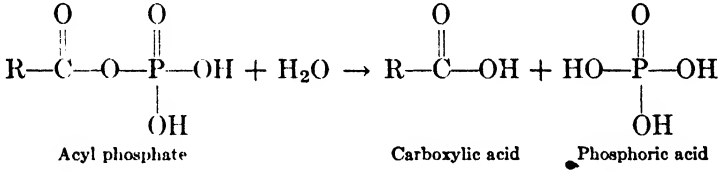
The anaerobic breakdown of glucose is, at present, much better understood than the subsequent oxidative processes. The elucidation of the various steps of the glycolytic process and the mechanism by which energy is derived from this process is indeed a significant achievement of biochemistry. This does not mean that our knowledge is complete, for there is yet much to learn. The facts concerning glycolysis have come mainly from two sources, the study of yeast fermentation and of glucose utilization by muscle tissue. The biochemical studies of yeast fermentation were greatly facilitated by the classic experiments of Buchner about 1897, which showed that fermentation could be obtained with cell-free preparations from yeast. Subsequent studies have been made largely with extracts of cells, although ample evidence has been obtained that many of the reactions found are applicable to the intact cell or organism.

Phosphate Bond Energy in Metabolism. The energy of the glucose molecule is not released in the animal body by the complete oxidation of a molecule at one time, but rather by smaller steps through a series of reactions. *The one known mechanism for the utilization of energy liberated by the metabolism of foodstuffs is through the formation of labile high-energy phosphate bonds.*[2] The term *high-energy* has been applied to such bonds to indicate that they release a relatively large amount of energy when they are broken, and not that there is a strong bonding energy between the phosphate and the group to which it is attached. That the bond is a labile one is indicated by the spontaneous or rapid hydrolysis of many compounds containing such bonds. The free energy

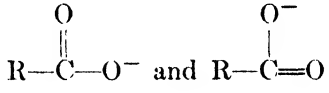
[2] F. Lipmann, *Advances in Enzymol.*, **1**, 99 (1941).

liberated by the hydrolysis of a high-energy phosphate bond has been estimated as about 12,000 calories per mole as contrasted to about 2,000 calories from an ordinary ester phosphate bond.[2]

Four types of high-energy phosphate linkages are known in biological systems. These are bonds between phosphate and (*a*) carboxyl groups, (*b*) other phosphate groups, (*c*) amino groups of guanidine derivatives, and (*d*) the enolic group of enol-phosphopyruvic acid. Other high-energy phosphate linkages may exist. The structural basis for the properties of these labile phosphates has been discussed by Kalckar.[3] The hydrolysis of a high-energy phosphate bond results in the formation of two relatively stable substances from an unstable compound with consequent liberation of a large amount of free energy. Carboxyl or acyl phosphates, as occur in 1,3-diphosphoglyceric acid and acetyl phosphate, and pyrophosphates, as occur in adenosine di- and triphosphate, are acid anhydrides. The reaction



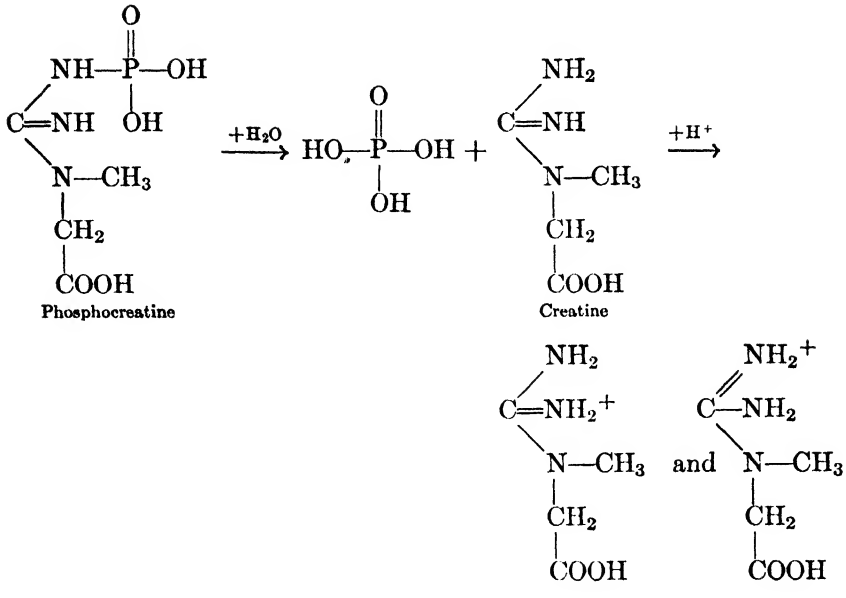
gives rise to the resonance of the carboxylate ion, as indicated by the structures



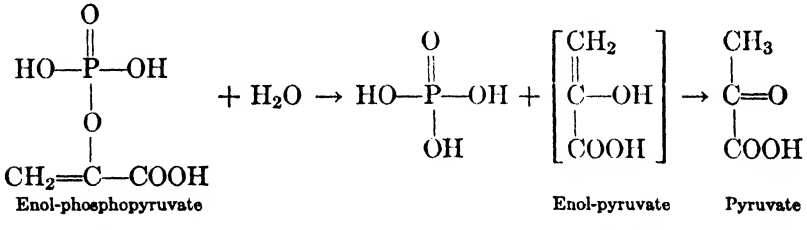
When two of the hydroxyls of phosphate are in combination as in adenosine diphosphate (ADP) the resonance of the phosphate is considerably decreased. This possibility of the formation of more stable resonating structures makes the phosphate bond labile. The guanidine phosphate linkage of creatine or arginine phosphate is analogous to the acyl phosphate, its hydrolysis giving rise to the stable resonating guanidinium ion as follows:[4]

[3] H. M. Kalckar, *Chem. Revs.*, **28**, 71 (1941); *Biol. Revs. Cambridge Phil. Soc.*, **17**, 28 (1942); *Ann. N. Y. Acad. Sci.*, **45**, 395 (1944).

[4] In these and subsequent formulas of this chapter, acidic compounds are shown for convenience as undissociated acids, although it will be recognized that many of the compounds occur as anions at physiological *p*H.



The hydrolysis of enol-phosphopyruvate allows the formation of the more stable pyruvate from the enol form.



The mechanism by which the energy-rich phosphate compounds are formed in glycolysis is fairly well understood, but the manner of their formation during the breakdown of glucose to carbon dioxide and water is not at all clear. It is known that a large portion of the energy from the oxidative steps is not released as heat but is stored in phosphate linkages. The energy-rich phosphate groups formed by various processes may be transferred in the presence of the requisite transphosphorylases to the *adenylic system, i.e.*, the compounds adenosine mono-, di-, and triphosphate, which are interconvertible through the addition or removal of energy-rich phosphate groups. The adenosine polyphosphates apparently serve as a primary energy source for cellular functions. For example, the contraction of muscle has been shown to be intimately related to the breakdown of adenosine triphosphate. The enzyme *adeno-*

sine triphosphatase of muscle has been demonstrated to be linked with the contractile protein *myosin*.[5] The adenosine polyphosphates in muscle are in equilibrium with creatine phosphate, by the reversible reaction ATP + creatine $\rightleftharpoons$ ADP + phosphocreatine.[6] The function which has been tentatively assigned to phosphocreatine in muscle is to serve as a store for high-energy phosphate. However, the manner in which phosphate bond energy is used for metabolic processes and syntheses is poorly understood.

Whether living cells have mechanisms other than the formation of phosphate bonds for the utilization of the energy of foodstuffs is an open question. Nearly all the energy liberated during the anaerobic breakdown of glucose to lactic acid may be accounted for in the formation of energy-rich phosphate bonds; thus phosphate bond energy may be regarded as the only form in which energy is made available through glycolysis. The fact that the yeast cell may carry out the metabolic processes necessary for growth and cell division, utilizing only the energy from anaerobic glycolysis, indicates that the energy of the phosphate bond may be used for the many and varied reactions requisite for maintenance, growth, and reproduction.

Digestion and Absorption of Carbohydrates in Animals. The simple hexoses when ingested are absorbed as such, but the di- and polysaccharides are hydrolyzed in the digestive tract before absorption. The hydrolysis of starch usually begins in the mouth through the action of salivary amylase (ptyalin). Salivary amylase catalyzes the hydrolysis of starch, glycogen, and dextrins into simpler molecules, the ultimate end product being maltose. The amylase ceases to act when the ingested food is acidified through mixing with the gastric juice, but thorough mixing of the stomach contents may be somewhat delayed after ingestion and the action of the salivary amylase thus prolonged. Although the amount of carbohydrate breakdown by salivary amylase may be appreciable its action is not essential for carbohydrate digestion.

The pancreas secretes a powerful amylase which acts similarly to salivary amylase. It is responsible for the breakdown of most ingested starches. Intestinal juice contains carbohydrases, including sucrase, maltase, and lactase, which convert the corresponding disaccharides to monosaccharides. Nearly all the carbohydrate which is digested is absorbed as monosaccharides.

There is evidence that hexoses are *phosphorylated* in the intestinal wall when they are absorbed. This accounts in part for the more rapid

[5] A. Szent-Györgyi, *Chemistry of Muscular Contraction*, Academic Press, New York, 1947.

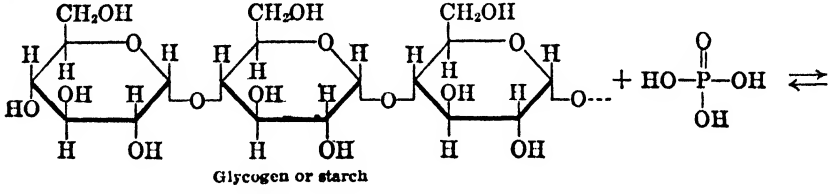
[6] ATP = adenosine triphosphate; ADP = adenosine diphosphate.

absorption of hexoses than of pentoses. Also this active mechanism of absorption, as contrasted to passive diffusion through a membrane, accounts for the absorption of glucose from an intestinal concentration which may be lower than that of blood.

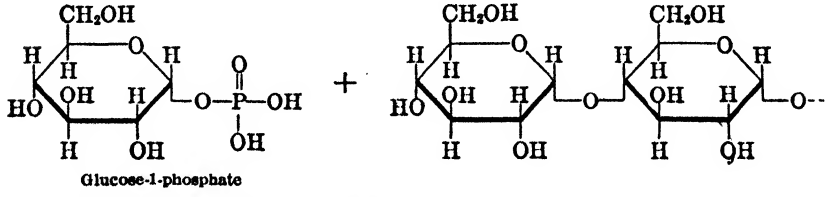
Enzymes capable of digesting cellulose are not secreted by higher animals. However, some animals are able to make use of a considerable portion of ingested cellulose through the action of microorganisms, both bacteria and protozoa, in their digestive tracts. This is particularly striking in the ruminants, where the ingested cellulose is broken down by microorganisms in the rumen. In the horse the cecum has a similar function.

The Conversion of Glucose to Glycogen and Starch. Well over fifty years ago the observations of the French physiologist, Claude Bernard, had established the interconversion of glycogen and glucose in the liver. The mechanism of this reaction long remained obscure, although it was assumed by many to be a simple hydrolytic cleavage. In 1937 Parnas showed that in animal tissues glycogen may undergo phosphorolysis to form a glucose monophosphate.[7] Comparatively recently other investigations, particularly those of the Coris, demonstrated that glucose-1-phosphate is an intermediate in the conversion of starch or of glycogen to glucose. The compound glucose-1-phosphate is frequently called the Cori ester after its discoverer. However, with this and other similarly named phosphate esters of carbohydrate metabolism the use of the chemical name is preferred to avoid confusion.

The breakdown of glycogen by animal tissues involves an addition of *phosphate*, not of water, to the glucosidic linkage, so that the C—O—C bond of the 1-4-glucosidic chain is replaced by the C—O—P bond of glucose-1-phosphate.



Glycogen or starch



Glucose-1-phosphate

[7] J. K. Parnas, *Ergeb. Enzymforschung*, 6, 57 (1937).

This reaction is *readily reversible*—an important property as it provides an explanation for the synthesis of glycogen as well as its breakdown. Cori and co-workers, in classical experiments using a purified muscle phosphorylase,[8] have demonstrated this reversibility. At physiological *p*H the equilibrium for glycogen-glucose interconversion is at about 77 per cent glycogen and 23 per cent glucose-1-phosphate. From the position of the equilibrium it is evident that the free energy change in the reaction is small. Thus the ester link in glucose-1-phosphate and the glucosidic linkage in the large carbohydrate molecule may be considered to have about the same energy content. The same mechanism functions in the breakdown and synthesis of starch by plants. For example, Hanes has demonstrated the synthesis of starch from glucose-1-phosphate by a phosphorylase from potato.[9] A recent symposium on the formation of disaccharides, polysaccharides, and nucleosides may be consulted for further reference.[10]

The synthesis of either starch or glycogen by isolated phosphorylases does not proceed unless a small amount of polysaccharide is added initially. The degree of activation by the added polysaccharide depends to a large extent on the number of terminal glucose units present in the substance added. The observed facts support the hypothesis that the synthesis consists of the addition of glucose-1-phosphate molecules to the end groups of the added polysaccharides to form longer chains. A branched polysaccharide with a relatively high number of terminal groups, such as glycogen or amylopectin, is a much more effective activator than amylose or the linear fraction of starch. In the presence of excess glucose-1-phosphate the lengthening of the side chain may continue until chain length itself becomes a factor limiting the rate. When large chains are formed they may be broken off to form separate units. The polysaccharides synthesized by purified muscle and potato phosphorylase are apparently linear polymers containing only 1,4-glucosidic linkages. However, impure phosphorylase preparations from potato, liver, yeast, and other sources have been found to give rise to a branched type of polysaccharide. This observation is considered to be due to the presence of a second enzyme which is able to bring about the formation of 1,6-glucosidic linkages and thus start branches which may be lengthened by the phosphorylase. The relative activity of the two enzymes may determine the nature of the polysaccharide formed (see p. 643).

The reversibility of the cleavage of glycogen or starch to glucose-1-phosphate is in sharp contrast to that found when glycogen or starch is

[8] G. T. Cori and C. F. Cori, *J. Biol. Chem.*, **148**, 117 (1943).

[9] C. S. Hanes, *Proc. Roy. Soc. (London)*, **B129**, 174 (1940).

[10] *Federation Proc.*, **4**, 226 (1945).

hydrolyzed by amylase and maltase to maltose and glucose. The products of the latter reactions, even when added to the enzymes in large amounts, are not polymerized to polysaccharides. The phosphorylation of glucose thus provides a mechanism for the reversibility of the reaction and, in addition, is essential for the metabolic breakdown of glucose.

Some Relationships of Glucose and Glycogen in the Animal Body. Glycogen occurs in animals principally in the liver and muscle. Other tissues contain relatively small amounts. The normal level of muscle glycogen is about 0.7 per cent, representing about 250 grams of muscle glycogen in an average man. The liver may contain about 6 per cent glycogen, or about 110 grams. The total of about 350 grams of glycogen normally present accounts for nearly all the carbohydrate stores of the body and represents about 1,500 Calories or roughly one-half of the total daily energy requirement.

The muscle glycogen level remains fairly constant; it is not decreased extensively by ordinary muscular contraction, but is considerably lowered by exhaustive contraction. Muscle glycogen is not reconvertible to blood sugar in significant amounts since the phosphatase for the conversion of glucose-6-phosphate to glucose and phosphate is apparently lacking in muscle. On the other hand, the liver glycogen level may vary quite widely with change in dietary conditions, energy output, and physiological disturbances. Liver glycogen and blood glucose are readily interconvertible by the mechanisms outlined above. This relationship functions to maintain a normal blood sugar level, which, in man, is about 90 to 120 mg. per 100 ml. Liver glycogen may be formed from the glycogenic amino acids, from the glycerol of fats, and from other metabolites. When muscular work is exhaustive, lactic acid formed from glucose breakdown may be liberated into the blood stream and carried to the liver where it is converted to glycogen. The conversion of lactic acid to glycogen during a period of aerobic recovery from anaerobic contraction may also occur. About four-fifths of the lactic acid is reconverted to glycogen while the remaining one-fifth is further oxidized to provide energy for the glycogen synthesis.

The principal mechanism by which various metabolites are converted to glycogen is probably through a reversal of the glycolytic reactions outlined in the following sections. Such a reversal would occur when the requisite metabolites are present in sufficiently large amounts together with an adequate supply of high-energy phosphate. That the liver glycogen is in a dynamic state, being constantly broken down and resynthesized, is indicated by studies using isotopes of carbon and hydro-

gen as tracers. From such studies it has been calculated that the "half life" of liver glycogen is approximately 1 day.

The Breakdown of Glucose to Pyruvate. There are a number of reactions which glucose and its breakdown products have been demonstrated to undergo in various cells. However, the scope of this text does not permit their inclusion here, and the following discussion will be limited principally to what is probably the main pathway of glucose breakdown by animals, plants, and microorganisms.

Glucose must be present as a phosphate ester before it can be converted into pyruvate by the cellular enzymes. Glucose-1-phosphate may arise from the breakdown of glycogen or starch, or glucose-6-phosphate (Robison ester) may be formed by the direct phosphorylation of glucose. The latter reaction involves an intermolecular transfer of phosphate from the terminal high-energy phosphate group of adenosine triphosphate to glucose (see p. 688). The utilization of a high-energy phosphate for this reaction is apparently necessary even though the ester group has a comparatively low energy content. As would be anticipated the reaction is essentially irreversible. However, glucose-6-phosphate may be readily hydrolyzed by *phosphatases* to give glucose and inorganic phosphate. The phosphorylation of glucose by adenosine triphosphate is catalyzed by the enzyme *hexokinase* and has been studied in detail in muscle extracts by Colowick and Kalckar.[11]

Glucose-1-phosphate is readily converted by the enzyme *phosphoglucomutase* to glucose-6-phosphate, a reaction involving the intramolecular migration of phosphate groups. The reaction was first regarded as irreversible, but more recent studies [12] with purified enzyme systems have shown that at equilibrium 94 per cent of added glucose-1-phosphate is converted to glucose-6-phosphate. The same equilibrium may be reached by starting with either glucose-1-phosphate or glucose-6-phosphate; this is convincing evidence for the existence and position of the equilibrium.

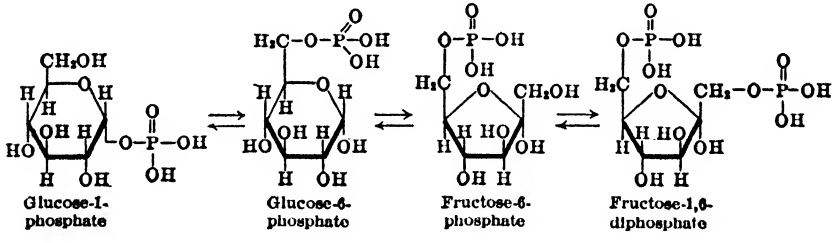
The glucose-6-phosphate in the cell may be reversibly transformed into its isomer fructose-6-phosphate (Neuberg ester) as a result of the action of the enzyme *phosphohexose isomerase*. The interconvertibility of glucose and fructose in animal tissues is mediated through the formation of their phosphate esters. The equilibrium in this reaction shows that there is little difference in the free energy content of the two compounds, as the equilibrium mixture contains about 70 per cent glucose-6-phosphate and 30 per cent fructose-6-phosphate.

[11] S. P. Colowick and H. M. Kalckar, *J. Biol. Chem.*, **116**, 119, 129 (1941).

[12] E. W. Sutherland, S. P. Colowick, and C. F. Cori, *J. Biol. Chem.*, **140**, 309 (1941).

From a consideration of the above equilibria it is evident that a muscle cell which has a relatively high equilibrium concentration of glycogen must contain considerable glucose-6-phosphate and fructose-6-phosphate. This is in accord with the fact that an equilibrium mixture of these hexose monophosphates is a normal constituent of resting muscle.

Fructose-6-phosphate is converted to fructose-1,6-diphosphate (Harden-Young ester) in the presence of ATP, Mg^{++}, and the requisite enzyme *phosphohexokinase*. As in the phosphorylation of glucose, this reaction requires the high-energy phosphate of the adenylic system. Fructose-1,6-diphosphate was isolated as early as 1905 by Harden and Young [13] in their pioneer work on yeast metabolism. Fructose-1,6-diphosphate may under appropriate conditions form fructose-6-phosphate, but in the reverse reaction the phosphate is liberated as inorganic phosphate, the energy content of the phosphate bond of the fructose ester being too small for phosphorylation of ADP. There is evidence that in the intact cell the breakdown of fructose-1,6-diphosphate is not a rapid reaction, and it has been postulated that this reaction may be limiting in the formation of pyruvate.[14] The reactions and structure of the biologically important hexose monophosphates may be summarized as follows:



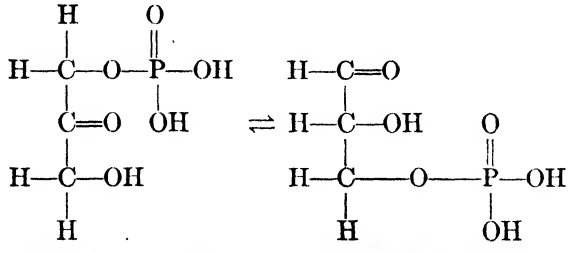
The fructose-1,6-diphosphate formed by the above reactions is further broken down to two triose residues. Meyerhof and co-workers [15] have shown that hexose diphosphate is reversibly split in the presence of *aldolase* into 1-phosphodihydroxyacetone and 3-phosphoglyceraldehyde. Aldolase has the general property of catalyzing the condensation of 1-phosphodihydroxyacetone and various aldehydes in typical aldol condensations to give keto sugars or related compounds. The two phospho-

[13] A. Harden, *Alcoholic Fermentation*, 4th ed., Longmans, Green & Co., New York, 1932.

[14] C. F. Cori, in *Symposium on Respiratory Enzymes*, University of Wisconsin Press, p. 175, 1942.

[15] O. Meyerhof, K. Lohmann, and P. Schuster, *Biochem. Z.*, **286**, 301 (1936).

trioses formed from hexose diphosphate exist in the following equilibrium, their interconversion being catalyzed by *phosphotriose isomerase.*

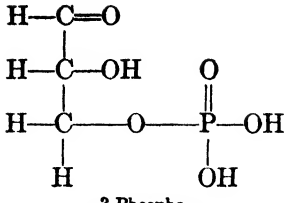


1-Phosphodihydroxyacetone 3-Phosphoglyceraldehyde

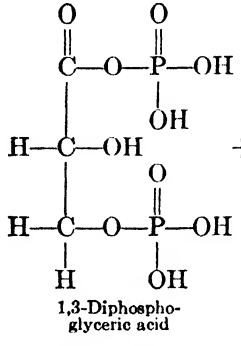
Only the glyceraldehyde phosphate undergoes the oxidative reaction leading to the subsequent steps of glycolysis. Thus the interconversion of the two triose residues provides a mechanism for utilization of both components. The equilibrium of the reaction strongly favors 1-phosphodihydroxyacetone, about 95–96 per cent of this component being present in an equilibrium mixture.[16] The concentration of phosphotriose isomerase in muscle has been estimated to represent about 4 per cent of the protein of muscle extract, and in addition this enzyme has been shown to be exceptionally active.[17] This high concentration and activity have been suggested by Meyerhof as mechanisms developed by cells to deal with the metabolically unfavorable triose phosphate equilibrium.

The oxidation of 3-phosphoglyceraldehyde is a key reaction in the glycolytic process. The elucidation of the chemistry of this step by Warburg and his co-workers was a discovery of prime importance because it represented the first demonstration of how the energy obtained from oxidation of metabolites may be made available for cellular processes. In 1939 Warburg and Christian [18] purified and isolated a *phosphoglyceraldehyde dehydrogenase* which, in the presence of reduced coenzyme I (see p. 995), catalyzed the oxidation of 3-phosphoglyceraldehyde. Neglein and Bromel,[19] in Warburg's laboratory, demonstrated that the product of the reaction was 1,3-diphosphoglyceric acid. The participation of inorganic phosphate in this reaction is obligatory, the overall reaction being:

[16] O. Meyerhof and R. Junowicz-Kocholaty, *J. Biol. Chem.*, **149**, 71 (1943).
[17] O. Meyerhof and L. V. Beck, *J. Biol. Chem.*, **156**, 109 (1944).
[18] O. Warburg and W. Christian, *Biochem. Z.*, **303**, 40 (1939).
[19] E. Neglein and H. Bromel, *Biochem. Z.*, **303**, 132 (1939).

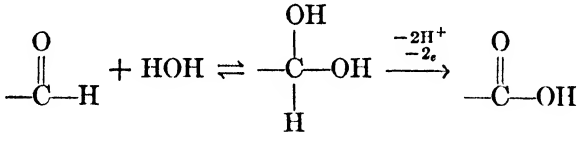


+ coenzyme I + phosphate ⇌

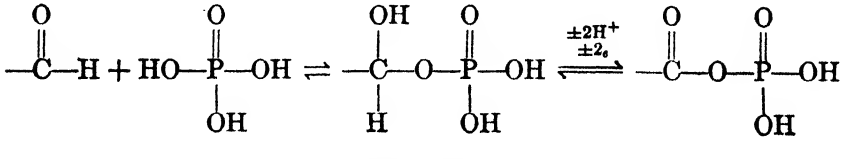


+ reduced coenzyme I

In a detailed study of this reaction with a purified enzyme system Meyerhof and Oesper [20] demonstrated that the requirements of the law of mass action for this equilibrium are fulfilled for every component of the system. This is further evidence for the validity of the reaction mechanism. In this reaction it is probable that phosphate is first added to 3-phosphoglyceraldehyde to give small amounts of 1,3-diphosphoglyceraldehyde; thus phosphate, instead of water, adds to form the electron donor proper. Instead of the reaction



which involves the formation of the stable carboxyl group with a large liberation of free energy, the participation of phosphate

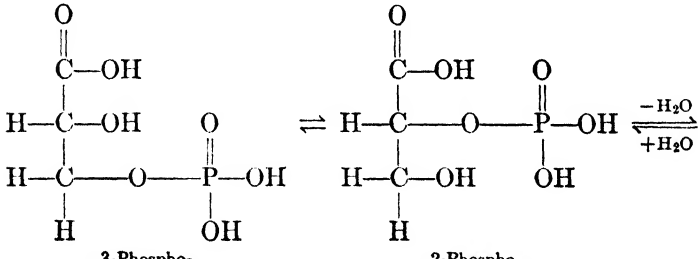


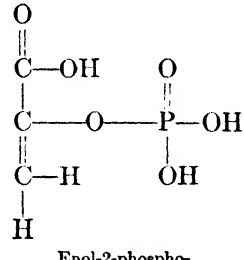
results in the formation of a labile, high-energy phosphate group with little release of free energy. The carboxyl phosphate formed is re-

[20] O. Meyerhof and P. Oesper, *J. Biol. Chem.*, **170**, 1 (1947).

versibly transferable to the adenylic system in the presence of the requisite transphosphorylase. The obligatory participation of phosphate in this oxidation thus makes it possible for the energy of the oxidation to be utilized and provides a mechanism for carrying out the reverse reaction.

The 3-phosphoglyceric acid formed when 1,3-diphosphoglyceric acid loses its high-energy phosphate group may be reversibly converted to 2-phosphoglyceric acid by an intramolecular migration of the phosphate group in the presence of *phosphoglyceromutase*. The 2-phosphoglyceric acid in the presence of *enolase* and Mg^{++} is converted to 2-phosphopyruvic acid. This reaction is inhibited by fluoride ions. In an appropriate muscle extract an equilibrium mixture of these three has been found to consist of 58.5 per cent 3-phosphoglycerate, 12.5 per cent 2-phosphoglycerate, and 29 per cent 2-phosphopyruvate. Their interconversion occurs as follows:





The formation of 2-phosphopyruvic acid is a reaction of considerable importance. The phosphate linkage in 2-phosphoglycerate is of the relatively stable ester type, whereas that in the 2-phosphopyruvate is a labile high-energy phosphate linkage. Although the total energy content of the two compounds in question is about equal, a larger portion of the energy in the 2-phosphopyruvate is concentrated in the phosphate group. The phosphate group of 2-phosphopyruvate can be transferred to the adenylic system, thus making the energy of the compound avail-

able for biological processes. This reaction has been shown to be accelerated by K^+ ions,[21] an observation which led to the demonstration of reversibility of the transfer of phosphate from phosphopyruvate to adenosine diphosphate.[22] This was the last of the reactions involved in the formation of pyruvate from glycogen to be shown to be reversible; it strengthens the possibility that the synthesis of glycogen from pyruvate may occur by a reversal of the glycolytic process.

In most living cells under aerobic conditions the reduced coenzyme I formed in the oxidation of glyceraldehyde may be oxidized by eventual interaction with the cytochrome system and oxygen. Under these conditions the pyruvate formed by glycolysis is further metabolized without being converted into lactate. However, under anaerobic conditions coenzyme I is oxidized by reaction with pyruvate in the presence of *lactic dehydrogenase*, the pyruvate being reduced to lactate. This is the basis of the formation of lactic acid under conditions of stress or oxygen lack.

The reactions outlined above for the conversion of glycogen to lactic acid in anaerobic metabolism and their coupling with the oxidation and reduction of coenzyme I and with the phosphorylation and dephosphorylation of the adenylic system are summarized in Fig. 106. As

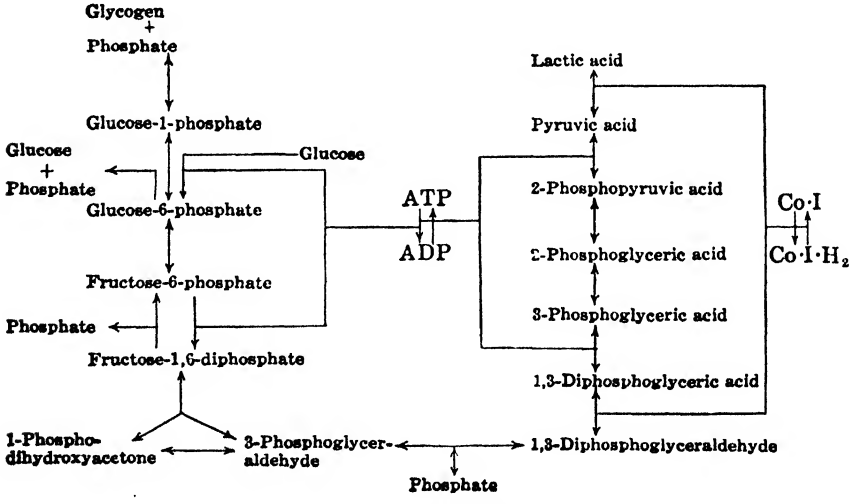


FIG. 106. Schematic presentation of the reactions of anaerobic glycolysis.

illustrated in this figure, the adenylic system has a central role in glycolysis. In the formation of pyruvate from glucose, two high-energy phosphate groups from adenosine triphosphate are required for the forma-

[21] P. D. Boyer, H. A. Lardy, and P. H. Phillips, *J. Biol. Chem.*, **149**, 539 (1943).
[22] H. A. Lardy and J. A. Ziegler, *J. Biol. Chem.*, **159**, 343 (1945).

tion of glucose-6-phosphate and fructose-1,6-diphosphate. For each glucose molecule metabolized four high-energy phosphate groups are made available, two from the oxidative step and two from each of the phosphopyruvate molecules formed. Thus there is a net gain of two high-energy phosphate groups. When glycogen serves as the glucose source there is a net gain of three high-energy phosphate groups. The economy of glycogen as a storage carbohydrate is thus apparent. The above deductions follow from reactions which have been studied in tissue extracts. However, there is definite evidence that in intact muscle each glucose molecule from glycogen, when completely metabolized to two lactic acid molecules, gives rise to a net generation of four high-energy phosphate bonds.[23] The mechanism by which the extra phosphate bond is formed is not established. The free energy change in the formation of lactic acid from one mole of glucose as contained in glycogen has been estimated at −58,000 gram calories. This amounts to less than 10 per cent of the total free energy which may be liberated by the complete oxidation of glucose. Most of the energy originally in the glucose is still present in the lactate molecule. The generation of four phosphate bonds in adenosine triphosphate represents a utilization of about 48,000 calories. Thus roughly 80 per cent of the free energy released in the anaerobic splitting of glycogen to lactic acid may be utilized for adenosine triphosphate synthesis.

The Use of Enzyme Inhibitors in the Study of Metabolism. The use of inhibitors for elucidating the mechanism of certain reactions is well illustrated by studies on the effect of *fluoride* and *iodoacetate* on the glycolytic process. With suitable inhibitors the effects of blocking one reaction while other reactions continue may be evaluated. Iodoacetate is known to inhibit the action of 3-phosphoglyceraldehyde dehydrogenase, whose activity depends on the integrity of the —SH groups of the enzyme protein. Sulfhydryl groups are known to combine with iodoacetate as follows:

$$RSH + CH_2ICOOH \rightarrow RSCH_2COOH + HI$$

The observation by Lundsgaard [24] that muscle poisoned with iodoacetate could contract several times without formation of lactic acid demonstrated that the contraction process and lactic acid formation are not inseparable. Fluoride ions are known to inhibit enolase, which catalyzes the conversion of 2-phosphoglycerate to enol-2-phosphopyruvate. The fluoride ion in the presence of phosphate forms a magnesium-fluorophosphate complex which, by removal of magnesium from com

[23] O. Meyerhof, *Ann. N. Y. Acad. Sci.*, **47**, 815 (1947).
[24] E. Lundsgaard, *Biochem. Z.*, **227**, 51 (1930).

bination with the enzyme, results in the inhibition of the reaction. The use of fluoride as an inhibitor of glycolysis allows isolation and identification of the precursors of phosphopyruvate that accumulate, in particular 3-phosphoglycerate and hexose phosphates.

These enzyme inhibitors, in common with most inhibitors, are not entirely specific. Fluoride will inhibit many enzymes, particularly those requiring Mg^{++}. Iodoacetate and other —SH reagents will, in general, inhibit enzymes which require intact sulfhydryl groups for their activity. The use of inhibitors often depends on the relative sensitivity of the various reactions that may be occurring. Other inhibitors whose use in the study of metabolic reactions has been particularly valuable are cyanide and malonate (see p. 1000).

Relations between Glycolytic and Oxidative Processes. When cells which possess the mechanisms for oxidative metabolism have access to oxygen, the breakdown of glucose is much slower than under anaerobic conditions. This inhibition of glycolysis during aerobic metabolism is commonly referred to as the *Pasteur effect*.[25] It allows the organism to conserve carbohydrate by utilization of the more efficient aerobic process. Although a number of postulates have been made to explain this observation, an accepted understanding of the effect is lacking.

One point of view is that products produced or utilized by the respiratory process bring about an inhibition of the glycolytic reactions. The maintenance of a high concentration of ATP and consequent lower concentration of ADP or adenylic acid by the oxidative reactions has been advanced as an explanation. Glycolysis would thus be inhibited by the lack of the adenylic system as a phosphate acceptor. The reduction of inorganic phosphate to a low level by its uptake during oxidative reactions would likewise inhibit glycolysis, since inorganic phosphate is necessary for the oxidation of 3-phosphoglyceraldehyde. These or similar possible explanations based on a dynamic balance between glycolysis and oxidative processes and the energy-consuming functions of the cell are in harmony with the concept that synthesis of glucose or polysaccharides involves a reversal of the glycolytic reactions. Such a reversal might occur by an increase or continuation of the processes bringing about inhibition. In accord with this concept, when the rate of consumption of high-energy phosphate is low the glycolytic and oxidative processes are held in check even though glucose and oxygen are present. An increased utilization of phosphate bond energy requires that the formation of energy-rich phosphate be likewise increased. If oxidizable substrates and oxygen are present, the oxidative reactions may

[25] F. Lipmann, in *Symposium on Respiratory Enzymes*, University of Wisconsin Press, p. 48, 1942; D. Burk, *Cold Spring Harbor Symposia Quant. Biol.*, **7**, 129 (1939).

provide the energy; if not, the anaerobic mechanisms must be more extensively utilized.

Alternative explanations for the Pasteur effect are based upon an inhibition produced by the action of oxygen independent of the respiratory process; *i.e.*, in the presence of oxygen an enzyme or some other substance necessary for glycolysis is converted to an inactive form, or an inhibitor of glycolysis is produced. These explanations receive support from (*a*) observations that the inhibition of glycolysis by respiration may be overcome by some agents without decrease in the magnitude of the respiration, (*b*) the inhibition of glycolysis by some oxidizing agents and the release of this inhibition by reducing agents, and (*c*) the possible occurrence of the Pasteur effect with a very low level of respiration. The inhibition of some of the enzymes of glycolysis that require intact —SH groups by oxidation of these groups to —S—S— linkages, with subsequent reactivation of the enzymes by glutathione, may be mentioned as one possibility of reversible oxidative inactivation.

The Oxidative Metabolism of Carbohydrates. Although the reactions of the anaerobic conversion of glucose to pyruvate are much better understood than the subsequent oxidative processes, sufficient knowledge is available to give us some understanding of the biological oxidation of carbohydrates. In the following discussion some of the metabolic reactions involved in the breakdown of pyruvate to carbon dioxide and water will be considered.

All animal tissues and many plants and microorganisms can metabolize pyruvate. It may be regarded as a *common intermediate in the catabolic and anabolic reactions of carbohydrate, fats, and proteins.* Pyruvate is a biologically reactive substance which may participate in a number of enzymic reactions involving condensation, decarboxylation, carboxylation, dismutation, oxidation, or reduction.[26] All reactions of pyruvate appear to require diphosphothiamine (cocarboxylase).

As examples of some of the reactions of pyruvate the following are cited. In *acetic fermentations*, as well as in animal tissues, pyruvate may undergo the following dismutation:

$$\underset{\text{Pyruvic acid}}{2CH_3COCOOH} + H_2O \rightarrow \underset{\text{Lactic acid}}{CH_3CHOHCOOH} + \underset{\text{Acetic acid}}{CH_3COOH} + CO_2$$

Pyruvate may be anaerobically decarboxylated by yeast to form acetaldehyde, and by bacteria and animal tissues to form acetoin, $CH_3COCHOHCH_3$, or it may undergo oxidative decarboxylation with the formation of carbon dioxide and acetyl phosphate (see p. 709). In

[26] E. Stotz, *Advances in Enzymol.*, **5**, 129 (1945); E. S. G. Barron, *Advances in Enzymol.*, **3**, 149 (1943).

addition pyruvate may give rise to acetate or an active two-carbon compound related to acetate in a reaction not yet completely understood. An important reaction of pyruvate, discussed elsewhere (p. 697), is its addition of carbon dioxide to form oxalacetate. Also pyruvate may participate in transamination reactions (see p. 455) with the formation of alanine. The reaction by which pyruvate is decarboxylated and coupled with oxalacetate to form aconitate is probably, quantitatively, the most important reaction for the metabolism of pyruvate, and it is discussed more fully below.

Of various mechanisms which have been proposed for the oxidation of breakdown products of glucose to carbon dioxide and water, the series of reactions for the oxidation of pyruvate known as the *tricarboxylic acid cycle* has received the most widespread consideration. This series of reactions is also referred to as the Krebs cycle, or, in its earlier form, as the citric acid cycle. The experimental basis for the tricarboxylic acid cycle has been the subject of a number of recent reviews,[27] and it will be presented only briefly here. The cycle was first proposed as the result of studies on pigeon breast muscle, but with some modification it has been suggested to apply not only to muscle but also to other animal tissues, to yeasts and other microorganisms, and possibly to plants.

The reactions comprising the cycle are shown in Fig. 107. Pyruvic acid undergoes oxidative decarboxylation and coupling with oxalacetic acid to form the tricarboxylic acid *cis-aconitic acid*, which through addition of water reversibly forms *isocitric acid*. The isocitric acid is oxidized to *oxalsuccinic acid*, which is subsequently decarboxylated to form *α-ketoglutaric acid*. This undergoes another decarboxylation to form *succinic acid* and carbon dioxide. The succinic acid is oxidized to *fumaric acid*, which, by addition of water, is converted to *malic acid*. Malic acid is then oxidized to form oxalacetic acid, which may again combine with pyruvic acid for repetition of the cycle. The net effect of the reactions is that pyruvate has been oxidized to carbon dioxide and water according to the equation

$$CH_3COCOOH + 2\tfrac{1}{2}O_2 \rightarrow 3CO_2 + 2H_2O$$

Some of the principal experimental results which serve as a basis for the cycle are:

1. *The stimulation of tissue respiration by small amounts of several compounds which participate in the cycle.* Additions of citrate and of the four-carbon dicarboxylic acids (succinate, fumarate, malate, and oxal-

[27] H. G. Wood, *Physiol. Revs.*, **26**, 198 (1946); H. A. Krebs, *Advances in Enzymol.*, **3**, 191 (1943); E. A. Evans, Jr., *Harvey Lectures Ser.*, **39**, 273 (1943-44).

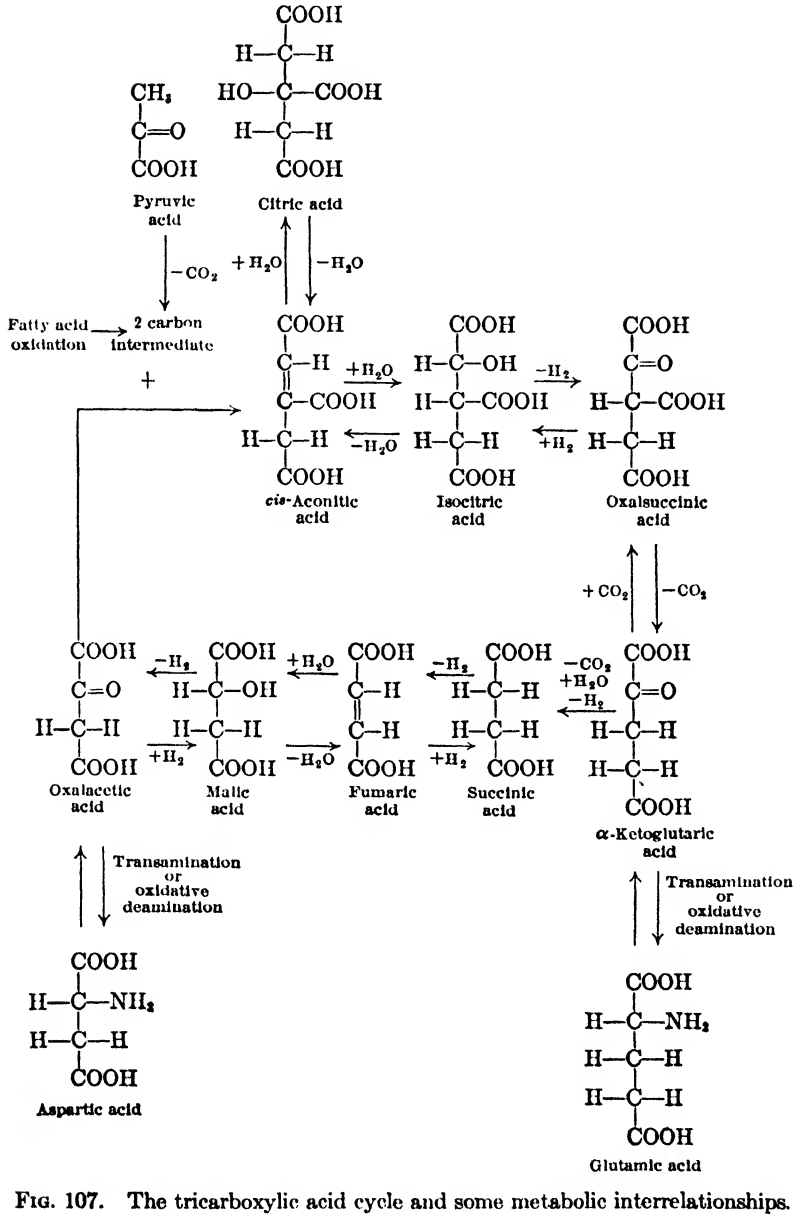


FIG. 107. The tricarboxylic acid cycle and some metabolic interrelationships.

acetate) to pigeon breast muscle in small amounts are known to accelerate greatly the utilization of pyruvate with increased oxygen uptake and carbon dioxide output. Inasmuch as these substances act in small amounts it seems probable that they are regenerated by the reactions involved.

2. *The utilization by tissue preparations of different components of the cycle at rates which will account for the rate of pyruvate utilization.* The reversible enzymic interconvertibility of succinate, fumarate, malate, and oxalacetate and the rapid metabolism of these substances are well established. Isocitrate, *cis*-aconitate, citrate, and α-ketoglutarate are known to be rapidly metabolized by tissue preparations, and the oxidation of these substrates is known to lead to the formation of succinate. The synthesis of citrate and α-ketoglutarate from pyruvate has been demonstrated.

3. *The oxidative formation of succinate from fumarate or oxalacetate in the presence of malonate.* Malonate is known to inhibit competitively succinic dehydrogenase, which catalyzes the interconversion of fumarate and succinate. Addition of malonate will block the reductive formation of succinate from fumarate. Hence there must be a second mechanism of succinate formation from fumarate or oxalacetate, and this formation occurs only under aerobic conditions. The oxidation of pyruvate by minced muscle is also inhibited by malonate and, in the presence of malonate, succinate accumulates. These facts are adequately explained by the proposed tricarboxylic acid cycle.

4. *The agreement between experiments with isotopic tracers and the proposed cycle.* The results of various studies with metabolites labeled with isotopes are in harmony with the existence and prominent function of the cycle. Proposed intermediates of the cycle isolated after utilization of labeled metabolites contain enough of the isotopes to establish the cycle as a prominent metabolic pathway. Also the position of the labeled atoms in the isolated compounds is readily explained by the proposed reactions.

Although the above and related evidence strongly suggests the existence of the proposed metabolic cycle or some closely analogous mechanism, the entire reaction scheme may not be regarded as definitely established. One of the steps which is least understood is the initial condensation of pyruvate and oxalacetate to form a tricarboxylic acid. Whether the pyruvate combines with oxalacetate to form an intermediate or whether the pyruvate is decarboxylated to form an active two-carbon compound which subsequently condenses with oxalacetate is yet to be determined. Evidence that a two-carbon intermediate, derived from either pyruvate or from fat metabolism, may be involved is presented

by Wood.[27] Direct elucidation of the condensation mechanism and of the necessity of the tricarboxylic acids for the oxidation of various two- and three-carbon metabolic intermediates would greatly strengthen the position of the tricarboxylic acid cycle.

The cycle as originally proposed by Krebs contained citrate as one of the intermediates. Experimental evidence for the exclusion of citrate from the main chain of reactions represents an interesting application of tracer technique. α-Ketoglutarate formed by pigeon liver homogenate from pyruvate in the presence of carbon dioxide labeled with heavy carbon has been found to contain labeled carbon in only the carboxyl group of the α-ketoglutarate adjacent to the carbonyl group. Formation of the α-ketoglutarate from citrate would require that the labeled carbon be present equally in both carboxyl groups. Citrate is a symmetrical molecule and would thus be equally susceptible to oxidation at the carbon atoms adjacent to either terminal carboxyl group. However, isocitrate as well as aconitate is not symmetrical, the presence of the hydroxyl group or double bond at the carbon adjacent to one carboxyl apparently being the site of further oxidation. The inclusion of these compounds in the cycle is thus in harmony with the experimental facts. Citrate and isocitrate are enzymically interconvertible; this property accounts in part for the metabolic effects of citrate.

The oxidations in the various steps of the cycle are linked, by mechanisms as yet only partly understood, to the cytochrome system and oxygen. In the oxidation of isocitrate to oxalsuccinate, triphosphopyridine nucleotide (TPN or coenzyme II) is known to transfer hydrogen to a flavoprotein which may subsequently reduce cytochrome *c*. Similarly the oxidation of malic acid is known to be coupled with the reduction of diphosphopyridine nucleotide (DPN or coenzyme I), which is subsequently re-oxidized by cytochrome through the mediation of a flavoprotein enzyme.

It has been amply demonstrated that the oxidation of pyruvate may lead to the formation of energy-rich phosphate. Although the mechanism of such formation is only vaguely understood, some observations in this regard are of sufficient interest to warrant their inclusion here. Ochoa's experiments[28] with heart muscle extracts indicate that the complete oxidation of a glucose molecule to carbon dioxide and water by 6 molecules of oxygen results in the formation of 6 energy-rich phosphate bonds per molecule of oxygen consumed. This would mean that 36 energy-rich phosphate bonds may be formed for each glucose molecule oxidized, which is equivalent to about 430,000 calories, or about 60 per

[28] S. Ochoa, *Ann. N. Y. Acad. Sci.*, **47**, 835 (1947).

cent of the total free energy of 688,000 calories derivable from the complete oxidation of glucose. It has been found that in man muscular work may be performed with an efficiency as high as 20 per cent. If about 60 per cent of the energy of glucose may be converted to phosphate bond energy, the subsequent utilization of this energy for muscular contraction evidently results in a considerable wastage of energy.

The formation of more than one high-energy phosphate for each two hydrogens removed from the substrate would necessitate a mechanism other than that analogous to the oxidation of 3-phosphoglyceraldehyde, which yields only one high-energy phosphate for removal of two hydrogens. The formation of as many as three or more high-energy phosphate groups per atom of oxygen is thermodynamically possible, but the means by which this is accomplished is obscure. The thermodynamic possibility for the formation of several high-energy phosphate groups by the various metabolic steps may be deduced from measured values of the oxidation-reduction potentials of the systems involved.[29]

The Assimilation of Carbon Dioxide. An important phase of research has been the elucidation of some of the mechanisms by which animals, plants, and bacteria may bring about the incorporation of carbon dioxide into organic compounds. Wood and Werkman in 1936 first demonstrated carbon dioxide fixation by heterotrophic bacteria.[30] In such studies the use of isotopes has been of particular value. Thus the fixation of carbon dioxide labeled with heavy or radioactive carbon may be demonstrated and the location of the assimilated carbon in certain compounds made feasible. For example, it has been shown that radioactive carbon dioxide administered to rats as bicarbonate, together with glucose or lactate, gives rise to glycogen containing radioactive carbon.[31] Elucidation and further understanding of the reactions of carbon dioxide fixation may help clarify the problems of photosynthesis for, as stated by Wood,[30] "It is quite probable that the underlying reactions may be similar in the two processes, the essential difference being that the required energy is radiant in photosynthesis and chemical in the case of heterotrophic metabolism."

There are four enzymic reactions in which the primary fixation of carbon dioxide is known to occur. It is very probable that other fixa-

[29] E. G. Ball, *A Symposium on Respiratory Enzymes*, University of Wisconsin Press, p. 16 (1942); *Ann. N. Y. Acad. Sci.*, **45**, 363 (1944).

[30] For recent reviews see H. G. Wood, *Physiol. Revs.*, **26**, 198 (1946), and S. Ochoa, in D. E. Green, *Currents in Biochemical Research*, Interscience Publishers, New York, 1946, p. 165.

[31] J. B. Conant, R. D. Cramer, A. B. Hastings, F. W. Klemperer, A. K. Solomon, and B. Vennesland, *J. Biol. Chem.*, **137**, 557 (1941).

tion reactions will be demonstrated. In animals, plants, and microorganisms the reaction

$$\underset{\text{Pyruvic acid}}{CH_3COCOOH} + CO_2 \rightleftharpoons \underset{\text{Oxalacetic acid}}{COOHCOCH_2COOH}$$

has been demonstrated. In animal tissues the decarboxylation of oxalsuccinate to α-ketoglutarate has been shown to be reversible. This reaction, it will be recalled, is one of the steps in the tricarboxylic acid cycle. The reversibility of the other decarboxylations of this cycle has not yet been demonstrated. A third mechanism has been demonstrated in bacteria, *i.e.*, the reversal of the breakdown of pyruvate to acetate and formate. This has been demonstrated to result in the fixation of carbon dioxide, the carbon dioxide being previously reduced to formate (see p. 709). A fourth fixation reaction is the reversible oxidative decarboxylation of malic acid coupled with coenzyme II as follows:[32]

$$\text{Malate} + TPN_{ox.} \rightleftharpoons \text{pyruvate} + CO_2 + TPN_{red.}$$

Carbon dioxide which is fixed by these or other possible primary fixation reactions may occur in a number of different compounds in the living cell because of the presence of mechanisms for interconvertibility of these compounds.

The fixation of carbon dioxide is an *endergonic* reaction, *i.e.*, a reaction which involves an increase in free energy. From the equilibrium constant and the free energies of formation of the reactants involved in the decarboxylation of oxalacetate to pyruvate, the free energy change has been estimated as about $-5{,}000$ calories.[33] The free energy released in the oxidative decarboxylations has been estimated as 8,000 to 9,500 calories. Thus, in order to bring about carboxylation reactions the cell must provide some mechanism for putting energy into the system. Suggested possibilities for bringing about carboxylation are (*a*) the nearly complete removal of the product formed, which makes the reaction feasible in spite of the unfavorable equilibrium, and (*b*) the participation of phosphorylated intermediates with transfer of the high-energy phosphate from the adenylic system to provide the requisite energy.[30]

The significance of carbon dioxide fixation in animal tissues is at present obscure. No known essential metabolite has been shown to be formed only by carbon dioxide fixation. It has been suggested that the fixation may serve to provide the requisite amounts of oxalacetate necessary for the tricarboxylic acid cycle. In plants, carbon dioxide fixation

[30] S. Ochoa, A. Mehler, and A. Kornberg, *J. Biol. Chem.*, **167**, 871 (1947).

[33] E. A. Evans, Jr., B. Vennesland, and L. Slotin, *J. Biol. Chem.*, **147**, 771 (1943).

by processes not yet well defined is an essential part of the photosynthetic process.

Interrelationships of Carbohydrate, Fat, and Protein Metabolism. From the studies of intermediary metabolism of various compounds it has become increasingly evident that the metabolism of carbohydrates, proteins, and fats results in the formation of a number of common intermediates. The older distinctions among the metabolism of these three principal components of foodstuffs are disappearing. The formation of the same intermediates not only provides a common pathway for the oxidation of carbohydrates, fats, and proteins, but may also provide means for their interconversion.

The amino acids alanine, glutamic acid, and aspartic acid may by oxidative deamination in presence of an amino acid oxidase give rise to pyruvate, α-ketoglutarate, and oxalacetate, respectively. The latter substances are components of the tricarboxylic acid cycle, thus providing a rational mechanism for integration of the amino acids in question with carbohydrate metabolism. Also these amino acids may be transformed to the various keto acids by the action of transaminases (see p. 455). The mechanisms by which energy may be obtained from other amino acids are not so clearly defined. However, it is known that roughly 50 per cent of the protein fed may give rise to carbohydrates under some conditions. For example, animals treated with the glycoside *phlorizin* show a diabetes because of lack of reabsorption of glucose in the tubules of the kidney; various amino acids when fed to such animals may give rise to extra glucose which is excreted in the urine. These and other experiments have been interpreted to mean that the amino acid in question may be converted to glucose, or is a *glycogenic* amino acid; the phenomenon is sometimes referred to as *gluconeogenesis*. However, it is possible that the various amino acids may give rise to extra glucose or glycogen, not by direct conversion, but by other mechanisms such as sparing the utilization of carbohydrate. The field is in need of further investigation. Amino acids other than glutamic acid may be converted to glutamic or α-ketoglutaric acids; this has been suggested for arginine, citrulline, ornithine, proline, and histidine,[34] although the metabolic pathways are not entirely established. For example, Stetten and Schoenheimer [35] fed proline labeled with both heavy N and deuterium to rats and isolated glutamic acid as well as ornithine containing N^{15} and deuterium, which demonstrates the conversion of proline to these substances.

Recent advances in the knowledge of the oxidation of fatty acids has demonstrated important interrelationships between carbohydrate and fat

[34] H. B. Lewis and R. L. Garner, *Ann. Rev. Biochem.*, **9**, 277 (1940).

[35] M. R. Stetten and R. Schoenheimer, *J. Biol. Chem.*, **153**, 113 (1944).

metabolism.[30, 36] The metabolism of the glycerol portion of the fats has been known for sometime to be closely akin to that of carbohydrate metabolism. One feasible metabolic pathway is by phosphorylation to form α-glycerophosphate, which may be oxidized to dihydroxyacetone phosphate or to phosphoglyceraldehyde and thereby enter the normal path for glucose utilization.

The oxidation of the fatty acid molecule is somewhat more obscure. A body of evidence has indicated that fatty acids are broken down or synthesized by two-carbon atom fragments [36] (see p. 826). Recent work by Lehninger [37] has yielded important contributions to this field. He has succeeded in obtaining liver preparations which will oxidize saturated fatty acids, giving rise to acetoacetate. The oxidation apparently requires phosphorylation of the fatty acid and the participation of coenzyme I. These and other experiments [38] have shown that the oxidation gives rise to a two-carbon compound which may condense to form acetoacetate. The two-carbon fragments from the fatty acids may also condense with oxalacetate and thus be oxidized via the tricarboxylic acid cycle. There is evidence that this is the mechanism by which acetoacetic acid and its reduction product, β-hydroxybutyric acid, are metabolized. The nature of the two-carbon fragment, as well as the mechanism of condensation, is at present obscure although acetyl phosphate has been given prominence in various hypotheses.

Experiments with feeding of isotopic acetate to rats [39] have provided evidence that acetate metabolism involves the intermediates of the tricarboxylic acid cycle. In addition the metabolic significance of acetate or closely related compounds is evident from the demonstration that acetate may be incorporated into compounds such as cholesterol, hemin, fatty acids, and glycogen, as well as glutamate and aspartate. Assigning to acetate itself the role of the hypothetical two-carbon intermediate has not been done primarily because of the relatively slow rate of utilization of acetate by tissues. It is probable that the hypothetical two-carbon compound may give rise to acetate if not used in other reactions.

Other additional evidence makes fairly convincing the hypothesis that fats may be oxidized by the tricarboxylic acid cycle. For example, Buchanan, *et al.*,[40] found that intermediates of the tricarboxylic acid cycle stimulated the oxidation of acetoacetate by kidney preparations.

[36] H. A. Lardy and C. A. Elvehjem, *Ann. Rev. Biochem.*, **14**, 1 (1945).

[37] A. L. Lehninger, *J. Biol. Chem.*, **165**, 131 (1946).

[38] S. Weinhouse, G. Medes, and N. F. Floyd, *J. Biol. Chem.*, **155**, 143 (1944); **158**, 411 (1945); **161**, 745 (1945).

[39] D. Rittenberg and K. Bloch, *J. Biol. Chem.*, **157**, 749 (1945).

[40] J. M. Buchanan, W. Sakami, S. Gurin, and D. W. Wilson, *J. Biol. Chem.*, **157**, 747 (1945); **159**, 695 (1945).

When isotopic acetoacetate was used the fumarate and α-ketoglutarate isolated contained a sufficient excess of C^{13} to indicate that the acetoacetate was oxidized via the tricarboxylic acid cycle.

The reader will recognize the inadequacy of the preceding discussion for explaining all the relations between fat, carbohydrate, and protein metabolism. Even though incomplete, our present knowledge of metabolic interrelationships serves to re-emphasize the remarkable anabolic and catabolic powers possessed by living cells and gives an insight to the way in which some of the complicated metabolic processes may occur.

Hormonal Control of Carbohydrate Metabolism. Several of the internal secretions are known to have an active role in the regulation of carbohydrate metabolism, in particular insulin from the pancreas and the hormones of the anterior pituitary gland and the adrenals. However, the biochemical mechanisms by which the hormones produce their effects are poorly understood.

Insulin action is concerned primarily with the utilization of glucose. In the absence of insulin the glucose level of the blood rises to a value above the renal threshold (about 140 mg. glucose per 100 ml. of blood), and glucose appears in the urine. If an excess of insulin is administered, as may occur in the treatment of diabetes, the blood sugar may be lowered to a level which results in coma or even in death. The hormones from the anterior pituitary act in part antagonistically to insulin in the pancreas. The pituitary also plays a role through its action on the adrenal cortex in the regulation of carbohydrate metabolism. In addition the thyroid hormone increases the rate of oxidation of carbohydrate in all tissues and the rate of liver glycogen formation.

The demonstration by Price, Cori, and Colowick [41] of the relation of insulin and a principle from the anterior pituitary to the activity of *hexokinase* has raised considerable interest since it is a pioneer observation of the control of a specific enzymatic reaction by hormones. Price's experiments showed that anterior pituitary hormones will inhibit the hexokinase reaction which is essential for glycolysis or glycogen synthesis. This inhibition is overcome by insulin which, however, has no effect on the reaction in the absence of the anterior pituitary extract. These observations may serve as a partial or complete basis for the function of insulin, although it must be emphasized that the importance of the results of these *in vitro* experiments in explaining the action of insulin in the intact animal awaits further experimentation.

Some Aspects of Carbohydrate Metabolism in Plants. The knowledge of carbohydrate metabolism in plants is limited in comparison to that of animal metabolism. For an excellent survey of the bio-

[41] W. H. Price, C. F. Cori, and S. P. Colowick, *J. Biol. Chem.*, **160**, 633 (1945); S. P. Colowick, C. F. Cori, and M. W. Slein, *J. Biol. Chem.*, **168**, 583 (1947).

chemical aspects of plant respiration the reader is referred to a recent review by James.[42] In some respects the carbohydrate metabolism of plants is similar to that of animals, and this similarity has led to assumptions, often unjustified, that other metabolic relationships found for animals or bacteria may be applicable to plants.

The storage of starch in plants and its breakdown into sugar is affected markedly by environmental conditions, particularly temperature. For example, it has been observed that needles of conifers accumulate starch during the actively growing season but that with the onset of winter temperatures the starch is converted into soluble sugars, chiefly glucose. Thus in the northern portion of the United States the needles become quite starch-free early in the winter period. Similarly, at the onset of winter, starch is present in relatively large quantities in the woody stems of various plants, but during the winter months the starch almost completely disappears, being replaced by oils or fats. In potatoes stored at temperatures near 0°C., both glucose and sucrose are formed from the starch. During growth, sweet potatoes contain very little sugar, but after harvest the starch is rapidly converted to sugars.

The factors which govern the transformations, as well as the mechanisms, are not completely understood. They may depend in part on the physical state of the starch as it occurs in the cell. Apparently the respiratory processes which utilize sugars are depressed by low temperatures more than the reactions leading to starch breakdown. If glucose-1-phosphate is formed from starch by the action of phosphorylase some mechanism for its removal would be necessary to allow further starch breakdown. The dephosphorylation of glucose-1-phosphate to give glucose, or its coupling with fructose to yield sucrose, would allow continued starch breakdown and explain the accumulation of these sugars in the stored potato. Amylase present in potatoes may also have a significant function in the accumulation of sugars.

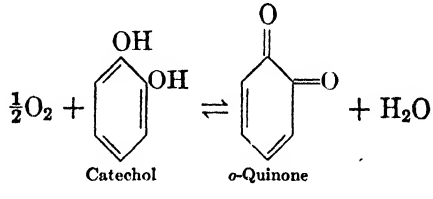
Evidence from various investigators has demonstrated that higher plants may glycolyze sugars to form pyruvate by a series of reactions closely analogous to those outlined above. Several observations which support this conclusion may be cited. Plants are known to phosphorylate sugars, and their respiration rate is dependent on an adequate inorganic phosphate supply. Hexose-6-phosphates, phosphoglycerate, pyruvate, and lactate have been identified as normal plant constituents. An enzyme has been obtained from potatoes and pea meal which catalyzes the formation of mixed triose phosphates. The formation of pyruvate from hexose diphosphate or phosphoglycerate by cell-free barley saps is preventable by fluoride. Lactate and alcohol may accumulate in plant tissues in the absence of air. The functioning of the

[42] W. O. James, *Ann. Rev. Biochem.*, **15**, 417 (1946).

adenylic system is indicated by the presence of labile phosphate compounds closely related to adenosine triphosphate in plants [43] and by catalysis of pyruvate formation by adenylic acid. Sugars commonly utilized by plants are glucose, fructose, and sucrose. Starch, and to a lesser extent, hemicelluloses, fructose anhydrides, and other carbohydrates may serve as respirable materials. Starch and sucrose constitute the principal available carbohydrate reservoirs of most plants. Their relationship has been clarified by experiments demonstrating that both starch and sucrose may undergo phosphorolysis to form glucose-1-phosphate.[44]

As with animal tissues, our knowledge of the oxidative mechanisms in plant tissues is fragmentary. The important observations of Keilen [45] establishing that *cytochromes* exist in nearly all forms of living matter, including plants, have been extended by a number of workers. The detection and isolation of cytochromes in plants is made difficult by the low concentrations present. This makes important the isolation of purified cytochrome *c* from wheat germ together with the preparation of an active *cytochrome oxidase*.[46] Although there is considerable evidence that the cytochrome system may function in many plants, it seems probable that this system does not have a universal role in plant respiration. Cytochrome has been reported to be absent in the leaves and stems of the tea plant. The respiration of the leaves of the mature wheat and barley plants has been found to be uninhibited by cyanide; therefore it can hardly depend on cytochrome oxidase.

In addition to the cytochrome system, there is evidence that *catechol oxidase* (tyrosinase), although not universally distributed in plants, may be responsible for the greater part of the oxygen uptake of potatoes and similar tissues.[47] Catechol or substituted catechols may be capable of acting as reversible oxidation-reduction systems in association with catechol oxidase, as indicated by the reaction



[43] H. G. Albaum and M. Ogur, *Arch. Biochem.*, **15**, 158 (1947).
[44] M. Doudoroff, H. A. Barker, and W. Z. Hassid, *J. Biol. Chem.*, **168**, 725 (1947).
[45] D. Keilen, *Ergeb. Enzymforsch.*, **2**, 239 (1933).
[46] D. R. Goddard, *Am. J. Botany*, **31**, 270 (1944).
[47] D. Baker and J. M. Nelson, *J. Gen. Physiol.*, **26**, 269 (1943); J. G. Boswell, *Ann. Botany*, **9**, 55 (1945).

The oxidized catechols have been shown to be reduced by tissue preparations. Catechol itself may produce inhibition of the respiration of potato tissue slices after a brief stimulation, but such catechol compounds as caffeic acid, protocatechuic acid, dihydroxyphenylalanine (dopa), and gallic acid do not have inhibitory effects. The ability of the catechol oxidase system to oxidize reduced coenzyme I [48] and ascorbic acid affords a possible link with carbohydrate metabolism.

Ascorbic acid oxidase may also play a role in the respiration of plant tissues through its reversible formation of dehydroascorbic acid. In this regard it has been shown that barley saps will catalyze the oxidation of lactic to pyruvic acid by means of the ascorbic system.[49] In the presence of lactate, pyruvate is formed, and the ascorbic acid is maintained in the reduced form. In addition, the oxidation of triose phosphate in glycolysis by barley preparations may occur by means of transfer of hydrogen through the system coenzyme I, dehydrogenase, ascorbic acid, ascorbic acid oxidase, and oxygen.[50]

The copper-containing enzymes catechol oxidase and ascorbic acid oxidase, like the iron-containing enzyme cytochrome oxidase, are inhibited by cyanide. Since, as mentioned above, the respiration of some plants is not inhibited by cyanide, other mechanisms than those mediated by the aforementioned enzyme systems are probably operative in the oxidative processes.

The possible occurrence of a tricarboxylic acid cycle or similar mechanism has received considerable attention by workers studying plant metabolism. Many have postulated that the tricarboxylic acid cycle is operative in plants. For example, Steward and Street,[51] state, "Indirect evidence that something like the Krebs carboxylic acid cycle is a widespread feature of plant metabolism is now impressive." Although the oxidation of pyruvic acid is known to occur in higher plants, the mechanisms for this oxidation remain unestablished, and the evidence does not yet warrant definite conclusions. Several components of the tricarboxylic acid cycle, namely, malate, succinate, fumarate, isocitrate, α-ketoglutarate, and oxalacetate, are known to occur in plant tissues. Malate, succinate, and isocitric dehydrogenases, together with fumarase and aconitase, are also known to be present. Acetate is known to be readily utilized by plants.

However, Bennet-Clark and Bexon [52] have presented evidence that

[48] F. Kubowitz, *Biochem. Z.*, **299**, 32 (1938).
[49] W. O. James and J. M. Cragg, *New Phytologist*, **42**, 28 (1943).
[50] W. O. James, C. R. C. Heard, and G. M. James, *New Phytologist*, **43**, 62 (1944)
[51] F. C. Steward and H. E. Street, *Ann. Rev. Biochem.*, **16**, 471 (1947).
[52] T. A. Bennet-Clark and D. Bexon, *New Phytologist*, **42**, 65 (1943).

in beet root slices a tricarboxylic acid cycle is not operative. Although malate, citrate, and succinate increased the respiration rate of beet root, their action was definitely not catalytic, and the extra respiration was not inhibited by malonate. This lack of inhibition by malonate may possibly be explained by studies showing that malonate at neutral *p*H does not penetrate the carrot root cells.[53] Chibnall [54] and others have considered that in plants the tricarboxylic acid cycle may be operative in the production of aspartic and α-ketoglutaric acids. These may be converted to their respective amino acids and give rise to the amides *asparagine* and *glutamine*.

It is well established, as mentioned in the preceding discussion, that plants will carry out reactions which lead to the production or the utilization of carbon dioxide. However, the knowledge of the reaction mechanisms is meager. Many higher plants possess a carboxylase which will decarboxylate pyruvate with the formation of acetaldehyde and carbon dioxide, and acetaldehyde, which is present in plant tissues, has been suggested as a likely intermediate in the oxidation of pyruvate. In addition to pyruvate carboxylase, oxalacetate carboxylase has been shown to occur widely in plants, and the decarboxylation of oxalacetate to pyruvate has been demonstrated to be reversible.[55]

The Utilization of Sugars by Microorganisms.[56] Sugars can be used as a source of energy by many microorganisms, and in many cases they constitute a preferred source. There is, however, a vast difference among the various microorganisms as to the type of sugar that can be utilized. Some use only one, whereas others use many. When only a single sugar is used it is usually glucose; if several sugars are used, glucose is nearly always one of them. Also, some organisms may be able to break down complex carbohydrates such as glycogen, starch, or even cellulose to usable sugars through the action of enzymes secreted by the organisms. As mentioned previously, the reactions given for glycolysis as well as some of the oxidative mechanisms are applicable to the metabolism of sugar by microorganisms which can use sugars as energy sources. In addition there is evidence that autotrophic organisms, *i.e.*, those that require only inorganic material and carbon dioxide, have an internal metabolism identical or similar to heterotrophic microorganisms. Thus, the autotrophic organism *Thiobacillus thiooxydans* has been shown to

[53] J. S. Turner and V. Hanly, *Nature*, **160**, 46 (1947).

[54] A. C. Chibnall, *Protein Metabolism in the Plant*, Yale University Press, 1939.

[55] B. Vennesland and R. Z. Felsher, *Arch. Biochem.*, **11**, 279 (1946); B. Vennesland, J. Ceithame, and M. J. Gollub, *J. Biol. Chem.*, **171**, 445 (1947).

[56] This section is from material prepared by Professor H. O. Halvorson, Department of Bacteriology, University of Minnesota.

contain an adenosine triphosphate and various phosphate esters of the glycolytic scheme.[57]

Microorganisms vary greatly in the extent to which they break down sugar. In some cases sugar is converted to lactic acid (as occurs with the homofermentative lactic acid bacteria), whereas in other cases it may be oxidized by aerobic processes to carbon dioxide and water. In still other cases sugars may be oxidized to gluconic, glucuronic, or uronic acids. Between these extremes there exist many varying degrees of breakdown; however, in all cases reactions occur that provide the organisms with energy for growth.

The utilization of sugars in the absence of oxygen is incomplete, and there remain products that result from intermolecular oxidation-reduction reactions. In most fermentations the end products are a mixture of alcohols, ketones, and various organic acids together with carbon dioxide or a mixture of carbon dioxide and hydrogen. With the most efficient anaerobic process possible, which may be obtained with some mixed cultures of bacteria, the ultimate end products of sugar breakdown will be CH_4 and CO_2, in accordance with the equation

$$C_6H_{12}O_6 \rightarrow 3CH_4 + 3CO_2$$

In aerobic processes the ultimate end products in the most efficient utilization will be carbon dioxide and water. The mechanism of the aerobic carbohydrate utilization of microorganisms is not so well understood as the fermentation processes. It may be mentioned briefly that the cytochrome system is known to function in many microorganisms, and that it is probable that a tricarboxylic acid cycle or closely similar reaction mechanism participates in the oxidative processes.

The widespread occurrence and importance in microorganisms of the reactions of glycolysis, as outlined in the preceding sections, are well established.[58] Indeed, much of the knowledge of the glycolytic process has evolved from study of the metabolism of microorganisms, particularly of the yeast fermentation process.

Although the reactions leading to the formation of pyruvate by microorganisms are the same as those occurring in plants and animals, the products subsequently formed may be quite varied. The homofermentative lactic acid bacteria utilize sugar by what is perhaps the simplest of all fermentative processes, the sole product being lactic acid formed by reduction of the pyruvate. Each mole of glucose metabolized results in the formation of approximately two moles of lactic acid. This group of bacteria includes *Streptococcus lactis* (the common milk-souring

[57] G. A. Le Page and W. W. Umbreit, *J. Biol. Chem.*, **147**, 263 (1943).

[58] C. H. Werkman and H. G. Wood, *Advances in Enzymol.*, **2**, 135 (1942).

organism), *S. faecalis*, *Lactobacillus bulgaricus*, and *L. delbrueckii*, as well as other species less well known.

In yeast, pyruvate undergoes decarboxylation to form acetaldehyde, and the acetaldehyde is reduced to ethyl alcohol through the mediation of coenzyme I. This reduction is coupled with the oxidative reaction of fermentation, *i.e.*, the oxidation of 3-phosphoglyceraldehyde. In the ordinary yeast fermentation small amounts of glycerol may be produced through reduction of 1-phosphodihydroxyacetone by reduced coenzyme I, followed by hydrolysis of the α-glycerophosphate formed by a phosphatase. In Germany during World War I glycerol was produced on a large scale by addition of bisulfite to block the reduction of acetaldehyde. The bisulfite adds to the aldehyde group, preventing its reduction, and as a consequence the production of glycerol by reduction is markedly increased. The reactions of yeast fermentation and some interrelationships are given in Fig. 108. In addition to the indicated products of yeast fermentation, others, such as formic and acetic acids, fusel oil, and 2,3-butyleneglycol, may be formed in small amounts.

Yeast ferments only the D- or naturally occurring forms of glucose, mannose, fructose, and galactose. The first three hexoses are readily utilized, but galactose is fermented slowly by most strains. However, if the yeast is grown in the presence of galactose it may be adapted to the rapid utilization of this sugar. None of the ordinary yeasts ferments pentoses.

Heterofermentative lactic acid bacteria utilize sugar in a manner which produces ethyl alcohol, acetic acid, carbon dioxide, and glycerol, in addition to lactic acid.[59] Several species of the genera *Leuconostoc* and *Lactobacillus* belong to this group. In the heterofermentation pyruvic acid may be reduced to produce lactic acid or decarboxylated to give acetaldehyde and carbon dioxide. The acetaldehyde formed may be reduced to ethyl alcohol or oxidized to acetic acid.

The propionic acid fermentation by the genus *Propionibacterium* is similar to that of yeast and the lactic acid bacteria. This fermentation occurs in the curing of silage and the ripening of certain types of cheese. The end products are chiefly propionic acid, formed by the reduction of lactic acid, and acetic acid, formed by oxidative decarboxylation of pyruvic acid. The oxidations of glyceraldehyde phosphate to 1,3-diphosphoglyceric acid and of pyruvic acid to acetic acid and carbon dioxide are coupled with the reductions of pyruvic acid to lactic acid and of lactic acid to propionic acid through the mediation of coenzyme I.

Under anaerobic conditions *Escherichia coli* and *Aerobacter aerogenes* possess enzyme systems which catalyze the formation of oxidation, re-

[59] M. E. Nelson and C. H. Werkman, *Iowa State Coll. J. Sci.*, **14**, 359 (1940).

duction, and condensation products from glucose. Again, the principal pathway resulting in the formation of pyruvic acid is the same as that for yeast. The two organisms have quite similar metabolic pathways and may be appropriately discussed together. Both have lactate,

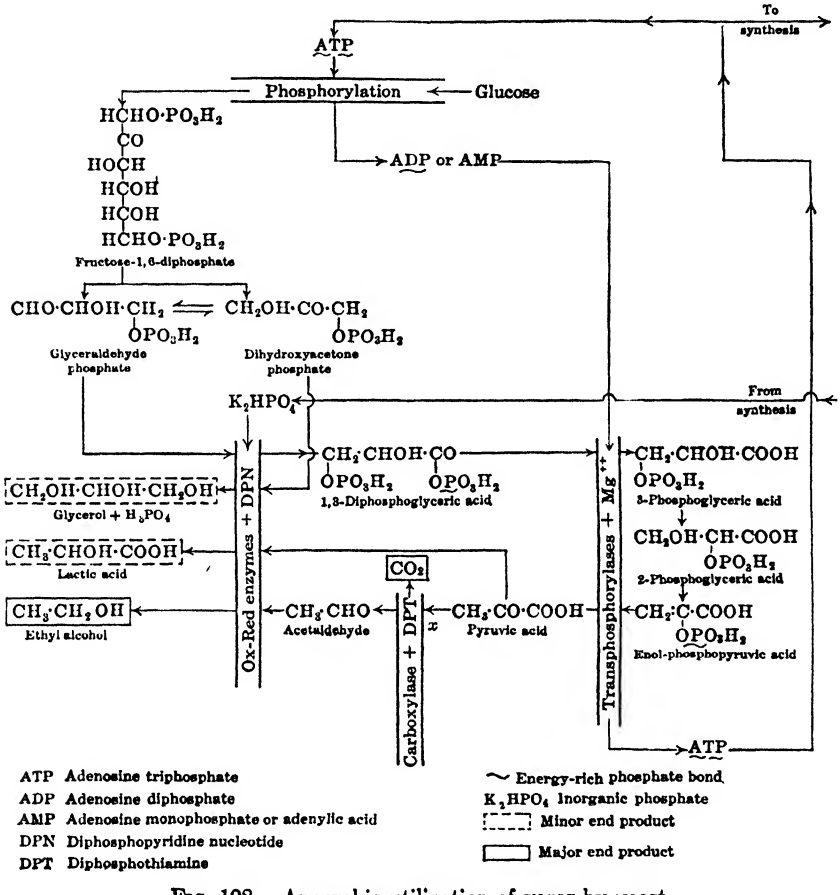


FIG. 108. Anaerobic utilization of sugar by yeast.

acetate, formate, ethyl alcohol, carbon dioxide, and hydrogen as metabolic products. The *Aerobacter* fermentation includes condensation reactions which produce neutral products from potential acid precursors; such condensations are absent in *Escherichia*. As a result, the latter produces more acid per unit sugar utilized. Both form carbon dioxide and hydrogen from formic acid, but *Aerobacter* has an additional source of carbon dioxide from the decarboxylation of pyruvic acid, a reaction absent in *Escherichia*. As a result, the carbon dioxide-hydrogen ratio

is greater than unity in *Aerobacter* and close to unity in *Escherichia*. The principal reactions of the *Aerobacter* fermentation are summarized in Fig. 109. The production of 2,3-butyleneglycol by *Aerobacter* is of industrial importance since the butyleneglycol can serve as a precursor of butadiene, $CH_2{=}CH{-}CH{=}CH_2$, which is used in the manufacture of synthetic rubber.

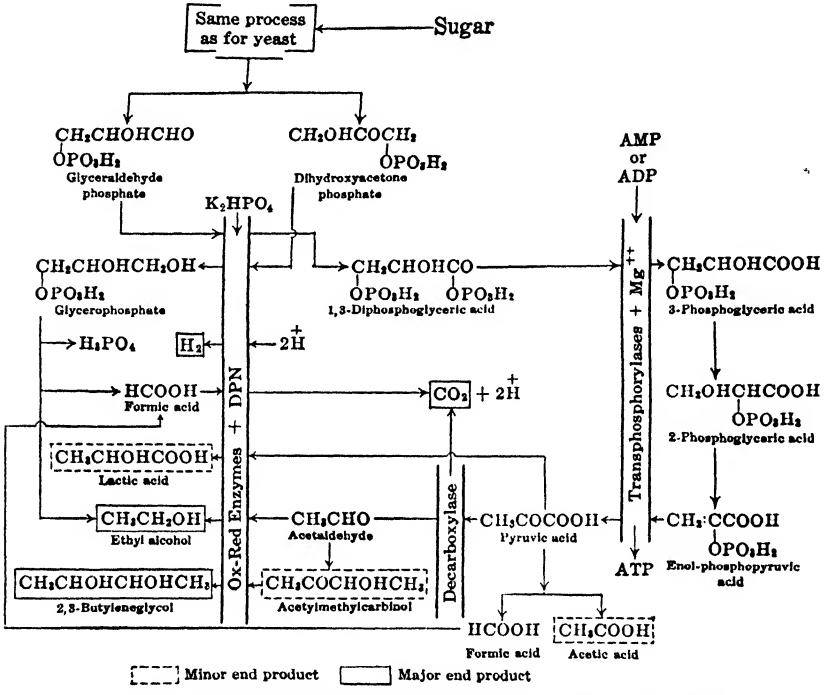


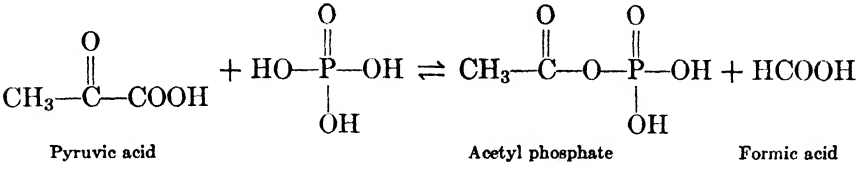
FIG. 109. Anaerobic utilization of sugar by *Aerobacter aerogenes*.

Recent studies [60] have cast some doubt on the role of acetaldehyde as the precursor of acetylmethylcarbinol as indicated in the charts. Adding acetaldehyde tagged with C^{13} on both carbons to active preparations of *Aerobacter* gave no excess of C^{13} in the acetylmethylcarbinol formed. Similar extracts from yeast and pig heart, however, produced acetylmethylcarbinol with C^{13} in all four positions when similarly tagged aldehyde was added.

Both organisms contain a hydrolytic enzyme which splits pyruvic acid into acetic and formic acids. Such a hydrolysis splits the bond between adjacent carbonyl and carboxyl groups, resulting in the release of considerable energy ($-\Delta F$ = 15,000 calories). This energy is avail-

[60] N. H. Gross and C. H. Werkman, *J. Bact.*, **51**, 576 (1946).

able for the binding of inorganic phosphate as acetyl phosphate in the phosphoroclastic reaction:



This reaction is not shown in the chart for reasons of simplification, but it is very important from theoretical standpoints. The demonstration of acetyl phosphate in bacterial reactions has stimulated the study of the role of acetate in many anabolic processes, *e.g.*, synthesis of fatty acids, sterols, hemin, and glycogen. Furthermore, the phosphoroclastic reaction provides another source of energy-rich phosphate which may be transferred to the adenylic acid carrier system for use in various syntheses.[61]

In both *Escherichia* and *Aerobacter* the formic acid formed is broken down to carbon dioxide and hydrogen by "formic hydrogenlyase." Evidence obtained by Ordal and Halvorson [62] indicates that formic hydrogenlyase is actually made up of three separate components, *i.e.*, formic dehydrogenase, hydrogenase, and a mediator between the two. The formic dehydrogenase step proceeds as follows:

$$H_2O + HCOO^- \rightarrow HCO_3^- + 2H^+ + 2e$$

Hydrogenase promotes the reaction

$$2H^+ + 2e \rightarrow H_2$$

The overall reaction, then, is expressed by

$$H_2O + HCOO^- \rightarrow HCO_3^- + H_2$$

Evidence that a mediator is necessary for the overall reaction to occur is obtained by observations on variants of *Escherichia*. Some variants contain only formic dehydrogenase, some only hydrogenase, and some contain both enzymes. As may be expected, no gas is formed by variants containing one enzyme alone, but under certain growth conditions cells containing both enzymes may also be incapable of gas formation. Therefore, it appears that a mediator is necessary to transfer electrons between the two enzyme systems so the overall reaction may go to completion.

[61] F. Lipmann, *Advances in Enzymol.*, **6**, 231 (1946).
[62] E. J. Ordal and H. O. Halvorson, *J. Bact.*, **38**, 199 (1939).

It should be remembered that the oxidation products must be in exact balance with the reduction products. The necessity for such balance is responsible for many interrelationships among the various products of glucose metabolism. Some of these relationships are discussed in this and the following paragraphs. For example, as in yeast, *Aerobacter* and *Escherichia* may form glycerol from glyceraldehyde phosphate to compensate for a deficiency of reduction products necessary for balancing the oxidative formation of diphosphoglyceric acid. Although these organisms can reduce pyruvic acid to lactic acid, the utilization of part of the pyruvate for other reactions makes it necessary that some other molecule be reduced. These organisms contain enzymes which catalyze the breakdown of glycerol thus formed into formic acid and ethyl alcohol. The formic acid is broken down by the formic hydrogenlyase, and the ethyl alcohol remains as a principal end product. Similarly, when the phosphoroclastic reaction forms acetic and formic acids, pyruvic acid is removed from the chain of reactions which yield reduction products of sugar. Therefore a molecule of glycerol is formed for each molecule of acetic acid produced from this reaction. In terms of end products, one molecule of ethyl alcohol is formed for each molecule of acetic acid produced.

The formation of acetylmethylcarbinol by *Aerobacter* removes either reducible products or precursors of reducible products; therefore, for every molecule of acetylmethylcarbinol produced, two potential reducible molecules are removed. However, acetylmethylcarbinol may be reduced to 2,3-butyleneglycol, leaving a deficiency of only one reducible molecule. Thus, one molecule of glycerol will be formed for each molecule of 2,3-butyleneglycol or, in the final analysis, one molecule of ethyl alcohol for each molecule of 2,3-butyleneglycol formed.

In *Escherichia*, there may also occur an enzymatic condensation of pyruvic acid and formic acid to yield malic acid. Malic acid thus formed is reduced to succinic acid through fumaric acid. This condensation removes some of the pyruvic acid from the phosphoroclastic reaction, and one less molecule each of acetic acid and ethyl alcohol is found for each molecule of succinic acid produced. It may also be noted that, if no succinic acid were formed, there would be two molecules of hydrogen for every molecule of ethyl alcohol.

Another interesting and important fermentation is the butanol-acetone fermentation, the mechanism of which has not been so accurately worked out as for the fermentations previously discussed. The organisms responsible are chiefly members of the genus *Clostridium*. The diagram in Fig. 110 lists the major and minor products (dotted lines) of the fermentation. Starch, in the form of either corn mash or

potatoes, is usually the starting material in this fermentation. Starch is converted to sugar, and the sugar in turn is converted to a variety of compounds: acetic acid, ethyl alcohol, 2,3-butyleneglycol, formic acid, and butyric acid. The acetic acid probably combines to form aceto-acetic acid which is decarboxylated to form acetone and carbon dioxide.

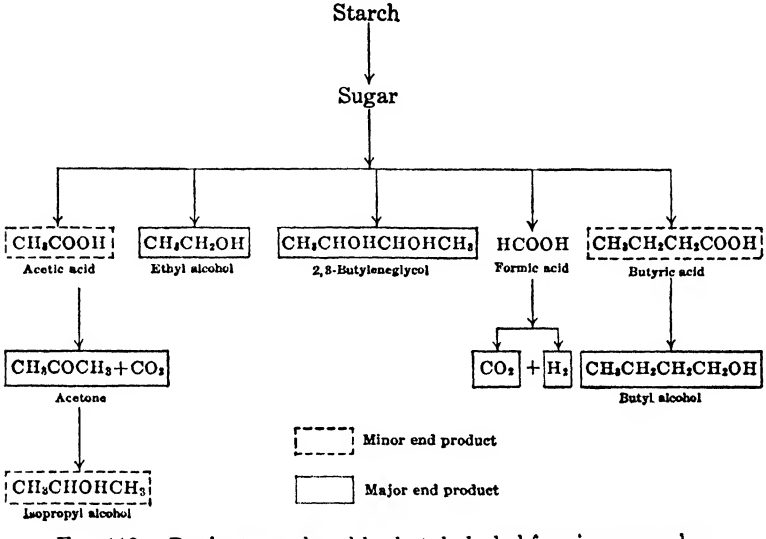


FIG. 110. Products produced by butyl alcohol-forming anaerobes.

A fraction of the acetone may be reduced to form isopropyl alcohol. The formic acid is broken down to carbon dioxide and hydrogen, probably in a manner analogous to that observed in the *E. coli* fermentation. The butyric acid is reduced to produce butyl alcohol. The major products of this fermentation are butyl alcohol, acetone, ethyl alcohol, 2,3-butyleneglycol, carbon dioxide, and hydrogen.

CHAPTER 27

The Glycosides and Saponins

The synthetic and the naturally occurring glycosides are similar in structure, and the formula may be represented by Sugar —O—R, where R represents the foreign (non-sugar or *aglycone*) group attached to carbon-1 of the sugar. This may be coniferyl alcohol as in the glucoside *coniferin*, or salicyl alcohol as in *salicin*, or aniline, or benzaldehyde, etc., as in various glycosides.

The glycosides are widely distributed in nature but usually occur in small amounts. The non-sugar residue is in most instances an aromatic compound. Only sugars capable of forming a pyranose ring form glycosides in nature, and most of the naturally occurring glycosides are compounds of glucose, although arabinose, xylose, ribose, rhamnose, galactose, mannose, and fructose have been isolated from glycosides.

The α- and β-forms of the glucosides have already been discussed. The methyl and ethyl glucosides and galactosides and the methyl xylosides are known in both α- and β-forms, but *all naturally occurring glucosides, so far as is known, occur only in the β-form* and are hydrolyzed only by emulsin.

The rotatory power of an unknown glucoside (*e.g.*, the α-form) can be calculated [1] from the optical rotation of the known β-form because

$$\text{Rotation of } \frac{\alpha + \beta \text{ glucosides}}{2} = \text{rotation of } \frac{\alpha + \beta \text{ glucose}}{2}$$

when R is not optically active.

The role of the glycosides in the plants has been interpreted as a mechanism whereby the substances which have great physiological activity are held inert until they are needed in the metabolism of the plant or in rendering poisonous substances inert so that they will not injure the plant cells.[2,3,4] Glycosides are most commonly found in the fruit, bark,

[1] C. S. Hudson, *J. Am. Chem. Soc.*, **31**, 66 (1909).
[2] R. Combes, *Rev. gén. botan.*, Vols. 29 and 30 (1917–18).
[3] R. Combes, *Compt. rend.*, **167**, 275 (1918).
[4] E. D. Clark, *J. Biol. Chem.*, **21**, 645 (1915).

or roots of plants, although they frequently occur in the leaves. They are usually levorotatory, crystalline, colorless, bitter, soluble in water or in alcohol. In order to isolate them without hydrolysis, it is desirable to destroy their accompanying enzyme by heat. They are easily hydrolyzed by emulsin or by dilute mineral acids.

The glycosides may be classified according to the nature of the non-sugar part of the molecule, under phenols, alcohols, aldehydes, acids, etc. A list of a few of the more important of the naturally occurring glycosides is shown in Table 56.

Certain glycosides contain two hexose molecules, *e.g.*, amygdalin which contains 2 glucose + benzaldehyde + HCN (as the nitrile). By appropriate technic one carbohydrate group can in some instances be hydrolyzed off, yielding a new glycoside containing only *one* hexose.

Both on account of the very small amount of glycosides present in plant tissues and the fact that, as a rule, glycosides do not form characteristic insoluble derivatives which allow for their isolation and identification, it is difficult to discover new glycosides and still more difficult to determine their constitution. Certain biochemical methods have assisted materially in this work:

1. *Bourquelot's method* for examining a plant for glycosides consists in determining the reducing sugar and the optical rotation of an extract of the plant tissue. Then emulsin is added, and after a period of incubation the reducing sugars and the optical rotation are redetermined. An increase in optical rotation indicates the presence of β-glycosides, and the amount of change may give a rough estimation of the amounts which are present.

2. The nature of the sugar can be more or less accurately determined, according to ter Meulen, by adding the different sugars to the glycoside + enzyme mixture. The sugar which retards the enzyme action is probably the one in the glycoside, inasmuch as the hydrolysis is an equilibrium reaction.

$$\text{Glycoside} + \text{enzyme} + H_2O \rightleftharpoons \text{enzyme} + \text{sugar} + \text{non-sugar residue}$$

This consideration possibly led to Bourquelot's *enzyme synthesis* work which established definitely that enzyme actions were equilibria and were reversible. The first demonstration of synthesis by enzyme action was carried out in the case of glycosides. Emulsin was added to a concentrated solution of the split products of the glycoside, and the mixture was incubated at 37°C. It was found that the equilibrium

$$\text{Glycoside} + H_2O \rightleftharpoons \text{sugar} + \text{non-sugar residue}$$

was shifted toward the left, *i.e.*, *the amount of reducing sugar decreased*

TABLE 56. CERTAIN NATURALLY OCCURRING GLYCOSIDES

Name	Hydrolytic Products	Source
	I. PHENOLS	
Arbutin	Glucose + hydroquinone	*Arbutus uva ursi*
Phlorizin	Glucose + phloretin	Bark of *Rosaceae*
	II. ALCOHOLS	
Coniferin	Glucose + coniferyl alcohol	Bark of fir tree
Populin	Glucose + saligenin + benzoic acid	Bark of *Populus*
Salicin	Glucose + saligenin	Bark of willow (*Salix* sp.)
	III. ALDEHYDES	
Amygdalin	2 Glucose + *dextro*-mandelonitrile	Seeds of bitter almond
Dhurrin	Glucose + *p*-oxymandelonitrile	Leaves of *Sorghum vulgare*
Linamarin	Glucose + acetonecyanhydrin	Leaves of young flax plants and flaxseed
Prulaurasin	Glucose + racemic mandelonitrile	Leaves of *Prunus laurocerasus*
Prunasin	Glucose + *dextro*-mandelonitrile	Young twigs of *Prunus padus*
Sambunigrin	Glucose + *levo*-mandelonitrile	Leaves of common elder (*Sambucus niger*)
	IV. ACIDS	
Gaultherin	Glucose + methylsalicylate	Bark of *Betula lenta* or *Gaultheria procumbens*
Jalapin	Glucose + jalapinolic acid	Roots of *Jalapa orizabensis*
	V. OXYCUMARIN DERIVATIVES	
Aesculin	Glucose + aesculetin	Bark of horse-chestnut (*Aesculus hippocastanum*)
Fraxin	Glucose + fraxetin	Bark of the ash (*Fraxinus* sp.)
	VI. OXYANTHRAQUINONE DERIVATIVES	
Ruberythric acid	Glucose + alizarin	Madder, the roots of *Rubia tinctorum*
	VII. OXYFLAVONE DERIVATIVES	
Apiin	Apiose + apigenin	Leaves of parsley, celery, etc.
Quercitrin	Rhamnose + quercetin	Bark of oak
Rutin	Glucose + rhamnose + quercetin	Petals of the violet, leaves of *Ruta graveolens*, etc.
Xanthorhamnin	2 Rhamnose + galactose + rhamnetin	Fruits of various species of *Rhamnus*
	VIII. MUSTARD OILS	
Sinigrin	Glucose + allyl isothiocyanate + $KHSO_4$	Black mustard seeds
	IX. ANTHOCYANINS	
	A. Cyanidin Derivatives	
Asterin (chrysanthemin)	Glucose + cyanidin	Scarlet-red winter aster
Cyanin	2 Glucose + cyanidin	Blue cornflower, red rose, etc.
Idaein	Galactose + cyanidin	Mountain cranberries
Keracyanin	Rhamnose + glucose + cyanidin	Black cherries
Mekocyanin	2 Glucose + cyanidin	*Papaver rhoeas*, L.
Sambucin	Glucose + cyanidin	Elderberries (*Sambucus nigra*)
	B. Delphinidin Derivatives	
Delphinin	2 Glucose + delphinidin	*Delphinium consolida*, L.
Gentianin	Glucose + delphinidin	*Gentiana acaulis*, *Gentiana vulgaris*
Violanin	Rhamnose + glucose + delphinidin	*Viola tricolor*, L.
	C. Hirsutidin Derivatives	
Hirsutin	2 Glucose + hirsutidin	*Primula hirsuta*
	D. Malvidin (Syringidin) Derivatives	
Malvin	2 Glucose + malvidin	*Primula viscosa*
Oenin	Glucose + malvidin	Blue grape

TABLE 56. CERTAIN NATURALLY OCCURRING GLYCOSIDES (*Continued*)

Name	Hydrolytic Products	Source
	IX. ANTHOCYANINS (*Continued*)	
	E. Pelargonidin Derivatives	
Callistephin	Glucose + pelargonidin	*Callistephus chinensis*, Nees
Pelargonin	2 Glucose + pelargonidin	Scarlet geranium, red cornflower
	F. Peonidin Derivatives	
Peonin	2 Glucose + peonidin	Red peony
	X. CARDIAC GLYCOSIDES	
	A. In Order Apocynaceae	
Adynerin	Oleandrose + adynerigenin	*Nerium oleander*, L.
Cymarin	Cymarose + strophanthidin	*Strophanthus kombé*, S.
Neriantin	Glucose + neriantogenin	*Nerium oleander*, L.
Oleandrin	Oleandrose + acetic acid + gitoxigenin	*Nerium oleander*, L.
Sarmentocymarin	Sarmentose + sarmentogenin	*Strophanthus sarmentosus*, L.
Thevetin	2 Glucose + digitalose	*Thevetia neriifolia*, S.
	B. In Order Asclepiadaceae	
Periplocymarin	Cymarose + periplogenin	*Periploca graeca W.*, B.
Periplocin	Cymarose + glucose + periplogenin	*Periploca graeca*, B.
Uzarin	2 Glucose + anhydrouzarigenin	Uzara tree
	C. In Order Liliaceae	
Convallatoxin	Rhamnose + convallatoxigenin	*Convallaria majalis*, F, L.
Proscillaridin A	Rhamnose + scillaridin A.	*Scilla maritima*
Scillaren A	Rhamnose + glucose + scillaridin A.	*Scilla maritima*
	D. In Order Moraceae	
α-Antiarin	Antiarose + antiarigenin	*Antiaris toxicaria*, La.
β-Antiarin	Rhamnose + anhydroantiarigenin	*Antiaris toxicaria*, La.
	E. In Order Scrophulariaceae	
Digilanide A	2 Digitoxose + acetyldigitoxose + glucose + digitoxigenin	*Digitalis lanata*, L.
Digilanide B	2 Digitoxose + acetyldigitoxose + glucose + gitoxigenin	*Digitalis lanata*, L.
Digilanide C	2 Digitoxose + acetyldigitoxose + glucose + digoxigenin	*Digitalis lanata*, L.
Digitoxin	3 Digitoxose + digitoxigenin	*Digitalis purpurea*, L.
Digoxin	3 Digitoxose + digoxigenin	*Digitalis lanata*, L.
Gitoxin	3 Digitoxose + gitoxigenin	*Digitalis purpurea*, L.
	XI. DIGITALIS SAPONINS	
Amolonin	3 Glucose + galactose + 2 rhamnose + tigogenin	*Chlorogalum pomperidianum*
Digitonin	4 Galactose + xylose + digitogenin	*Digitalis purpurea*
Dioscin	Rhamnose (?) + diosgenin	*Dioscorea tokoro*
Gitonin	3 Galactose + pentose + gitogenin	*Digitalis purpurea*
Sarsasaponin	2 Glucose + rhamnose + sarsasapogenin	*Radix sarsaparillae*
Tigonin	2 Glucose + 2 galactose + rhamnose + tigogenin	*Digitalis purpurea*, *Digitalis lanata*
Trillarin	2 Glucose + diosgenin	*Trillium erectum*
Trillin	Glucose + diosgenin	*Trillium erectum*
	XII. VARIOUS	
Dibenzoylglucoxylose	Glucoxylose + benzoic acid	Leaves of *Dariesia latifolia*
Indican	Glucose + indoxyl	Leaves of *Indigofera*

The synthesis of glycosides was thus definitely proved. Upon dilution of the system, the reaction again shifted to the right, and hydrolysis of the glycoside took place.

Bridel and Béguin [5] used this method of enzyme synthesis for the preparation of ethyl-L-arabinoside by adding emulsin to a solution of pure L-arabinose in ethanol and allowing the mixture to stand for 173 days, adding more emulsin on the twenty-seventh and forty-third days. They stated that the compound isolated was α-ethyl-L-arabinoside and suggested that possibly the commercial emulsin contained α-L-arabinosidase as a specific enzyme.

Cyanogenetic or Cyanophoric Glycosides. These glycosides have received a considerable amount of attention not only because of their practical importance (they are present in such substances as bitter almonds, linseed cake, sorghum, and other fodder plants), but also because of theoretical interest in their function in plants. They contain nitrogen as HCN, and because of that fact can be more or less easily estimated. HCN is fairly common among higher plants and occurs, in the majority of cases, combined in glycosides.

Although the cyanogenetic glycosides are somewhat rare as compared with saponins, they have a much wider distribution than was once supposed, and are present in many economic plants. Flax contains a small amount of HCN in the dry seed (0.008 per cent HCN), but as much as 0.135 per cent has been found in the germinated seedlings. The same increase occurs in the germination of bitter almonds. The stage of development at which the maximum amount is reached may be different for different plants. Thus, in flax the maximum occurs when the seedlings are 4 to 5 inches high, while in *Lotus arabicus* the maximum is at the time of flowering. In sorghum there appears to be no HCN in the seed, but Willaman found the maximum amount in seedlings about 3 inches high, decreasing to practically zero in the mature plant.

Uncombined HCN has been found in plants in a few instances, but it is extremely rare, and the presence of HCN may be considered as evidence of the presence of a cyanophoric glycoside. Therefore, the quantitative determination of the cyanophoric glycosides consists of the determination of HCN, formed by the hydrolysis of the glycosides, (*a*) by autolyzing the plant tissue, or (*b*) by addition of acid, or (*c*) by incubating, after the addition of emulsin, and distilling off the HCN into alkali. Autolysis was the first method employed; in fact the autolysis of crushed bitter almonds was the first means by which HCN was identified as a plant constituent.

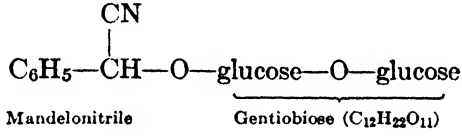
A great deal of attention has been given to methods for the determina-

[5] M. Bridel and C. Béguin, *Compt. rend.*, **182**, 812 (1926).

tion of HCN, and the various factors governing its quantitative liberation, quantitative distillation, and measurement have been exhaustively studied. There is still much room for improvement in the present methods. In most cases one deals with quantities ranging from a fraction of a milligram to one or two milligrams of HCN, and the measurement depends on the conversion of ferrous salts to Prussian blue. The conditions of formation must be such that the pigment will remain suspended in a colloidal sol, so that the measurement can be made colorimetrically.

Amygdalin, $C_{20}H_{27}O_{11}N$, is probably the best known of the glycosides. It is obtained from the kernels of the bitter almond, *Prunus amygdalus*. It also occurs in smaller amounts in the kernels of the peach, plum, apricot, and most fruits belonging to the *Rosaceae*.

Liebig and Wöhler, in 1837, found that amygdalin was hydrolyzed to glucose, benzaldehyde, and HCN. Caldwell and Courtauld showed that the reaction took place in two stages when acids were used as the hydrolyzing agents, one glucose radical being split off before the second one was attacked. The structure has been shown to be



Emulsin hydrolyzes the two glucose bonds at about the same rate, so that the end products of emulsin hydrolysis are two molecules of glucose and mandelonitrile, which further hydrolyzes into benzaldehyde and HCN. Amygdalase hydrolyzes off only one glucose radical, forming a new glycoside, *prunasin*, which in turn is hydrolyzed by the enzyme *prunase*. The enzyme prunase was first found in the leaves of the cherry laurel, hence the name. "Emulsin" therefore contains two enzymes, amygdalase and prunase. Prunase cannot act until the amygdalase has split off one glucose radical. This is another example of enzyme biological specificity and indicates that one glucose radical shields the rest of the molecule from the enzyme.

Three glycosides of the same formula and composition as prunasin are known:

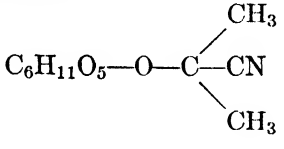
1. Prunasin *dextro*-mandelonitrile-glucoside
2. Prulaurasin *racemic*-mandelonitrile-glucoside
3. Sambunigrin *levo*-mandelonitrile-glucoside

This is one of the few instances where the *dextro*, *levo*, and *racemic* forms of a compound all occur in nature.

Dhurrin occurs in *Sorghum vulgare;* it was isolated by Dustan and Henry in 1902. Its structure is *p*-hydroxymandelonitrile-glucoside, hydrolyzing to glucose, HCN, and *p*-hydroxybenzaldehyde. It also occurs in the forage crop, "Sudan grass."

Vicianin, $C_{19}H_{25}O_{10}N$, occurs in the seeds of the wild vetch, *Vicia angustiafolia*, and yields on hydrolysis HCN, benzaldehyde, and a disaccharide, vicianose (glucose + arabinose). It may be noted that this glycoside is similar to amygdalin except that one glucose radical of amygdalin has been replaced by arabinose. The wild vetch seed contains a relatively small amount of glycoside (HCN = 0.0033 per cent of the seed), but a small amount may produce undesirable effects. The vetch is one of the so-called "non-separable" weed seeds in grain, and the presence of vetch in wheat will produce a pronounced benzaldehyde odor in bread doughs made from the flour.

Linamarin or *phaseolunatin*,



hydrolyzes to β-glucose, acetone, and HCN. It occurs in flaxseed, flax plants, etc.

Phlorizin, yielding, as the products of acid hydrolysis, glucose, phloroglucinol, and *p*-hydroxyhydratropic acid,

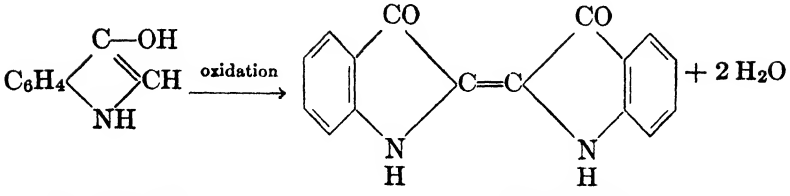
$$(p)HO—C_6H_4—CH(CH_3)—COOH$$

is found in the bark of the apple, pear, and other rosaceous trees. Phlorizin possesses the remarkable property of causing "artificial diabetes" or glycosuria when taken internally, or better when injected subcutaneously (in oil). Much of the experimental work on diabetes has been suggested by experiments made on phlorizinized dogs.

Salicin, a glycoside hydrolyzing to β-glucose and saligenin (*o*-hydroxybenzyl alcohol) occurs in the bark of most species of *Salix* (willows). This glycoside is assuming increased importance because of the action of saligenin as a local anesthetic. Certain salicin derivatives show abnormal optical rotations.[6]

Indican, $C_{14}H_{17}O_6N$, on hydrolysis yields glucose and indoxyl. Indoxyl is the leuco base of indigo blue to which it passes on oxidation. The glycoside occurs in the indigo plants and can be extracted by acetone. A specific enzyme is present in the leaf which hydrolyzes the

[6] A. Kunz, *J. Am. Chem. Soc.*, **48**, 262 (1926).



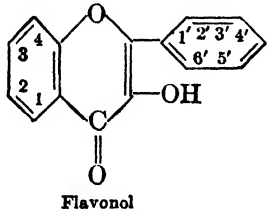
Indoxyl (colorless) Indigo blue (indigotin)

glycoside, and the indoxyl in the presence of air and of an *oxidase* is converted to the indigo blue. The sugar radical must be split off before oxidation can take place. This is one of the best-known cases where one enzyme (glucosidase) must act to produce a compound before a second enzyme (oxidase) can act to produce a third substance.

The production of indigo from indican was at one time one of the world's great industries, and the indigo planters of India were practically ruined by Baeyer's synthesis of indigo. In carrying out the successful synthesis, the Badische Company spent nearly $5,000,000 during seventeen years of research before a pound of synthetic indigo had been sold. Natural indigo is no longer a serious competitor of synthetic indigo except in the Orient, where cheap labor is available and an inferior product is acceptable. The crude natural indigo often contains only 5 to 7 per cent of dyestuff.

Although indigo occurs exclusively as a glycoside in the plant kingdom, and in the animal body as a decomposition product of tryptophan formed by bacterial action, a bromo-substituted indigo occurs in a gland in the mollusks *Murex brandaris* and *Murex trunculus*. This compound, 6,6′-dibromoindigo, was known to the ancients as "Tyrian purple." Friedländer [7] isolated 1.5 grams of the dyestuff from 12,000 mollusks. Research has shown that the synthetic 5,5′-dibromoindigo dyes cloth a clearer purple and is much to be preferred to the "Tyrian purple" as a dyestuff. 6,6′-Dibromoindigo is one of the few naturally occurring organic compounds which contain bromine.

The hydroxyflavone glycosides contain flavonol, or hydroxyflavone, or some derivative of this, as the non-sugar radical:



Flavonol

[7] P. Friedländer, *Ber.*, **39**, 1060 (1906); *Ber.*, **42**, 765 (1909).

All the flavonol derivatives are yellow dyes. In many of these glycosides the carbohydrate is rhamnose. Quercitrin occurs in oak bark. On hydrolysis it yields quercetin (1,3,3′,4′-tetrahydroxyflavonol) and rhamnose. Such glycosides are usually the raw materials for the preparation of rhamnose.[8] Quercetin is widespread as a plant pigment, occurring in red onion skins, cotton flower, etc.

The various shades of blue, purple, violet, and mauve and most of the red colors of flowers, fruits, leaves, and stems of plants are due to a group of pigments known as *anthocyanins* which exist, for the most part, in solution in the cell sap. The sugar-free pigments, or aglycones, are called anthocyanidins. The hydroxyflavones and anthocyanins are discussed in Chapter 35.

Cardiac Glycosides.[9] The widespread group of cardiac glycosides are so classified because of their common characteristic action in decreasing the frequency and increasing the intensity of the heart beat of individuals with damaged heart function; an overdose produces vomiting, and the heart ceases beating (in systolic standstill). Aside from their therapeutic uses, certain members of this group (particularly those obtained from the bark of various species of *Stropanthus*) have been used by the aborigines of Africa in preparing arrow poisons. The aglycones of the cardiac glycosides are designated as *genins* and are convulsive poisons of no medicinal value.

Plants belonging to the orders *Apocynaceae* and *Scrophulaceae*, particularly to certain genera of *Digitalis* (foxglove) of the latter order, are the principal sources of glycosides of therapeutic value. The glycosides are obtained by extracting the plant tissues (particularly the seeds and leaves) and are difficult to obtain in pure condition because of the partial hydrolysis by enzymes present in the tissues. Although the majority of the members of this group listed in Table 56 are probably partial degradation products, it has been shown that the digilanides A, B, and C, the purpurea glycosides, and scillaren A are true plant glycosides.[10] They may be extracted from the dried leaves with alcohol, the tannins removed, and the glycosides purified by precipitation and crystallization or by partition between such solvents as chloroform and water or aqueous methanol. The sap of fresh leaves of *Digitalis purpurea* will degrade the

[8] C. F. Walton, Jr., *J. Am. Chem. Soc.*, **43**, 127 (1921).

[9] For a comprehensive treatment of the cardiac glycosides see R. C. Elderfield, *Chem. Revs.*, **17**, 187 (1935). For a more recent review of their chemistry see the chapter on the steroids by W. H. Strain, in H. Gilman, *Organic Chemistry*, 2nd ed., Vol. II, John Wiley & Sons, New York, 1945.

[10] A. Stoll and W. Kreis, *Helv. Chim. Acta*, **16**, 104, 1390 (1933); A. Stoll and J. Renz, *ibid.*, **22**, 1193 (1939).

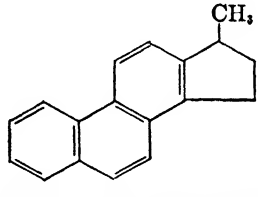
purpurea glycoside A to digitoxin, purpurea glycoside B to gitoxin, and digilanide C to digoxin.[10]

$$\underset{\text{Purpurea glycoside A}}{C_{47}H_{74}O_{18}} + H_2O \rightarrow \underset{\text{Digitoxin}}{C_{41}H_{64}O_{13}} + \underset{\text{Glucose}}{C_6H_{12}O_6}$$

$$\underset{\text{Purpurea glycoside B}}{C_{47}H_{74}O_{19}} + H_2O \rightarrow \underset{\text{Gitoxin}}{C_{41}H_{64}O_{14}} + \underset{\text{Glucose}}{C_6H_{12}O_6}$$

The native cardiac glycosides contain up to four sugar residues per molecule. These comprise the common sugars glucose and rhamnose and a number of other sugars which have not been found elsewhere. The rare sugars [11] may be classified into two broad groups: the one comprising *hexomethyloses, viz.*, digitalose and probably antiarose, and the other, *2-desoxyhexomethyloses, viz.*, digitoxose, cymarose, diginose, oleandrose, and sarmentose, of which all but digitoxose contain methyl ether groupings.

Within the past several years our knowledge of the chemical structure of the aglycones, or genins, of the cardiac glycosides has greatly increased, and it is now known that they fall into the ever-expanding list of substances related to the sterols and bile acids, *i.e.*, they are derivatives of the reduced cyclopentenophenanthrene nucleus. *Diel's hydrocarbon*,



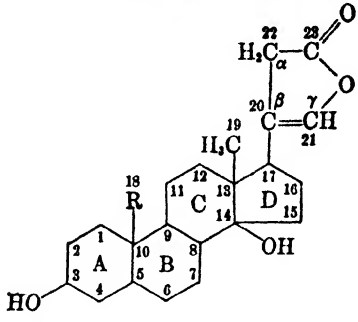
Methyl-cyclopentenophenanthrene
(Diel's hydrocarbon)

is obtained by the dehydrogenation of cholesterol, ergosterol, the sex hormones, the toad poisons (bufotoxin, bufotalin, and cinobufagin), and the non-sugar constituent of the cardiac glycosides and of the saponins. Likewise the non-sugar residue of the poisonous glycosides present in squills contains this nucleus, and it is also the characteristic nucleus of the carcinogenetic compounds. Therefore this nucleus has not only a wide distribution in nature but is the basis of a great many physiologically active compounds.

While the cyclopentenophenanthrene nucleus is the characteristic nucleus of these compounds, the actual aglycones contain the hydrogenated nucleus (cyclopentanoperhydrophenanthrene) with a lactone

[11] For a complete discussion of the occurrence, preparation, and chemistry of the rare sugars found in the cardiac glycosides, see R. C. Elderfield, *Advances in Carbohydrate Chem.*, **1**, 147 (1945).

side chain from carbon-17. With the exception of scillaridin A, the cardiac aglycones have the following ring system:



Rings A/B: *cis* or *trans*, generally *cis*.
Carbon-3 hydroxyl: *trans* or *cis*, generally *trans*.

All these genins have a hydroxy group at carbon-3 to which the sugar residues of the glycoside are apparently attached, and most of them also have a hydroxy group at carbon-14.[12]

Saponins. The saponins comprise a rather widely distributed group of compounds which are so named because, like soap, they have the property of markedly lowering surface tension. As a result, their aqueous colloidal solutions foam readily when shaken, and they stabilize emulsions of fats and oils so that the emulsions are very difficult to "break." The cardiac glycosides also produce foams, and hence are saponins, but because of their characteristic action on the heart they have been treated here as a separate class. The saponins of the digitalis group, such as digitonin, gitonin, and tigonin, which occur with the cardiac glycosides, are designated *digitalis saponins* to differentiate them from other members of this class. Like the cardiac glycosides, the digitalis saponins, upon hydrolysis, yield aglycones or *sapogenins* which contain the cyclopentanoperhydrophenanthrene nucleus and yield Diel's hydrocarbon when dehydrogenated; thus, when hydrolyzed, digitonin, gitonin, and tigonin yield the sapogenins, digitogenin, gitogenin, and tigogenin. Most of the other saponins have a different ring system and give rise to 1,2,7-trimethylnaphthalene (sapotalene) when they are hydrogenated. Digitonin is the major component of the glycosides of digitalis. It possesses the valuable property of forming an insoluble compound with cholesterol and other sterols and is used for the quantitative extraction of these compounds.[13] The digitonin

[12] For a detailed discussion of the chemistry of the genins see Chapter 19 on the steroids by W. H. Strain in H. Gilman, *Organic Chemistry*, 2nd ed., Vol. II, John Wiley & Sons, New York, 1945.

[13] A. Windaus, *Z. physiol. Chem.*, **65**, 110 (1910); A. Windaus and S. V. Shah, *Z physiol. Chem.*, **151**, 86 (1926).

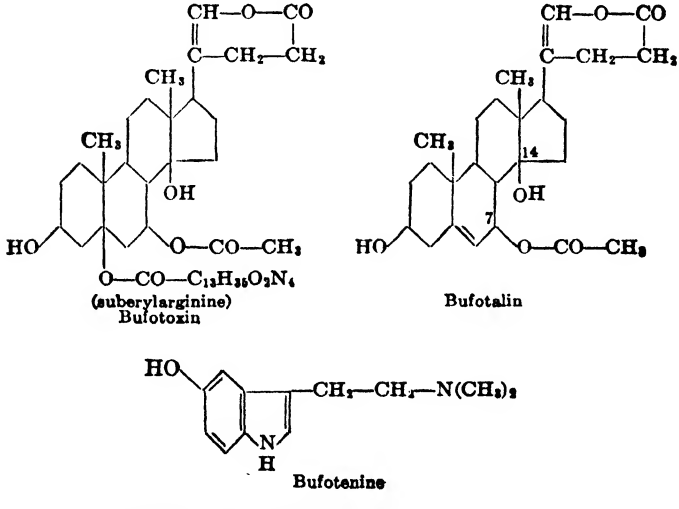
may be recovered from the sterol-digitonin precipitate for further use.[14]

When injected intravenously, the saponins are poisonous, but they are not toxic to the higher animals when taken orally. They are more poisonous toward lower than toward higher forms of animals and have been used by primitive races of Africa and South America to catch fish. Thus saponin-containing plants are macerated and thrown into the stream, and the poisoned fish, which are edible by humans, rise to the surface. The saponins have the ability to hemolyze red blood corpuscles in very low concentrations.

Toad Poisons. The toad poisons, secreted by the parotid (salivary) gland of certain toads along with other substances, are also derivatives of the phenanthrene nucleus. The poisons proper are suberylarginine derivatives of the genins (called *bufagins*) which vary with the species. In *Bufo vulgaris* the poison glands secrete a milky fluid containing bufotoxin, bufotalin, bufotenine, and epinephrine.

Bufotoxin is a cardiac poison more or less similar in its action to digitalis extract but lacking the persistency of action which characterizes the digitalis glycosides. Bufotenine is unusual in that it is a derivative of 5-hydroxyindole. It is a cardiac stimulant and shows pressor action. Bufotoxin shows the unusual grouping of an esterification with suberylarginine on the hydroxyl group situated on carbon-5 of the phenanthrene nucleus.

The formulas for bufotoxin, bufotalin, and bufotenine are [15]



[14] T. Panzer, *Z. physiol. Chem.*, **78**, 414 (1912).

[15] Formulas from H. Wieland and G. Hesse, *Ann.*, **517**, 22 (1935).

CHAPTER 28

The Pectic Substances [1]

The pectic substances are colloidal carbohydrates of high molecular weight and rather complex composition.[2] Galacturonic acid, galactose, arabinose, xylose, methanol, and acetic acid have been identified as hydrolytic products. The word "colloid" should be emphasized here at the beginning. All pectic substances which can be properly so designated are highly hydrophilic; they are almost always precipitated by alcohol or similar dehydrating agents; and this results in impurities, inadvertent changes in molecular weight and properties, and in a confusion in identification that has confounded pectin students for a hundred years.

Beginning in 1824 with Braconnot and Payen, a host of workers throughout the nineteenth century studied these substances, but mostly from the botanical viewpoint. The true chemistry of the pectins began in 1917 when Ehrlich [3] identified galacturonic acid as accounting for 67 per cent of beet pectin. The next definite step was the identification of methyl esters by von Fellenberg [4] in 1918. To be sure, arabinose and galactose had, after various tribulations, been established as hydrolytic products of "pectin preparations"; but, as we shall see, it is still questionable whether their polymers, araban and galactan, are parts of the true pectin molecule and whether they actually contribute anything to the one outstanding property of pectin itself, namely gel formation in a 65 per cent solution of sugar.

Galactose has been recognized for a long time as a constituent of the pectins. Its presence was at first merely surmised from the fact that mucic acid is formed on treatment of pectins with nitric acid. Since, however, galacturonic acid also yields mucic acid, it was the isolation

[1] This chapter is contributed by Dr. J. J. Willaman, Eastern Regional Research Laboratory, United States Department of Agriculture, Philadelphia; formerly Professor of Agricultural Biochemistry in the University of Minnesota.

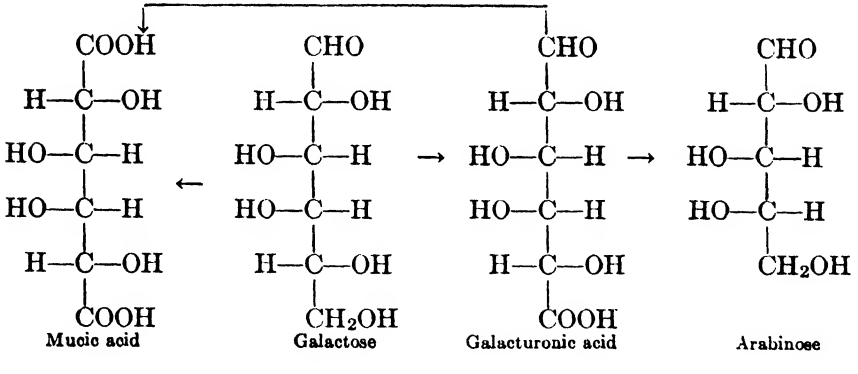
[2] For a recent review of the chemistry of pectic substances, see M. A. Joslyn and H. J. Phaff, *Wallerstein Labs. Communs.*, **10**, 39 (1947).

[3] F. Ehrlich, *Chem. Ztg.*, **41**, 197 (1917).

[4] Th. von Fellenberg, *Biochem. Z.*, **85**, 118 (1918).

of the sugar from the hydrolytic products of the pectin that furnished the necessary proof. It is unlikely that it is produced from any other substance during hydrolysis by weak acids. It occurs in pectin as the anhydride galactan.

On the other hand, there has been real reason to question the presence of arabinose. It was at first surmised from the fact that the pectic substances give very strong reactions for pentoses; in fact, by the phloroglucide method of pentose estimation, a very high percentage of pentose in the pectin can be demonstrated. However, in this method furfural is formed by boiling the material with 12 per cent hydrochloric acid. This treatment converts the galacturonic acid to arabinose, and then the latter to furfural. Hence, this phloroglucide method is not necessarily either a qualitative or a quantitative indication of pentose. Furthermore, although arabinose has been identified as such in the acid hydrolysate of pectins, it could easily be an artifact, due to decarboxylation of some of the galacturonic acid. Myers and Baker [5] have apparently succeeded in avoiding these difficulties and have established the presence of arabinose in citrus pectin.



Acetic acid has been definitely identified in the pectin of flax and of sugar beet by Ehrlich,[6] of sugar beet by Nelson,[7] and of lemons by Myers and Baker.[5] Nelson could not find it in the pectin from the fruit of apple, tomato, or lemon.

Nomenclature. Pectin chemistry has always been plagued by a surfeit of terms. The American Chemical Society Committee on Nomenclature of the Pectic Substances [8] lists over 30 terms in the literature.

[5] P. B. Myers and G. L. Baker, *Delaware Agr. Exp. Sta. Bull.* 187 (1934).

[6] F. Ehrlich and R. von Sommerfeld, *Biochem. Z.*, **168**, 263 (1926); F. Ehrlich and F. Schubert, *ibid.*, **169**, 13 (1926).

[7] E. K. Nelson, *J. Am. Chem. Soc.*, **48**, 2945 (1926).

[8] Report, *Chem. Eng. News*, **22**, 105 (1944).

The first committee in 1926 [9] conceded but three pectic substances, *protopectin*, *pectin*, and *pectic acid*. The second committee added to these *pectinic acid*. Within our present knowledge these are sufficient.

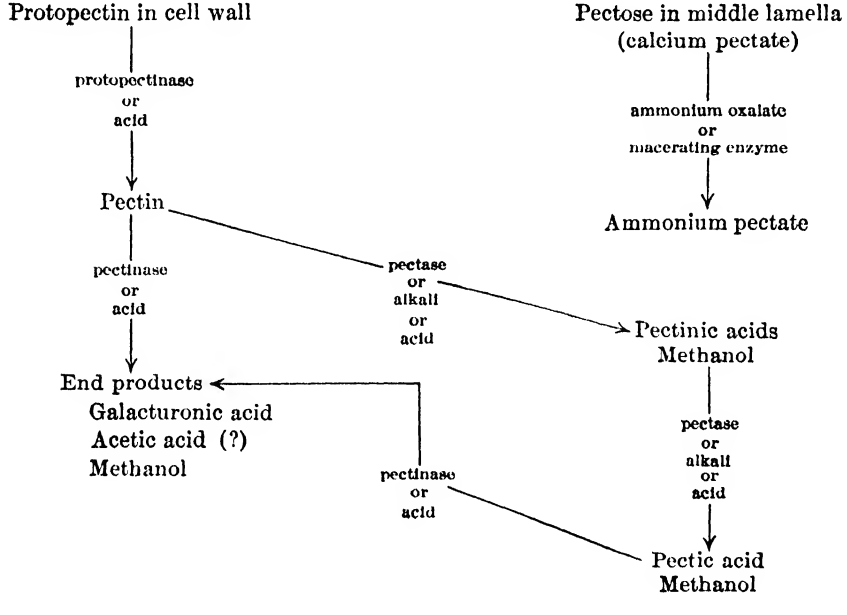
Before defining and describing these substances, it is necessary to anticipate the later discussion by stating here that any pectic compound is essentially a long chain of galacturonic acid units, probably several hundred, combined by glycosidic linkages, and hence with the carboxyl groups as branches from the chain.

Protopectin, as the name implies, is the mother substance of the group. It occurs most abundantly in the cell walls of parenchyma, such as fleshy tissues of fruits and roots. But there are two distinct places in these cell walls where pectins occur, and there is abundant evidence that the pectins in these two sites are not identical. Pectin compounds form the middle lamella, where they act as a cementing material between cells, and they likewise occur as incrustations or thickenings on the cell wall. There has been considerable shifting of opinion concerning the nature of the pectin compounds in these two regions. In the early history of pectin chemistry their identity was kept separate, and the existence of still other pectic compounds was hypothecated. Later it was thought simpler to consider all insoluble cell-wall pectin as pectose. Still later this name was changed to protopectin. And at the present time there is very good evidence to show that the cell-wall thickenings and the middle lamella are quite different in nature.[10,11] Carré and Haynes call the material forming the thickenings "pectose," and the other "middle-lamella pectin."

Both are highly insoluble in water. Their essential differences are these: The material of the wall thickenings is readily converted to soluble pectin by long-continued boiling with water, or by much shorter heating with dilute acid. The material of the middle lamella is high in calcium, is not dissolved by acid, but it is by weak alkali and by ammonium oxalate, properties very suggestive of calcium pectate. When the wall thickenings are removed, the tissue remains intact. When the middle lamella is removed, the cells separate, causing a maceration of the tissue. This brings us to a dilemma. Protopectinase, by the very sound of it, should be the enzyme that hydrolyzes protopectin to soluble pectin, and Carré [12] has shown that it does. But it has always been known as the enzyme that causes maceration—in other words, that dis-

[9] Report, *J. Am. Chem. Soc. Proc.*, **49**, 38 (1927).
[10] M. H. Carré and D. Haynes, *Biochem. J.*, **16**, 60 (1922).
[11] M. H. Carré, *Ann. Botany*, **39**, 811 (1925).
[12] M. H. Carré, *Biochem. J.*, **19**, 257 (1925).

solves the middle-lamella pectin.[13] Bonner [14] argues, and with good reason, that the two pectins in cell walls are different; that the thickenings are the precursors of soluble pectin; and that "It would seem at present that the calcium pectate theory of the middle lamella is supported by the most evidence." Perhaps the best way out of the dilemma is to think of the wall thickenings as protopectin, and of its enzyme as protopectinase, but to find another name for the macerating enzyme. The term "macerating enzyme" has been used in the accompanying scheme depicting interconversions of the pectic substances.



Pectin is the best known of all the pectic substances and is the pectin of commerce. It is a material of high molecular weight; it disperses in water to a viscous colloidal sol; it is readily precipitated from this sol by alcohol, which acts as a dehydrating agent, by lead, iron, and other heavy metal salts, but *not by salts of calcium.* Its most outstanding property is its ability to form sugar-acid-pectin gels, and such fruit jellies have been long known.

The purest preparations of pectin contain from 10 to 12 per cent of methyl alcohol combined as methyl ester groups. Demethoxylation occurs slowly when the pectin is boiled with water or dilute acid; but it occurs within a few minutes at room temperatures when pectin is treated

[13] F. R. Davison and J. J. Willaman, *Botan. Gaz.*, **83**, 329 (1927).
[14] J. Bonner, *Botan. Rev.*, **2**, 475 (1936).

with dilute alkali. In fact, this saponification constitutes a simple method for determining the methoxy content.

This deesterified pectin is *pectic acid.* Probably the purest preparations of pectic substances so far described have been of this pectic acid, because it is a simpler substance and its solubilities enable it to be purified more readily. The usual procedure is to prepare pectin by repeated precipitations with alcohol; saponify the pectin with dilute sodium hydroxide, producing soluble sodium pectate; acidify with hydrochloric acid, producing the gelatinous, insoluble pectic acid; and dry this with alcohol and ether. Or the saponification may be brought about by limewater, producing the gelatinous, insoluble calcium pectate. This gel may be freed from calcium by treating with ammonium oxalate, inasmuch as the ammonium pectate is soluble.

The *pectinic acids* are partially deesterified pectins. They thus lie between pectin and pectic acid and can have any methyl ester content between 10 and 0 per cent. It is most useful, however, to use the term for those having about 7 per cent or less of ester, since it is at about this level that pectin ceases to have normal gel-forming characteristics with sugar and acquires the capacity to *form a gel with calcium,* irrespective of sugar content.

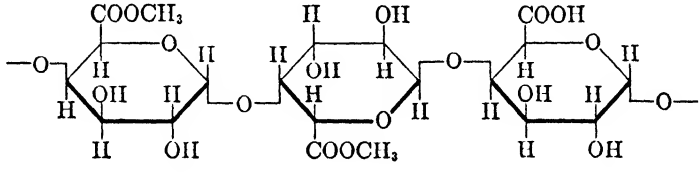
Structure of Pectic Substances. Views on the structure of pectic acid have undergone a series of changes. In 1916, Schryver and Haynes prepared pectic acid from four different sources and found that all four agreed with the empirical formula $(C_{17}H_{24}O_{16})_x$. Nanji, Paton, and Ling isolated what they consider to be the basic unit of the pectins, a substance with the empirical formula $(C_{35}H_{50}O_{33})_x$. The specific order of the units in the ring was not suggested. Ehrlich, after careful work on beet and flax pectin covering several years, found the same constituents, and in the same proportions. However, he isolated several tetragalacturonic acids of closed-ring structure, which led him to the conclusion that the sugar acids and sugars are not in the same ring.

More recently Myers and Baker have deduced evidence that pectic acid consists of two closed rings of tetragalacturonic acid joined together; or a total of eight acid groups minus nine molecules of water, with all carboxyl groups exposed. To conclude this historical summary, we will bring in two oversize words. Ehrlich believes that the pectin from sugar beets is a triacetylarabinogalactodimethoxytetragalacturonic acid. Myers and Baker submit that lemon pectin is monoarabinomonogalactodiacetylheptamethoxyloctagalacturonic acid.

It appears now that the above views are of historical interest only. Bonner [14] is convinced of a long cellulose-like chain of galacturonic acid units arranged more or less as the galactose units are arranged in galac-

tan except that carbon-6 is in a carboxyl group in the pectins instead of carrying the normal primary alcohol group characteristic of the sugars. Occasionally a galactose or an arabinose residue may replace a galacturonic acid residue in the chain, but such replacement would not disturb the normal linkage relationships. Bonner bases his argument for the "cellulose" or "galactan" chain type of structure for the pectins on (*a*) the x-ray diffraction studies of Meyer and Mark [15] and of Van Iterson [16] who have shown that the galacturonic residues are arranged in chains, and (*b*) the viscosity studies of Henglein and Schneider [17] which indicate that at least 200 galacturonic residues are present in the average chain. From viscosity measurements Owens, *et al.*,[18] deduced molecular weights of 23,000 to 71,000, and chain units from 120 to 370 for various pectinic acid preparations.

Schneider and Bock [19] suggest the following chain structure for pectin, one entirely analogous to that for cellulose. His illustration shows a portion of a chain with one free carboxyl and two others esterified.



Bonner [20] further points out that pectic acid possesses a very high charge density and that its negative charge is due almost entirely to the free carboxyl groups. In calcium pectate one calcium ion is shared between two carboxyl groups, the calcium thus serving, through electrostatic attraction, to hold two chains of the pectic acid more or less parallel to each other, so that dried calcium pectate shows orientation and in many instances is doubly refracting.

Araban and Galactan. It may be noticed that in the above discussion of the present conception of pectin structure no allowance was made for arabinose and galactose, although between them they may account for up to 29 per cent of the pectin preparation, and they have always been prominent constituents in pectin research. They occur as araban

[15] K. H. Meyer and H. Mark, *Der Aufbau der hochpolymeren organischen Naturstoffe*, Akademische Verlagsgesellschaft, Leipzig, 1930.

[16] G. van Iterson, Jr., *Chem. Weekblad*, **30**, 2 (1933).

[17] F. Henglein and G. Schneider, *Ber.*, **69B**, 309 (1936).

[18] H. S. Owens, H. Lotzkar, T. H. Schultz, and W. D. Maclay, *J. Am. Chem. Soc.*, **68**, 1628 (1946).

[19] G. G. Schneider and H. Bock, *Angew. Chem.*, **51**, 94 (1938).

[20] J. Bonner, *Proc. Acad. Sci. Amsterdam*, **38**, 346 (1935).

and galactan, and practically always are found in pectin preparations. That is not the same as saying, however, that they are true parts of the polygalacturonide molecule. This idea is currently expressed by referring to them as the "ballast" substances. Part of the ballast can be removed by repeated precipitation of the sample in 60 per cent ethanol, or by washing the sample exhaustively with the solvent. Table 57 shows the range in arabinose and galactose in five pectin preparations by Ehrlich and in lemon albedo by Myers and Baker.

TABLE 57. COMPOSITION OF SOME PECTIN PREPARATIONS

Source of Pectin	Ash, per cent	Galacturonic Acid, per cent	CH_3O, per cent	CH_3COO, per cent	L-Arabinose, per cent	D-Galactose, per cent	L-Xylose, per cent
Sugar beet	0.31	67.5	5.5	10.4	13.1	14.8	....
Flax stalk	....	61.2	4.1	8.6	10.9	13.9	10.9
Strawberry	0.53	79.7	9.7	10.5	6.6	....	....
Raspberry	0.42	75.7	11.2	10.7	14.9	....	....
Orange (albedo)	0.33	67.3	6.0	10.9	14.2	15.6	....
Lemon (albedo)	1.29	89.4	11.6	2.3	6.1	3.1	....

When pectin is deesterified to pectinic acids by hydrochloric acid, ballast is lost rapidly; when deesterified by enzyme, it is hardly lost at all.[21] Speiser and co-workers conclude that "a substantial portion of the ballast is attached to the polygalacturonide chain by primary covalent bonds (probably ester linkage) having an activation energy of 18,500 ± 4,000 calories. The remainder of the ballast is held by secondary valence forces or merely included as a physical mixture." Hirst and Jones [22] think that only secondary valences are involved. However the attachments are, their removal so far has been from pectinic and pectic acids only. Apparently no one has removed ballast to any great degree from pectin and still have the latter retain its typical gel-forming properties with 65 per cent of sugar. An indication that it may be done was furnished by Baker and Goodwin [23] who, by removing some ballast, increased the galacturonide content 13 per cent and gained 34 per cent in jelly grade.

Commercial Pectin. The introduction of commercial pectin has made possible the production, either in the home or in the factory, of jellies from fruits which contain too little pectin of their own. It has

[21] R. Speiser, C. R. Eddy, and C. H. Hills, *J. Phys. Chem.*, **49**, 563 (1945).
[22] E. L. Hirst and J. K. N. Jones, *J. Chem. Soc.*, 454 (1939).
[23] G. L. Baker and M. W. Goodwin, *Delaware Agr. Exp. Sta. Bull.* 234 (1941).

made possible the standardization of jelly making; and it eliminates the long cooking of delicately flavored fruits.

Most commercial pectin is prepared either from apple pomace or cull lemons. Table 58 shows the quantities of pectin in these and in other

TABLE 58. PERCENTAGE OF PECTIN IN RAW MATERIALS

	Fresh Material	*Dry Matter Basis*
Apple pomace	1.5–2.5	15–18
Lemon pulp	2.5–4.0	30–35
Orange pulp	3.5–5.5	30–40
Beet pulp	1.0	25–30
Carrots	0.62	7.1

materials suggested as sources of pectin.[24] The pulp, either apple or lemon, is given a preliminary washing, and is then cooked with dilute acid, either mineral or organic, usually together with sulfurous acid, to bring the pectin into solution. The filtered extract may be evaporated to a sirup and sold as liquid pectin, or the pectin may be precipitated by various means. In lemon albedo extracts the pectin is precipitated with aluminum sulfate and ammonia.[24] This precipitate is dried and the aluminum removed by washing with acidified alcohol. The final product is sold as powdered pectin. Another method is the precipitation of the pectin by alcohol. Polyphosphates have recently come into use for solubilizing pectin.[25] Since they are effective at a much lower acidity, they spare pectin grade and are less corrosive on equipment.

Pectin is usually sold according to its "jelly grade," which is the number of pounds of sucrose that 1 pound of pectin can "carry" in a jelly of standard acidity and water content. As prepared commercially, pectin has a grade of 150 to 300, but is usually standardized to 100 for sale. A grade of 514 has been attained in the laboratory, with indications of 520 as a theoretical limit for lemon pectin.[26]

The quantitative extraction of pectin from lemon albedo (white portion of the rind), and the factors affecting the jelly grade of pectin, have been exhaustively investigated by Myers and Baker.[26] The *p*H, temperature, and time of extraction are closely interdependent in controlling the yield of pectin and the jelly grade. Thus, with 30 minutes of heating, 80°C. and *p*H 1.5 are optimal. With increasing time of heating the yield of pectin increases, but the jelly grade decreases, and so on through other combinations.

[24] C. P. Wilson, *Ind. Eng. Chem.*, **17**, 1065 (1925).
[25] G. L. Baker and C. W. Woodmansee, *Fruit Products J.*, **23**, 164 (1944).
[26] P. B. Myers and G. L. Baker, *Delaware Agr. Exp. Sta. Bull.* 168 (1931).

The relative viscosity of pectin solutions is a measure of their grade up to a grade value of 350, and a simple viscosity measurement with a pipet tells the sugar-holding capacity of a fruit juice.[27] The viscosity, and hence the jelly grade, is conditioned by the degree of polymerization of the polygalacturonide chain. The methoxyl content, and that of some of the other constituents, may be considerably reduced without impairing the viscosity, provided depolymerization does not occur. Schneider and Bock [28] studied the relations among molecular weight, gel strength, and viscosity, using for the latter the nitrated pectins in acetone. Their results are given in Table 59.

TABLE 59. MOLECULAR WEIGHT, GEL STRENGTH, AND VISCOSITY OF PECTINS

Molecular Weight	*Gel Strength, grams*	*Viscosity (Nitropectin in Acetone), $\eta_{sp/C}$*
180,000	220–300	110
140,000	180–220	85
115,000	130–180	70
90,000	100–130	55
50,000	20– 50	30
30,000	No gel	20

In recent years the manufacture of pectin has steadily increased to about 2 million pounds a year, and new uses are continually sought for it. Some proposed uses are: in case-hardening of metals, as non-hygroscopic agent in dried fruit juices, in sizing of textiles, as emulsifying agent, in bakery products to delay staling, in medicine.[29] Galacturonic acid can be the source material for ascorbic acid.[30]

A gel made with pectin requires three ingredients in proper concentration—pectin, sugar (usually sucrose), and acid. The sugar must be present to the extent of 65 to 70 per cent, and the acidity must be equivalent to a *p*H of 2.8 to 3.5. This acidity is attained by the use of approximately 0.05 per cent tartaric acid, or 0.07 per cent citric acid in the final gel. The quantity of pectin required depends, of course, on its quality, as noted above. It may be as low as 0.3 and as high as 0.7 per cent.

Pectinic Acids. If the high-ester pectin discussed above is deesterified until it has 6 per cent or less of CH_3O, a stage at which about 55 per

[27] P. B. Myers and G. L. Baker, *Delaware Agr. Exp. Sta. Bull.* 149 (1927).

[28] G. G. Schneider and H. Bock, *Ber.*, **71B**, 1353 (1938).

[29] "A Symposium on Pectin and Pectic Pastes," *Bull. Nat. Formulary Committee* **9**, No. 1 (1940).

[30] R. Pasternack and P. Regna, U. S. Pat. 2,188,777 (1940).

cent of all carboxyl groups are in the free state, the resultant pectinic acid will form a gel with calcium through a wide range of acidity and independently of the presence of sugar. This deesterification can be accomplished by acid,[31] ammonia,[32] or enzyme.[33] The calcium pectinate gels are used for fruit salads, vegetable aspics, and gelled desserts of various sorts.[34] These products can be processed in the containers, whereas the products obtained with gelatin as the gelling agent cannot be.

Quantitative Estimation of Pectic Substances. The commonest and simplest method of determining pectin is to precipitate the boiled and filtered solution with alcohol. The precipitate is then washed with alcohol, dried, and weighed. This eliminates most of the proteins, but not the polysaccharides, such as starch and gums.

The method devised by Carré and co-workers [35] for the estimation of pectin is the most definite of any yet proposed. The pectin is saponified with 0.1 *N* sodium hydroxide, acidified with acetic acid, and the pectic acid precipitated as the calcium salt by means of calcium chloride. The calcium pectate is highly insoluble and can be washed rather free from impurities. As prepared from apple pectin it has a composition corresponding to the formula $C_{17}H_{22}O_{16}Ca$.

Miss Carré has extended this determination to apply to all pectic substances and has used it in extensive investigations on apples.[36] The finely pulped tissue is first extracted with cold water to remove the pectin. The residue is boiled with 0.0133 *N* hydrochloric acid for 3 to 5 periods of 3 hours each. This dissolves the protopectin occurring as thickenings on the cell walls. The dissolved material is then saponified and converted into calcium pectate as outlined above. The residue, containing now only the pectic substance of the middle lamella, is boiled with 0.0133 *N* sodium hydroxide for several periods of ½ hour each. This dissolves the middle lamella and converts it into pectic acid, which is determined as before as calcium pectate.

Another method that has been suggested for determining the pectic

[31] G. L. Baker and M. W. Goodwin, U. S. Pat. 2,133,273 (1938); A. G. Olsen and R. Stuewer, U. S. Pat. 2,132,577 (1938).

[32] R. M. McCready, H. S. Owens, and W. D. Maclay, *Food Inds.*, **16**, 794, 864, 906 (1944).

[33] J. J. Willaman, H. H. Mottern, C. H. Hills, and G. L. Baker, U. S. Pat. 2,358,430 (1944).

[34] C. W. Kaufman, E. R. Fehlberg, and A. G. Olsen, *Food Inds.*, [12] **14**, 57 (Dec. 1942), and [1] **15**, 58 (Jan. 1943).

[35] A. M. Emmett and M. H. Carré, *Biochem. J.*, **20**, 6 (1926). *Cf. Biochem. J.*, **16**, 60 (1922).

[36] M. H. Carré, *Biochem. J.*, **16**, 704 (1922). *Cf. Ann. Botany*, **39**, 811 (1925).

substances as a whole consists in measuring the carbon dioxide produced by the decarboxylation of the uronic acids when the material is boiled with 12 per cent hydrochloric acid.[37] Although this is apparently a quantitative measure of the uronic acid content, these acids are found in other substances which may occur in association with the pectins.

Pectic Enzymes.[38] Three pectin enzymes have been generally recognized:[13] *protopectinase, pectinase,* and *pectase.* The uncertainties and ambiguities clouding the action and nomenclature of protopectinase were discussed above. Carré [36] found evidence that it occurs in apples and that it converts protopectin to pectin.

Pectase is an esterase which hydrolyzes the methyl ester groups of pectin. It is found abundantly in tomato juice, citrus albedo, many leaves, especially those of alfalfa, in many other parts of higher plants, and in some fungi.

Pectinase hydrolyzes all pectin compounds, possibly including protopectin, to their simple components, galacturonic acid and possibly sugars. There is evidence [39] that pectinase can act on pectin only when the carboxyl groups are free, and hence that this enzyme must be accompanied by pectase in order to exert its action. Pectinase action on pectin sols is manifested by the lowering of viscosity, by the formation of reducing substances, and by the appearance of precipitated material, probably lignin. Pectinase preparations from molds are used commercially for the destruction of pectin in fruit juices to enhance their filtrability.[40-42] Pectinase will probably be found to be a mixture of enzymes.

Biological Significance of Pectin. The pectic substance of the middle lamella apparently functions as a cementing material between cells. During the ripening of fruit this lamella is dissolved, the cells separate from each other, and the tissue disintegrates. The retting of flax has been shown to be a bacterial dissolution of the pectic material between fibers.

Carré and Horne [43] have followed in great detail the changes in the pectic substances of apples during the after-ripening period in cold

[37] W. H. Dore, *J. Am. Chem. Soc.*, **48**, 232 (1926).

[38] For a recent review of this subject, see H. J. Phaff and M. A. Joslyn, *Wallerstein Labs. Communs.*, **10**, 133 (1947); see also Z. I. Kertesz, *Ergeb. Enzymforsch.*, **5**, 233 (1936).

[39] E. F. Jansen, L. R. MacDonnell, and R. Jang, *Arch. Biochem.*, **8**, 113 (1945).

[40] Z. I. Kertesz, *N. Y. Agr. Exp. Sta. (Geneva) Bull.* 589 (1930).

[41] J. J. Willaman and Z. I. Kertesz, *N. Y. Agr. Exp. Sta. (Geneva) Tech. Bull.* 178 (1931).

[42] A. Mehlitz, *Konserven-Ind.*, **17**, 306, 321, 729, 742, 756 (1930).

[43] M. H. Carré and A. S. Horne, *Ann. Botany*, **41**, 193 (1927).

storage. They used the chemical methods outlined above for distinguishing pectin, pectose, and the middle-lamella substance, and also microchemical methods based on the specific staining of pectic substances by ruthenium red, $Ru_2(OH)_2Cl_4 \cdot 7(NH_3) \cdot 3H_2O$. The changes which they found are shown in Fig. 111.

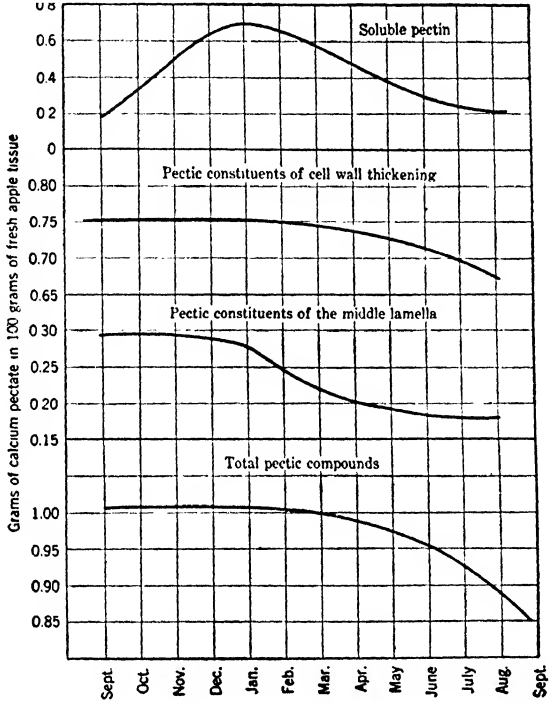


FIG. 111. The pectic changes which take place in apples from the time of picking to the last stages of senescence. (Data of Carré and Horne.)

Appleman and Conrad [44] found that during the ripening of peaches the total pectic substances remained constant, but that there was a decided decrease in protopectin and a corresponding increase in pectin. They found that in tomatoes this change is very rapid, and that the disintegration of the fruit during the canning process is the greatest when the ratio of pectin to protopectin is greatest. Kertesz [45] developed the addition of calcium to canned tomatoes to increase their firmness; the calcium probably reacted with the pectic materials of the fruit. Pectins

[44] C. O. Appleman and C. M. Conrad, *Maryland Agr. Exp. Sta. Bull.* 283 (1926); *Maryland Agr. Exp. Sta. Bull.* 291 (1927).

[45] Z. I. Kertesz, T. G. Tolman, J. D. Loconti, and E. H. Ruyle, *N. Y. Agr. Exp. Sta. Tech. Bull.* 252 (1940).

are also of significance in cucumber pickle fermentation,[46] in cotton fibers,[47] and in dehydrated fruits.[48]

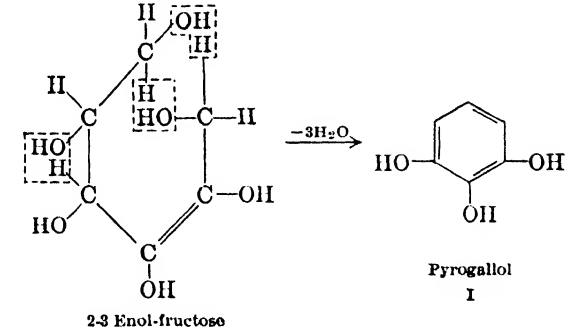
It is fitting that we conclude this discussion of pectin by quoting Bonner[49] who, writing in 1946, said, "While very considerable advances in the chemistry of pectic compounds have been made during the past eight years, no corresponding advance in our knowledge of their physiology has occurred."

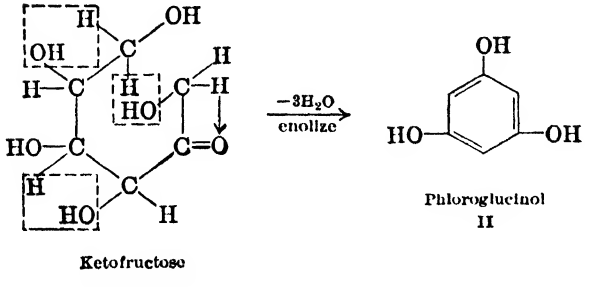
[46] F. W. Fabian and E. A. Johnson, *Michigan Agr. Exp. Sta. Tech. Bull.* 157 (1938).
[47] F. Leger and P. Larose, *Canadian J. Res.*, **19B,** 61 (1941).
[48] G. L. Baker and W. G. Murray, *Food Research*, **12,** 129 (1947).
[49] J. Bonner, *Botan. Rev.*, **12,** 535 (1946).

CHAPTER 29

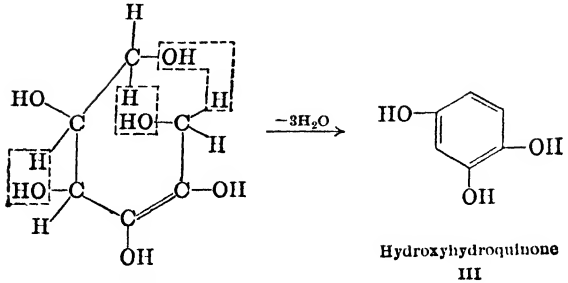
Lignin and the Tannins

Both lignin and the tannins contain polyhydroxy aromatic nuclei, so that at least in this respect they have certain features in common. Presumably both are derived in some way from the carbohydrate nucleus. Wislicenus [1] suggests that the phenolic groupings characteristic of lignin (and presumably also of the tannins) may arise by the dehydration of fructose as indicated in the following diagrams, where the end products of the dehydration are pyrogallol (I), phloroglucinol (II), and hydroxyhydroquinone (III):





[1] H. Wislicenus, *Papier-Fabr.*, **31** (Fest- und Auslandsheft), 65 (1933).



2-3-Enol-fructose

Hydroxyhydroquinone
III

LIGNIN

In 1838, Payen treated wood with nitric acid and potassium hydroxide and, besides obtaining a more or less pure cellulose, he noted that the treatment had removed some product which was richer in carbon than the residual cellulose. This material he designated as "matières encrustants," although the term *lignin* had been applied to this material in 1833 by de Candolle. Lignin is still often referred to as a cellulose-encrusting material. Since these early observations a voluminous literature has been built up in the field of lignin chemistry,[2] and there is still considerable controversy with regard to the actual structure of the lignin molecule or molecules, for it appears that there may be several lignins.

Color Reactions. Many reagents, both organic and inorganic, have been used for many years in testing for lignin. Of the organic reagents, mainly phenols and amines, the one most used is phloroglucinol which gives, in hydrochloric acid, a bright purple color with lignified material. The Mäule test gives a bright red color with deciduous wood and an indefinite brown with coniferous woods. A sensitive test for lignin consists in treating the lignified material with chlorine, followed by a

[2] The reader is referred to the following sources for essential literature: W. Fuchs, *Die Chemie des Lignins*, Julius Springer, Berlin, 1926; K. Hess, *Die Chemie der Zellulose und ihrer Begleiter*, Akademische Verlagsgesellschaft, Leipzig, 1928, pp. 129–199; E. Hägglund, *Holzchemie*, Akademische Verlagsgesellschaft, Leipzig, 1939, pp. 126–230; H. Hibbert, *et al.*, *Can. J. Research*, **2**, 357 (1930), **3**, 65, 130 (1930), **4**, 110, 119, 240 (1931), **5**, 1, 302 (1931), **13**, 28, 110 (1935), **14**, 12, 55, 115, 404 (1936), **15**, 38, 532 (1937), **16**, 54 (1938), *J. Am. Chem. Soc.*, **55**, 4720 (1933), **57**, 528 (1935), **58**, 340 (1936), **59**, 125, 2447 (1937), **60**, 565 (1938), **61**, 509, 725, 865, 1477, 2190 (1939), **62**, 1412, 2065, 2149, 2284, 2803 (1940), **63**, 2371, 3031 (1941), **65**, 1170, 1862 (1943), **66**, 26, 598 (1944), *Ber.*, **71B**, 734 (1938); A. G. Norman, *The Biochemistry of Cellulose, the Polyuronides, Lignin, etc.*, Clarendon Press, Oxford, England, 1937; E. Ott, *High Polymers*, Vol. V, Cellulose and Cellulose Derivatives, Interscience Publishers, Inc., New York, 1943, p. 447; L. E. Wise, *Wood Chemistry*, Reinhold Publishing Corp., New York, 1944, pp. 272, 634.

dilute solution of sodium sulfite. A pink to purple-red coloration is produced, the intensity of color varying approximately with the amount of lignin present.

The Extraction and Estimation of Lignin. The two methods which have been most generally used for the isolation and the quantitative estimation of lignin are the 72 per cent sulfuric acid method and the fuming hydrochloric acid method. When a woody tissue containing lignin is treated with 72 per cent sulfuric acid under rather definite experimental conditions [3-6] the cellulose and other structural elements are dissolved or rendered water-soluble, and on dilution with water the lignin remains as an insoluble precipitate which can be filtered, washed free of sulfuric acid, dried, and weighed, and thus the lignin content of the tissues can be estimated. An alternative method is the use of fuming hydrochloric acid.[7-9]

Alkalies, alkaline salts, certain acids and acid salts, and alcohols, on the contrary, dissolve the lignin and leave the cellulose as the insoluble residue. The pulping of wood is almost wholly concerned with bringing the lignin into a soluble form, so that the cellulose fibers can be recovered essentially lignin-free, although most paper pulps contain from 1 to 5 per cent of lignin. Sodium hydroxide, sodium sulfide, sodium sulfite, and lime, either alone or in various combinations, and sulfurous acid and acid bisulfites are the chemicals most used for the commercial production of cellulose pulps. The use of nitric acid for the production of high-grade cellulose from plant materials was the subject of many investigations since Payen's time;[10] during World War II this process was used on a large scale on beechwood in Germany, albeit under uneconomical conditions. Delignification of plant materials with alcohols and other organic solvents has aroused interest in recent years.[11-14] In pulp-

[3] G. J. Ritter, R. M. Seborg, and R. L. Mitchell, *Ind. Eng. Chem., Anal. Ed.*, **4**, 202 (1932).

[4] C. J. Peterson, A. W. Walde, and R. M. Hixon, *Ind. Eng. Chem., Anal. Ed.*, **4**, 216 (1932).

[5] E. C. Sherrard and E. E. Harris, *Ind. Eng. Chem.*, **24**, 103 (1932).

[6] E. C. Dryden, J. D. Reid, and S. I. Aronovsky, *Paper Trade J.*, [11] **119**, 119 (Sept. 14, 1944).

[7] E. Hägglund and C. B. Björkman, *Biochem. Z.*, **147**, 74 (1924).

[8] M. Phillips, *J. Assoc. Offic. Agr. Chem.*, **15**, 118 (1932).

[9] M. J. Goss and M. Phillips, *J. Assoc. Offic. Agr. Chem.*, **19**, 341 (1936).

[10] S. I. Aronovsky, J. D. Reid, E. C. Dryden, E. R. Whittemore, and D. F. J. Lynch, *Paper Ind. and Paper World*, **21**, 41, 151, 335 (1939).

[11] F. E. Brauns and H. Hibbert, *Can. J. Research*, **13**, 28 (1935).

[12] S. I. Aronovsky and R. A. Gortner, *Ind. Eng. Chem.*, **28**, 1270 (1936).

[13] H. Erbring and H. Geinitz, *Kolloid-Z.*, **84**, 215 (1938).

[14] H. Lau, *Paper Ind. and Paper World*, **23**, 247 (1941).

ing wood with a mixture of *n*-butanol and water, practically all the extracted lignin is found in the alcohol phase of the binary liquor and can be recovered upon removal of the alcohol.

It is generally agreed that lignin is probably changed during the process of isolation. It has even been claimed that lignin is not present as such in the plant material, but that it is a secondary reaction product of the action of strong mineral acids on partly methylated carbohydrates.[15,16] The aromatic nature of lignin, however, and the extraction of a small amount of "native lignin" [17] from wood by neutral solvents such as alcohol and dioxane, do not substantiate such a claim. The difficulty with problems involving the chemistry of lignin lies in the fact that lignin is not a chemical entity, that it is amorphous, that it forms colloidal solutions in those media in which it is dispersible, and that until the lignin is pretty thoroughly broken up it does not yield crystalline compounds which can be readily characterized.

Lignin Derivatives. Lignin, as isolated, contains methoxyl groups, the methoxyl value ranging from 9.9 to 21.0 per cent, depending in part on the botanical source from which the lignin is derived but still more on the method used for its isolation.[18] Lignin can be further alkylated up to a maximum of at least 32.4 per cent —OCH_3 groups by appropriate chemical methods.[19]

Lignin, as isolated, can be acetylated with either acetic anhydride or acetyl chloride, and the literature reports acetylated lignin containing acetyl groups in amounts ranging from 19.85 to 37.85 per cent. Similarly it can be benzoylated to yield benzoyl derivatives, and various other aromatic and aliphatic acid chlorides have been used to introduce the desired groupings. Stearic esters prepared from isolated, hardwood alkali lignin have been proposed as mold lubricants for plastics and as flow modifiers for inks and paints.[20] The monobasic esters of alkali lignin are more or less soluble in a number of organic solvents, whereas esters prepared from lignin isolated with sulfuric acid are fairly insoluble in these solvents.[21] This is one more indication that isolated lignins dif-

[15] R. S. Hilpert, *Cellulosechemie*, **16**, 92 (1935).

[16] R. S. Hilpert and H. Hellwage, *Ber.*, **68B**, 380 (1935); *Cellulosechemie*, **17**, 25 (1936).

[17] F. E. Brauns, *J. Am. Chem. Soc.*, **61**, 2120 (1939); *Paper Trade J.*, [14] **111**, 33 (Oct. 3, 1940).

[18] L. E. Wise, *Wood Chemistry*, Reinhold Publishing Corp., New York, 1944, p. 297.

[19] F. E. Brauns and H. Hibbert, *Can. J. Research*, **13**, 78 (1935).

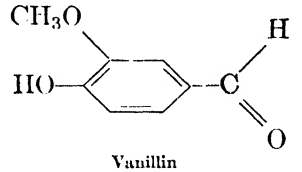
[20] H. F. Lewis, F. E. Brauns, M. A. Buchanan, and E. B. Brookbank, *Ind. Eng. Chem.*, **35**, 1113 (1943).

[21] J. J. McNair and E. C. Jahn, *Paper Trade J.*, [8] **117**, 29 (Aug. 19, 1943).

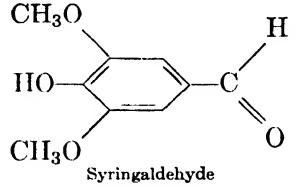
fer according to the methods used in their preparation. Lignin ethers have also been prepared.[22]

Chlorine reacts with lignin to form chlorolignins which are soluble in alkali, glacial acetic acid, and alcohol. The chlorination of lignin was first studied by Bevan and Cross,[23] and it is the basis of the Pomilio process [24] for pulping plant materials. Lignin may also be brominated. Halogenation of lignin results in some loss of methoxyl. Lignin is nitrated very easily to form nitro-derivatives, one "nitrolignin" having the approximate composition $C_{42}H_{37}N_3O_{24}$.[25] In the pulping of wood by the sulfite process, lignin is sulfonated to yield lignosulfonic acid. α- and β-Sulfonic acids have been described which agree more or less with the formula $C_{26}H_{30}SO_{12}$. In all probability, however, the compounds which have been described are mixtures of several substances.

On oxidation lignin yields a great variety of products, mostly aliphatic and aromatic acids. Among the compounds which have been reported are formic, acetic, oxalic, succinic, fumaric, malonic, adipic, benzoic, phthalic, isophthalic, mellitic, hemimellitic, trimellitic, and anisic acids. Vanillin (IV) has been obtained in yields as high as 25 per cent of the lignin in coniferous wood, isolated lignin, or lignosulfonic acid.[26] Under similar conditions beechwood and other deciduous woods yield syringaldehyde (V) in addition to vanillin.[27] On reduction with hydriodic acid



Vanillin
IV



Syringaldehyde
V

and phosphorus a mixture of liquid and solid hydrocarbons, ranging in molecular weight from 167 to 842, has been obtained. No pure hydrocarbon was identified; however, no *n*-hexyliodide was obtained, indicating that the straight 6-carbon chain of the sugars was absent.

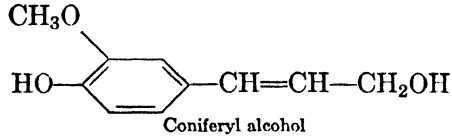
Catalytic hydrogenation of lignin, with a copper chromite catalyst, in dioxane at 250° to 260° and under 170 to 400 atmospheres of hydrogen produced methanol, 4-*n*-propyl-1-cyclohexanol, 4-*n*-propyl-1,2-

[22] F. E. Brauns, H. F. Lewis, and E. B. Brookbank, *Ind. Eng. Chem.*, **37**, 70 (1945).
[23] E. J. Bevan and C. F. Cross, *J. Chem. Soc.*, **38**, 666 (1880).
[24] U. Pomilio, *Chimie & industrie*, **11**, 1091 (1924); *Ind. Eng. Chem.*, **31**, 657 (1939).
[25] F. Fischer and H. Tropsch, *Ges. Abhandl. Kenntnis Kohle*, **6**, 279 (1921).
[26] K. Freudenberg, W. Lautsch, and K. Engler, *Ber.*, **73B**, 167 (1940).
[27] R. H. J. Creighton, J. L. McCarthy, and H. Hibbert, *J. Am. Chem. Soc.*, **63**, 312 (1941).

cyclohexanediol, and 3-(4-hydroxycyclohexyl)-1-propanol,[28,29] With Raney nickel as the catalyst the hydrogenation can be carried out in aqueous or alkaline solution or suspension, yielding similar products.[30,31] With nickel oxide as the catalyst and dioxane as the solvent, 1-*n*-propyl-3-methoxy-4-hydroxybenzene, protocatechuic acid, pyrocatechol, and *p*-hydroxybenzoic acid [32] were identified in the ether-soluble portion of the hydrogenated product.

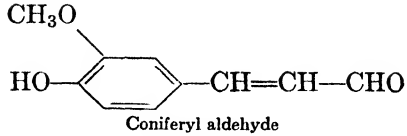
Distillation with zinc dust has yielded guaiacol, catechol, and 1-*n*-propyl-3-methoxy-4-hydroxybenzene. Fusion with alkalies has yielded oxalic acid and other aliphatic acids, catechol, protocatechuic acid, vanillic acid, and "lignic acid," a dark, amorphous, humin-like product. Dry distillation has yielded acetone, methyl alcohol, acetic acid, eugenol, 5-hydroxymethylfurfural, phenol, *o*-cresol, 1-vinyl-3-methoxy-4-hydroxybenzene, and 1-*n*-propyl-3-methoxy-4-hydroxybenzene. Dry distillation has also yielded a series of saturated and unsaturated hydrocarbons. The saturated hydrocarbons isolated had empirical formulas $C_{13}H_{26}$, $C_{14}H_{26}$, $C_{16}H_{30}$, $C_{24}H_{44}$, and $C_{30}H_{60}$. The unsaturated hydrocarbons had empirical formulas $C_{11}H_{16}$, $C_{12}H_{16}$, and $C_{13}H_{16}$.

The Structure of Lignin. In the final determination of the structure of lignin, all the above compounds must be taken into consideration. A number of structural formulas for lignin have been proposed by Schrauth,[33] Kürschner,[34] Pavolini,[35] Fuchs,[36] and others, but they are not in complete harmony with the known chemical reactions of lignin. Most of the recent workers agree that coniferyl alcohol (VI), coniferyl aldehyde (VII), or some closely related compound, as postulated by Klason,[37] is probably the essential building unit in lignin.



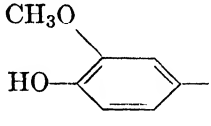
Coniferyl alcohol
VI

[28] E. E. Harris, J. D'Ianni, and H. Adkins, *J. Am. Chem. Soc.*, **60**, 1467 (1938).
[29] E. E. Harris and H. Adkins, *Paper Trade J.*, [20] **107**, 38 (Nov. 17, 1938).
[30] E. E. Harris, J. Saeman, and E. C. Sherrard, *Ind. Eng. Chem.*, **32**, 440 (1940).
[31] J. F. Saeman and E. E. Harris, *J. Am. Chem. Soc.*, **68**, 2507 (1946).
[32] Y. Hatihama, S. Zyodai, and M. Umezu, *J. Soc. Chem. Ind. Japan*, **43** (Suppl.), 127 (1940).
[33] W. Schrauth, *Z. angew. Chem.*, **36**, 149 (1923).
[34] K. Kürschner and W. Schramek, *Tech. Chem. Papier- u. Zellstoff-Fabr.*, **29**, 35 (1932).
[35] T. Pavolini, *Industria chimica*, **6**, 1367 (1936).
[36] W. M. Fuchs, *Z. angew. Chem.*, **44**, 111 (1931); *J. Am. Chem. Soc.*, **58**, 673 (1936).
[37] P. Klason, *Svensk Kem. Tid.*, **29**, 5, 47 (1917); *Ber.*, **53B**, 706, 1864 (1920).

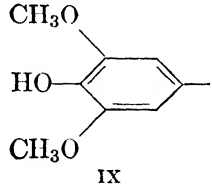


Coniferyl aldehyde
VII

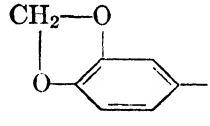
Recent work by Freudenberg and by Hibbert and their co-workers indicates that lignin may be considered a polymerized product derived from phenyl propane building units, R—C—C—C—, where R is a guaiacyl (VIII) or syringyl (IX) radical. Freudenberg and his co-workers [38-42] consider that R may also be a piperonyl (X) group. He postulates that the three-carbon chain may be (XI), (XII), or (XIII),



VIII

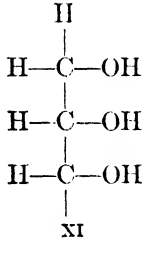


IX

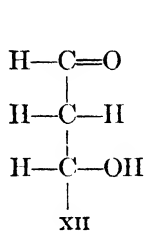


X

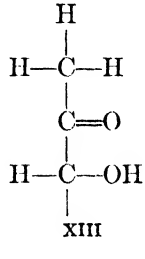
and that these are attached to the aromatic nuclei, (VIII), (IX), or (X). Only (VIII) and (X) are present in spruce lignin, whereas hardwood lignin also contains (IX).



XI



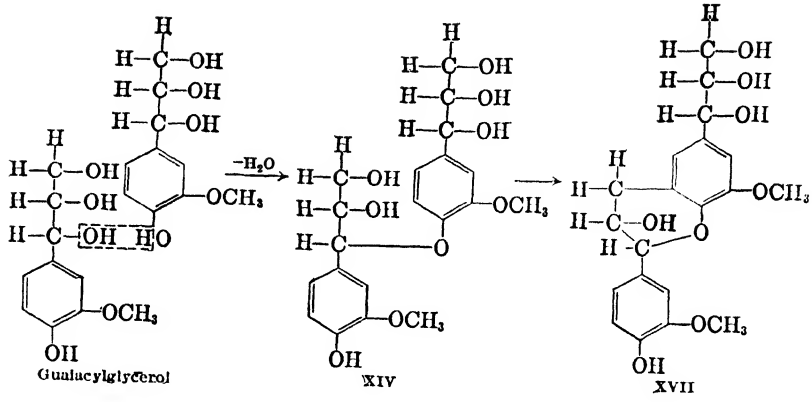
XII

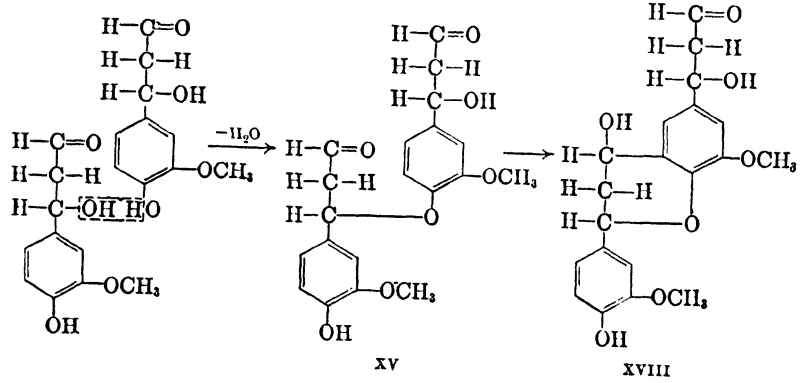


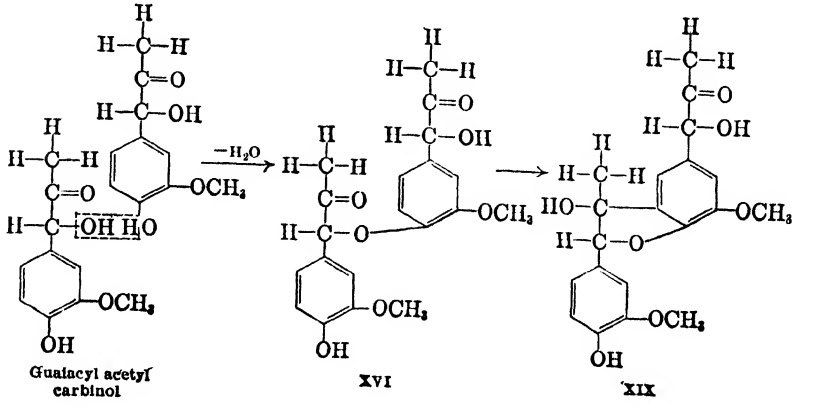
XIII

Lignin may be regarded as the product resulting from the condensation of two of these units in such a way that ether linkages (XIV, XV, XVI) are formed. These compounds then undergo nuclear condensation, with formation of chroman rings (XVII, XVIII) and furan rings (XIX).

[38] K. Freudenberg, M. Harder, L. Markert, and E. Spiess, *Ber.*, **61B**, 1760 (1928).
[39] K. Freudenberg, H. Zocher, and W. Dürr, *Ber.*, **62B**, 1814 (1929).
[40] K. Freudenberg, *J. Chem. Education*, **9**, 1171 (1932); *Tannin, Cellulose, Lignin*, J. Springer, Berlin, 1933; *Ann. Rev. Biochem.*, **8**, 81 (1939).
[41] K. Freudenberg, M. Meister, and E. Flickinger, *Ber.*, **70B**, 500 (1937).
[42] K. Freudenberg, F. Klinck, E. Flickinger, and A. Sobek, *Ber.*, **72B**, 217 (1939).

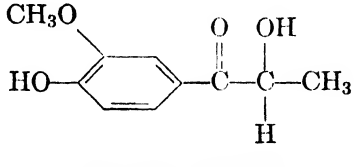

−H2O
Guaiacylglycerol
XIV
XVII


−H2O
XV
XVIII

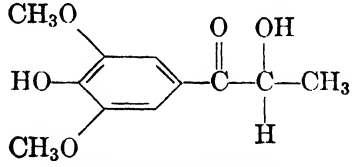

−H2O
Guaiacyl acetyl carbinol
XVI
XIX

The condensation products (XVII, XVIII, XIX) may undergo further condensation with other building units in a similar manner to form complex polymers of high molecular weight. Building units containing the syringyl (IX) and piperonyl (X) nuclei may undergo condensation in a manner analogous to that illustrated for the guaiacyl (VIII) containing unit, but the structure of the condensation products, obviously, could not be the same.

Hibbert and his co-workers [43-47] isolated phenyl propane derivatives from the aqueous solutions of the ethanolysis products of wood, including vanillin, syringaldehyde, vanilloyl methyl ketone, syringoyl methyl ketone, α-ethoxypropiovanillone, and α-ethoxypropiosyringone. The first four compounds may be true scission products of lignin, and the last two must have been formed by ethylation of the corresponding hydroxy compounds. The isolation of these products lends experimental support to Freudenberg's views on lignin structure. According to Hibbert,[48] α-hydroxypropiovanillone (XX) and α-hydroxypropiosyringone (XXI) are very sensitive compounds and tend to undergo,



α-Hydroxypropiovanillone
XX



α-Hydroxypropiosyringone
XXI

under the influence of mild chemical reagents, intramolecular dismutation transformations involving hydrogen migrations and oxidation-reduction systems, the intermediate product being an ene-diol (see p. 746).

These derivatives, containing a phenol group and a carbonyl group or an unsaturated ethylene linkage, have a strong tendency to undergo para- and ortho-nuclear condensations under very mild conditions, with the formation of condensation polymers.

It can be seen that the condensation product (XXIII) contains a

[43] M. J. Hunter, A. B. Cramer, and H. Hibbert, *J. Am. Chem. Soc.*, **60**, 2815 (1938); **61**, 516 (1939).

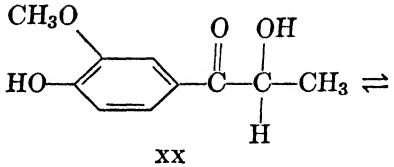
[44] J. J. Pyle, L. Brickman, and H. Hibbert, *J. Am. Chem. Soc.*, **61**, 2198 (1939).

[45] A. S. MacInnes, E. West, J. L. McCarthy, and H. Hibbert, *J. Am. Chem. Soc.*, **62**, 2803 (1940).

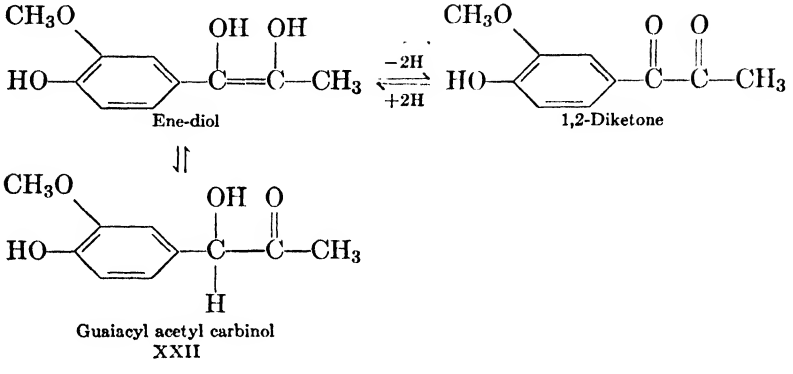
[46] M. Kulka, W. L. Hawkins, and H. Hibbert, *J. Am. Chem. Soc.*, **63**, 2371 (1941).

[47] J. H. Fisher, W. L. Hawkins, and H. Hibbert, *J. Am. Chem. Soc.*, **63**, 3031 (1941).

[48] H. Hibbert, *J. Am. Chem. Soc.*, **61**, 725 (1939); *Tech. Assoc. Papers*, **24**, 492 (1941); *Ann. Rev. Biochem.*, **11**, 183 (1942).

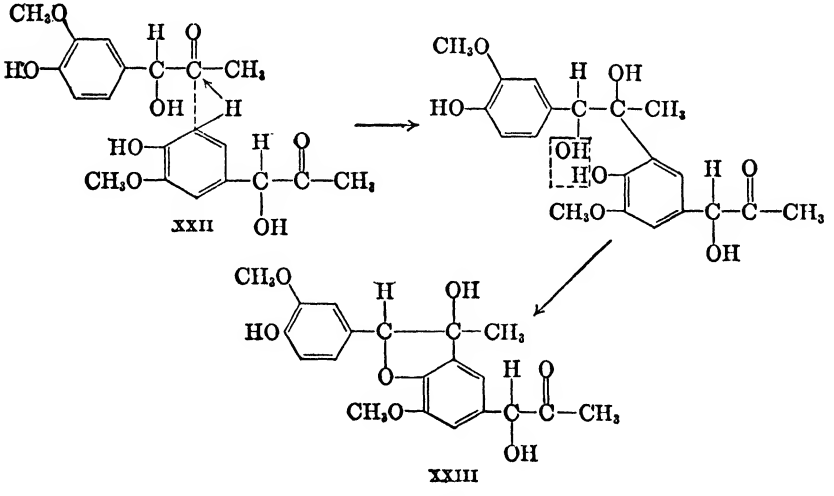


XX



XXII

furan ring and is structurally similar to the product (XIX) postulated by Freudenberg. The Hibbert mechanism differs from Freudenberg's only in the order in which the various steps are assumed to occur. An-



XXII XXIII

other type of mechanism must be postulated for the formation of lignin containing the syringyl-building unit, since both positions ortho to the phenol hydroxyl are blocked.

Phillips [49] suggests that the fundamental units are guaiacol and *n*-

[49] M. Phillips, *Science*, **73**, 568 (1931).

hydroxypropylguaiacol, "where the hydrogens in the *n*-propyl sidechain are substituted by alcohol hydroxyl groups." Ehrlich [50] does not give a structural formula for lignin but suggests that pectin is the precursor of lignin. He notes that lignified tissues contain cellulose and hemicelluloses but little or no pectin, and that non-lignified tissues contain relatively little hemicelluloses but large amounts of pectin.[51] He proposes the following equation as a working hypothesis for the formation of lignin in the flax plant:

$$C_{46}H_{68}O_{40} \rightarrow C_{45}H_{48}O_{16} + CO_2 + 10H_2O + 6O_2$$

Brauns [17] suggests that the "native lignin"-building unit is a condensation product of five phenylpropane building stones. The building unit has four methoxyl groups, four hydroxyl groups one of which is phenolic, and an easily enolizable carbonyl group. The empirical formulas for "native lignin," which give all of the typical lignin reactions, are:

$$C_{41}H_{32}O_6(CO)(OCH_3)_4(OH)_4$$

for the keto form, and

$$C_{41}H_{31}O_6(COH)(OCH_3)_4(OH)_4$$

for the enol form, with a molecular weight of about 840.

It is evident from the foregoing discussion that no clear-cut picture of the molecular structure of lignin is available as yet. This is due, largely, to the difficulties in isolating lignin in a chemically unmodified form and consequently to the incompleteness of our knowledge of the chemistry of this material. The weight of evidence, however, seems to favor the phenylpropane structures advocated by Freudenberg, Hibbert, and their co-workers.

The Function of Lignin. The function of lignin in the plant is apparently to give strength and rigidity to the cell wall. The unlignified tissue is usually soft, contains a higher percentage of water, and has a much lower breaking strength. Hibbert believes that the ene-diol, 1,2-diketone, oxidation-reduction systems postulated by him may function in plant respiration, according to the Szent-Györgyi theory that oxidation-reduction systems of the ene-diol type must be present irrespective of any other systems which may function in plant respiration.[52] Thus lignin may be considered to be "derived from a series of mono-

[50] F. Ehrlich, *Cellulosechemie*, **11**, 161 (1930).

[51] Pectic substances having some of the properties of pectic acid have been found recently in five softwoods, in amounts up to 0.5 per cent, according to E. Anderson, *J. Biol. Chem.*, **165**, 233 (1946).

[52] A. Szent-Györgyi, *Ber.*, **72A**, 53 (1939).

molecular hydrogen-transporting plant catalysts, which during the life of the plant and in the subsequent post-mortal stage, undergo conversion into more complex condensation polymer products." [48]

Lignin is extremely resistant to decay and bacterial action; woods which have been buried for at least 500,000 years still contain the great bulk of the original lignin in an apparently unaltered state.[53] Weathering and decay of plant material by microorganisms result in an apparent increase in lignin content, owing to preferential decomposition of the carbohydrate constituents of the material.[54] In fact, lignin in the natural plant material has a retarding effect on the bacterial decomposition of the carbohydrates.[55] Certain fungi, however, such as *Trametes Pini, Polystictus hirsutus, Fomes Pini, Polyporus juniperinus,* and a few other *Basidiomycetes,* attack lignin, either preferentially or along with the other wood constituents, cellulose and hemicellulose. The high resistance of lignin to bacterial decomposition, especially under anaerobic conditions, led to a suggestion that the resistant lignin, accumulated in the decomposition of plant residues, forms the source of peat and coal.[56]

We have already referred to the fact that Payen called lignin an encrusting material. For many years there has been an extremely lively controversy as to whether lignin is simply an encrustant surrounding the cellulose fibers and cementing them together or whether it is actually combined with the cellulose. It appears [12, 57] that both views are correct. Extraction with aqueous *n*-butanol solutions, at temperatures sufficiently high to produce complete miscibility of the water and alcohol, removed practically all the lignin from aspen wood, but only about two-thirds of that present in jack pine wood. It thus appears that all the lignin in aspen and two-thirds of the lignin in jack pine may be considered as encrustant, whereas the more resistant portion of the jack pine lignin may be chemically combined. Therefore, the question of the absence or presence of a chemical union between lignin and the other constituents of plant material seems to depend on the botanical species of the lignified tissue under investigation.

Lignin in Soil. Natural humus originates from the decomposition of plant and animal residues in the soil by microorganisms, under aerobic and anaerobic conditions. Since lignin is more resistant to this type of

[53] W. A. Gortner, *J. Am. Chem. Soc.*, **60**, 2509 (1938).

[54] S. I. Aronovsky, J. O. Younger, and G. H. Nelson, *Paper Trade J.*, [8] **120**, 124 (Feb. 22, 1945).

[55] S. A. Waksman and T. C. Cordon, *Soil Sci.*, **45**, 199 (1938).

[56] F. Fischer and H. Schrader, *Brennstoff-Chem.*, **2**, 37 (1921).

[57] A. J. Bailey, *Paper Trade J.*, [6] **111**, 27 (Aug. 8, 1940).

attack than the plant and animal carbohydrates, it follows that lignin and lignin complexes form a major portion of the humus. These constituents are also predominant in animal manures.[58] The base-exchange ability of the soil organic matter is largely due to the ability of lignin to adsorb cations [59] and is a linear function of the lignin content of the soil.[60] Recent experiments on the addition of isolated lignin to soil used for potato culture resulted in an increased yield of tubers with higher starch content.[61] The addition of calcium lignosulfonate (from sulfite waste liquor) to soil increased the yield of lima beans.[62] Clarification of the chemistry and structure of the lignin molecule will, therefore, go a long way toward increasing our knowledge of the chemistry of soils.

Lignin in Animal Metabolism. We have already indicated that lignin is extremely resistant to decay and bacterial action, and it seems probable that it contributes little, if anything, to the energy requirements of animals. When it is fed to dogs or cows, there is a partial breakdown of the lignin in the animal body, as evidenced by an increase in the amount of benzoic acid eliminated as hippuric acid [63] and also a loss in methoxyl content of the lignin as re-isolated from the feces.[64] Lignin in untreated straw and in straw pulp was not digested by sheep.[65] Accordingly, such breakdown of lignin as does occur is probably not due to bacterial action.

THE TANNINS

Tannin is a generic name for a group of substances widely distributed in the higher plants and showing certain characteristic physical and chemical properties. Tannins may be found in leaves, tea (15 per cent), sumac, *Rhus coriaria* (13–15 per cent), rhododendron, etc. They occur also in wood and stems, especially in the bark of the oaks, hemlock, etc.; in fruits, especially unripe fruits, *e.g.*, persimmons, plums, hulls of walnuts and of butternuts and hickory nuts; in seeds (more rarely), but especially in the seed coats; in algae, fungi, and pathological plant growths. Plant galls may contain as much as 75 per cent tannin and rarely contain less than 25 per cent.

[58] S. A. Waksman, *Humus*, The Williams and Wilkins Co., Baltimore, 1936, p. 185.

[59] G. H. Nelson and S. I. Aronovsky, *Paper Trade J.*, [19] **113**, 31 (Nov. 6, 1941).

[60] W. T. McGeorge, *J. Am. Soc. Agron.*, **26**, 575 (1934).

[61] S. Dunn, J. Seiberlich, and D. S. Eppelsheimer, *Northeastern Wood Utilization Council* (New Haven, Conn.) *Bull.* 7, 21 (Oct. 1945).

[62] V. Sauchelli, *Am. Fertilizer*, [3] **104**, 7 (1946).

[63] F. A. Csonka, M. Phillips, and D. B. Jones, *J. Biol. Chem.*, **85**, 65 (1929).

[64] M. Phillips, H. Weihe, D. B. Jones, and F. A. Csonka, *Proc. Soc. Exptl. Biol. Med.*, **26**, 320 (1929).

[65] W. S. Ferguson, *Biochem. J.*, **36**, 786 (1942).

Only one instance is recorded of a tannin from an animal source. Three per cent of a substance was extracted from corn weevils which, on hydrolysis with dilute acids, yielded dextrose, gallic acid, and a red "phlobaphene." The existence of a true animal tannin may well be doubted, for this product may have been derived from undigested food remains.

Tannins may be characterized by the following properties:

1. They are, in general, amorphous, rarely crystalline.
2. They have an astringent taste.
3. They give colors (inks) with ferric salts.
4. They are precipitated from solution by potassium dichromate, lead acetate, and by alkaloids.
5. They precipitate gelatin from solution, which property enables them to convert hide into leather.
6. Their sols develop a deep red color on the addition of potassium ferricyanide.
7. They are readily soluble in hot water to form solutions which are in reality colloidal sols.
8. They all contain polyhydroxyphenols or derivatives of polyhydroxyphenols, in many instances in complex, condensed ring structures. A part of the tannins is hydrolyzable with acids to yield phenolic derivatives and in some instances sugars, usually D-glucose. Sugars, however, are not an essential part of the tannin molecule, and, when the sugar is removed by appropriate means, the residue which remains still shows the properties characteristic of tannins as a class. Alkali fusion and dry distillation of the tannins give decomposition products that are principally phenolic in character—catechol, pyrogallol, phloroglucinol, resorcinol, and hydroquinone, or their corresponding acids protocatechuic acid, pyrogallic acid, resorcylic acid, etc.
9. They may act as a chromogen for oxidases, *e.g.*, the green walnut hulls when broken open darken at once, as the result of oxidation of tannin by oxidases or even by exposure to air.
10. When tannins are heated with dilute acids, insoluble amorphous anhydrides or *phlobaphenes* (apparently anhydrides of the tannin) are produced in addition to other hydrolysis products. These phlobaphenes are produced by any process tending to cause the tannin to lose water. They are red or brown substances, practically insoluble in water, chemically relatively inert, and they occur in nature only associated with the tannins.

Various classifications have been proposed for the tannins. The classifications based on the blue or green colors produced by the tannins with ferric salts are without meaning; such colors depend mainly on the

orientation of the free phenolic hydroxyl present. Perkin and Everest [66] divided the tannins into three groups.

I. Tannins related to depsides (the carboxyl group of one phenolic acid combined with the hydroxyl group of another to form an ester linkage).

II. Tannins related to diphenyldimethylolid (tannins derived from ellagic acid).

III. Phlobaphene-producing tannins (phlobatannins).

A more logical classification, originally proposed by Freudenberg,[67] has been used by Nierenstein: [68]

I. Condensed tannins
   A. Catechin tannins
      Acacatechin tannins
      Isoacacatechin tannins
      Gambir catechin tannins
   B. Maclurin tannins
II. Hydrolyzable tannins
   A. Gallotannins
   B. Ellagitannins
   C. Caffetannins
III. Unclassified tannins

The Condensed Tannins. The condensed tannins all contain the *phloroglucinol nucleus*. They cannot be hydrolyzed by acids or enzymes, and they apparently represent polymers of relatively simple polyhydroxy compounds containing aromatic nuclei. The chief tannins of commerce which fall in this class are Indian cutch, cube gambir, and quebracho tannin. Indian cutch is the tannin extracted from *Acacia catechu*, *A. catechuoides*, and *A. sundra*. Cube gambir is the tannin in the extract from *Uncaria gambir*, *U. acida*, *U. dasyoneura*, *U. bernaysii*, and *U. lanosa*. Quebracho tannin is extracted from the heartwood of the quebracho tree, *Quebrachia lorentzii*. Indian cutch and cube gambir have been articles of commerce and used in medicine since the dawn of European history. They were introduced into Europe from the Far East and are referred to by Pliny in his *Historia Naturalis*. However, only within the last thirty years have the structural problems presented by these tannins been elucidated by the brilliant researches of Freudenberg and Nierenstein. Indian cutch is the condensation product of

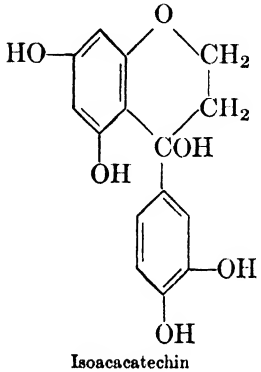
[66] A. G. Perkin and A. E. Everest, *The Natural Organic Colouring Matters*, Longmans, Green and Co., London, 1918.

[67] K. Freudenberg, *Die Chemie der natürlicher Gerbstoffe*, J. Springer, Berlin, 1920; *Tannin, Cellulose, Lignin*, J. Springer, Berlin, 1933.

[68] M. Nierenstein, *The Natural Organic Tannins*, The Sherwood Press, Cleveland, Ohio, 1935.

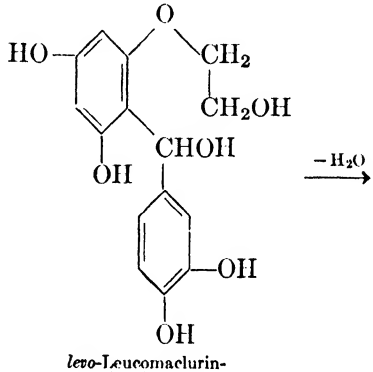
acacatechin (XXIV) with some *isoacacatechin* (XXV), whereas gambir contains the *d-catechin* (XXVI) nucleus.

It will be noted that, in the formulas of various catechins, there are two asymmetric carbon atoms which can account for a number of stereoisomers. Nierenstein notes that eight such isomers have been identified as occurring in nature. He observed that in *Acacia catechu* the catechin occurs in the heartwood, so he concluded that it must be a final product of plant metabolism. In the young twigs he discovered *levo*-leucomacluringlycol ether (XXVII), and he shows how this com-



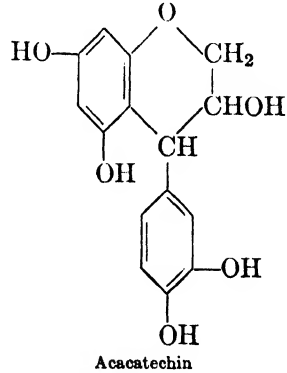
Isoacacatechin

XXV



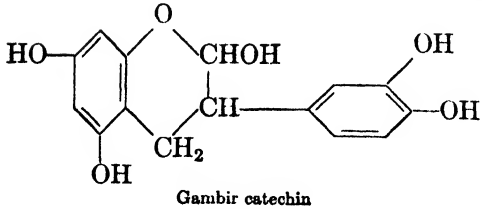
levo-Leucomaclurin-glycol ether

XXVII



Acacatechin

XXIV



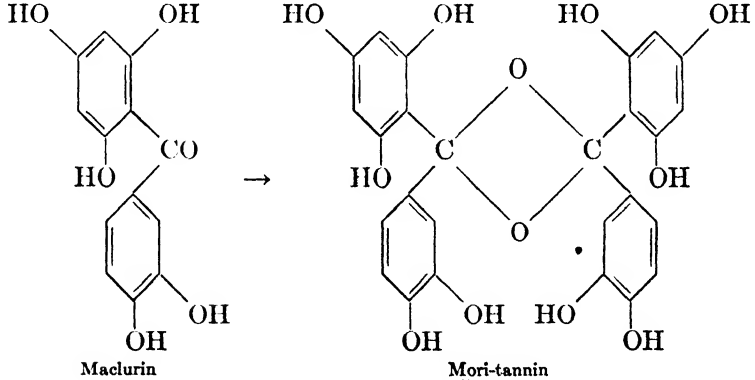
Gambir catechin

XXVI

pound condenses, by the loss of water, to form acacatechin and isoacacatechin.

Although Indian cutch, cube gambir, and quebracho tannin are the principal catechin tannins of commerce, catechin tannins also occur in rhubarb, the cacao bean, the cola nut, in mahogany wood, in canaigre root (*Rumex hymenosepalus*), and in a considerable number of the species of eucalyptus, as well as in a number of other less common plant families. In certain of the tannins of this group the catechin residue is present in the form of a carboxylic acid which allows for its condensation with the phenolic group of another catechin residue. In some of the tannins, glucose is attached by a glycosidal linkage to one of the hydroxyl groups.

The maclurin tannins are apparently condensation products of the maclurin nucleus. Nierenstein suggests that the mori-tannin (XXVIII), which occurs in the wood of *Chlorophora tinctoria*, is the condensation product of two molecules of maclurin (XXIX) brought about first by the addition and then by the subtraction of two molecules of water.



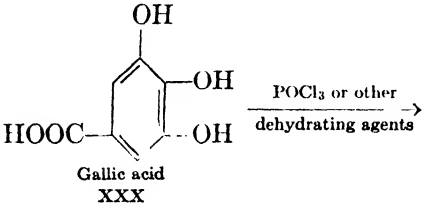
Maclurin
XXIX

Mori-tannin
(Nierenstein)
XXVIII

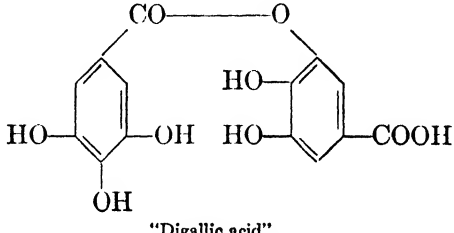
The Hydrolyzable Tannins. These may be divided into the *gallotannins*, the *ellagitannins*, and the *caffetannins*, yielding respectively gallic acid or *m*-digallic acid, ellagic acid, or caffeic acid and quinic acid.

The gallotannins are very widely distributed in plants, and the "tannin" of commerce is one of their decomposition products, gallotannic acid or digallic acid. This is an amorphous powder, soluble in water, acid to litmus, soluble in alcohol and glycerol, but only very slightly soluble in other organic solvents such as ether, benzene, and carbon bisulfide. It possesses a very astringent taste. The chief commercial source of the gallotannins is Chinese nutgalls.

Fischer and Bergmann [69] in their studies on the tannins came to the conclusion that they were glucose derivatives in which one molecule of glucose was combined with ten molecules of gallic acid (XXX). They accordingly concluded that Chinese nutgall tannin corresponded to a pentadigalloyl glucose where one molecule of digallic acid (XXXI) is esterified on each of the five hydroxyl groups of the glucose molecule in the same way as glucose forms a pentaacetyl derivative. Fischer and



Gallic acid
XXX



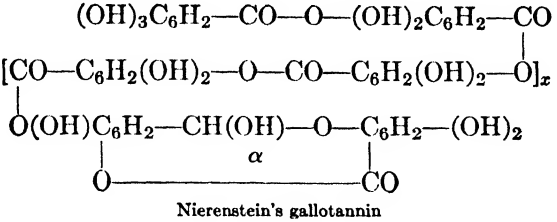
"Digallic acid"
"Gallotannic acid"
"Tannic acid"
XXXI

Freudenberg [70] synthesized a pentagalloyl glucose and considered that it was in all probability closely related to, if not identical with, the tannin of the Chinese nutgall. In the course of their work they prepared hepta(tribenzoylgalloyl)-*p*-iodophenylmaltosazone, $C_{220}H_{142}O_{58}N_4I_2$, which has a molecular weight of 4,021. This compound has one of the highest molecular weights of any compound which has so far been synthesized and of which the structure is definitely known. The iodine was deliberately introduced into the molecule in order to provide for an accurate analysis, since in such a large molecule the analysis for carbon and hydrogen alone would not be sufficiently accurate to determine structural relationships.

Later, however, Nierenstein showed that Fischer's idea of the natural tannin being related to a pentadigalloyl glucose was incorrect, inasmuch as he demonstrated that only the α-glycoside linkage of glucose was united to the gallic acid residue, and he was able to hydrolyze off the

[69] E. Fischer and M. Bergmann, *Ber.*, **51**, 1760 (1918); **52B**, 829 (1919).
[70] E. Fischer and K. Freudenberg, *Ber.*, **45**, 915, 2709 (1912); **46**, 1116 (1913).

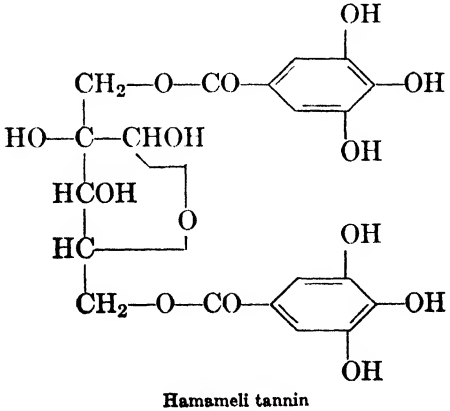
glucose and still retain the characteristic properties of the tannin. Nierenstein [71] believes that the nutgall is a polydigalloylleucodigallic acid anhydride (XXXII) which may sometimes occur in the form of an α-glycoside where the glucose residue is attached to the —OH group designated by α.



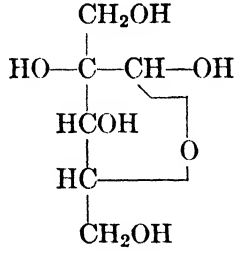
Nierenstein's gallotannin
XXXII

The tannins of *Acer ginuala* and of Chinese rhubarb are apparently much simpler compounds than the tannin of nutgalls. The acertannin seems to be digalloylaceritol, in which the carboxyl group of gallic acid is esterified to hydroxyl groups of the aceritol. The tannin of Chinese rhubarb appears to be simply gallic acid-β-glucoside.

The tannin (XXXIII) [72] of witch hazel, *Hamamelis virginica*, is apparently the digalloyl derivative of the very unusual hexose, hamamelose (XXXIV),[73] in which the gallic acid is combined with the two primary alcohol groups of the sugar.



Hamameli tannin
XXXIII



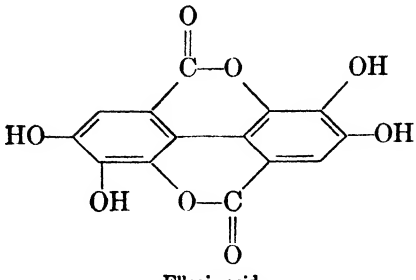
Hamamelose
XXXIV

[71] M. Nierenstein, C. W. Spiers, and A. C. Hadley, *J. Am. Chem. Soc.*, **47**, 1726 (1925).

[72] K. Freudenberg and F. Blümmel, *Ann.*, **440**, 45 (1924).

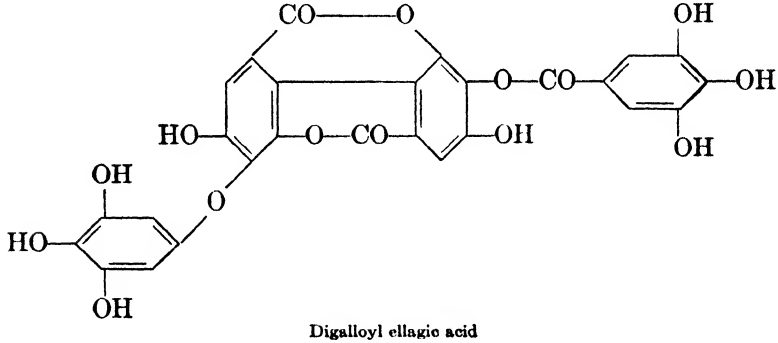
[73] Otto Th. Schmidt, *Ann.*, **476**, 250 (1929).

The ellagitannins are derivatives of ellagic acid (XXXV), and Nier-



Ellagic acid
(the dilactone of the diphenyl derivative corresponding to gallic acid)
XXXV

enstein suggests that they may be in many instances the mono-, di-, tri, or tetragalloyl derivatives of ellagic acid (XXXVI).

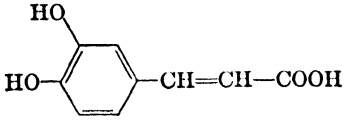


Digalloyl ellagic acid
XXXVI

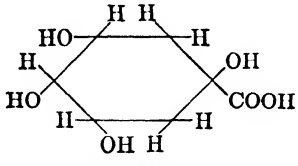
These tannins include the tannin of chestnut wood, which is apparently a monogalloylellagic acid with a quercetin residue linked to one of the remaining hydroxyl groups of ellagic acid. Dividivitannin (pods of *Caesalpinia coriaria*) is apparently similar to the tannin of chestnut wood except that a $C_{14}H_{14}O_{11}$ residue replaces the $C_{15}H_9O_6$ residue of quercetin.

The caffetannins yield, on hydrolysis, caffeic acid (XXXVII) and quinic acid (XXXVIII) and usually a residue of unidentified constitution. Possibly in most instances the caffeic acid and quinic acid are united to form chlorogenic acid, and Nierenstein suggests that the caffetannins are probably condensation products of chlorogenic acid (XXXIX). The caffetannins occur in a great variety of plants, in-

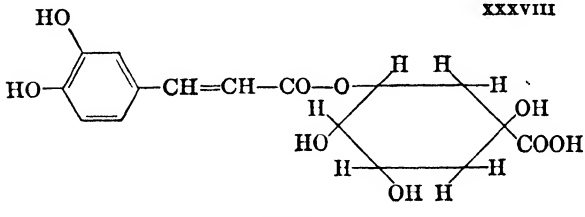
cluding such important species as *Pinus laricio, Larix europaea, Coffea arabica, Papaver somniferum, Digitalis purpurea, Nicotiana tabacum,* etc.



Caffeic acid
XXXVII



Quinic acid
XXXVIII

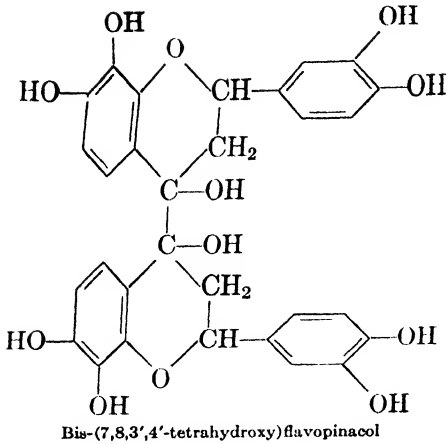


Chlorogenic acid
XXXIX

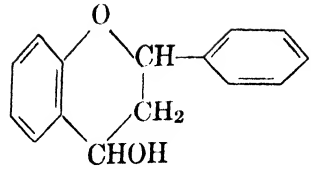
Aside from the classes of tannins noted, Nierenstein observes that there are a great number of tannins, the investigation of which has not yielded sufficient information to provide for even a hypothesis as to the structural elements which constitute the tannin. He suggests that in this group of "unclassified tannins" many will probably fall in line with some one of the groups noted above, and that perhaps other groupings may have to be formed when more extended investigation of the various tannins has been conducted. In this unclassified group are the oak tannins, the larch tannins, the fir tannins, the sequoia tannins, the spruce tannins, the willow-bark tannins, the rhododendron tannins, etc.

Russell,[74] who favors the Perkin and Everest classification, suggests that the most important tannins belong to the phlobatannin group, which represents the "true tannin class, the others being exceptional members that happen to have tanning properties." From his work on the polyhydroxychalcones Russell concludes that the phlobatannins are phenolic hydroxy flavopinacols (XL) which are derivatives of 4-hydroxyflavan (XLI). He synthesized a number of flavopinacols which were qualitatively indistinguishable from the natural phlobatannins.[75]

[74] A. Russell, *Chem. Revs.*, **17**, 155 (1935).
[75] A. Russell and J. Todd, *J. Chem. Soc.*, 421 (1937).



Bis-(7,8,3′,4′-tetrahydroxy)flavopinacol
XL



4-Hydroxyflavan
XLI

The Tanning Process. The process of tanning may be regarded as the conversion of a relatively hydrophilic colloid gel into a relatively non-hydrophilic gel which we call "leather." The manufacture of leather through the action of tannins upon animal hide is an art which dates from the earliest antiquity. It was developed exclusively as an art, and only in recent years has chemical control come into the larger tanneries. It is only natural that such a basic industry should have built up a relatively enormous literature. Wilson [76] has covered the pertinent literature to 1928, and this was brought up to date by McLaughlin and Theis.[77] Wilson and Thomas and co-workers [78] regard the tanning process as a stoichiometrical combination of the tannin with the hide substance. However, this is not the view of Gustavson,[79] McLaughlin, *et al.*,[80] nor apparently of Dorothy Jordan Lloyd,[81] for the tanning process may be completely reversed by shifting the hydrogen-ion concentration, whereby the tannin (or chrome in chrome tanning) can be completely

[76] J. A. Wilson, *The Chemistry of Leather Manufacture*, Vols. I and II, 2nd ed., Chemical Catalog Co., New York, 1928–1929.

[77] G. D. McLaughlin and E. R. Theis, *The Chemistry of Leather Manufacture*, Reinhold Publishing Corp., New York, 1945.

[78] A. W. Thomas and S. B. Foster, *Ind. Eng. Chem.*, **15**, 707 (1923); A. W. Thomas and A. Frieden, *ibid.*, **15**, 839 (1923); A. W. Thomas and M. W. Kelly, *ibid.*, **15**, 928, 1148, 1262 (1923); **16**, 800 (1924); **17**, 41 (1925).

[79] K. H. Gustavson, *J. Am. Chem. Soc.*, **48**, 2963 (1926).

[80] G. D. McLaughlin, D. H. Cameron, and R. S. Adams, *J. Am. Leather Chemists' Assoc.*, **29**, 657 (1934); **32**, 98 (1937).

[81] D. J. Lloyd, "The Theory of the Tanning Process with Particular Reference to Vegetable Tanning and Chrome Tanning," *Proc. Intern. Congr. Pure and Applied Chem.*, Madrid (1934), separate, 13 pp.

removed with the regeneration of the hide substance, which can then be retanned. McLaughlin points out that in the chrome tanning process the taking up of chrome by the hide substance follows the typical Freundlich adsorption isotherm and that, when the logarithms of the chrome oxide taken up by the hide substance are plotted against the logarithms of the chrome oxide remaining in solution, a straight line is obtained, as one would anticipate from an adsorption reaction. Gustavson notes that the maximum amount of chrome is combined with the collagen when the collagen is isoelectric. Miss Lloyd stresses in particular the water relationships and points out that the relatively non-hydrophilic tannin molecule replaces water molecules associated with the hydrated collagen and that this process of dehydration is the essential stage in the tanning process. She adds that the association between the collagen and the tannin is probably that of a covalency, for the tannin can be stripped out of the tanned skin by alkaline solutions. It appears, therefore, as though colloidal adsorption and perhaps a coacervation may account for the tanning process, for tanning proceeds most rapidly at a low *p*H where the collagen would be positively charged and the micelles of the tannin would possess a negative charge.

The technic employed for the evaluation of a tannin extract gives a definition of tannin from a practical viewpoint, *i.e., that portion of the water-soluble matter of certain vegetable materials which will precipitate gelatin from solution and which will combine with hide fibers to form compounds which are resistant to washing.* The remaining portion of water-soluble materials is called non-tannins. The tannin-containing "solution" is shaken with purified hide powder or small squares of hide [80] until all the tannin has been removed from solution. This point is ascertained by testing filtered portions with gelatin to see whether a precipitate still forms.

Synthetic "Tannins." Several substitutes for tannin in leather manufacture have been suggested. These have been called synthetic "tannins," [82,83] *i.e.*, substances which tan hides but which may or may not have a chemical structure analogous to natural tannins. Stiasny [84] has prepared such a synthetic "tannin" from a phenol treated with sulfuric acid and formaldehyde in the proportion of one molecule of formaldehyde to two molecules of phenol. The method of preparation suggests Bakelite, but this tanning substance is water-soluble, owing to its sulfonic acid groups. It is claimed to make a good leather and to require a shorter tanning period than is necessary when natural tannins

[82] H. Diedorf, *J. Am. Leather Chem. Assoc.*, **8**, 394 (1913).
[83] E. Niehoul, *Sci. Am. Monthly*, **1**, 326 (1920).
[84] E. Stiasny, *Leather World*, **12**, 227 (1913); *J. Soc. Chem. Ind.*, **32**, 775 (1913).

are used. The lignosulfonates of waste sulfite liquor (from cellulose production) also have tanning properties.

Meunier and Seyewetz [85] used quinone. They found that when gelatin was treated with phenols under conditions suitable for oxidation the resulting precipitate became insoluble in boiling water. They state that "quinonated gelatin constitutes the most stable form of insoluble gelatin thus far known," because it resists not only boiling water but also dilute acids and alkalies. Only 2 hours at 15°C. are required to render the gelatin completely insoluble. Thomas and Kelly [86] note that the optimum conditions for quinone tanning require an alkaline solution, approximately *p*H 8–10.

Physiological Significance of the Tannins. Plant physiologists differ widely in their views as to the function of tannins in plants. Pfeffer believed that they served some useful purpose and were not merely by-products. He did not consider that they were necessary, however, to furnish benzene rings from which other compounds, such as proteins, etc., might be synthesized. Neither did he consider them essential for cell activity. Moore suggested that the plant builds up tannins as a means of neutralizing acids by esterification. He also believed that tannins may play an important role in the opening and closing of the stomata of growing leaves and in the lignification of plant cell walls. Kraus found tannins to be a direct synthetic product formed only in sunlight and in carbon dioxide supply, and translocated to stems, bark, and roots. Sachs concluded that tannins occur (are produced) in the regions of most intense metabolic activity, such as in active leaves, in rapid tissue formation as in galls and in other pathological growths, and as a consequence of some particular stimulation. Drabble and Nierenstein [87] concluded that cork formation is associated with tannins, and that in the plant cells the complex phenols are acted upon by acids and formaldehyde and are thus precipitated in the "cork" cell.

Other workers [88] believe that cell wall formation or spore formation is associated, to a greater or lesser extent, with tannins. It has even been suggested that tannins are a special protection of the plant against animals! Cook and Taubenhaus [89] suggest that they act as an anti-

[85] L. Meunier and A. Seyewetz, *Compt. rend.*, **146**, 987 (1908); *Collegium*, **195**, 202 (1908); *Mon. Sci.*, **23**, 91 (1909).

[86] A. W. Thomas and M. W. Kelly, *Ind. Eng. Chem.*, **16**, 925 (1924); **18**, 383 (1926).

[87] E. Drabble and M. Nierenstein, *Biochem. J.*, **2**, 96 (1907).

[88] C. van Wisselingh, *Koninkl. Akad. Wetenschap. Amsterdam*, **9**, 685 (1910); *Pharm. J.*, **91**, 571 (1913); *Botan. Centr., Beihefte*, **32A**, 155 (1914); *Pharm. Weekblad*, **52**, 1349 (1915).

[89] M. T. Cook and J. J. Taubenhaus, *Delaware Agr. Exp. Sta. Bull.* 91 (1911); *Bull.* 97 (1912).

septic in case of wound formation, especially as a protection against fungi. They point out that fungi are quite sensitive to tannins and that, apparently, parasitic fungi are more sensitive than the saprophytic forms. It is true that tannin collects in wounded tissue in abnormally large amounts.

In green fruits tannin is more or less abundant but apparently disappears as the fruit ripens. It may not actually disappear but instead may be "locked up" in giant cells in a more or less insoluble form. This seems to be the mechanism in the persimmon.[90] The Japanese "process" the persimmons by placing the unripe fruit in casks from which their rice wine (sake) has just been drawn. In 5 to 15 days the casks are opened, and the astringent taste will have wholly disappeared. Gore used anesthetics, such as ether or chloroform, to produce the same result. It was found that the tannin was localized in giant cells, some large enough to be seen with the naked eye. In the partly ripe persimmon, these cells do not break as soon as taken into the mouth but swell and eventually burst, emptying their thick tannin-bearing contents on the tongue. In the *processed* or *ripe* persimmons the cell contents have undergone a change, becoming more refractive and hard and losing their imbibitional power, so that they swell but slightly and do not burst. The loss of astringency presumably is due to the hardening of the contents of these giant cells.

In view of the wide distribution of tannins in the plant world and of the differences in their chemical and physical properties, it is highly improbable that all the tannins have the same physiological significance.

[90] H. C. Gore, *Bur. Chem., U. S. Dept. Agr., Bull.* 141 (1911); *Bull.* 155 (1912).

IV

THE LIPIDS AND ESSENTIAL OILS

The fatty acid compounds and cholesterol, because of their insolubility in water, probably are of the greatest importance in regulating transference of water-soluble substances, including proteins. By their employment in combinations with the water-soluble substances and their presence in membranes, the passage of materials about the body may be controlled.

WALTER R. BLOOR (1948)

CHAPTER 30

The Simple Lipids: The Fatty Acids, Glycerides, Waxes, Sterols, and Bile Acids

The natural fats and oils may be divided into two major groups, (1) the fatty or non-volatile fats and oils, and (2) the essential or volatile oils. Chemically the two groups are entirely distinct, group (1) being composed of esters of fatty acids (mainly glycerol esters), and group (2) being those plant products volatile with steam and separating as an oily layer in the distillate. Chemically the essential oils are aldehydes, alcohols, acids, hydrocarbons, terpenes, etc. Rarely are they the esters of the ordinary fatty acids. Oil of cloves, wintergreen, and turpentine are typical examples. Because of the great difference in chemical nature, the essential oils are not included in the term "lipids."

True fats are composed, in general, of only carbon, hydrogen, and oxygen, but certain derivatives or fat-like compounds, the "phospholipids," "lecithins," etc., contain nitrogen or phosphorus or both.

There is no generally adopted system for the classification of the lipids. However, the classification proposed by Bloor [1] is as satisfactory as any which has been proposed. His classification, with minor modifications, follows.

A CLASSIFICATION OF THE LIPIDS

Substances having the following characteristics:

(*a*) Insolubility in water and solubility in the fat solvents, such as ether, chloroform, benzene.
(*b*) Relationship to the fatty acids as esters, either actual or potential.
(*c*) Utilization by living organisms.

I. Simple lipids—esters of fatty acids with various alcohols.

(*a*) Fats—esters of the fatty acids with glycerol, solid at room temperature.

[1] W. R. Bloor, *Chem. Revs.*, **2**, 243 (1925); *Biochemistry of the Fatty Acids and their Compounds, the Lipids*, Reinhold Publishing Corp., New York, 1943.

(*b*) Oils—esters of the fatty acids with glycerol, liquid at room temperature.

(*c*) Waxes—esters of the fatty acids with alcohols other than glycerol, alcohol commonly monoatomic.

II. Compound lipids—compounds of the fatty acids with alcohols, but containing other groups in addition to the alcohol.

(*a*) Phospholipids—substituted fats containing phosphoric acid and nitrogen—lecithins, cephalins, sphingomyelins.

(*b*) Glycolipids—compounds of the fatty acids containing both a carbohydrate and a nitrogen base but containing neither glycerol nor phosphoric acid—the cerebrosides.

(*c*) Phosphatidic acids—compounds similar to the phospholipids but with the organic base removed.

(*d*) Sulfolipids—lipids containing sulfuric acid.

(*e*) Miscellaneous—plasmalogens, aminolipids, etc.

III. Derived lipids—substances derived from the above groups by hydrolysis.

(*a*) Fatty acids.

(*b*) Sterols.

(*c*) Alcohols.

(*d*) Nitrogen bases.

(*e*) Fatty aldehydes.

A CLASSIFICATION OF THE FATTY ACIDS[1]

I. The saturated fatty acids, $C_nH_{2n}O_2$, $C_nH_{2n+1}COOH$

A. *With straight carbon chain* (only those with an even number of carbon atoms are commonly found in natural fats)

Name	*Formula*	*Occurrence*
Butyric	$C_4H_8O_2$	Milk fat
Caproic	$C_6H_{12}O_2$	Butter, coconut, and palm nut oils
Caprylic	$C_8H_{16}O_2$	Butter, coconut, and palm nut oils
Capric	$C_{10}H_{20}O_2$	Butter, coconut, and palm nut oils
Lauric	$C_{12}H_{24}O_2$	Laurel oil, spermaceti
Myristic	$C_{14}H_{28}O_2$	Nutmeg butter
Palmitic	$C_{16}H_{32}O_2$	Animal and vegetable fats
Stearic	$C_{18}H_{36}O_2$	Animal and vegetable fats

[1] Many of the aliphatic acids occurring in plants or animal tissues, such as oxalic pyruvic, and citric, are not included, since they do not occur in fats or waxes.

Name	*Formula*	*Occurrence*
Arachidic	$C_{20}H_{40}O_2$	Peanut oil
Behenic	$C_{22}H_{44}O_2$	Oil of ben, from seeds of *Moringa pterygosperma*
Lignoceric	$C_{24}H_{48}O_2$	Arachis oil, glycolipids
Cerotic	$C_{26}H_{52}O_2$	Beeswax, Chinese wax, opium wax, wool fat
Montanic	$C_{28}H_{56}O_2$	Montan wax
Melissic	$C_{30}H_{60}O_2$	Beeswax, Carnauba wax
B. *With branched carbon chain*		
14-Methyl palmitic	$C_{17}H_{34}O_2$	Wool fat
Tuberculostearic (10-methyl)	$C_{19}H_{38}O_2$	Human tubercle bacillus
Phytomonic (10- or 11-methyl)	$C_{20}H_{40}O_2$	Crown gall bacillus
Phthioic	$C_{26}H_{52}O_2$	Human tubercle bacillus
II. The unsaturated fatty acids		
A. *Monoethenoid series*, $C_nH_{2n-1}COOH$		
Crotonic ($\Delta^{2:3}$)	$C_4H_6O_2$	Croton oil
Tiglic (2-methyl, $\Delta^{2:3}$)	$C_5H_8O_2$	Croton oil
Myristoleic ($\Delta^{9:10}$)	$C_{14}H_{26}O_2$	Fish and whale oils
Palmitoleic ($\Delta^{9:10}$)	$C_{16}H_{30}O_2$	Fish and whale oils
Oleic ($\Delta^{9:10}$)	$C_{18}H_{34}O_2$	Animal and vegetable fats
Petroselinic ($\Delta^{6:7}$)	$C_{18}H_{34}O_2$	Seed fats of *Umbelliferae*
Vaccenic ($\Delta^{11:12}$)	$C_{18}H_{34}O_2$	Butterfat
Gadoleic ($\Delta^{9:10}$)	$C_{20}H_{38}O_2$	Fish oils
Erucic ($\Delta^{13:14}$)	$C_{22}H_{42}O_2$	Rapeseed oil
Cetoleic ($\Delta^{11:12}$)	$C_{22}H_{42}O_2$	Marine animal and fish oils
Nervonic ($\Delta^{15:16}$)	$C_{24}H_{46}O_2$	Cerebrosides
B. *Diethenoid series*, $C_nH_{2n-3}COOH$		
Linoleic ($\Delta^{9:10,\ 12:13}$)	$C_{18}H_{32}O_2$	Vegetable oils, such as linseed and cottonseed
Eicosadienoic	$C_{20}H_{36}O_2$	Human milk fat
C. *Triethenoid series*, $C_nH_{2n-5}COOH$		
Linolenic ($\Delta^{9:10,\ 12:13,\ 15:16}$)	$C_{18}H_{30}O_2$	Linseed oil
Eleostearic ($\Delta^{9:10,\ 11:12,\ 13:14}$)	$C_{18}H_{30}O_2$	Chinese wood oil
D. *Polyethenoid fatty acids*		
With four double bonds:		
Parinaric ($\Delta^{9:10,\ 11:12,\ 13:14,\ 15:16}$)	$C_{18}H_{28}O_2$	Seed oil of *Parinarium laurinum*
Arachidonic ($\Delta^{5:6,\ 8:9,\ 11:12,\ 14:15}$)	$C_{20}H_{32}O_2$	Lecithin, cephalin
With five double bonds:		
Clupanodonic ($\Delta^{4:5,\ 8:9,\ 12:13,\ 15:16,\ 19:20}$)	$C_{22}H_{34}O_2$	Japanese sardine oil
E. *Monoethinoid series*		
Tariric (triple bond at 6:7)	$C_{18}H_{32}O_2$	Fats of *Picramnia* sp.

III. Hydroxy fatty acids

A. *Saturated monohydroxy acids*, $C_nH_{2n}O_3$

Name	*Formula*	*Occurrence*
β-Hydroxybutyric	$C_4H_8O_3$	Metabolism, animal
α-Hydroxy-*n*-decanic	$C_{10}H_{20}O_3$	Brain phospholipids
Sabinic (12-hydroxy)	$C_{12}H_{24}O_3$	Conifer wax
Juniperic (16-hydroxy)	$C_{16}H_{32}O_3$	Conifer wax
Lanopalmic	$C_{16}H_{32}O_3$	Wool fat
Cerebronic (2-hydroxy)	$C_{24}H_{48}O_3$	Phrenosin (brain tissue)
Cocceric	$C_{31}H_{62}O_3$	Cochineal wax

B. *Unsaturated monohydroxy acids*, $C_nH_{2n-2}O_3$

Ricinoleic (12-hydroxy, $\Delta^{9:10}$)	$C_{18}H_{34}O_3$	Castor oil
Oxynervonic (2-hydroxy, $\Delta^{15:16}$)	$C_{24}H_{46}O_3$	Cerebrosides

C. *Saturated dihydroxy acids*, $C_nH_{2n}O_4$

Dihydroxystearic (9,10-dihydroxy)	$C_{18}H_{36}O_4$	Castor oil
Lanoceric	$C_{30}H_{60}O_4$	Wool fat

IV. Keto acids

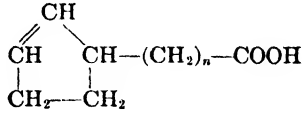
Licanic (4-keto, $\Delta^{9:10,\ 11:12,\ 13:14}$)	$C_{18}H_{30}O_3$	*Rosaceae* seed fats

V. Saturated dibasic acids, $C_nH_{2n-2}O_4$

Thapsic (α,ω-di-COOH)	$C_{16}H_{30}O_4$	Conifer wax
Japanic (α,ω-di-COOH)	$C_{21}H_{40}O_4$	Japan wax

VI. Chaulmoogric series

A. *Cyclic acids with one double bond*



Aleprolic (n = 0)	$C_6H_8O_2$	Chaulmoogra oil
Aleprestic (n = 4)	$C_{10}H_{16}O_2$	Chaulmoogra oil
Aleprylic (n = 6)	$C_{12}H_{20}O_2$	Chaulmoogra oil
Alepric (n = 8)	$C_{14}H_{24}O_2$	Chaulmoogra oil
Hydnocarpic (n = 10)	$C_{16}H_{28}O_2$	Chaulmoogra oil
Chaulmoogric (n = 12)	$C_{18}H_{32}O_2$	Chaulmoogra oil

B. *Cyclic acids with two double bonds*

Gorlic (5:6 dehydrochaulmoogric)	$C_{18}H_{30}O_2$	Chaulmoogra oil

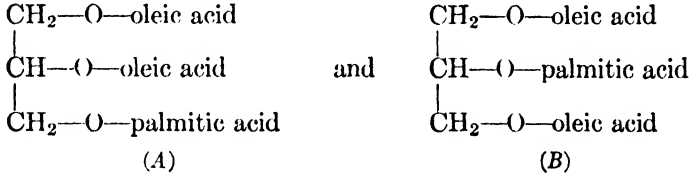
THE GLYCERIDES

Esters of Glycerol. Inasmuch as glycerol is a trihydroxy alcohol, we have possible a number of esters. Thus, we may have monoglycerides, diglycerides, and triglycerides. In the case of the di- and tri-

glycerides we may have the same or different fatty acids united in the ester, resulting in the formation of simple or mixed glycerides.

Mono- and diglycerides apparently never occur in nature, nor are they present in freshly prepared oils and fats. Their presence in a fat or oil indicates that the material has undergone partial saponification. The presence of free fatty acid likewise generally indicates that hydrolysis of the neutral triglycerides has taken place, although the free acids occasionally occur in appreciable amounts.[3]

Most naturally occurring fats are composed of mixed glycerides. The milk fat of the monotreme or Australian spiny anteater is a notable exception, since it appears [4] to be pure triolein. When mixed glycerides occur, the problem of space relationships enters to complicate a study of the chemical constitution of the glyceride. Thus, for example, the two fats



will have the same empirical formula, will yield the same quantities of the same products upon hydrolysis, but, because of the space configuration of the molecule, they will exhibit somewhat different physical properties. For example, *A* will contain an asymmetric carbon atom, whereas *B* will not. When such facts are kept in mind, together with the fact that the fats are relatively inert chemically and practically always occur as mixtures which are exceedingly difficult to separate, some appreciation can be gained of the difficulties involved in a study of the organic chemistry of the fats and oils.

Hilditch [5] has compiled a wealth of data on the fatty acid constituents of many natural fats. The reader will find this volume a useful reference book in this field.

Synthetic Glycerides. King [6] notes that prior to 1929 seven different methods had been proposed for the synthesis of simple and mixed

[3] W. A. Gortner, *J. Biol. Chem.*, **159**, 135 (1945).

[4] H. R. Marston, *Australian J. Exptl. Biol. Med. Sci.*, **3**, 217 (1926).

[5] T. P. Hilditch, *The Chemical Constitution of Natural Fats*, John Wiley & Sons, New York, 1940.

[6] H. P. Averill, J. N. Roche, and C. G. King, *J. Am. Chem. Soc.*, **51**, 866 (1929); *cf.* also *ibid.*, **52**, 365 (1930); H. E. Robinson, J. N. Roche, and C. G. King, *ibid.*, **54**, 705 (1932); D. T. Jackson and C. G. King, *ibid.*, **55**, 678 (1933); O. E. McElroy and C. G. King, *ibid.*, **56**, 1191 (1934); B. F. Stimmel, *ibid.*, **56**, 1724 (1934).

glycerides. Four of these general methods were devised by Grün, but the difficulty with most of them is that they result in an unpredictable rearrangement of the acyl groups which had been substituted in the glycerol molecule. Thus, for example, starting with the monohalogen-hydrin, the halogen may be on one of the α-hydroxyl groups of glycerol or on the β-hydroxyl group. It seems to be a general rule that an acyl group substituted in the β-position on glycerol tends to wander to the more stable α-position, so that most of the compounds which have been described as β-monoglycerides actually have turned out to be α-monoglycerides owing to the wandering of the acyl group. Similarly most of the compounds which were earlier described as α,β-diglycerides have later turned out to be α,α-diglycerides, again because of the wandering of the group which had been substituted initially in the β-position. This problem of the wandering of the acyl groups has been stressed by several workers.[7] Daubert and Longenecker [8] successfully utilized this tendency for the acyl groups to migrate in preparing synthetic mixed diglycerides.

Bergmann and Carter [9] used 1,3-benzylidene glycerol to prepare β-monoglycerides and studied the properties of a number of these synthetic compounds. This appears to be one of the best methods by which the β-glyceride can be prepared. Even in this case unusual care must be taken not to heat the reaction mixture appreciably, particularly in the presence of acids, since wandering of the group from the β- to the α-position may take place in the presence of heat and acids.

In Bergmann's method glycerol is condensed with benzaldehyde to form the stable crystalline 1,3-benzylidene derivative. This compound is then treated in the presence of pyridine or quinoline with the acid chloride of the fatty acid which it is desired to esterify on the β-hydroxyl group, the pyridine or quinoline acting as a reservoir for the hydrochloric acid formed in the reaction. The benzylidene group is then removed by catalytic hydrogenation, leaving the β-monoglyceride as the residual product.

Shortly before his death in 1920, Fischer [10] devised methods for the preparation of unsymmetrical triglycerides and also improved the methods for the synthesis of symmetrical di- and triglycerides. Fischer's methods, in combination with Bergmann's method noted above, are

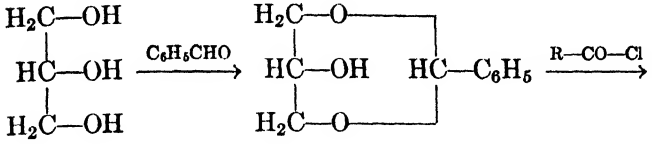
[7] H. S. Hill, M. S. Whelen, and H. Hibbert, *J. Am. Chem. Soc.*, **50**, 2235 (1928); H. Hibbert and N. M. Carter, *ibid.*, **51**, 1601 (1929); A. Fairbourne and G. W. Cowdrey, *J. Chem. Soc.*, 129 (1929); B. F. Daubert and C. G. King, *J. Am. Chem. Soc.*, **61**, 3328 (1939).

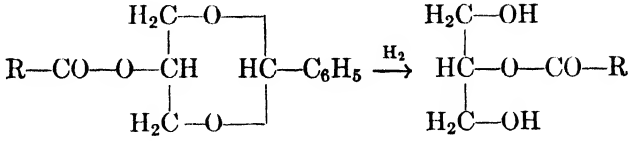
[8] B. F. Daubert and H. E. Longenecker, *J. Am. Chem. Soc.*, **66**, 53 (1944).

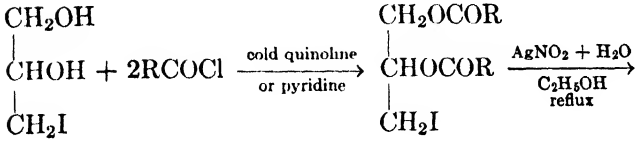
[9] M. Bergmann and N. M. Carter, *Z. physiol. Chem.*, **191**, 211 (1930).

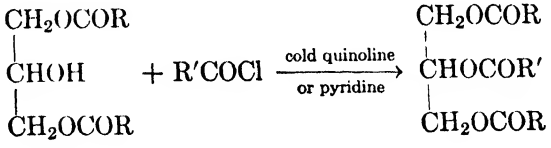
[10] E. Fischer, M. Bergmann, and H. Bärwind, *Ber.*, **53**, 1589 (1920).

the ones generally used in modern glyceride synthesis. The reactions by the Fischer method can be indicated as follows:

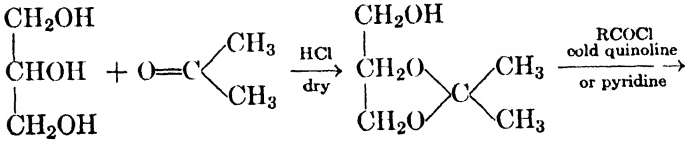


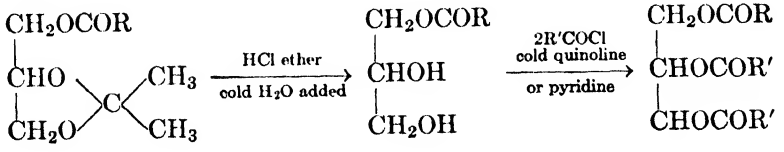






and





It will be noted that in the first series of reactions using a halohydrin the first product formed is the α,β-diacyl derivative and that this rearranges when the halogen is removed to form the α,α-derivative. This more stable compound can then be treated with the acid chloride of a different fatty acid which will condense and add the acyl radical in the β-position, thus forming a symmetrical mixed triglyceride.

In the second series of reactions the resultant is an unsymmetrical mixed triglyceride. The α,β-acetone derivative of glycerol is formed, and only one acyl group is added to this in the α-position. Then, when the acetone derivative is saponified, a different acyl group can be added in the α,β-positions.

The success of these two methods of synthesis lies in the fact that all the reactions can be carried out at temperatures not exceeding 35°C., excepting in that series of reactions where the $\beta \rightarrow \alpha$ shift is brought about. Low temperature and low acidity stabilize the β-compounds.

A very large number of simple and mixed glycerides, both saturated and unsaturated, have been synthesized. Daubert and his co-workers,[11] working in Longenecker's laboratory, have been particularly active in this field.

The physical properties of the synthetic glycerides have been studied and compared with the physical properties of fractions isolated from natural fats or oils. In this way the structure of certain of the naturally occurring compounds has been definitely ascertained.

The Volatile and Non-Volatile Fatty Acids. In the analysis of fats and oils the fatty acids may be classified into the volatile and non-volatile groups. The fatty acids containing ten or less carbon atoms are classified as the *water-soluble fatty acids*. All fatty acids containing more than ten carbon atoms are classified as insoluble fatty acids, although lauric acid is slightly soluble in boiling water. As would be expected, the solubility of the fatty acids in water decreases as the number of carbon atoms increases. This has already been indicated in the discussion of the phenomena involved in the molecular orientation at interfaces. The *volatile fatty acids* are the same six fatty acids which are classified as soluble fatty acids. These are the only ones which can be distilled at atmospheric pressure. They possess a fairly high boiling point, but because of the high vapor tension they can be readily removed from an aqueous solution by steam distillation. Lauric acid is slightly volatile under the same conditions. Most natural fats contain some of the volatile fatty acids.

The titration value of a steam distillate of a saponified fat is the means of determining the *Reichert-Meissl number* of a fat, *i.e.*, the number of

[11] B. F. Daubert, H. H. Fricke, and H. E. Longenecker, *J. Am. Chem. Soc.*, **65**, 1718, 2142 (1943); B. F. Daubert, C. J. Spiegl, and H. E. Longenecker, *ibid.*, **65**, 2144 (1943); B. F. Daubert and H. E. Longenecker, *ibid.*, **66**, 53 (1944); F. L. Jackson, B. F. Daubert, C. G. King, and H. E. Longenecker, *ibid.*, **66**, 289 (1944); B. F. Daubert, *ibid.*, **66**, 290 (1944); B. F. Daubert and A. R. Baldwin, *ibid.*, **66**, 997, 1507 (1944).

milliliters of 0.1 *N* KOH required to neutralize the steam distillate obtained from 5 grams of saponified fats. Butterfat, because of its butyric acid content, has a very high (for fats) Reichert-Meissl number. Jensen [12] notes that this process accounts for 85 to 88 per cent of the total butyric acid, 85 to 100 per cent of the caproic acid, and 24 to 25 per cent of the caprylic acid present in the glycerides of butterfat. From 10 to 13.6 per cent of the fatty acids of butterfat are volatile.

The water-insoluble or non-volatile fatty acids differ somewhat in chemical structure, which fact permits their classification into the saturated and unsaturated fatty acids, the unsaturated fatty acids being further subdivided according to the type and degree of unsaturation, as has already been noted in the early part of this chapter. Oleic, stearic, palmitic, linoleic, and linolenic acids constitute the principal insoluble fatty acids occurring in nature.

Solubility data for fatty acids in various organic solvents over a wide range in temperature have been reported by Hoerr and Ralston [13] and by Foreman and Brown.[14]

The Structure of the Fatty Acid Molecule. Oleic acid is the most important of the unsaturated fatty acids, occurring as it does rather generally throughout the vegetable and animal kingdoms. The structure of oleic acid was established by the following steps:

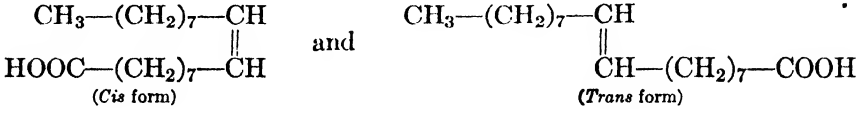
1. An elementary analysis showed only carbon, hydrogen, and oxygen to be present in a proportion corresponding to the empirical formula $(C_9H_{17}O)_x$.
2. A molecular weight determination showed that the formula must be $C_{18}H_{34}O_2$.
3. Esterification and titration with bases indicated the presence of a single carboxyl group.
4. The addition of iodine indicated the presence of one double bond, forming a diiodide (bromine could not be used, inasmuch as it substituted in the compound).
5. Stearic acid was obtained upon the reduction of oleic acid (the addition of two hydrogen atoms).
6. When oleic acid undergoes oxidation, a dihydroxystearic acid is first formed, the molecule then breaking at the double bond upon further oxidation into a saturated C_9 monobasic acid and a saturated C_9 dibasic acid. Both of these C_9 acids were found to be straight-chain carbon compounds. Therefore, the double bond of oleic acid was

[12] O. Jensen, *Z. Nahr. Genussm.*, **10**, 265 (1905).
[13] C. W. Hoerr and A. W. Ralston, *J. Org. Chem.*, **9**, 329 (1944).
[14] H. D. Foreman and J. B. Brown, *Oil & Soap*, **21**, 183 (1944).

in the center of the carbon chain, and the formula could be written

$$CH_3—(CH_2)_7—CH{=}CH—(CH_2)_7—COOH$$

This, however, does not represent the true formula, for such a compound can exist in two stereoisomeric forms,



The *cis* modification is oleic acid (m.p. 14°); the other is elaidic acid (m.p. 45°). Elaidic acid does not occur in nature. Oleic acid is transformed into elaidic acid in the presence of nitrogen trioxide, N_2O_3.

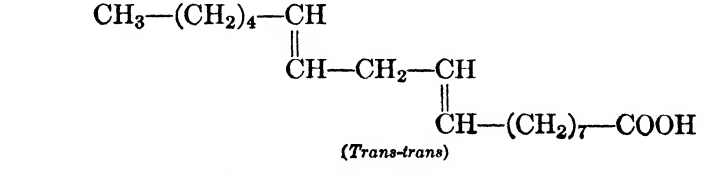
It is relatively easy to determine the structure of the molecule of a fatty acid which contains only a single double bond. However, the determination of the structure of fatty acids containing two or more double bonds is much more difficult.

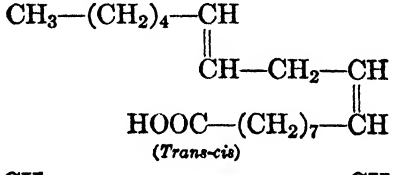
Two general observations on structure apply to the great majority of the naturally occurring unsaturated fatty acids. In only rare instances, such as for eleostearic acid, are the double bonds conjugated; generally they are separated by a methylene group. Furthermore, the first double bond is usually in the 9,10-position.

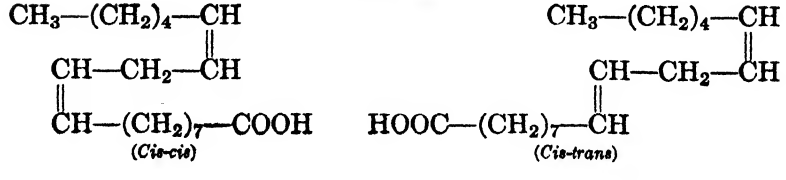
Linoleic acid has the formula

$$CH_3—(CH_2)_4—CH{=}CH—CH_2—CH{=}CH—(CH_2)_7—COOH$$

Such a formula has four possible *cis-trans* stereoisomers.







Similarly, eleostearic acid,

$$CH_3—(CH_2)_3—CH=CH—CH=CH—CH=CH—(CH_2)_7—COOH$$

has eight possible stereoisomeric forms corresponding to those noted for linoleic acid. Eleostearic acid occurs in Chinese wood oil,[15] from 90 to 95 per cent of the glycerides of Chinese wood oil being present as the esters of this acid.

Linolenic acid, with three double bonds at the ninth, twelfth, and fifteenth carbon atoms, follows the general rule for unsaturated fatty acids. Arachidonic acid appears to be 5,8,11,14-eicosatetraenoic acid.[16] Thus the last 10 carbon atoms of this C_{20} acid are the same as on linoleic acid. Sixteen stereoisomeric forms are possible.

Oleic,[17] linoleic,[18] linolenic,[19] ricinoleic,[20] and arachidonic [21] acids have been isolated in high states of purity by fractional crystallization of the free acids or their methyl esters from acetone at very low temperatures. All appear to have exclusively *cis*-ethylenic linkages in the natural state.[22]

The problem of *cis-trans* isomerism has taken on added significance since the study of Cooper and Edgar [23] of the comparative biological effects of *cis-trans* isomers. These workers studied maleic, fumaric, dibromosuccinic, succinic, citraconic, mesaconic, and itaconic acids. They found that the *trans*-acids were superior to their *cis*-isomerides (*a*) in regard to bactericidal action, (*b*) as protein precipitants, and (*c*) as "activating" enzymatic reactions. They found the *cis*-acids to be more strongly adsorbed by proteins than the *trans*-acids. They note that fumaric acid is a stronger disinfectant than phenol and is much less toxic to higher animals, and suggest that *trans*-derivatives may prove to be useful germicides. The *cis*-acids and *trans*-acids also show quite different behavior when their adsorption on charcoal is studied.[24] This adsorption appears to be apolar, *i.e.*, the undissociated molecule, and not the ions, is adsorbed.[25]

[15] F. H. Rhodes and C. J. Welz, *Ind. Eng. Chem.*, **19**, 68 (1927).
[16] C. L. Arcus and I. Smedley-Maclean, *Biochem. J.*, **37**, 1 (1943).
[17] J. B. Brown and G. Y. Shinowara, *J. Am. Chem. Soc.*, **59**, 6 (1937).
[18] J. B. Brown and G. G. Stoner, *J. Am. Chem. Soc.*, **59**, 3 (1937); J. B. Brown and J. Frankel, *ibid.*, **60**, 54 (1938); J. Frankel and J. B. Brown, *ibid.*, **63**, 1483 (1941).
[19] G. Y. Shinowara and J. B. Brown, *J. Am. Chem. Soc.*, **60**, 2734 (1938).
[20] J. B. Brown and N. D. Green, *J. Am. Chem. Soc.*, **62**, 738 (1940).
[21] G. Y. Shinowara and J. B. Brown, *J. Biol. Chem.*, **134**, 331 (1940).
[22] J. W. McCutcheon, M. F. Crawford, and H. L. Welsh, *Oil & Soap*, **18**, 9 (1941).
[23] E. A. Cooper and S. H. Edgar, *Biochem. J.*, **20**, 1060 (1926).
[24] E. R. Linner and R. A. Gortner, *J. Phys. Chem.*, **39**, 35 (1935).
[25] V. L. Frampton and R. A. Gortner, *J. Phys. Chem.*, **41**, 567 (1937).

The Properties of Natural Fats and Oils. All the naturally occurring fats and oils have a specific gravity less than 1.0. Many of the fats can be obtained in a crystalline form. Both fats and oils are readily soluble in the so-called fat solvents, *i.e.*, ether, petroleum ether, chloroform, carbon bisulfide, and carbon tetrachloride, etc., but they are, in general, only sparingly soluble in alcohol, and, because of this fact, ethyl alcohol is usually chosen as a solvent from which attempts are made to prepare crystalline fats. The fats themselves, the fatty acids derived from them, and especially the alkali salts of the fatty acids or soaps greatly reduce the surface tension of water. Accordingly, the soaps are used generally as emulsifying agents or detergents. The fats are excellent solvents for other compounds which have similar nature and which have similar solubilities. As we shall have occasion to note later, pure lard is often used as a solvent to gather the essence of flowers in the manufacture of perfumes. An analogous instance is the phenomenon which occurs when onions and butter are placed side by side in the icebox, or when cows eat wild garlic and the butterfat is tainted with the garlic.

The Hydrogenation of Oils. The pioneer work in this field was due to the activities of Sabatier,[26] who discovered that certain metals catalyzed the reaction $H_2 + R—CH=CH—R \rightarrow R—CH_2—CH_2—R$. The principal metals which have been employed are iron, cobalt, copper, nickel, platinum, and palladium. The reaction is carried out at an elevated temperature in the presence of a hydrogen atmosphere and the catalyst.

An excellent history of the commercial application of the process of hydrogenation to oils is presented in the court decision of the case of Procter and Gamble Company *versus* Berlin Mills Company [27] (*cf.* also Richardson [28]). The lard substitutes, such as "Crisco," "Snowdrift," etc., do not represent anywhere nearly a complete hydrogenation of the vegetable oils. If the vegetable oils were completely hydrogenated, the resulting product would be brittle and similar to stearin or tallow. "Crisco" contains 20 to 25 per cent of saturated fats, 65 to 75 per cent of oleins, and 5 to 10 per cent of linoleins.

Selective hydrogenation can be utilized as an aid in the analysis of a fat. Thus, Hilditch and Stainsby [29] studied the hydrogenation of the body fat of the pig, following the degree of hydrogenation with the

[26] P. Sabatier, *Ind. Eng. Chem.*, **18**, 1005 (1926).

[27] "Hydrogenation of Oils," *J. Ind. Eng. Chem.*, **9**, 1146 (1917).

[28] A. S. Richardson, C. A. Knuth, and C. H. Milligan, *Ind. Eng. Chem.*, **16**, 519 (1924); **17**, 80 (1925); A. S. Richardson and A. O. Snoddy, *ibid.*, **18**, 570 (1926).

[29] T. P. Hilditch and W. J. Stainsby, *Biochem. J.*, **29**, 90 (1935).

iodine number of the fat. Certain of their data are reproduced in Table 60. It will be noted that in the original fat having an iodine number of 63 there is 13.5 per cent of linoleic acid. When this fat is hydrogenated to an iodine number of 48.9, only 1 per cent of linoleic acid remains, and the increase in oleic acid practically accounts for the loss in the linoleic acid fraction. The stearic acid at this stage has increased only a little more than 1 per cent. A further increase in hydrogenation causes linoleic acid to disappear completely and reduces sharply the oleic

TABLE 60. THE SELECTIVE EFFECT OF HYDROGENATION ON PIG FAT

(Data of Hilditch and Stainsby)

Iodine Number of Fat	Fatty Acids in Fat				
	Palmitic, mole per cent	Stearic, mole per cent	Oleic, mole per cent	Linoleic, mole per cent	Myristic, mole per cent
63.0 *	27.3	14.4	40.9	13.5	2.8
48.9	27.4	15.7	52.1	1.0	2.8
43.0	27.4	21.4	47.3	None	2.9
35.8	27.4	29.6	39.2	None	2.8
28.7	27.5	36.2	32.5	None	2.8
18.7	27.5	47.5	21.2	None	2.8
10.0	27.5	57.4	11.3	None	2.8

* Original fat.

fraction with a corresponding increase in the stearic acid fraction. Further hydrogenation progressively causes the oleic acid fraction to decrease and the stearic acid fraction to increase. It is evident, therefore, that those fatty acids containing two double bonds are preferentially hydrogenated so far as one of the double bonds is concerned. Apparently the double bond farthest from the carboxyl group is the one that hydrogenates the most readily. Furthermore an unsaturated acid in the α-position on the glycerol molecule is preferentially hydrogenated over a similar unsaturated acid on the β-position.

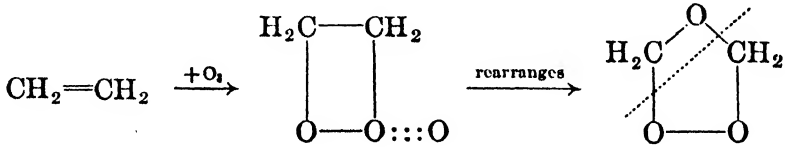
Waterman, *et al.*,[30] have particularly investigated conditions under

[30] H. I. Waterman and J. A. van Dijk, *Rec. trav. chim.*, **50**, 679 (1931); H. I. Waterman, M. J. van Tussenbroek, and J. A. van Dijk, *ibid.*, **50**, 793 (1931); H. I. Waterman and J. A. van Dijk, *ibid.*, **50**, 279 (1931); H. I. Waterman and M. Zaayer, *ibid.*, **51**, 401 (1932); H. I. Waterman, J. A. van Dijk, and C. van Vlodrop, *ibid.*, **51**, 953 (1932); and H. I. Waterman and C. van Vlodrop, *ibid.*, **52**, 9 (1933).

which selective hydrogenation takes place. At high temperatures and low pressures and in the presence of a sluggish catalyst, there is a maximum of selective hydrogenation with respect to the molecules of the fats and oils. Under these conditions a given iodine value will yield larger amounts of unsaturated fatty acids. High pressure and low temperature, on the contrary, yield more of the completely saturated fatty acids. These relationships of temperature and pressure appear to be peculiar to the fatty acid molecules inasmuch as van Dijk, *et al.*,[31] found that, at high pressure and low temperature, palm oil could be hardened without destroying the carotenoids and that cod liver oil could be hardened without destruction of vitamin D and without any great destruction of the compound which gives the color reaction with antimony trichloride and which is taken as a measure of vitamin A (Chapter 36, p. 892). They noted, however, that when tested biologically the vitamin A content of the hydrogenated palm oil had been reduced.

The opposite of hydrogenation, *i.e.*, dehydrogenation, in which hydrogen is abstracted from the oils, has been more or less successfully accomplished on a small scale by heating the oils with catalysts. Dehydrogenation does not appear to have been practiced on a large commercial scale comparable to the hydrogenation industry. At least there is no general knowledge of such large-scale operation. If dehydrogenation on a large scale could be readily accomplished, the *drying oils* for paints could be readily made from the natural non-drying oils.

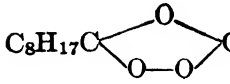
Ozonides. Many organic compounds containing unsaturated linkages add ozone at the double bond. Thus, ethylene adds one molecule of ozone to form an ozonide,



When an ozonide is treated with water, it breaks up into aldehydes, ketones, or peroxides of aldehydes or ketones, and from a study of the resulting decomposition products we can decide at what position the unsaturated double bond occurred in the original organic compound. Thus, oleic acid forms an ozonide in acetic acid solution, and on treatment of the oleic acid ozonide with water it decomposes into

[31] J. A. van Dijk, R. T. A. Mees, and H. I. Waterman, *Biochem. Z.*, **245**, 25 (1932).

hydrogen peroxide, pelargonic acid, and azelaic acid semialdehyde,



$$C_8H_{17}C\begin{matrix}O\\ \diagup\quad\diagdown\\ O\text{—}O\end{matrix}CH\text{—}(CH_2)_7\text{—}COOH + H_2O \rightarrow$$

$$H_2O_2 + C_8H_{17}\text{—}CHO + CHO\text{—}(CH_2)_7\text{—}COOH$$

$$C_8H_{17}CHO + H_2O_2 \rightarrow C_8H_{17}COOH + H_2O$$

This confirms the original structure proposed for oleic acid, *i.e.*, that the double bond is in the center of the molecule.

Spontaneous Changes in Fats and Oils. Rancidity. The term rancidity is used in two different industries to represent two entirely different changes which take place in fats and oils: (1) the hydrolysis of the glycerides, with the liberation of free fatty acids; and (2) the oxidation of fats and oils containing unsaturated acids, resulting in the formation of aldehydes, ketones, and acids having a lower molecular weight than the acids which were naturally present. The term rancidity, as employed in the oil and fat industry, usually applies to the oxidative process, whereas in the dairy industry it applies to hydrolysis with the corresponding liberation of butyric acid which can be detected by its characteristic odor. The workers in the dairy industry refer to the oxidative process as producing "tallowy butter" in contrast to the "rancid butter" produced by the hydrolytic process. As a general rule, oxidation and hydrolysis occur simultaneously, so that probably from the industrial standpoint no sharp line of demarcation can be drawn, although from the chemical standpoint the two mechanisms are sharply differentiated.

The strong odor characteristic of the lower fatty acids, especially of butyric acid, is readily produced in milk fat upon slight hydrolysis. Most other fats contain relatively small amounts of the lower fatty acids and much larger amounts of the higher fatty acids, stearic, palmitic, etc., which are almost odorless. Accordingly, hydrolysis of such fats produces comparatively little off-odor. The tallowy odor is produced by the oxidation of the unsaturated fatty acids and the formation of aldehydes and ketones. The oxidation of oleic acid is mainly responsible for the intense tallowy odor. Oxidation of linoleic acid produces less off-odors, whereas the oxidation of linolenic acid produces very slight amounts of off-odors. Milk fat accordingly contains the particular composition of fatty acids which makes possible an intense odor resulting from either the hydrolytic or oxidative type of rancidity.

In the process of hydrolysis there is always an increase in titratable acidity. The presence of any trace of lipase (the fat-splitting enzyme) hastens and promotes this type of rancidity.

Oxygen is necessary in order to produce the oxidative type of rancidity. Heat, light, and moisture, together with the presence of certain metals which catalyze the reaction, hasten the oxidative process. Greenbank and Holm [32] noted that the greatest effect of light occurs in the ultraviolet at a wave length of approximately 3,600 Å. Coe and LeClerc [33] have stated that in cottonseed oil and corn oil oxidative rancidity appears to be due primarily to the photochemical action of light on a compound which probably exists simultaneously in the oil or is produced from compounds which give rise to peroxides. The wave lengths of light most active in catalyzing rancidity seemed to correspond with the light absorption regions of the oil.[34]

Among the metals affecting the reactions, the prooxidant activity of copper and iron salts is particularly great. Small traces in the order of one-fifth part per million of these catalysts exert a strong effect in promoting fat rancidity. The existence of iron in various complexes and in different valence states greatly affects the degree of catalytic activity.[35]

The character of the fat and the environmental conditions determine whether one obtains aldehydes, acids, ketones, etc. The Kreis color test (HCl + phloroglucinol + ether + the suspected fat) is often used for detecting deterioration due to oxidative rancidity, although a positive Kreis test in the crude oil may be due, not to decomposition products of the fat, but to substances derived from the seed from which the fat or oil was obtained.[36]

The Kreis color test is due to the presence of epihydrin aldehyde, $CH_2(-O-)CH-CHO$, and unless this aldehyde is present the product will not give a positive Kreis test. Accordingly the absence of color development in the Kreis test is not an invariable proof of the absence of rancidity, although Triebold [37] found a straight-line relationship between the color intensity of the Kreis test and the amount of oxygen absorbed per unit weight of the fat. This was true for a considerable variety of fats.

The oxidative process can be divided into two periods: (1) the period of induction, and (2) the period of active oxygen absorption. During the period of induction there is a negligible absorption of oxygen, and the susceptibility of a fat to oxidation may be determined by ascertain-

[32] G. R. Greenbank and G. E. Holm, *Ind. Eng. Chem.*, **25**, 167 (1933).
[33] M. R. Coe and J. A. LeClerc, *Ind. Eng. Chem.*, **26**, 245 (1934).
[34] M. R. Coe, *Oil & Soap*, **18**, 241 (1941).
[35] J. F. Lingenfelter and W. A. Gortner, *Federation Proc.*, **4**, 97 (1945).
[36] W. B. Smith, *Ind. Eng. Chem.*, **12**, 764 (1920).
[37] H. O. Triebold, *Cereal Chem.*, **8**, 518 (1931).

ing under specified conditions the relative length of the induction period. Figure 112 [38] shows the variation of the induction period, in terms of rate of oxygen absorption plotted against time, for several samples of lard. There was a marked correlation between the ease with which these various samples of fats became rancid when used in baked products and the shortness of the induction period.

The form of the curves for oxygen absorption is that of an autocatalytic reaction. Most workers agree that autoxidation is a chain

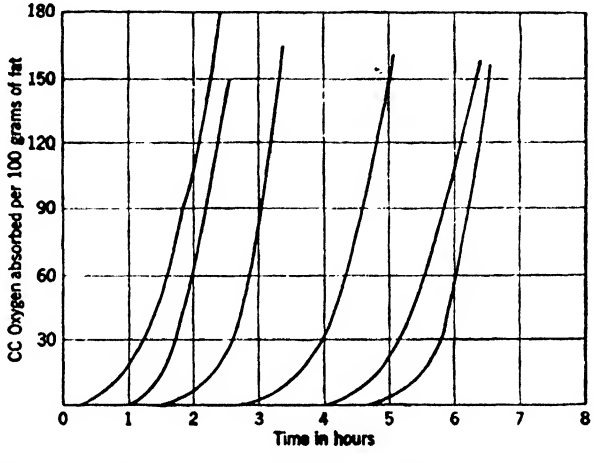


FIG. 112. Showing the variation in the length of induction period of oxygen absorption for six commercial samples of lard. Temp. 95°C. (Data of Triebold.)

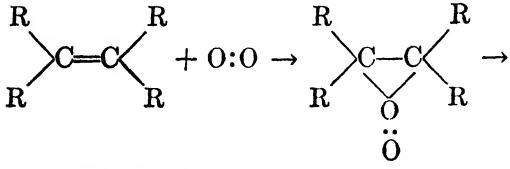
reaction. Such a mechanism explains why a trace of an oxidized fat may catalyze oxidative rancidity in a large mass of neutral fat. The chain reaction mechanism also explains the nature of the induction period [39] and the high rate of reaction which follows the close of the induction period.

At the end of the induction period, peroxide formation increases considerably. Simultaneously, linoleic acid decreases, the total unsaturation decreases, the mean length of the carbon chain of the fatty acids is shortened, and there is an increase in the conjugation of the unsaturated linkages.[40]

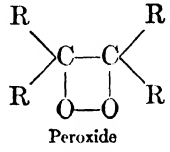
Staudinger [41] has considered that the primary attack of the oxygen is at the double bond of the fatty acid. The reactive *moloxide* thus

[38] H. O. Triebold and C. H. Bailey, *Cereal Chem.*, **9**, 50 (1932).
[39] H. N. Stephens, *J. Am. Chem. Soc.*, **58**, 219 (1936).
[40] L. J. Filer, Jr., K. F. Mattil, and H. E. Longenecker, *Oil & Soap*, **22**, 196 (1945)
[41] H. Staudinger, *Ber.*, **58B**, 1075 (1925).

formed later rearranges to form a more stable *peroxide,* which in turn may break down to form scission products or may polymerize.

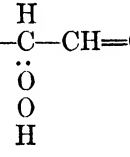


Unsaturated fat Moloxide



Peroxide

→ Polymerization or decomposition products

Another group of workers, led by Farmer and Sutton,[42] postulate the initial formation of a *hydroperoxide,*  , by addition of oxygen at a carbon in the α-position to the double bond. This is offered as an explanation for the frequent lack of parallelism in oxygen absorption and decrease in unsaturation. Atherton and Hilditch [43] have reported that, at 20°, the hydroperoxide was formed on methyl oleate, but at high temperatures (120°) the oxidation proceeded differently, attacking the double bond without hydroperoxide formation. Such observations would appear to reconcile the divergent theories as to the initial steps in oxidative rancidity.

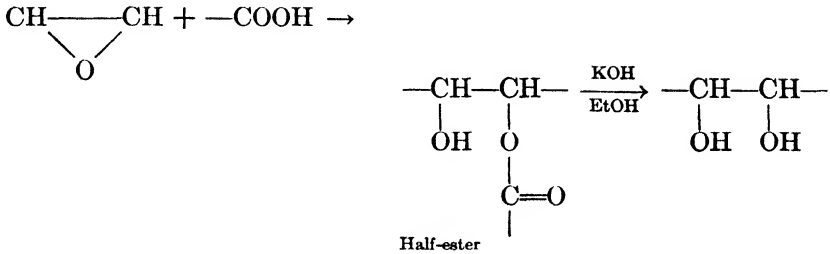
Deatherage and Mattill,[44] by measuring the oxygen consumption and analyzing the oxidation residues of four mono-unsaturated fatty compounds, have obtained evidence on the course of the rancidity reactions. Their data indicate that one of several changes may follow peroxide formation at the double bond. The carbon chain may cleave to yield aldehydes, and these in turn may autoxidize further to form peracids and acids. Or the peroxide may react with another double bond to form two moles of an ethylene oxide derivative. The dihydroxystearic acid

[42] E. H. Farmer, *Trans. Faraday Soc.,* **38,** 340 (1942); E. H. Farmer, G. F. Bloomfield, A. Sundralingam, and D. A. Sutton, *ibid.,* p. 348; E. H. Farmer and D. A. Sutton, *J. Chem. Soc.,* 119 (1943); D. A. Sutton, *ibid.,* 242 (1944).

[43] D. Atherton and T. P. Hilditch, *J. Chem. Soc.,* 105 (1944).

[44] F. E. Deatherage and H. A. Mattill, *Ind. Eng. Chem.,* **31,** 1425 (1939).

formed when oleic acid is oxidized is believed to arise from such a reaction:



Antioxidants. Moureu and Dufraisse [45] were among the pioneers in the study of antioxygenic catalysts. They noted that one molecule of hydroquinone as an antioxidant could protect 40,000 molecules of acrolein from autoxidation, and they accordingly regarded antioxidants as catalysts.

The mechanism which they proposed for the reaction was that the oxidant A unites with oxygen to form a peroxide, AO_2. This peroxide oxidizes the antioxidant B with the formation of a peroxide, BO, the oxidant being transformed at the same time to a lower oxide, AO. AO and BO are antagonistic and mutually react to regenerate the three original molecules, A, B, and O_2 in their original state.

$$A + O_2 \rightarrow AO_2$$

$$AO_2 + B \rightarrow AO + BO$$

$$AO + BO \rightarrow A + B + O_2$$

Thus, in this scheme the original molecule of the antioxidant is regenerated over and over again.

Of course, the mechanism of Moureu and Dufraisse would require some external source of energy, since some free energy would have to be lost for each of the first two reactions to take place. Accordingly, energy equal to the sum of these free energy losses would have to be gained for the final reaction to proceed.

Mattill and Olcott and co-workers [46] have made very careful studies

[45] C. Moureu and C. Dufraisse, *Chem. Revs.*, **3**, 113 (1926); *cf.* also S. Berkman, J. C. Morrell, and G. Egloff, Booklet 206, Universal Oil Products Co., Chicago, 1936.

[46] H. A. Mattill, *J. Biol. Chem.*, **90**, 141 (1931); H. S. Olcott, *J. Am. Chem. Soc.*, **56**, 2492 (1934); R. B. French, H. S. Olcott, and H. A. Mattill, *Ind. Eng. Chem.*, **27**, 724 (1935); H. S. Olcott and H. A. Mattill, *Oil & Soap*, **13**, 98 (1936); L. A. Hamilton and H. S. Olcott, *ibid.*, **13**, 127 (1936); H. S. Olcott and H. A. Mattill, *J. Am. Chem. Soc.*, **58**, 1627 (1936); *ibid.*, **58**, 2204 (1936); and L. A. Hamilton and H. S. Olcott, *Ind. Eng. Chem.*, **29**, 217 (1937).

of the oxidative process with particular reference to the nature of the reaction and the chemical mechanism of the antioxidants. They suggest that inhibitors act by breaking the chain and in particular by destroying the moloxides and peracids and that the inhibitors have little or no effect on the peroxides. This probably explains why determination of the amount of peroxides present does not conclusively measure either oxidative rancidity or the amount of oxygen which has been absorbed by fat. For example, lard is in a stage of incipient rancidity when the peroxide number (milliequivalents of peroxide oxygen per kilogram of fat) is in the low range of 10–15. Incipient rancidity for hydrogenated vegetable shortenings is generally at a slightly higher level of peroxides, and for such vegetable oils as peanut and cottonseed oils, rancidity is not approached until the peroxide value is up to 40–80.

During the induction period, the antioxidants or inhibitols are breaking the chains almost as fast as they are formed, but in this process they are themselves being used up or converted into other inactive products. When all the inhibitor has been transformed into inert products, the chains which form are no longer broken, and the oxidative reaction proceeds at its normal but very high rate. Most investigators now agree that the inhibitor (or inhibitol) has been destroyed at the close of the induction period. Gallic acid, a very effective antioxidant for vegetable fats, has completely disappeared when the induction period is ended.[47] Tocopherol also is oxidized during the induction period.[48] Thus, the antioxidant cannot be a true anticatalyst in the sense of Moureu and Dufraisse.

Mattill and Olcott studied the inhibiting effect of a great variety of organic compounds. They found that *o*- and *p*-diphenols are extremely effective as antioxidants. *m*-Diphenols, on the other hand, are inactive. In order to be effective, the hydroxyl groups must be on the benzene ring and not on aliphatic side chains. Inositol is completely ineffective. Only one hydroxyl group is required on the naphthalene nucleus. α-Naphthol is much more effective than β-naphthol. Quinone is a very effective antioxidant. β-Naphthoquinone is very effective. α-Naphthoquinone is essentially inert.

Many of the vegetable oils contain a natural inhibitor which Mattill and Olcott designated by the term inhibitol. This natural inhibitor occurs particularly in wheat germ and cottonseed oil. The available evidence points to inhibitol and vitamin E as being identical. Rather interestingly, the inhibitol concentrates are very potent antioxidants, so far as lard and other animal fats are concerned. They also protect

[47] L. J. Filer, Jr., K. F. Mattil, and H. E. Longenecker, *Oil & Soap*, **21**, 289 (1944).
[48] C. Golumbic, *Oil & Soap*, **20**, 105 (1943).

purified fatty acids and fatty acid esters, but they are less effective in protecting the vegetable oils from which they were obtained. These observations are in line with data which indicate that animal fats are generally low in vitamin E (tocopherol) content and that higher levels of the tocopherols are decreasingly effective as antioxidants. Gum guaiac,[49] N.D.G.A. or nordihydroguaiaretic acid,[50] and soybean lecithin have proved successful in prolonging the shelf-life of lard and are being used commercially. Indeed, appreciable benefit in an increased stability for lard may be obtained merely by the incorporation of relatively small amounts of vegetable oils,[51] apparently owing to the tocopherol content of the latter. Animal fats have very little natural antioxidants.

Fatty acid monoesters of L-ascorbic and D-isoascorbic acids have been prepared [52] and tested as antioxidants for fats.[53] They proved to be markedly synergistic with phospholipids and α-tocopherol in their antioxidant effect. Their ability to inhibit rancidity was appreciably greater for the vegetable oils than for lard.

Phosphoric and citric acids have proved to be effective synergists when used in combination with N.D.G.A.[54] and other antioxidants.

Drying Oils. Natural fats and oils exhibit two kinds of changes in the presence of air or oxygen. One type is the oxidative rancidity which we have just discussed. The other type is exhibited by the more highly unsaturated or drying oils which absorb oxygen and polymerize to stable insoluble films. These drying oils are valuable in the paint and varnish industry.

The drying oils include particularly those in which the glycerides are at least in part esterified with acids belonging to the linoleic or linolenic series. Chief among the drying oils are tung oil, linseed oil, poppyseed oil, sunflower seed oil, hempseed oil, walnut oil, etc. Intermediate between the true drying oils and non-drying oils are a group of semi-drying oils including such oils as soybean oil and to a lesser degree corn oil, rapeseed oil, cottonseed oil, mustard oil, etc. These oils are more unsaturated than such non-drying oils as castor oil, olive oil, and peanut oil. In general, the iodine number of an oil may be taken as a probable indication of whether it is a drying oil, semi-drying oil, or a non-drying oil. The higher the iodine number, the greater is the probability that

[49] H. S. Mitchell and H. C. Black, *Ind. Eng. Chem.*, **35**, 50 (1943).

[50] W. O. Lundberg, H. O. Halvorson, and G. O. Burr, *Oil & Soap*, **21**, 33 (1944).

[51] R. W. Riemenschneider, J. Turer, and W. C. Ault, *Oil & Soap*, **21**, 98 (1944).

[52] D. Swern, A. J. Stirton, J. Turer, and P. A. Wells, *Oil & Soap*, **20**, 224 (1943).

[53] R. W. Riemenschneider, J. Turer, P. A. Wells, and W. C. Ault, *Oil & Soap*, **21**, 47 (1944).

[54] K. F. Mattil, L. F. Filer, Jr., and H. E. Longenecker, *Oil & Soap*, **21**, 160 (1944)

it will be a drying oil. However, the actual test for a drying oil is the exposure to air of a film of the oil on a glass plate. A drying oil should form a dry film which is insoluble in acetone in 2 to 6 days. A semi-drying oil will be somewhat sticky after even a week's exposure. A non-drying oil will still be fluid after 18 or 20 days of exposure to air.

As might be anticipated where such a vast industry as the oil and paint trade is concerned, there is an enormous literature with regard to the drying properties of oils. However, there is still no unanimity of opinion about what happens when a drying oil changes into an insoluble and resistant film. Either oxygen or heat or both are essential to the process.

The modern view and the historical background are excellently summarized by Bradley [55] who considers that the predominant reaction is polymerization by a bimolecular addition of unsaturated fatty acids. The theory of Scheiber [56] that isomerization leading to conjugation of the double bonds in fatty acids precedes polymerization has gained considerable support and experimental confirmation.[55, 57-59] It is interesting to note that tung oil, an excellent drying oil, is rich in eleostearic acid, the polyethylenic linkages of which are already conjugated. The rate of polymerization is much slower for linseed oil, owing to the limiting effect of the conjugating reaction.

It appears that the chemical changes occurring in drying oils under the influence of heat and oxygen consist of (1) a shift in the unsaturated linkages to form conjugated esters, followed by (2) a 1,4-diene addition leading to the polymer film.

The phenomena of polymerization and condensation were the subject of a general discussion before the Faraday Society in September, 1935.[60] Although all the papers of this symposium are pertinent to the question of the transformations taking place in the drying of oils, the paper by Carothers [61] is particularly recommended.

Both Carothers and Bradley stress the fact that the free fatty acids of linseed and tung oil do not show drying properties. The esters of these acids with monoatomic alcohols do not show drying properties.

[55] T. F. Bradley, *Ind. Eng. Chem.*, **29**, 440 (1937); *ibid.*, **29**, 579 (1937); T. F. Bradley and W. B. Johnston, *ibid.*, **32**, 802 (1940); T. F. Bradley and D. Richardson, *ibid.*, **32**, 963 (1940).

[56] J. Scheiber, *Farbe u. Lack*, 585 (1929).

[57] J. S. Brod, W. G. France, and W. L. Evans, *Ind. Eng. Chem.*, **31**, 114 (1939).

[58] W. C. Ault, J. C. Cowan, J. P. Kass, and J. E. Jackson, *Ind. Eng. Chem.*, **34**, 1120 (1942).

[59] L. J. Filer, Jr., K. F. Mattil, and H. E. Longenecker, *Oil & Soap*, **22**, 196 (1935).

[60] *Trans. Faraday Soc.*, **32**, 1 (1936).

[61] W. H. Carothers, *Trans. Faraday Soc.*, **32**, 39 (1936).

The monoglycerides of these acids do not show drying properties. In general, the diglycerides of these acids do not show drying properties, although through heat treatment the diglycerides may acquire drying properties. During the heat treatment there is a loss of volatile constituents which may amount to as much as 10 per cent of the weight of the material, and Bradley believes that the heat treatment converts a considerable fraction of the diglycerides into triglycerides. The mono- and diglycol esters of the fatty acids of linseed and tung oils do not show drying properties prior to a heat treatment. It remains therefore for the triglycerides to be relatively specific in exhibiting the drying characteristics.

It appears that in the particular fatty acids characteristic of the drying oils and in their particular arrangement relative to one another when they are esterified in the form of triglycerides, the various functional groups have acquired space relationships to favor isomerization, intermolecular condensations, and polymerizations, the resultant of which is the insoluble dry film which makes these triglycerides such valuable industrial products.

Saponification of Fats. The Formation of Soaps. Saponification of fats by alkalies brings about hydrolysis, with the formation of glycerol and the salts of the fatty acids. The metallic salts of the higher fatty acids are soaps. The behavior of soaps as colloidal systems in themselves and as emulsifying agents has already been discussed under colloids.[62] Calcium, magnesium, and iron soaps are insoluble, the former being the scum which forms from sodium soaps in hard water.

Six general methods have been used to saponify fats:

1. The acidification process, in which the fat is heated with dilute sulfuric acid or hydrochloric acid under pressure at a temperature exceeding 100°C.

2. The autoclave process, in which the fats are treated with steam under pressure in the presence of a small amount of calcium hydroxide.

3. Saponification with sodium hydroxide or potassium hydroxide solutions at the boiling temperature.

4. The cold saponification process, in which the required amount of

[62] For further references see Martin H. Fischer, with the collaboration of G. D. McLaughlin and M. O. Hooker, *Soap and Proteins*, John Wiley & Sons, New York, 1921 (out of print); J. W. McBain, "Colloidal Chemistry of Soap," Third Report on Colloid Chemistry, *Brit. Assoc. Advancement Sci.*, 2 (1920); J. W. McBain and E. Walls, Fourth Report on Colloid Chemistry, *ibid.*, p. 244 (1922); J. W. McBain, Chap. V of *Colloid Chemistry, Theoretical and Applied*, Vol. I, edited by Jerome Alexander, Chemical Catalog Co., New York, 1926; and W. H. Nuttall, Fifth Report on Colloid Chemistry, *Brit. Assoc. Advancement Sci.*, 38 (1923).

concentrated alkali is added and the mixture is allowed to stand for several days.

5. The Twitchell process [63] in which saponification is carried out in the presence of sulfobenzenestearic acid which acts as a catalyst.

6. The biological method, in which lipase or lipolytic enzymes are added to the fat, resulting in the formation of the free fatty acids and glycerol.

A discussion of the chemistry or technology of the soap manufacturing process lies outside of the scope of this volume. Here again there is a voluminous literature. Soaps have been known since almost the dawn of history. Perhaps, however, it is pertinent to make a mere mention of the recent development which has resulted in the production of the products known to the housewife under such names as "Dreft" and "Dreen." Someone has said that these products represent the only major change in the manufacture of detergents since the days of Cleopatra. These products are not soaps in the ordinary sense of the term but instead are the sodium salts of the sulfonic acids of the higher fatty alcohols, principally lauryl alcohol, cetyl alcohol, and octodecyl alcohol. The corresponding fatty acids are reduced to the alcohol by catalytic hydrogenation. The alcohols are then converted into the corresponding sulfonic acids. The sodium salts of these sulfonic acids possess the extremely desirable property of being marked surface-tension depressants, excellent detergents, and likewise possess the property of forming soluble salts with calcium and magnesium. Accordingly they can be used as detergents in the hardest of water, even in saturated limewater. Furthermore the sulfonic acids are stronger acids than acetic acid, so that the sodium salts are not decomposed in weak solutions of the organic acids and exert their detergent properties even under acid conditions. Because of the high solubility of the salts of the alkaline earths of these products, they are readily rinsed from the fiber after the washing process, and, because their solutions are not alkaline, there is less damage to wool and silk goods in the washing process than when somewhat alkaline soap solutions are used.

We have noted that fats are hydrolyzed by lipase or lipolytic enzymes. In general, no marked specificity is exhibited by the lipases from various biological sources. However, the particular mixture of glycerides present in any given fat or oil does have a marked influence on the *rate* at which the various fats or oils are hydrolyzed by a given lipase. This is excellently demonstrated by the action of ricinus lipase [64] on a series

[63] *Ind. Eng. Chem.*, **9**, 192 (1917).

[64] H. E. Longenecker and D. E. Haley, *J. Am. Chem. Soc.*, **57**, 2019 (1935).

of fats and oils. The order of hydrolysis of the oils in this particular series of experiments was peanut > castor > corn > cottonseed > soybean > rape > olive > linseed > neat's foot > peach kernel > coconut > whale > fish > sperm. At the end of 10 hours, 97.2 per cent of the peanut oil had been hydrolyzed and only 19.5 per cent of the sperm oil. The lipase has no action on cholesterol esters.[65]

Glycerol, the Alcohol Present in Fats. Glycerol is a triatomic alcohol having the formula $CH_2OH—CHOH—CH_2OH$. A major commercial source of glycerol has been the waste water of soap factories. Stimulated by the necessity of finding new sources of glycerol for the preparation of nitroglycerin, attention has been directed to the biological synthesis of this important substance. It has been prepared by bacterial and yeast fermentations of carbohydrates.[66] In yeast fermentation, certain strains of yeast were found to be resistant to sodium sulfite. When they were added to solutions containing dextrose and sodium sulfite, alcoholic fermentation was largely inhibited, whereas glycerol fermentation was stimulated.

Recently, synthetic glycerol has been manufactured on a commercial scale. Propylene obtained from petroleum is converted to allyl chloride, and this in turn is converted to glycerol.

Glycerol is miscible in water in all proportions, and when strongly heated either alone or in the presence of a dehydrating agent, such as potassium bisulfate, it decomposes to form acrolein, $CH_2{=}CH—CHO$. Acrolein possesses a very penetrating, acrid odor. The formation of acrolein may be regarded as a test for the presence of glycerol. The odor of acrolein is noticeable when a tallow candle is blown out or when a fat is heated to too high a temperature.

The oxidation products of glycerol depend on the oxidizing agent employed, *i.e.*, the oxidation potential. It may be oxidized to glyceric aldehyde, or completely oxidized to carbon dioxide and water. As one might expect, inasmuch as it is closely related to sugars, glycerol is relatively easily oxidized. Glycerol may be regarded as the alcohol formed by the reduction of the triose, glycerose. It shows the sugar character, inasmuch as it reduces many metallic salts, forms glycerolates with alkalies, is esterified by acids, *e.g.*, glycerol phosphoric acid ester, and in fact has all the properties which might be expected to accompany a trihydroxy alcohol.

[65] F. E. Kelsey, *J. Biol. Chem.*, **130**, 187 (1939).

[66] *Chem. Age, London*, **28**, 352 (1920); *cf.* also G. Barger, *Ann. Repts. on Progress Chem., Chem. Soc. London*, **16**, 166 (1919); A. Guilliermond, *The Yeasts* (translated and revised by F. W. Tanner), John Wiley & Sons, New York, 1920 (out of print); W. Connstein and K. Lüdecke, *Ber.*, **52B**, 1385 (1919).

THE WAXES

Waxes differ from fats in that they are esters of mono- (or in some instances di-) hydroxy saturated alcohols or of sterols with certain of the higher fatty acids. Some of the higher alcohols of the $C_nH_{2n+2}O$ series which occur in waxes are shown in Table 61. Most of the alcohols shown in this table occur in nature only in the waxes. The waxes occur as

TABLE 61. SOME HIGHER ALCOHOLS OF THE C_nH_{2n+2} SERIES OCCURRING IN WAXES

Compound	*Empirical Formula*	*Occurrence*
n-Dodecanol (lauryl alcohol)	$C_{12}H_{26}O$	Cuticle wax of *Cascara sagrada*
n-Tetradecanol (myristyl alcohol)	$C_{14}H_{30}O$	Sperm head oil
n-Hexadecanol (cetyl alcohol)	$C_{16}H_{34}O$	Spermaceti
n-Octadecanol (stearyl alcohol)	$C_{18}H_{38}O$	Spermaceti
Carnaubyl alcohol	$C_{24}H_{50}O$	Wool fat
Neoceryl alcohol	$C_{25}H_{52}O$	Beeswax (*Apis mellifera*)
n-Hexacosanol (ceryl alcohol)	$C_{26}H_{54}O$	Beeswax, wool fat, cuticle wax
n-Octacosanol	$C_{28}H_{58}O$	Apple cuticle wax, wax of wheat blades
Montan alcohol	$C_{29}H_{60}O$	Beeswax
n-Triacontanol (myricyl alcohol)	$C_{30}H_{62}O$	Carnauba wax, beeswax, sugar cane wax
Cocceryl alcohol	$C_{30}H_{62}O_2$	Cochineal wax (*Coccus cacti*)
Melissyl alcohol	$C_{31}H_{64}O$	Beeswax
Laccerol	$C_{32}H_{66}O$	Lac wax (*Coccus lacca*)
Incarnatyl alcohol	$C_{34}H_{70}O$	Beeswax, clover wax, especially *Trifolium incarnatum*

insect secretions and as protective coatings on the cuticle of the leaves or fruits of plants. They rarely occur as cell constituents.

Waxes are much more difficultly saponified than are those fats and oils which are esters of glycerol. As a rule, saponification of waxes requires rather long boiling with alcoholic potassium hydroxide. The fatty acids of the waxes, of course, pass into the alkali as "soaps," but the higher alcohols are not water-soluble and appear in the "unsaponifiable matter" fraction. Thus, whereas the unsaponifiable matter of a true fat may amount to only 1 or 2 per cent of the material taken, the corresponding fraction for the liquid waxes may range from 31 to 43 per cent, and for the solid waxes may reach as much as 55 per cent of the material which was saponified. Chibnall, *et al.*,[67] have studied the

[67] A. C. Chibnall, S. H. Piper, A. Pollard, J. A. B. Smith, and E. F. Williams, *Biochem. J.*, **25**, 2095 (1931); and A. Pollard, A. C. Chibnall, and S. H. Piper, *ibid.*, **25** 2111 (1931).

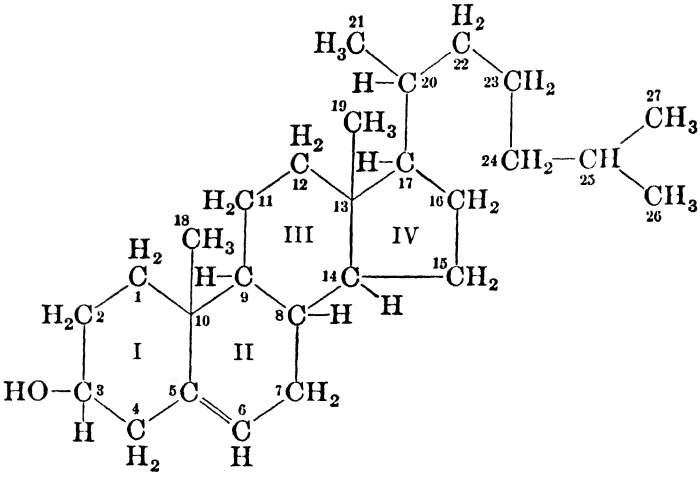
wax constituents of apple cuticle and of certain of the grasses. Their papers may be consulted for some of the technics which can be applied in studies of this type.

THE STEROLS

In discussing the saponins we have already indicated that those compounds, as well as the toad poisons, the sex hormones, and the sterols are derivatives of the cyclopentenophenanthrene nucleus. The structure of the sterols has been a very active subject of investigation for many years, but it was in the period 1932–1934 that the structure of cholesterol was finally ascertained.

The name sterol is derived from the Greek *stereos*, meaning solid, with the *ol* ending, and therefore means literally a solid alcohol. The Greek prefix *chole* means bile, and therefore cholesterol means literally the solid alcohol derived from bile. Cholesterol has been known for many years as the most common constituent of human gallstones. It is present in all cells of the animal organism, and in the higher animals it occurs in large amounts in the brain and nerve tissues. It is also present in very appreciable amounts in egg yolk and in the suprarenal gland. Cholesterol contains one hydroxyl group on carbon-3 and one double bond between carbon-5 and carbon-6. It occurs in animal tissues in association with the hydrogenated compound, dihydrocholesterol, where the double bond between carbon-5 and carbon-6 has been saturated.

In animal feces there is excreted a sterol known as coprosterol, from the Greek *kopro*, meaning dung. Coprosterol is a dihydrocholesterol isomeric with the dihydrocholesterol occurring in normal cells and tissues. Apparently the isomerism is on carbon-5 of the sterol nucleus. It will be noted that, if the double bond between carbon-5 and carbon-6 becomes hydrogenated, carbon-5 becomes an asymmetric carbon atom. There is thus the possibility of a *cis-trans* isomerism revolving around the groups on carbon-5. The —OH group at carbon-3 is not involved in this particular isomerism, since, when cholesterol is reduced to the hydrocarbon *cholestane* (CH_2 at carbon-3 and the double bond at carbon-5 and carbon-6 saturated), a different hydrocarbon is obtained from that which results from the reduction of coprosterol. The hydrocarbon from coprosterol is known as *coprostane* and is isomeric with cholestane. In cholestane ring I and ring II correspond to a *trans*-decalin. In coprostane we have *cis*-decalin. Another way of expressing this is that in cholestane the hydrogen on carbon-5 is directed back of the plane of the diagram, whereas in coprostane the hydrogen is directed in the "near" position.



Cholesterol

Cholesterol can be conveniently prepared from sheep or hog brains by extracting the macerated, dehydrated tissue with ether. Crude cholesterol dissolves and can be crystallized directly by concentrating the ether extract, or the ether extract can be saponified and the cholesterol extracted and recrystallized from the non-saponifiable residue.

A series of color reactions is produced by the sterols when they are treated with strong acids under dehydrating conditions. In the *Liebermann* or *Liebermann-Burchard reaction* the sterol is treated with chloroform, acetic anhydride, and concentrated sulfuric acid to produce a green color. In the *Salkowski reaction* a solution of the sterol in chloroform is shaken with concentrated sulfuric acid. A cherry-red color appears in both layers. The *Steinle-Kahlenberg reaction* produces a clear purple solution changing to cobalt blue on exposure to light when the sterols-in-chloroform solution is treated with antimony pentachloride. In the *Tschugajeff reaction* a glacial acetic acid solution of the sterol is boiled after the addition of zinc chloride and acetyl chloride. Whitby [68] and Schoenheimer, Dam, and von Gottberg [69] have outlined the use of the color reactions for the colorimetric estimation of the sterols.

The color reactions are apparently specific for the unsaturated sterols, for they are not given by dihydrocholesterol, coprosterol, or the hydrocarbons derived from the sterols. The *Rosenheim reaction*, where the sterol is dissolved in chloroform and a few drops of a strong aqueous solution of trichloracetic acid is added, appears to be more or less specific for ergosterol and for sterols having a double bond between carbon-4

[68] G. S. Whitby, *Biochem. J.*, **17**, 5 (1923).

[69] R. Schoenheimer, H. Dam, and K. von Gottberg, *J. Biol. Chem.*, **110**, 659 (1935).

and carbon-5. Bloor's [70] colorimetric method for cholesterol is an adaptation of the Liebermann-Burchard reaction.

Cholesterol may be estimated gravimetrically [71] as the cholesteride of digitonin inasmuch as cholesterol forms an insoluble precipitate with digitonin as has already been noted.

Cholesterol inhibits the reaction of saponins and prevents the hemolysis of red cells in the presence of saponins.[72] The esters of cholesterol do not inhibit the action of saponins. It thus appears probable that cholesterol forms a definite compound with all saponins as it does with digitonin.

In 1934, Bergmann [73] isolated a new sterol from the nonsaponifiable matter of oysters, *Ostrea virginica*. The same compound was isolated from the clam, *Venus mercenaria*, and was named ostreasterol. It appears to be the only sterol present in the oyster and replaces cholesterol in this organism and perhaps in all members of the *Lamellibranchiata*. It has the formula $C_{29}H_{48}O$, possesses two double bonds, and is isomeric with the plant sterol, stigmasterol, since on catalytic reduction it yields the same saturated alcohol, sitostanol, as do sitosterol and stigmasterol. Bergmann points out that the above facts are very significant inasmuch as this represents the first case in which a sterol derived from an animal source has been shown to be convertible into a typical sterol of plant origin, and he raises the question whether there is a sharp line of demarcation between the plant sterols (phytosterols) and animal sterols (zoösterols).

Some of the sterols identified from natural materials are listed in Table 62. All these sterols are characterized by a hydroxyl group on carbon-3 and in a *cis*-position to the methyl group on carbon-10.

Digitonin thus precipitates these sterols, but will not form insoluble compounds with their *epimers* having the 3-hydroxy group in the *trans*-position. Furthermore, with the exception of coprosterol, all these sterols appear to have a *trans*-orientation of ring I to ring II along the line between carbon-5 and carbon-10.

It will be noted that in most of the naturally occurring sterols the double bond between carbon-5 and carbon-6 of the cholesterol nucleus is present. The only known exceptions are zymosterol and spinasterol. The C_{28} and C_{29} sterols are generally cholestane derivatives in which a methyl or ethyl group replaces a hydrogen on carbon-24.

[70] W. R. Bloor, *J. Biol. Chem.*, **24**, 227 (1916).

[71] R. Girardin and E. Spach, *Bull. soc. chim. biol.*, **8**, 813 (1926).

[72] F. Ransom, *Deut. med. Wochschr.*, **27**, 194 (1901); *cf.* also W. Hausmann, *Beitr. chem. Physiol.*, **6**, 567 (1905); E. Abderhalden and E. R. Le Count, *Z. exptl. Path Therap.*, **2**, 199 (1906).

[73] W. Bergmann, *J. Biol. Chem.*, **104**, 316, 553 (1934).

It appears likely that a number of sitosterols exist in plants. β-Sitosterols and γ-sitosterols probably differ only in the configuration of the ethyl group on carbon-24.

Cinchol, found in cinchona bark, appears to be identical with β-sitosterol.[74] Lanosterol [75] and cafesterol [76] have been shown not to be true sterols, and hence incorrectly named.

TABLE 62. SOME NATURALLY OCCURRING STEROLS

Name	*Formula*	*Double Bonds*	*Occurrence*
Dihydrocholesterol	$C_{27}H_{48}O$	0	Animal tissues
Coprosterol	$C_{27}H_{48}O$	0	Feces
Cholesterol	$C_{27}H_{46}O$	1 ($\Delta^{5:6}$) *	Animal tissues, gallstones
Microcionasterol	$C_{27}H_{46}O$	1 †	Sponges
Actiniasterol	$C_{27}H_{44}O$	2 ‡	Sea anemones
7-Dehydrocholesterol	$C_{27}H_{44}O$	2 ($\Delta^{5:6,\ 7:8}$) §	Horned snail
Zymosterol	$C_{27}H_{44}O$	2 ($\Delta^{9:11\text{ or }8:14,\ 24:25}$) ‖	Yeast
Campesterol	$C_{28}H_{48}O$	1 ($\Delta^{5:6}$) ¶	Soybean oil, wheat germ oil
Brassicasterol	$C_{28}H_{46}O$	2 ($\Delta^{5:6,\ 22:23}$) **	Rape seed oil
Ergosterol	$C_{28}H_{44}O$	3 ($\Delta^{5:6,\ 7:8,\ 22:23}$) ††	Yeast, ergot
β,γ-Sitosterols	$C_{29}H_{50}O$	1 ($\Delta^{5:6}$) ‡‡	Higher plants
Clionasterol	$C_{29}H_{50}O$	1 ($\Delta^{5:6}$) §§	Sponges
Stigmasterol	$C_{29}H_{48}O$	2 ($\Delta^{5:6,\ 22:23}$) ‖‖	Soybean, calabar bean
Spinasterol	$C_{29}H_{48}O$	2 ($\Delta^{8:9,\ 22:23}$) ¶¶	Spinach, cabbage
Fucosterol	$C_{29}H_{48}O$	2 ($\Delta^{5:6,\ 24:28}$) ***	Algae
Ostreasterol	$C_{29}H_{48}O$	2 †††	Oysters

* A. Windaus, *Z. physiol. Chem.*, **213**, 147 (1932).
† W. Bergmann and T. B. Johnson, *Z. physiol. Chem.*, **222**, 220 (1933).
‡ M. Deffner, *Z. physiol. Chem.*, **278**, 165 (1943).
§ A. Windaus and F. Bock, *Z. physiol. Chem.*, **245**, 168 (1937).
‖ H. Wieland, F. Rath, and W. Benend, *Ann.*, **548**, 19 (1941); B. Heath-Brown, I. M. Heilbron, and E. R. H. Jones, *J. Chem. Soc.*, 1482 (1940).
¶ E. Fernholz and W. L. Ruigh, *J. Am. Chem. Soc.*, **63**, 1157 (1941).
** E. Fernholz and H. E. Stavely, *J. Am. Chem. Soc.*, **62**, 428 (1940).
†† A. Windaus, H. H. Inhoffen, and S. Reichl, *Ann.*, **510**, 248 (1934).
‡‡ W. Dirscherl and H. Nahm, *Ann.*, **555**, 57 (1943).
§§ F. R. Valentine, Jr., and W. Bergmann, *J. Org. Chem.*, **6**, 452 (1941).
‖‖ E. Fernholz, *Ann.*, **507**, 128 (1933).
¶¶ H. E. Stavely and G. N. Bollenback, *J. Am. Chem. Soc.*, **65**, 1600 (1943).
*** H. B. MacPhillamy, *J. Am. Chem. Soc.*, **64**, 1732 (1942).
††† W. Bergmann, *J. Biol. Chem.*, **104**, 553 (1934).

[74] W. Dirscherl, *Z. physiol. Chem.*, **257**, 239 (1939).
[75] H. Schulze, *Z. physiol. Chem.*, **238**, 35 (1936).
[76] A. Wettstein and K. Miescher, *Helv. Chim. Acta*, **26**, 631 (1943).

The Sterols and Vitamin D. A great stimulus to reasearch in the field of the sterols followed the discovery of Rosenheim and Webster,[77] Steenbock and Black,[78] and Hess, Weinstock, and Helman,[79] who noted independently and at approximately the same time that apparently pure samples of cholesterol, when irradiated with ultraviolet light, acquired the properties of replacing the antirachitic vitamin, vitamin D, in the diet of an animal. After these discoveries had been made, many workers irradiated oils and fats from various plant and animal sources, irradiated various foodstuffs and various sterols, and found rather generally that the sterol fraction acquired antirachitic properties by irradiation. As the study progressed, however, it was noted that all samples of sterols did not acquire the antirachitic property to the same degree, and a number of workers found almost simultaneously that cholesterol which had been purified by chemical methods had lost the property of becoming antirachitic when irradiated.

Prior to this time, Hess, Weinstock, and Sherman [80] had noted that irradiated cholesterol lost its antirachitic properties when it was recrystallized, and that the apparently pure cholesterol, which could be obtained from the recrystallization of irradiated cholesterol, could not again be activated by irradiation. They also noted [81] that after irradiation all the sterol could not be precipitated by digitonin. At about the same time Rosenheim and Webster [82] observed that only the sterols obtained directly from plant or animal tissues, such as cholesterol, sitosterol, and ergosterol, could be activated by irradiation. The group of "excretory" sterols, including coprosterol, were not activated by irradiation—an indication that the presence of an unsaturated linkage was necessary in order that the substance should acquire antirachitic properties.

Hess and Anderson,[83] in a study of Anderson's α-, β-, and γ-sitosterols, found that the α-sitosterol became strongly antirachitic after irradiation, whereas the β- and γ-sitosterols acquired no antirachitic properties. At the time their paper was published the suggestion had been made that ergosterol was the precursor of vitamin D. They noted that their β- and γ-sitosterols had been purified by chemical process (bromination) and suggested that possibly this process had destroyed any precursor of the antirachitic factor which might originally have been

[77] O. Rosenheim and T. A. Webster, *Lancet*, **208**, 1025 (1925).
[78] H. Steenbock and A. Black, *J. Biol. Chem.*, **64**, 263 (1925).
[79] A. F. Hess, M. Weinstock, and F. D. Helman, *J. Biol. Chem.*, **63**, 305 (1925).
[80] A. F. Hess, M. Weinstock, and E. Sherman, *J. Biol. Chem.*, **66**, 145 (1925).
[81] A. F. Hess, M. Weinstock, and E. Sherman, *J. Biol. Chem.*, **67**, 413 (1926).
[82] O. Rosenheim and T. A. Webster, *Biochem. J.*, **20**, 537 (1926).
[83] A. F. Hess and R. J. Anderson, *J. Biol. Chem.*, **74**, 651 (1927).

present in these fractions as an impurity. Rosenheim and Webster further noted that, when cholesterol was purified by chemical means and had lost the property of becoming antirachitic by irradiation, it likewise no longer possessed a characteristic absorption spectrum band in the ultraviolet region [84, 85] which is characteristic for cholesterol purified only by recrystallization. It appeared, therefore, as if some impurity were present in cholesterol purified by physical means and that the presence of this impurity accounted for the characteristic ultraviolet absorption band and for the acquiring of antirachitic properties by the cholesterol preparation.

Accordingly Rosenheim, Webster, Hess, and Windaus began an intensive investigation of the nature of the impurity which might be present in the various sterol preparations and which might act as the provitamin of vitamin D. These workers were attacking the problem in different laboratories, but through a friendly agreement they were all kept informed of the progress of the work in their several institutions. Almost simultaneously Windaus and Hess [86] and Rosenheim and Webster [87] announced the finding that neither cholesterol which contains one double bond, sitosterol containing one double bond, nor stigmasterol containing two double bonds acquired antirachitic properties on irradiation after their purification by chemical methods. They did observe, however, that ergosterol, which contains three double bonds and which Rosenheim and Webster [88] had previously found to be after irradiation "highly protective even in doses of 1 mg.," was apparently present in small amounts in all the sterol preparations which could be activated by irradiation, and they announced that the provitamin of vitamin D is ergosterol.

It seemed at this point as though one of the major problems associated with the chemistry of vitamin D had been solved. However, in later work it was found that the product (calciferol) derived from irradiated ergosterol and thought for a time to be the naturally occurring vitamin D, did not possess the same antirachitic properties for chicks as are evidenced by the vitamin D of cod liver oil. The problems involved were attacked by a great many workers, and Bills [89] has been particularly active in this work.

The antirachitic potency of the naturally occurring vitamin D is not due to the presence of a single compound, but there are apparently a

[84] F. W. Schlutz and M. R. Ziegler, *J. Biol. Chem.*, **69**, 415 (1926).

[85] R. A. Morton, I. M. Heilbron, and E. D. Kamm, *J. Chem. Soc.*, 2000 (1927).

[86] A. Windaus and A. Hess, *Nachr. Ges. Wiss. Göttingen, Math.-physik. Klasse*, 175 (1926).

[87] O. Rosenheim and T. A. Webster, *Lancet*, **212**, 306 (1927).

[88] O. Rosenheim and T. A. Webster, *Biochem. J.*, **20**, 537 (1926); **21**, 389 (1927)

[89] C. E. Bills, *Physiol. Revs.*, **15**, 1 (1935).

number of antirachitic substances more or less closely related. When ergosterol is irradiated with ultraviolet light, ring II of the sterol is opened [90] between carbon-9 and carbon-10. Carbon-9 has two hydrogens attached to it, and a new double bond is formed [91] between carbon-10 and carbon-19, the methyl group at carbon-19 being converted to a methylene group. Somewhat later Windaus, Lettré, and Schenck [92] prepared 7-dehydrocholesterol by forming a double bond between carbon-7 and carbon-8 of the cholesterol molecule, and found that this compound not only gave the same ultraviolet absorption spectrum as ergosterol but that it, like ergosterol, could be converted by ultraviolet light into an antirachitic substance which was more potent than calciferol when tested on chicks.

A considerable number of compounds have been shown to have provitamin D activities. A list compiled mainly by Heilbron and Jones [93] is shown in Table 63. On the basis of later data indicating that 7-dehydrocampesterol [94] and 7-hydroxycholesterol [95] also have provitamin potency, these compounds have been added to Heilbron's list.

TABLE 63. PROVITAMIN D ACTIVITIES OF VARIOUS STEROLS

Positive	*Negative*
Ergosterol	Epi-lumisterol
Lumisterol	Pyrocalciferol
Epi-ergosterol	Isopyrocalciferol
Ergosterol-22-oxide	7-Dehydrostigmasterol
22-Dihydroergosterol	3-Hydroxy, $\Delta^{5:6,\ 7:8}$ choladienic acid
7-Dehydrocholesterol	3,17-Dihydroxy, $\Delta^{5:6,\ 7:8}$ androstadiene
7-Dehydrositosterol	
Epi-7-dehydrocholesterol	
7-Hydroxycholesterol	
7-Dehydrocampesterol	
22-Dihydrotachysterol	

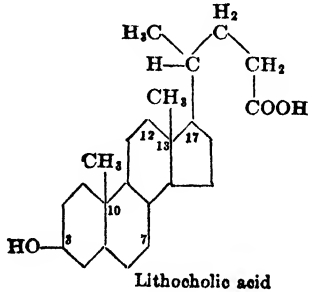
All provitamin D materials known thus far have two double bonds on carbons-5:6 and -7:8. However, not all sterols possessing unsaturated linkages at these points can acquire antirachitic properties under the influence of ultraviolet light, the lack of activity of 7-dehydrostigmasterol being an example. In some cases epimerization does not affect provitamin activity (epi-ergosterol), whereas in other cases the stereoisomer is inactive (epi-lumisterol).

[90] H. Lettré, *Z. angew. Chem.*, **47**, 736 (1934).
[91] R. Kuhn and E. F. Möller, *Z. angew. Chem.*, **47**, 145 (1934).
[92] A. Windaus, H. Lettré, and F. Schenck, *Ann.*, **520**, 98 (1935).
[93] I. M. Heilbron and E. R. H. Jones, *Ann. Rev. Biochem.*, **9**, 135 (1940).
[94] W. L. Ruigh, *J. Am. Chem. Soc.*, **64**, 1900 (1942).
[95] S. Lassen and E. Geiger, *Proc. Soc. Exptl. Biol. Med.*, **53**, 181 (1943).

The Isolation of Sterols. The sterols are contained in the unsaponifiable matter of fats and oils and are usually isolated by saponifying the fats, drying the soaps which are formed, and extracting these dry soaps with dry ether. The sterols are extracted by the ether, whereas the soaps are insoluble. The sterols are not water-soluble, but in the saponified fat are dispersed as an emulsion in the presence of the soap solution. Cottonseed oil and corn oil yield about 1 per cent of unsaponifiable matter. Olive oil yields somewhat less than 1 per cent, and wheat oil about 2.5 per cent. Fish and liver oils may contain as much as 10 to 20 per cent of unsaponifiable matter. In such oils, however, other substances, largely hydrocarbons, are present, so that the unsaponifiable matter does not represent the sterol content.

THE BILE SALTS

The bile of the mammals is a golden brown to greenish fluid with an alkaline reaction, secreted by the liver and stored in the gall bladder. From the gall bladder it passes in small amounts into the intestines, where its chief function apparently is to emulsify fats and oils and to promote their hydrolysis and the absorption of the fatty acids by the intestinal mucosa. Bile contains inorganic salts, cholesterol, bile pigments, and salts of the bile acids. Although the bile acids are not found in fats and oils and are characteristic of bile, nevertheless their structure is so similar to that of the sterols that it seems justifiable to consider them in connection with the chemistry of the sterols. All the common bile salts have the sterol nucleus with a hydroxyl group on carbon-3 as in cholesterol, but the aliphatic side chain on carbon-17 is shortened, the break occurring between carbon-24 and carbon-25 with the conversion of carbon-24 into a carboxyl group, as shown in the formula for *lithocholic acid*. *Desoxycholic acid* has an additional hydroxyl group on carbon-12. *Chenodesoxycholic acid* is isomeric with desoxycholic acid but has its two hydroxyl groups on carbon-3 and carbon-7.



Lithocholic acid

Cholic acid has three hydroxyl groups situated at carbon-3, carbon-7, and carbon-12. These are the four common bile acids, although many others have been found in bile of various animals. In bile they occur as the sodium salts and are conjugated with either glycine or taurine.

The sodium salts of the bile acids, and particularly sodium desoxycholate, possess the remarkable property of forming association complexes [96] with many water-insoluble materials which thereby become water-soluble. Thus, hydrocarbons, such as naphthalene and xylol, are readily "dissolved" by aqueous solutions of the bile salts. Camphor, the natural fats and oils, cholesterol, etc., are likewise rendered water-soluble and brought into a form in which they are rather readily diffusible through membranes. The association complexes which are formed by the bile salts and compounds of the type that we have been discussing are relatively stable. For example, the stearic acid-desoxycholic acid complex dissolves without change in alkali, and the stearic acid is so firmly bound that it can be split off completely only by drastic oxidation or dehydration of the bile acid.

ANALYSIS OF FATS AND OILS

Various special technics have been devised for the study of fats and oils. Some of the methods are essentially empirical but have nevertheless been retained by the industry.

We can give only a brief outline of some of the more important of the technics. For the detailed methods, some of the analytical handbooks should be consulted.[97]

Extraction of the Fats or Oils from the Plant or Animal Tissues. In order completely to extract the fats or oils from the tissues, the tissues are generally dried, and *the drying must take place without oxidation of the fat or oil.* Various drying methods have been proposed, *e.g.*, tissues have been dried by heat in an air oven, or by heat *in vacuo*, or by heat in a neutral gas, such as nitrogen. Drying by heat in air is not to be recommended, and whenever possible the drying should be done *in vacuo* or in the presence of a neutral gas. In some instances, it has been found desirable to dry material by adding absolute alcohol to abstract the water and, at a low temperature, distilling off the dilute

[96] H. Wieland and H. Sorge, *Z. physiol. Chem.*, **97**, 1 (1916).

[97] Association of Official Agricultural Chemists, *Official and Tentative Methods of Analysis*, 5th ed., Washington, D. C., 1940; A.C.S. Committee on Analysis of Commercial Fats and Oils, *Ind. Eng. Chem.*, **18**, 1346 (1926); and G. S. Jamieson, *Vegetable Fats and Oils*, 2nd ed., Reinhold Publishing Corp., New York, 1943; *cf.* Chap. VI, Methods.

alcohol so formed. In other instances, anhydrous calcium sulfate (plaster of Paris) or anhydrous sodium sulfate is used to combine with the water, the water being bound in the form of water of crystallization. In still other instances, notably in meat samples, the material has been frozen, and then the frozen material dried *in vacuo* over sulfuric acid, the desiccator being kept at a temperature below zero. Meat can be dried in this way with practically no oxidation of the fats, the meat sample retaining almost its original size and shape.

After drying, the fat or oil is extracted by some appropriate solvent. In general, dry neutral ether is chosen as the solvent, although chloroform or carbon tetrachloride has been used by some investigators. In the event that the material contains a high percentage of protein, it is sometimes necessary to follow the original ether extraction by extraction with absolute alcohol, and this in turn by a second ether extraction. This is markedly necessary where fat is adsorbed upon protein surfaces or where fats or fat-like compounds are combined with proteins to form the so-called lecithoproteins or lipoproteins. After extraction, the solvent is removed by evaporation at low temperature, leaving behind the sample of fat or oil for examination.

Physical Properties Which May Be Determined on the Fat or Oil Samples. A number of physical properties have been used to characterize fats and oils. These include:

1. The specific gravity, which may be determined by means of the hydrometer or preferably a pycnometer or a Westphal balance.

2. The melting point, which may be obtained by suspending a disc of the solid fat approximately midway in an aqueous alcohol solution which has been adjusted so as to have approximately the same specific gravity as the fat. This solution is then slowly heated, and the point is noted at which the disc of fat "rounds up" into a globule. This so-called melting point is more or less indefinite, because a fat is composed of a number of substances having different melting points.

3. The titer test, *i.e.*, the temperature at which the fatty acids, which have been liberated from the fat after the fat has been saponified by alkali, resolidify.

4. The refractive index, taken either with an Abbé butyro refractometer or a dipping refractometer. The refractive index is a very important and valuable index of the composition and purity of a fat or oil.

5. The optical rotation may be taken; sometimes it gives important data.

6. The viscosity, which is again an important criterion, is very largely used in the industry. The Saybolt or the MacMichael viscometers are more commonly employed.

7. The specific heat. In the event that the fat or oil is to be used for purposes of lubrication, it is sometimes desirable to obtain the value for the specific heat.[98]

8. The ultraviolet absorption characteristics of the material. When the polyunsaturated fatty acids are subjected to prolonged heating with

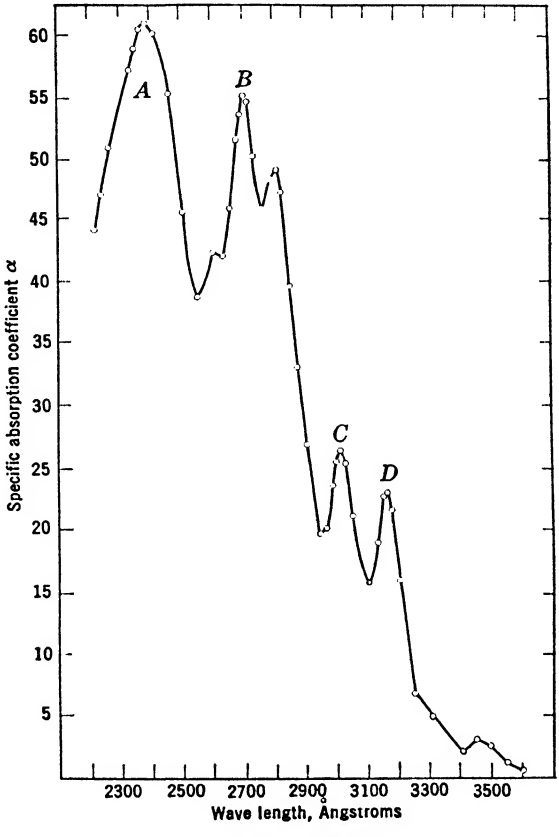


FIG. 113. Ultraviolet absorption spectrum of alkali-isomerized arachadonic acid. (After Beadle,[101] on methyl arachidonate sample prepared by J. B. Brown, Ohio State University.) The absorption maximum *A* at about 2,340 Å. characterizes two double bonds in conjugation. Three conjugated double bonds produce a peak *B* at about 2,680 Å., and four double bonds in conjugation produce absorption maxima *C* and *D* at about 3,010 and 3,160 Å.

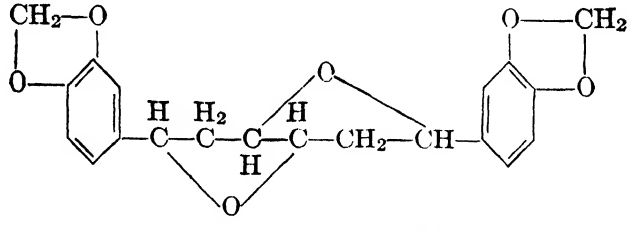
strong alkali [99] some of the double bonds rearrange to give conjugated systems. Since the latter give strong absorption bands at approximately 235 mμ (two double-bond conjugation), 270 mμ (three double-bond con-

[98] D. Wesson and H. P. Gaylord, *Cotton Oil Press*, **2**, 40 (1918).
[99] T. Moore, *Biochem. J.*, **31**, 138 (1937); **33**, 1635 (1939).

jugation), and 300 and 315 mμ (four double-bond conjugation), they can readily be detected when present in fats. Figure 113 shows the UV absorption spectrum of alkali-isomerized arachidonic acid.

Burr and Miller [100] and Beadle [101] have reviewed the applications of ultraviolet absorption spectrophotometry to the analysis and study of polyene fatty acid systems in edible fats and oils. Burr and co-workers have used this technique as a tracer in studying the biochemistry of fats *in vivo*.

Color Reactions. Numerous color reactions have been described by which an oil from one plant or animal source may be distinguished from similar oils from other plant or animal sources. For example, sesame oil gives a very characteristic color in the presence of a trace of furfural and concentrated hydrochloric acid. The color has been shown [102] to be due to the presence of a compound, sesamoline, $C_{20}H_{18}O_6$, which on hydrolysis yields sesamol, the 3,4-methylene ether of 1,3,4-tri-hydroxybenzene. Sesamoline has the structure



where the sesamol appears to be united in the molecule in an unusual glycosidal-like linkage.

Color reactions can be regarded only as *circumstantial evidence*, inasmuch as they depend on impurities carried along into the fat or oil from its natural source. The glycerides of one plant or animal species are, in general, not different from those of other species of plants or animals. Accordingly there is no color test which will distinguish a *pure* fat or oil from other fats and oils.

Qualitative Tests for Particular Classes of Fats. 1. As already noted, oleic acid is transformed into its isomer, elaidic acid, by treatment with nitric oxide. Accordingly the *elaidin test* is a qualitative test for the presence of oleins. An oil containing a large percentage of olein becomes a solid fat on treatment with nitric oxide.

[100] G. O. Burr and E. S. Miller, *Chem. Revs.*, **29**, 419 (1941).

[101] B. W. Beadle, *Oil & Soap*, **23**, 140 (1946); *cf.* also B. W. Beadle and H. R. Kraybill, *J. Am. Chem. Soc.*, **66**, 1232 (1944).

[102] W. Adriani, *Z. Untersuch. Lebensm.*, **56**, 187 (1928); *cf.* also J. Böeseken and W. D. Cohen, *Biochem. Z.*, **201**, 454 (1928).

2. The linoleic and linolenic acids of the drying oils, when treated with bromine, form characteristic tetrabromides and hexabromides. This reaction has given rise to the so-called *hexabromide test* for drying oils.

Chemical Methods. A number of chemical methods have been proposed for the characterization of fats and oils.

1. *The Acid Number.* The acid number may be defined as the milligrams of potassium hydroxide necessary to neutralize the free fatty acids present in 1 gram of a fat, oil, or wax.

2. *The Iodine Absorption Number.* This may be defined as the number of centigrams of iodine taken up by 1 gram of fat. In other words, it may be expressed as the percentage of iodine taken up by 1 gram of fat.

3. *The Saponification Number or the Koettstorfer Number.* This is the milligrams of potassium hydroxide necessary to saponify 1 gram of a fat or oil.

4. *The Total Fatty Acids.* Five grams of the fat or oil are saponified, and the unsaponifiable matter is extracted from the soaps. The soaps are then decomposed by shaking with 1:5 hydrochloric acid, liberating the free fatty acids. These are dissolved in anhydrous ether. The ethereal solution is dried, and an aliquot of the solution is evaporated and the residue weighed.

5. *The Reichert-Meissl Number.* This is the number of milliliters of 0.1 *N* potassium hydroxide required to neutralize the volatile acids from 5 grams of a saponified fat or oil.

6. *The Hehner Number.* This is the weight of the non-volatile (insoluble) fatty acids yielded by 5 grams of a saponified fat or oil.

7. *The Polenske Number.* This is the number of milliliters of 0.1 *N* potassium hydroxide required to neutralize the non-volatile fatty acids obtained from 5 grams of a saponified fat or oil. The residue remaining in the flask from the Reichert-Meissl determination is usually used to determine the Polenske number.

8. *The Acetyl Number.* This is the milligrams of potassium hydroxide necessary to combine with the acetic acid liberated by the saponification of 1 gram of acetylated fat or oil.

9. *Thiocyanogen Number.* Kaufmann [103] observed that whereas bromine and iodine add to *both* of the double bonds of linoleic acid, as well as to the single bond of oleic acid, *thiocyanogen*, $(SCN)_2$, *although it adds to the double bond of oleic acid, adds to only one of the double bonds of linoleic acid.* It therefore is possible to calculate the amount of linoleic acid in a mixture by determining the ratio between the iodine number

[103] H. P. Kaufmann, *Z. Untersuch. Lebensm.*, **51**, 15 (1926); *cf.* also L. Zeleny and C. H. Bailey, *Ind. Eng. Chem.*, **24**, 109 (1932).

or the bromine number and the thiocyanogen number. The thiocyanogen reagent is prepared by treating lead thiocyanate in glacial acetic acid with the calculated amount of bromine. Lead bromide is formed, and the thiocyanogen remains dissolved in the glacial acetic acid. The solution used should be about 0.05 *M*. The paper by Zeleny and Bailey may be consulted for details of the analysis and for methods of calculation. The use of the thiocyanogen number has demonstrated that the more reactive double bond of linoleic acid is preferentially hydrogenated before there is any appreciable hydrogenation of the double bond of oleic acid.

Separation, Identification, and Estimation of the Individual Constituents of an Oil, Fat, or Wax. No definite scheme for investigating fats and oils in general can be set up, since both the objectives of the research and the nature of the material under investigation influence the final choice of techniques.

Many of the physical and chemical tests listed above will provide both qualitative and quantitative information regarding the individual constituents of the fat or isolated fractions of the fat. Accordingly, a common procedure is to saponify the fat or oil and extract the unsaponifiable material from the soaps. The fatty acids may then be liberated, and the volatile fatty acids separated and studied.

Non-volatile fatty acids may be separated by any one of several methods, such as (*a*) the lead salt-ether method, (*b*) fractional precipitation of the barium or magnesium salts, (*c*) fractional distillation of the esters of the acids *in vacuo*, (*d*) precipitation from acetone at low temperatures, or (*e*) chromatographic adsorption—alumina, activated carbon, magnesium oxide, and silica gel having been used (alumina has thus far offered the most promise as an adsorbent).

The lead salts of the *saturated* fatty acids are insoluble in ether, whereas the lead salts of the *unsaturated* fatty acids are soluble in ether.[104] Accordingly the lead salt-ether method is, in general, used to separate the higher saturated fatty acids from the unsaturated fatty acids, although use of the lead salt-ether method in conjunction with one of the other technics may be necessary.[105] The hydroxamic acids [106] formed when hydroxylamine is added to a mixture of fatty acids ($R—COOH + NH_2OH = R—CO—NH—OH + H_2O$) may sometimes be used to advantage. The sodium salts of stearohydroxamic acid and palmito-

[104] Elaidic acid, although unsaturated, is solid at room temperature, and its lead salt also resembles the saturated fatty acids in being insoluble in ether.

[105] E. B. Holland, J. C. Reed, and J. P. Buckley, *J. Agr. Research*, **6**, 101 (1916); E. B. Holland and J. P. Buckley, *ibid.*, **12**, 719 (1918).

[106] A. H. Lewis, *Biochem. J.*, **20**, 1356 (1926).

hydroxamic acid are insoluble in alcohol, whereas the corresponding sodium salt of oleohydroxamic acid is freely soluble. The differences are so marked that Lewis states a quantitative separation to be possible. The sodium salts of the saturated fatty acids are insoluble when the chain contains twelve or more carbon atoms, and soluble for acids containing eight or less carbon atoms. Lauric acid (C_{12}) forms a somewhat soluble sodium hydroxamic salt.

The liquid, unsaturated fatty acids may be separated by several means.

(*a*) *As Bromo Derivatives.* The unsaturated fatty acids are brominated in glacial acetic acid. Organic solvents are then used to separate the bromine compounds into groups showing different solubilities. The oleic group (one double bond) of fatty acids yields dibrom derivatives which are soluble in petroleum ether. The linoleic group (two double bonds) yields tetrabrom derivatives, insoluble in petroleum ether but soluble in ethyl ether. The linolenic group (three double bonds) yields hexabrom derivatives, insoluble in petroleum ether, insoluble in ethyl ether, but soluble in hot benzene.

(*b*) *As Oxidation Products.* The separation of the liquid, unsaturated fatty acids through their oxidation products is carried out by oxidizing the fatty acids to their corresponding hydroxy acids. The linolenic series yields hexahydroxystearic acids which are soluble in cold water. The linoleic series yields tetrahydroxystearic acids, soluble only in large quantities of boiling water. The oleic series yields dihydroxystearic acids, insoluble in hot water, but soluble in ethyl ether. The influence of polar hydroxyl groups is very marked in the above solubility series.

(*c*) *By Fractional Crystallization.* This method [107] avoids the danger of formation of conjugated bonds which may arise under the influence of heat or by bromide formation.

The solid, saturated fatty acids are separated by the fractional distillation of their methyl and ethyl esters. In many instances, the esters have to be refractionated a number of times, provided any considerable number of the higher saturated fatty acids is present. The various fractions are then converted into their silver salts. These are crystallized and analyzed for silver as a test for the purity of the various fractions.

The non-saponifiable residue may be further studied for the presence of the higher alcohols, including the sterols. The methods are largely those of fractional crystallization or fractional distillation, combined with the formation of chemical derivatives which may serve for identification purposes.

[107] J. B. Brown, *Chem. Revs.*, **29**, 333 (1941).

(*d*) *By Molecular Distillation.* The use of molecular (short-path) distillation as a means of separating natural components of fats and oils has been reviewed by Embree.[108] This method appears to be of value in efficiently removing free fatty acids and unsaponifiable matter where the material under investigation may be altered by saponification, such as for the sterol esters and vitamin K. It is not well suited for separation of triglycerides. The rapid development of this technique is attested by the voluminous literature on the subject which has appeared in recent years. The reader is referred to bibliographies which are available [109] for publications on molecular distillation equipment and its application in a variety of fields.

[108] N. D. Embree, *Chem. Revs.*, **29**, 317 (1941).

[109] S. B. Detwiler, Jr., and K. S. Markley, *Oil & Soap*, **16**, 2 (1939); S. B. Detwiler, Jr., *ibid.*, **17**, 241 (1940); S. S. Todd, *ibid.*, **20**, 205 (1943).

CHAPTER 31

The Compound Lipids

The compound lipids are substances of a fat-like nature, yielding on hydrolysis fatty acids or derivatives of fatty acids and containing in the molecule either nitrogen or phosphorus, or both phosphorus and nitrogen.

Our knowledge of the organic phosphorus compounds occurring in both plants and animals is still very limited and incomplete. At least six groups of organic phosphorus compounds occur in nature.

1. *Phytin* or *phytic acid*, in which phosphoric acid is combined with inositol as inositol phosphoric acid, phytin being the calcium-magnesium salt of inositol phosphoric acid. Inositol is hexahydroxy-hexahydrobenzene, and it occurs very widely distributed throughout the plant kingdom, notably in the pericarp of seeds. The phosphoric acid combined with the inositol probably functions as a reserve phosphorus supply.

The inositol itself has become of particular interest owing to recent observations of its biochemical functioning. Eastcott [1] found that inositol is bios I, a substance necessary for the normal reproduction of the yeast cell.[2] Since then, inositol has been shown to be a dietary essential for the mouse [3] and a constituent of a group of compound lipids, and it has also been identified as one of the more effective materials for counteracting fatty livers.

2. *The hexose phosphates*, according to Harden [4] and Harden and Young,[5] are intermediary products in the alcoholic fermentation by yeast.

3. *Nucleic acids* have already been discussed. In them phosphoric acid is combined with carbohydrates and purine and pyrimidine bases.

4. *The phosphoproteins* have likewise already been discussed. The best-known examples are casein and vitellin. The phosphorus is present

[1] E. V. Eastcott, *J. Phys. Chem.*, **32**, 1094 (1928).

[2] E. Wildiers, *La Cellule*, **18**, 313 (1901); *cf.* also W. L. Miller, *J. Chem. Education*, **7**, 257 (1930).

[3] D. W. Woolley, *Science*, **92**, 384 (1940).

[4] A. Harden, *Alcoholic Fermentation*, 4th ed., Longmans, Green and Company, London, 1932.

[5] A. Harden and W. J. Young, *Proc. Roy. Soc. London*, **B77**, 405 (1906).

in the protein molecule as phosphoric acid, esterified on the hydroxyl group of the hydroxyamino acids. The phosphoproteins, so far as we are aware, occur only in the animal kingdom.

5. *Creatine-phosphoric acid* (phosphocreatine) and the *adenosine phosphates* were discussed when we considered the mechanism of muscle contraction. *Argeninephosphoric acid,* which replaces phosphocreatine in crustacean muscle, likewise has been noted.

6. There remain *the phospholipids* (the phosphatides of Thudichum). In contrast to the groups which we have noted, the phospholipids are soluble in fat solvents. The best-known example of this group is probably "lecithin," although the "lecithins" which have been most widely studied probably do not represent a pure compound. In lecithins and cephalins, the phosphoric acid occurs esterified on one of the hydroxyl groups of glycerol. In inositol-containing phospholipids, the phosphoric acid may be esterified on the inositol. In sphingomyelins, the phosphoric acid is esterified on the hydroxyl group of a nitrogenous base, sphingosine.

The Phospholipids. Few fields of biological chemistry offer more difficulties than the study of the separation and identification of the chemical structure of the compound lipids which comprise such a large fraction of the nervous tissue. The work in the field of the compound lipids dates from the publication by Thudichum, in 1874, of a paper entitled, "Researches on the Chemical Constitution of the Brain." [6] Thudichum continued to publish in this difficult field of chemical research for a number of years.

In recent years our knowledge of the chemistry of this group of lipids has seen great advancement. It is now evident that the phospholipids constitute a far more complex group of compounds than had been generally assumed.

Thannhauser and Schmidt [7] propose the following classes for grouping the phosphorus-containing lipids:

(1) *Monoamino phosphatides,* consisting of fatty acid esters of a phosphorylated polyvalent alcohol, combined with a nitrogen-containing group. Their ratio of P:N is 1:1.

(A) *Lecithins*—phosphoric acid esters of diglycerides and choline.
Lysolecithins—phosphoric acid diesters of saturated monoglycerides and choline.

[6] J. L. W. Thudichum, *Report of the Medical Officer of the Privy Council and Local Government Board, New Series,* **3,** 113 (1874).

[7] S. J. Thannhauser and G. Schmidt, *Physiol. Revs.,* **26,** 275 (1946).

(B) *Cephalins*, in which the nitrogen is in the form of a primary amino group (ethanolamine or serine).

(a) *Phosphatidyl ethanolamines*, phosphoric acid diesters of diglycerides and ethanolamine.

(b) *Phosphatidyl serines*, the hydrolysis products of which are fatty acids, phosphoric acid, polyvalent alcohols, and serine.

(2) *Plasmalogens* (*acetal-phosphatides*), consisting of phosphoric acid diesters of a higher fatty acid aldehyde acetal of glycerol and of ethanolamine.

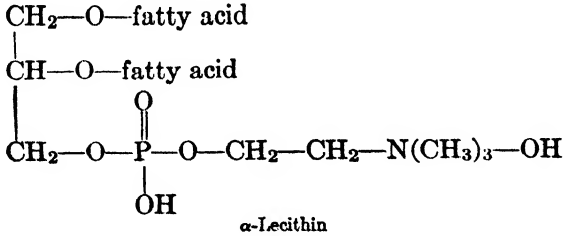
(3) *Inositol-phosphatides* from brain, soybeans, bacteria.

(4) *Phosphatidic acids*, *cardiolipins*, containing no nitrogenous base but otherwise similar to lecithin. The hydrolysis products are fatty acids, polyvalent alcohols, and phosphoric acid (bound as a monoester).

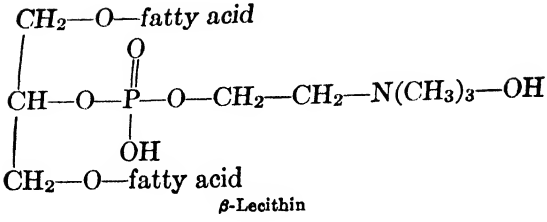
(5) *Phosphatides of acid-fast bacteria*, an indistinctly defined group, yielding on hydrolysis phosphoric acid, polyhydroxy compounds (such as carbohydrate, inositol), and fatty acids with both straight and branched chains.

(6) *Diamino phosphatides* (*Sphingomyelins*), consisting of acid amides of sphingosine with fatty acids in ester linkage with phosphoryl-choline.

The Lecithins. On hydrolysis, lecithins yield one molecule of glycerolphosphoric acid, two molecules of fatty acids, and one molecule of choline. The choline is attached to the glycerolphosphoric acid by esterification of the hydroxyl group of the oxyethyl radical with the phosphoric acid. The fatty acids are esterified on the two remaining hydroxyl groups of glycerol. It is evident that, if only one of the hydroxyl groups of glycerol is esterified with phosphoric acid, two isomeric glycerolphosphoric acids are possible. In one of these, phosphoric acid would be esterified in the α-position, and such a compound should be optically active. In the other, phosphoric acid would be esterified in the β-position and the glycerolphosphoric acid would not be optically active. Glycerolphosphoric acid, as it has been obtained from a number of lecithins, shows optical activity, but there is abundant evidence that both α- and β-lecithins occur in nature.

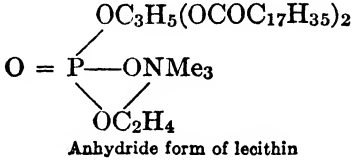


α-Lecithin

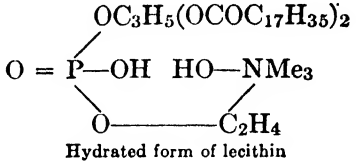


β-Lecithin

Lecithins can exist both in the hydrated form and in the anhydride form where water has been lost between the residual acid group of the phosphoric acid and the hydroxyl group of the substituted ammonium hydroxide.



Anhydride form of lecithin



Hydrated form of lecithin

In the analyses of various lecithins which have been reported by different investigators, we find records of the presence of stearic, palmitic, oleic, linoleic, and arachidonic acids. Since there are only two molecules of fatty acids in a lecithin, it is evident that a considerable number of isomeric "lecithins" are possible, particularly in view of the fact that both α- and β-forms offer additional opportunities for isomerism. It thus is apparent that the term lecithin does not mean a specific chemical compound but rather that it is the generic name of a group of compounds possessing similarity in structure but likewise possessing specific differences in the nature of the fatty acids esterified in the molecule. Thudichum stated that all true lecithins contain one molecule of an unsaturated fatty acid and one molecule of a saturated fatty acid. Levene's studies point in the same direction, and most workers agree that, in general, this appears to be so, although recently evidence has been brought forward for a distearyllecithin and dioleyllecithin. There seems to be no *a priori* reason why the fatty acids associated with the lecithin molecule should of necessity be different, and perhaps we are dealing here with the same general rule that by and large simple glycerides rarely occur in nature. The difficulty of definitely settling even this problem lies in the great ease with which lecithins hydrolyze and the still greater ease with which they oxidize. Lecithins, when pure, have a translucent, paraffin-like, colorless appearance, but they quickly turn yellow on exposure to air and even in a very few minutes may become an opaque brown. Therefore if a "pure" lecithin preparation is tested, the product must be prepared without exposure to air at any stage in the process.

Since lecithins are constituents of all cells they apparently have a very important function in life processes. It has been suggested that one of their functions is to assist in regulating the permeability of the cell membrane. The lecithins are excellent emulsifying agents and probably play a role in the maintenance of protoplasmic structure. Furthermore, they readily form coordination complexes with proteins, and in the form of lecithoproteins must play important physiological roles.

Grün and Limpächer [8] claim to have synthesized a lecithin by first treating distearin (or other diglycerides) with one molecule of phosphorus pentoxide and then treating the product with two molecules of choline bicarbonate or some similar salt of choline. The lecithin was isolated by neutralizing its benzene solution with alcoholic alkali, thus separating cholinephosphoric acid, and then precipitating the lecithin with acetone, separating it from residual distearin. The product precipitated with acetone analyzed for lecithin and had the characteristic physical and chemical properties of lecithin.

Lecithins are readily hydrolyzed by lipase. Contardi and Ercoli [9] point out that there are really four distinct lecithinases corresponding to the four types of ester linkages in lecithin.

Lecithinase-A liberates only one molecule of fatty acids. This enzyme occurs in cobra venom and rattlesnake venom.[10] When one fatty acid is split off from a lecithin, the residual product is known as lysolecithin. Lysolecithin is powerfully hemolytic, and there is evidence that the hemolytic action of cobra venom lies in the ability of the venom to form lysolecithin within the organism.

Lecithinase-B, which occurs in rice hulls, splits both fatty acid molecules from the lecithin molecule. Lecithinase-C can apparently hydrolyze off only the choline radical, and lecithinase-D is a true glycerolphosphatase and breaks the bond between the phosphoric acid and the glycerol molecule.

The Cephalins (*Kephalins*). These compounds, usually closely associated with the lecithins, are also insoluble in and precipitated by acetone. They have generally been distinguished from the lecithins on the basis of their insolubility in alcohol. This separation of the two groups is extremely difficult, for a mixture of the two does not sharply separate in absolute alcohol. Bull and Frampton [11] found that, in the presence of lecithins, the cephalins were fairly soluble in alcohol. Fur-

[8] A. Grün and R. Limpächer, *Ber.*, **59**, 1350 (1926); **60**, 147 (1927).

[9] A. Contardi and A. Ercoli, *Biochem. Z.*, **261**, 275 (1933).

[10] E. J. King and M. Dolan, *Biochem. J.*, **27**, 403 (1933); *cf.* also E. J. King, *ibid.*, **25**, 799 (1931); **28**, 476 (1934).

[11] H. B. Bull and V. L. Frampton, *J. Am. Chem. Soc.*, **58**, 594 (1936).

thermore, although a natural mixture of "cephalins" has a low solubility in ethanol, some of the purified lipids from this group are actually quite soluble in this solvent.[12]

For many years cephalins were considered to be identical with the lecithins except for the presence in the molecule of aminoethyl alcohol instead of choline. In 1941, Folch and Schneider [13] reported that a considerable percentage of the nitrogen of brain cephalin was in the form of a hydroxyamino acid, presumably serine. This finding has been confirmed,[14] and data are now available [15] which indicate that the serine-containing cephalin is widely distributed in animal tissues. Brain and lung phospholipids appear to be particularly rich in this lipid, whereas the lipids of egg yolk are entirely devoid of amino acid.

The crude cephalin obtained from the lipids of many tissues, both plant and animal in origin, contains still another group of phospholipids of which inositol is a constituent. These findings of the extreme unhomogeneity of "cephalin" account for the common observation that the yield of cephalin fatty acids is invariably far below the theoretical amount based on the formula for phosphatidyl ethanolamine.

The various cephalins are all strongly acidic, and they may be isolated as metallic salts.

Grün and Limpächer [16] claim to have synthesized an ethanolamine-containing cephalin, using a procedure analogous to that which they used with lecithin. The compound which they synthesized, however, contained two molecules of stearic acid, whereas the products isolated by Levene [17] contained one molecule of stearic acid with the other molecule either oleic or arachidonic acid. Magistris [18] prepared lysocephalin and found that it possessed very little hemolytic power. He suggested that in all probability lysocephalin does not possess hemolytic power and that the hemolysis noted in his preparations is due to the presence of small amounts of lysolecithin.

The Sphingomyelins. These phospholipids occur in large quantities in the brain and nerve tissue, and in smaller quantities can be isolated from most of the organs of the body. They yield as decomposition

[12] J. Folch, *J. Biol. Chem.*, **146**, 35 (1942).

[13] J. Folch and H. Schneider, *J. Biol. Chem.*, **137**, 51 (1941).

[14] J. Folch, *J. Biol. Chem.*, **139**, 973 (1941); E. Chargaff and M. Ziff, *ibid.*, **140**, 927 (1941).

[15] E. Chargaff, M. Ziff, and D. Rittenberg, *J. Biol. Chem.*, **138**, 439 (1941); C. Artom, *ibid.*, **157**, 595 (1945).

[16] A. Grün and R. Limpächer, *Ber.*, **60**, 151 (1927).

[17] P. A. Levene and I. P. Rolf, *J. Biol. Chem.*, **54**, 91 (1922).

[18] H. Magistris, *Biochem. Z.*, **210**, 85 (1929).

products [19] phosphoric acid, lignoceric acid, cerebronic acid, stearic acid, and two nitrogen bases, choline and sphingosine. Sphingosine appears to be an unsaturated diatomic amino alcohol having the structure [20] $C_{13}H_{27}$—CH=CH—CHOH—$CHNH_2$—CH_2OH.

Some of the problems involved in the constitution of the sphingomyelin molecule have not been solved. Levene proposes that the fatty acid is conjugated with the amino group and that the phosphoric acid is esterified on a hydroxy group and that the choline in turn is esterified on the phosphoric acid. Since sphingomyelins contain only one fatty acid radical, and, since at least three fatty acids have been isolated from sphingomyelins, it is evident that there are at least three sphingomyelins: a stearylsphingomyelin, a lignocerylsphingomyelin, and a nervonylsphingomyelin. It is interesting to note that Thudichum suggested that there were probably several sphingomyelins.

Nervonic acid, $C_{24}H_{46}O_2$, is an unsaturated fatty acid having one double bond between the fifteenth and sixteenth carbon atoms. The formula thus indicates that it is the singly unsaturated acid corresponding to the saturated C_{24} lignoceric acid. There is an accumulation of sphingomyelins in the brain, liver, and spleen of infants suffering from Niemann-Pick's disease. This disease of infancy is marked by an enlarged spleen and liver, the spleen showing lipid deposits, by anemia, and by a leukocytosis with a marked increase in lymphocytes. It is usually associated with idiocy and blindness where there is no demonstrable lesion in the structure of the eye. The cerebrosides appear to be absent in this disease and to be replaced by sphingomyelins.

Sphingomyelins are white, crystalline, non-hygroscopic substances relatively stable to light and air. They become hydrated and emulsified in water. They have been distinguished from the lecithins and cephalins by their insolubility in ether and also by the absence of glycerol among the hydrolytic products. The former criterion is not reliable, since other compound lipids, such as lecithin containing two saturated fatty acids,[7] are ether-insoluble.

Inositol-Containing Lipids. In 1930, Anderson [21] found inositol in the phospholipid fraction of human tubercle bacilli. Since then, its presence has been reported in the phospholipids of soybeans [22] and of

[19] P. A. Levene, *J. Biol. Chem.*, **15**, 153 (1913); **18**, 453 (1914); **24**, 69 (1916); *cf.* also P. A. Levene and F. A. Taylor, *ibid.*, **52**, 227 (1922).

[20] H. E. Carter, F. J. Glick, W. P. Norris, and G. E. Phillips, *J. Biol. Chem.*, **142**, 449 (1942).

[21] R. J. Anderson, *J. Am. Chem. Soc.*, **52**, 1607 (1930).

[22] E. Klenk and R. Sakai, *Z. physiol. Chem.*, **258**, 33 (1939); D. W. Woolley, *J. Biol. Chem.*, **147**, 581 (1943).

nervous tissue.[23] Despite similar inositol contents (*ca.* 16 per cent) in the isolated lipids from these different sources, they appear to be quite dissimilar in other respects. For example, Folch [24] found both serine and glycerol in the inositol-containing lipid in brain "cephalin" and reported [25] that metadiphosphate is a constituent of this compound. The hydrolyzed lipid yielded inositol, phosphoric acid, and fatty acid in the ratios 1:2:1. Woolley [22] found his "soybean lipositol" to be free of serine and glycerol, and analysis showed a molecular ratio of 1:1:2 for inositol, phosphoric acid, and fatty acids. Galactose and tartaric acid are also constituents of the inositol lipid from soybeans.

Sulfur-Containing Phospholipids. Thudichum isolated a sulfur-containing phospholipid from brain tissue which he called cerebrosulfatide. This was further studied by Koch,[26] who observed that the compound contained equal molecular amounts of sulfuric acid and phosphoric acid. Fränkel and Karpfen [27] made a further study of this compound. They gave the empirical formula as $C_{101}H_{152}N_3PSO_{26}$, having a ratio of S:P:N of 1:1:3, all the nitrogen being in the form of amino groups. They identified, as hydrolytic products, phosphoric acid, sulfuric acid, glycerol, and aminoethyl alcohol. All the nitrogen was present in the form of the aminoethyl alcohol, choline and sphingosine being completely absent. The only fatty acid which appeared to be present was a hitherto unknown hydroxy acid, α-hydroxy-*n*-decanic acid, $C_{10}H_{20}O_3$.

Fränkel and Gilbert [28] had previously isolated from human brain substance another phospholipid containing sulfur, the compound having the formula $C_{93}H_{191}N_3SPO_{18}$. The hydrolytic products of this compound were phosphoric acid, sulfuric acid, glycerol, aminoethyl alcohol, and cerebronic acid. Just how these various decomposition products are linked in these lipids remains to be determined. These sulfur-containing compounds of Fränkel have not been generally accepted by other workers. MacLean refers to them as "alleged lipids," and it may be that they represent a coacervate of a true phospholipid with other sulfur-containing compounds.

Phosphatidic Acids. Phosphatidic acids, in which the base of lecithin or cephalin has been replaced by a metallic ion, have been reported

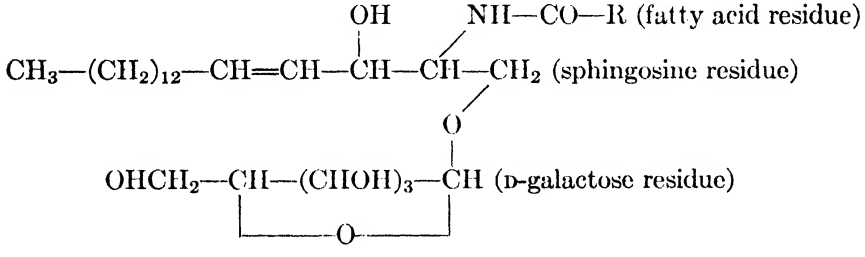
[23] J. Folch and D. W. Woolley, *J. Biol. Chem.*, **142**, 963 (1942).
[24] J. Folch, *J. Biol. Chem.*, **146**, 35 (1942).
[25] J. Folch, *Federation Proc.*, **5**, 134 (1946).
[26] W. Koch, *Z. physiol. Chem.*, **70**, 94 (1910).
[27] S. Fränkel and O. Karpfen, *Biochem. Z.*, **157**, 414 (1925).
[28] S. Fränkel and O. Gilbert, *Biochem. Z.*, **124**, 206 (1921).

as occurring in plants [29] and in heart tissue (cardiolipins).[30] Apparently their physical properties are quite similar to the common phospholipids. Palmitic, stearic, linoleic, and linolenic acids have been found in these compounds.

The Glycolipids. The glycolipids are variously referred to in the literature as cerebrosides (Thudichum), galactolipins (Leathes and MacLean), and galactosides (Rosenheim). They are characterized by containing one molecule of sphingosine, one molecule of D-galactose, and one molecule of a C_{24} fatty acid. The fatty acid is attached to the amino group of sphingosine through an amide ("peptide") linkage, since there is no free amino group present in the cerebroside and the D-galactose can be split off, leaving the sphingosine-fatty acid compound.

If the fatty acid is split off, there remains a non-reducing compound, *psychosine*, consisting of one molecule of sphingosine and one molecule of D-galactose with the amino group free. The D-galactose is combined with one of the hydroxyl groups of sphingosine through a glycosidal linkage. The D-galactose is the normal pyranose form.[31]

Three cerebrosides are known with certainty, and Klenk has produced evidence for a fourth. All have the same fundamental structure, the only difference being in the nature of the fatty acid residue.



In this formula of the cerebrosides it is not known with certainty which hydroxyl group is involved in the linkage between the sphingosine and D-galactose.

The three cerebrosides which are generally recognized are *phrenosin* (Thudichum, 1879) or *cerebron* (Thierfelder, 1905), *kerasin* (Thudichum, 1874; Rosenheim,[32] 1916), and *nervon* (Klenk, 1925). The fatty acid of phrenosin is cerebronic acid (phrenosinic acid), which according to

[29] A. C. Chibnall and H. J. Channon, *Biochem. J.*, **23**, 176 (1929).
[30] M. C. Pangborn, *J. Biol. Chem.*, **143**, 247 (1942); **161**, 71 (1945).
[31] J. Pryde and R. W. Humphreys, *Biochem. J.*, **20**, 825 (1926).
[32] O. Rosenheim, *Biochem. J.*, **10**, 142 (1916).

Klenk [33] is the α-hydroxy-n-C_{24} acid. The acid of kerasin is lignoceric acid, the n-C_{24} acid. The acid of nervon is nervonic acid,[34] a singly unsaturated n-C_{24} acid with the double bond between carbons 15 and 16. In his studies of the cerebrosides Klenk isolated a fourth acid which he calls oxynervonic acid.[35]

As might be anticipated, the various cerebrosides show very similar chemical properties. When dry, they are white and more or less waxlike and separate from alcohol in microscopic particles which give the superficial appearance of crystal balls. They apparently, however, are not truly crystalline, although claims have been made for obtaining phrenosin in true crystal form. Under proper experimental conditions they readily form "liquid crystals"; this property probably accounts for the great divergence of data in the literature regarding their "melting point." The cerebrosides are not limited to the brain or nervous tissue but have been reported to occur in the spleen, kidneys, liver, egg yolk, blood corpuscles, lungs, adrenals, the retina of the eye, fish sperm, and even in plant sources such as fungi, seeds, and the heart wood of oak. In Gaucher's disease the amount of cerebrosides in the liver and spleen is greatly increased. Aside from the isolation of cerebrosides from the spleen and liver in Gaucher's disease the general occurrence of these compounds outside of nerve tissue, especially from plant sources, should be looked upon with suspicion.

No cerebroside approaching in purity preparations made from brain and nerve tissue has as yet been isolated from plant sources, and it is possible that some of the old literature dealt with impure preparations of phospholipids contaminated with carbohydrate materials. Page [36] has brought together our knowledge of the brain lipids, and his book together with that of Thierfelder and Klenk can be consulted for further data.

Plasmalogens. Thannhauser recognizes the plasmalogens, or acetalphosphatides, as a distinct group of lipids in which a fatty acid aldehyde in acetal linkage replaces the two fatty acid groups of ethanolaminecephalin. The formulae suggested [37] for two possible forms of plasmalogens are:

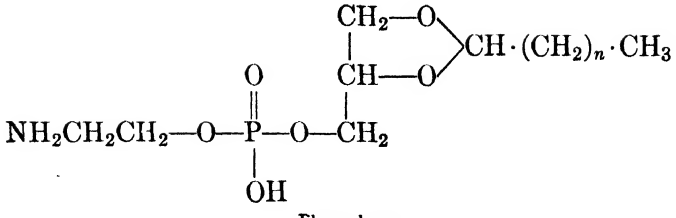
[33] E. Klenk, *Z. physiol. Chem.*, **157**, 291 (1926); **157**, 283 (1926); **166**, 268 (1927); **166**, 287 (1927); **174**, 214 (1928); **179**, 312 (1928).

[34] E. Klenk, *Z. physiol. Chem.*, **145**, 244 (1925).

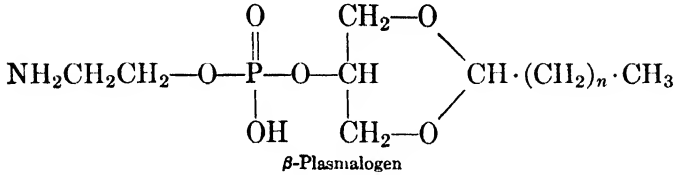
[35] E. Klenk, *Z. physiol. Chem.*, **174**, 214 (1928).

[36] I. H. Page, *Chemistry of the Brain*, C. C. Thomas Company, Springfield, Illinois, 1937.

[37] R. Feulgen and T. Bersin, *Z. physiol. Chem.*, **260**, 217 (1939).



α-Plasmalogen



β-Plasmalogen

The plasmalogens are thus the acetals of the cholamine ester of either α- or β-glycerophosphate.

CHAPTER 32

Metabolism of the Lipids [1]

Digestion. Ingested fat has a pronounced effect on the emptying time of the stomach. Although the length of time the food remains in the stomach is prolonged, and although the existence of a weak gastric lipase has been shown, very little splitting of fat occurs at this stage.

When the fat enters the small intestine, emulsification is effected, aided by the action of soaps formed in small amounts in the more alkaline environment of the duodenum, but principally by the dissolving action of the bile salts secreted into the intestine. The fats are thus brought into intimate contact with the powerful lipase of the pancreatic juice, which appears to be the major fat-digesting enzyme. The fats are hydrolyzed to the component fatty acids and glycerol, in which form they can be absorbed through the intestinal mucosa.

Aside from the individual and species characteristics of the animal, a number of factors play a role in determining the overall digestibility of fat. The melting point of the fat long has been considered of major importance. Many observations confirm the fact that as the melting point is raised to above 50°, the digestibility of the resulting fat is sharply lowered. For example, Evans and Lepkovsky [2] found that caprylin, caprin, and even laurin, melting at 46° [3] had high digestibilities of the order of 96–97 per cent. Myristin, melting at 57°,[3] had a slightly lower digestibility, whereas palmitin, with a melting point of 65°,[3] was only 73 per cent digestible, and only 40 per cent of stearin (m.p. 71.5° [3]) was absorbed.

There is strong evidence that the digestibility of a fat bears a closer relationship to its fatty acid make-up than to its melting point alone. The longer chain, saturated fatty acids in particular depress the nutritional availability of a fat. Hoagland and Snider [4] have reported

[1] This chapter contributed by Dr. Willis A. Gortner, Head, Department of Chemistry, Pineapple Research Institute, Honolulu; formerly Associate Professor of Biochemistry, Cornell University.

[2] H. M. Evans and S. Lepkovsky, *J. Biol. Chem.*, **96**, 179 (1932).

[3] T. P. Hilditch, *The Chemical Constitution of Natural Fats*, John Wiley & Sons, New York, 1940, p. 353.

[4] R. Hoagland and G. G. Snider, *J. Nutrition*, **26**, 219 (1943).

that as little as 5 to 10 per cent of stearic acid or stearin added to olive oil will markedly lower its digestibility. Palmitic acid or its glyceride will also cause a lowered digestibility, but myristic and lauric acids are practically completely utilized and at levels of 15 per cent do not affect the high availability of triolein.

That non-lipid constituents in the diet also may affect fat digestibility is emphasized by Barnes and his co-workers,[5] who reported that a greater digestibility of dietary fat is evident on a high-protein than on a low-protein intake.

With adequate intakes of the fat-soluble vitamins and of linoleic acid, the nutritional value of different fats appears to be solely a matter of digestibility. When margarine is fortified with vitamins A and D, it is the equivalent of butter in maintaining the nutritional status of the consumer.[6]

Absorption. The bile salts have the power to "dissolve" the free fatty acids or the fatty acid soaps formed from the action of lipase and bring them into a form which is freely diffusible into the intestinal mucosa. They are thus able to pass the epithelial walls and be absorbed into the epithelial cells of the mucous membrane. Here they are apparently resynthesized into fats. The mechanism of this resynthesis is still unknown.

The most widely held theory of the mechanism of fat absorption is that of Sinclair,[7] who proposed a key role for phospholipids in the intestinal mucosa. According to this concept, lipids are split in the gastrointestinal tract, and the emulsified fatty acids enter the villi lining the intestine. There the fatty acids become incorporated into phospholipids as an intermediate step to resynthesis of glycerides. His experiments suggest that the fatty acids first combine with the phospholipids, and the neutral fats are split off, and the phospholipid is regenerated. The mucosal phospholipid molecules thus serve as a bucket brigade passing the absorbed fatty acids from the intestine to the fat in the blood or lymph circulation. The make-up but not the quantity of the phospholipid was found to be affected by the absorption of fat.

Recent work indicates that other mechanisms may be involved in fat absorption. Absorption and transport across the intestinal mucosa via phospholipid formation are disputed to some extent by data of Barnes, *et al.*,[8] who used as a tracer fat the methyl esters of alkali-conjugated

[5] R. H. Barnes, M. F. Primrose, and G. O. Burr, *J. Nutrition*, **27**, 179 (1944).

[6] L. A. Maynard, H. E. Longenecker, G. O. Burr, C. A. Elvehjem, F. F. Elliott, and C. M. McCay, *Natl. Research Council U. S. Reprint and Circ. Ser.*, 118 (August, 1943).

[7] R. G. Sinclair, *J. Biol. Chem.*, **82**, 117 (1929).

[8] R. H. Barnes, E. S. Miller, and G. O. Burr, *J. Biol. Chem.*, **140**, 233 (1941).

corn-oil fatty acids. No parallelism was found between the transport of fatty acids and the rate of entrance of the acids into the intestinal phospholipids. Frazer [9] has presented evidence that lipolysis is only partially complete and that *unhydrolyzed* fat may be absorbed from the intestine, although it is acknowledged that this may not be the major means of handling dietary fat.

All fats are not hydrolyzed at the same rate, nor are all fatty acids absorbed at the same rate. Deuel [10] finds that the lower fatty acids (C_4, C_6, C_8) are absorbed most rapidly, that capric acid is absorbed at only half this rate, and that lauric acid is absorbed relatively slowly.

Fat absorption appears to be necessary for the optimal absorption of a number of fat-soluble compounds, among which are cholesterol, vitamins A, D, E, and K, and carotene.

The Transport of the Fatty Acids. In contrast to amino acids and the simple sugars, the bulk of the absorbed fat does not go directly through the intestinal mucosa into the blood stream but instead passes largely into the lymph circulation and through the thoracic duct (or some other lymphaticovenous communication [11]) into the venous circulation. It therefore does not pass first through the liver but rather short-circuits this organ and does not return to the liver except through the arterial circulation. In the blood stream the fats are present in the form of an extremely finely divided emulsion which has been called the chylomicron [12] emulsion, the chylomicrons ranging from 0.5 to 1.0 μ in diameter. The chylomicron emulsion increases enormously after the ingestion of fat and then gradually falls through a period of hours to the normal level.

The question as to how fat is removed from the blood and utilized by the various cells and tissues is even more obscure than are some of the problems relating to its transport into and across the intestinal mucosa. There is no evidence at the present time that lipase hydrolysis is involved during this process.

The possible role of cholesterol and of phospholipid in fatty acid transport is referred to later.

The Biological Synthesis of Fats. The excellent series of studies which Schoenheimer, Rittenberg, and their associates [13] have made with deuterium-containing fatty acids have shed considerable light on fat

[9] A. C. Frazer, *J. Physiol., London*, **102**, 306, 329 (1943).

[10] H. J. Deuel, Jr., L. F. Hallman, and A. Reifman, *J. Nutrition*, **21**, 373 (1941).

[11] J. M. Little and C. S. Robinson, *Am. J. Physiol.*, **134**, 773 (1941).

[12] S. H. Gage and P. A. Fish, *Am. J. Anat.*, **34**, 1 (1924).

[13] R. Schoenheimer, D. Rittenberg, and associates, *J. Biol. Chem.*, **111**, 175 (1935); **113**, 505 (1936); **114**, 381 (1936); **117**, 485 (1937); **120**, 155 (1937); **120**, 503 (1937); **121**, 235 (1937); **133**, 329 (1940).

anabolism by living organisms. It is now evident that the fat depots, in sharp contrast to brain or nerve tissue, are not static storage tissues but rather undergo a relatively rapid and continual turnover of glycerides. The greater part of the dietary fat is transported to and deposited in these fat tissues before it is utilized. Furthermore, there is a continual transformation of one fatty acid into another, even though the newly synthesized fatty acid is abundantly supplied in the diet.

Butyric and caproic acids are not stored in the fat depots, but rather are completely metabolized. Apparently they are not converted to higher fatty acids. On the other hand, the longer chain fatty acids are readily built up or degraded into fatty acids of different chain lengths, and *in vivo* desaturation of saturated fatty acids as well as saturation of the unsaturated acids is a normal phenomenon. Apparently desaturation only occurs to the extent of inserting one double bond in the 9:10 position. Schoenheimer's work indicates that fat anabolism is a dynamic equilibrium according to the scheme:

$$\text{Carbohydrates} \rightarrow \text{lower fatty acids} \leftrightharpoons \underset{\substack{\uparrow\downarrow \\ \text{palmitoleic acid}}}{\text{palmitic acid}} \rightleftharpoons \underset{\substack{\uparrow\downarrow \\ \text{oleic acid}}}{\text{stearic acid}}$$

Although the body can modify the absorbed fats, it is well known that the dietary fat may have a strong influence on the character of the storage fat. This is illustrated by work of Anderson and Mendel,[14] who fed soybean oil (iodine number 132), corn oil (I.N. 124), cottonseed oil (I.N. 108), peanut oil (I.N. 102), Crisco (I.N. 79), lard (I.N. 63), butter (I.N. 36), and coconut oil (I.N. 8) to rats and then studied the degree of unsaturation of the body fats. The latter closely paralleled the dietary fats, but fats having an unsaturation greater than oleic were partially saturated and those averaging less than one double bond per fatty acid molecule were desaturated to some extent, indicating a tendency and ability of the animal to modify the food fat.

The use of isotopes of carbon and hydrogen has also led to a better understanding of the mechanism whereby fats are synthesized by living organisms. That fat can readily be synthesized from products derived from the breakdown of carbohydrate is well known. This synthesis apparently takes place in the liver. There is some evidence that thiamine is necessary for this conversion,[15] although paired feeding experiments [16] indicate that the failure of fatty acid synthesis on a thiamine-deficient diet may be due to weight loss from a loss of appetite of the animal. Al-

[14] W. E. Anderson and L. B. Mendel, *J. Biol. Chem.*, **76**, 729 (1928).
[15] E. W. McHenry and G. Gavin, *J. Biol. Chem.*, **125**, 653 (1938); **128**, 45 (1939).
[16] G. E. Boxer and DeW. Stetten, Jr., *J. Biol. Chem.*, **153**, 607 (1944).

though proteins can give rise to fats in the animal body, it appears probable that this is brought about by, first, the formation of carbohydrate or carbohydrate metabolites from protein, and then the transformation into fat. Another interesting relationship of the vitamins is the observation [17] that pyridoxine apparently is necessary for the conversion of protein to the carbohydrate intermediates, after which thiamine is required for the body to complete the conversion to fat. This observation assumes added significance from the work of Gunsalus and his coworkers, who have demonstrated that a pyridoxine derivative functions as the coenzyme for amino acid metabolism (*cf.* p. 934).

The studies with deuterium demonstrated that there is a rapid and continuous conversion of carbohydrate to fat under normal dietary conditions. Furthermore, Rittenberg and Bloch,[18] using C^{13} and D as trace elements, have shown that *acetate* is a specific precursor for the synthesis of fatty acids. The most promising theory for synthesis of the long-chain fatty acids is that acetic acid is changed to acetaldehyde (a known breakdown product of carbohydrates) and that two molecules of acetaldehyde condense to form aldol. Successive aldol condensations would then build up the long carbon chain from two-carbon fragments. The role of acetate in fat anabolism also accounts for the observed ability of the body to change palmitic into stearic acid without first breaking down the fatty acid into small units.

The Origin of Milk Fat. The synthesis of milk fat has been studied by a great many investigators, and probably it is the easiest fat to study, since samples can be secured at frequent intervals without injury to the animal. Hilditch [19] has summarized our knowledge of the composition of butterfat with particular reference to the various sources from which the fatty acids characteristic of butterfat may be derived. One of the older theories was that the short-chain fatty acids of butterfat arose from the phospholipids of the blood. Hilditch rules out this theory and concludes that instead the milk fats arise from the glycerides of the blood. This conclusion is supported by evidence [20] that neutral fat is removed from the blood by the lactating mammary gland. Hilditch suggests that in the mammary gland there is the conversion of a part of the preformed oleoglycerides into glycerides containing the shorter chain fatty acids, such as butyric and caproic, and that at the same time some of the oleic acid is converted into stearic acid. He suggests that

[17] E. W. McHenry and G. Gavin, *J. Biol. Chem.*, **138**, 471 (1941).

[18] D. Rittenberg and K. Bloch, *J. Biol. Chem.*, **154**, 311 (1944).

[19] T. P. Hilditch, *Compt. rend.*, Vol. II, V^{e} Congrès International Technique et Chimique des Industries Agricoles, Scheveningue (1937), p. 367.

[20] L. Voris, G. Ellis, and L. A. Maynard, *J. Biol. Chem.*, **138**, 491 (1940).

this implies an enzymatic oxidation-reduction system operating from the alkyl end of the oleic acid group combined with the glycerol. He points out that such an enzymatic mechanism should be specific for C_{18}-fatty acids and should be blocked provided that the fatty acids available were largely of the C_{20}- or C_{22}-series.

Cod liver oil contains large proportions of the highly unsaturated acids of the C_{20}- and C_{22}-series. Accordingly Hilditch fed milch cows cod liver oil in one series of experiments and linseed oil in another series, and analyzed the milk and total milk fat, butter fatty acids, and the distribution of the individual fatty acids in the fat before, during, and after the feeding of the specific oils. Table 64 shows certain of the results obtained. It will be noted that the feeding of the cod liver oil containing the C_{20}- and C_{22}-series of unsaturated fatty acids did, in a very appreciable degree, block the synthesis of butyric, caproic, caprylic, and

TABLE 64. EFFECT ON THE COMPOSITION OF MILK FAT OF THE FEEDING OF COD LIVER OIL AND LINSEED OIL

(Data of Hilditch)

	Cod Liver Oil			Linseed Oil		
	Before	During	After 2 weeks	Before	During	After 2 weeks
Total milk-fat per day, lb.	2.22	1.30	2.04	2.80	2.46	2.74
Butter acids per day, lb.	2.10	1.24	1.95	2.66	2.33	2.60
Fatty Acids	Percentage Distribution of Fatty Acids in Butter Acids					
Butyric	4.28	1.61	4.62	4.51	4.29	4.23
Caproic	1.90	0.81	2.05	1.88	2.15	1.92
Caprylic	2.38	0.81	2.05	1.88	1.29	1.92
Capric	2.86	1.61	3.08	3.00	2.15	3.07
Lauric	3.81	4.03	3.59	3.76	3.00	3.85
Myristic	10.0	6.45	10.26	10.15	8.15	10.0
Palmitic	25.24	23.39	25.13	25.19	21.89	25.38
Stearic	12.38	8.06	12.31	12.41	9.87	12.31
Arachidic	0.95	0.81	0.51	0.75	0.43	0.77
Oleic	30.95	42.74	30.76	30.81	39.48	31.15
Octadecadienoic	4.28	4.83	4.61	4.51	6.00	4.23
C_{20}-C_{22}	0.95	4.83	1.03	1.13	1.29	1.15

capric acids. On the other hand, the C_{20}- and C_{22}-acids appeared in the butterfat in much larger quantities than normal. Hilditch notes that only 0.02 pound per day of butyric acid appeared in the milk fat during the cod liver oil feeding in contrast to the normal of approximately 0.09 pound.

That the short-chain fatty acids in butter are not characteristic of all milk fats is evident from a comparison of the fatty acid make-up of cow milk fat [21] and mature human milk fat [22] as shown in Table 65. Human

TABLE 65. FATTY ACIDS IN MILK FAT

Fatty Acids	*Mature Human Milk,* per cent*	*Cow Milk,† per cent*
Saturated		
Butyric	0.4	3.0
Caproic	0.1	1.4
Caprylic	0.3	1.5
Capric	2.2	2.7
Lauric	5.5	3.7
Myristic	8.5	12.1
Palmitic	23.2	25.3
Stearic	6.9	9.2
As arachidic	1.1	1.3
Total	48.2	60.2
Unsaturated		
Decenoic	0.1	0.3
Dodecenoic	0.1	0.4
Tetradecenoic	0.6	1.6
Hexadecenoic	3.0	4.0
Octadecenoic	36.5	29.6
Octadecadienoic	7.8	3.6
Octadecatrienoic	0.4	
Eicosatetraenoic	0.9	0.3
As eicosadienoic	2.4	
Total	51.8	39.8

* Data of Baldwin and Longenecker (1944).
† Data of Hilditch and Longenecker (1938).

milk fat contains only very small amounts of the volatile fatty acids and has a considerably higher percentage of the long-chain unsaturated fatty acids than does cow milk fat.

[21] T. P. Hilditch and H. E. Longenecker, *J. Biol. Chem.*, **122**, 497 (1938).
[22] A. R. Baldwin and H. E. Longenecker, *J. Biol. Chem.*, **154**, 255 (1944).

The Biological Synthesis of Sterols. It has been known for a long time that the vegetable sterols can be synthesized by both the lower and the higher plants. In the higher plants, such synthesis apparently can take place at any stage in the development of the plant. For many years it was assumed that the higher animals in some way transformed the plant sterols into cholesterol. However, in 1925, Channon [23] and Randles and Knudson [24] independently demonstrated that cholesterol could be synthesized in the animal body. Young rats were placed on a diet free from cholesterol and from any other sterols, at least so far as the absence of the Liebermann-Burchard reaction was concerned. These animals reproduced, and their young were continued on a sterol-free diet. The analysis of adult rats which had been born and raised on a sterol-free diet showed that relatively large quantities of cholesterol were present in the animal at birth and that larger amounts accumulated during the growth process.

Additional evidence that cholesterol is synthesized in the animal organism has been adduced by Schoenheimer,[25] who demonstrated (1) that cholesterol is readily absorbed from the intestinal tract; (2) that coprosterol and dihydrocholesterol are not absorbed from the intestinal tract; (3) that the plant sterols, sitosterol, stigmasterol, and even ergosterol, are not absorbed from the intestinal tract; (4) that the source of dihydrocholesterol in the feces is through an excretion from the body into the intestinal tract and that its origin is not due to bacterial action; (5) that coprosterol is not a normal constituent of the body sterols, nor is it a normal constituent of plant sterols, but is apparently formed in the intestine; (6) that the body synthesizes its cholesterol and probably its own sterol which acts as a precursor for vitamin D.

The fact that cholesterol is readily absorbed from the intestinal tract, whereas the other sterols are not absorbed, is a striking example of the specificity of biological functions and of the importance of chemical configuration in the regulation of such biological functions. The mechanism involved in this selective absorption still remains to be elucidated. Obviously it is not one of simple diffusion.

In 1937, Rittenberg and Schoenheimer [26] showed that cholesterol is synthesized *in vivo* by the coupling of a large number of small molecules. Later it was shown [27] that, as for fatty acid synthesis, acetate is a specific

[23] H. J. Channon, *Biochem. J.*, **19**, 424 (1925).

[24] F. S. Randles and A. Knudson, *J. Biol. Chem.*, **66**, 459 (1925).

[25] R. Schoenheimer, *Science*, **74**, 579 (1931).

[26] D. Rittenberg and R. Schoenheimer, *J. Biol. Chem.*, **121**, 235 (1937).

[27] K. Bloch and D. Rittenberg, *J. Biol. Chem.*, **143**, 297 (1942); **145**, 625 (1942); **155**, 243 (1944).

cholesterol precursor. The mechanism of this synthesis of cholesterol from such small units remains to be evolved.

The Columbia group have also demonstrated that cholesterol is the parent substance in the synthesis of other biological compounds having the cyclopentanoperhydrophenanthrene (steroid) nucleus, including the hormone metabolite, pregnanediol [28] (and hence, probably, progesterone), and the bile acid, cholic acid.[29]

Catabolism of the Fatty Acids. In 1905, Knoop [30] proposed the classical theory that fatty acid breakdown involves oxidation at the β carbon atom with the resultant loss of two carbon atoms from the fatty acid chain. By feeding phenyl-substituted fatty acids, he was able to demonstrate that fatty acids with an even number of carbon in the chain gave two-carbon residues (phenylacetic acid) whereas those with an odd number of carbons were oxidized to an odd-carbon product (benzoic acid). This theory also seemed to explain the fact that acetoacetic acid and other so-called "ketone bodies" are excreted when the body is faced with an abnormal demand for metabolism of fat.

A considerable accumulation of evidence has been built up to support this scheme of fat oxidation. However, certain apparently anomalous experimental observations have led to a number of modified theories. One which has received particular attention is that which holds that multiple, alternate oxidation occurs at every alternate carbon along the fatty acid chain, after which the chain breaks into four carbon fragments. Oxidation proceeding simultaneously from both ends of the carbon chain (ω-oxidation) has also been proposed.

The intermediary metabolism studies of Schoenheimer and Rittenberg gave results consistent with the β-oxidation theory of Knoop. In further support, Carter [31] has reviewed the data obtained from *in vivo* studies of phenyl-substituted and branched-chain fatty acids. This paper cites the strong case for β-oxidation, probably through desaturation, followed by oxidation to the β-keto acid without involving a β-hydroxy acid as an intermediate. Although ω-oxidation has been shown to occur to some extent in lower fatty acids,[32] Carter favors the theory of Flaschentrager [33] that ω-oxidation is an emergency mechanism not normally used in the metabolism of higher fatty acids.

[28] K. Bloch, *J. Biol. Chem.*, **157**, 661 (1945).
[29] K. Bloch, B. N. Berg, and D. Rittenberg, *J Biol. Chem.*, **149**, 511 (1943).
[30] F. Knoop, *Beitr. chem. Physiol. Path.*, **6**, 150 (1905).
[31] H. E. Carter, *Biol. Symposia*, **5**, 47 (1941).
[32] P. E. Verkade and J. van der Lee, *Biochem. J.*, **28**, 31 (1934).
[33] B. Flaschentrager, K. Bernhard, C. Lowenberg, and M. Schlapfer, *Z. physiol Chem.*, **225**, 157 (1934).

The Origin of Ketone Bodies. The primary ketone body is acetoacetic acid.[34] Apparently β-hydroxy-butyric acid is formed as a side reaction by reduction of the acetoacetic, while acetone arises from the decarboxylation of acetoacetic acid. The main site of ketone body formation has been shown to be the liver, although Medes, *et al.*,[35] showed that the kidney can also produce acetoacetate.

There is evidence [36] that the ketone bodies which arise from the incomplete combustion of fatty acids are formed by the condensation of two molecules of acetic acid which results from β-oxidation of the fatty acid. This scheme is consistent with the facts that some ketone bodies can arise from odd-carbon fatty acids,[34] that more than one molecule of acetoacetic acid may be produced from one molecule of fatty acid,[34] and that the longer chain fatty acids produce more ketone bodies than the acids with fewer carbons.[37]

MacKay and Barnes regard the production of ketone bodies as a physiological mechanism for supplying the extrahepatic tissues with an energy source when there is a deficiency of carbohydrate. Ketone bodies are produced from fatty acids in the liver and transported by the blood to tissues where they are utilized.

The recent evidence concerning the metabolic origin of acetoacetic acid has strengthened the β-oxidation theory of Knoop. The revised theory,[38] which seems to account adequately for known facts about fat catabolism, assumes that two carbon atoms are dropped from the fatty acid chain by successive β-oxidation. The acetic acid (or other two-carbon intermediary compound) thus formed then condenses to form acetoacetic acid.

Phosphorylation in Fatty Acid Oxidation. Recent work indicates that phosphorylation, an essential feature of *in vivo* oxidation of carbohydrates, is also a step in fatty acid catabolism. Munoz and Leloir [39] found that an enzyme in guinea pig liver will oxidize lower fatty acids in the presence of adenylic acid, cytochrome *c*, inorganic phosphate,

[34] M. Jowett and J. H. Quastel, *Biochem. J.*, **29**, 2159 (1935).

[35] G. Medes, S. Weinhouse, and N. F. Floyd, *J. Biol. Chem.*, **157**, 751 (1945).

[36] E. M. MacKay, R. H. Barnes, H. O. Carne, and A. N. Wick, *J. Biol. Chem.*, **135**, 157 (1940); E. M. MacKay, A. N. Wick, and C. P. Barnum, *ibid.*, **136**, 503 (1940); M. E. Swendseid, R. H. Barnes, A. Hemingway, and A. O. Nier, *ibid.*, **142**, 47 (1942); S. Weinhouse, G. Medes, and N. F. Floyd, *ibid.*, **155**, 143 (1944).

[37] J. S. Butts, C. H. Cutler, L. F. Hallman, and H. J. Deuel, Jr., *J. Biol. Chem.*, **109**, 597 (1935); H. J. Deuel, Jr., L. F. Hallman, J. S. Butts, and S. Murray, *ibid.*, **116**, 621 (1936).

[38] E. M. MacKay, *J. Clin. Endocrinol.*, **3**, 101 (1943).

[39] J. M. Munoz and L. F. Leloir, *J. Biol. Chem.*, **147**, 355 (1943); L. F. Leloir and J. M. Munoz, *ibid.*, **153**, 53 (1944).

magnesium ions, and an oxidizable substrate, such as α-ketoglutarate, glutamic acid, fumarate, succinate, malate, or citrate. Lehninger [40] found a similar enzyme to be present in rat liver. If adenosine triphosphate is present, coupled oxidation is unnecessary. A free carboxyl group on the fatty acid is necessary. The enzyme will oxidize straight-chain fatty acids from C_4 to C_{18}, as measured by oxygen uptake in the Warburg respirometer or by acetoacetate formation. Phosphorylation inhibitors, such as fluoride and iodoacetate, inhibit the reaction. These workers suggest that fatty acids are oxidized by going through a phosphorylated intermediate (acyl phosphate?) in the presence of adenosine triphosphate and the enzyme.

THE ROLE OF LIPIDS IN BODY PROCESSES

Carbohydrates and proteins metabolized by the body yield somewhat over 4 Calories per gram. Fats, on the other hand, have a fuel value of more than 9 Calories per gram. The very high caloric value of fat makes it valuable for the storage of energy. In addition to serving as an energy source for the carrying on of metabolic processes, the lipids appear to have a number of non-caloric functions in the body. The detailed accounting of the physiological role of the lipids is outside the scope of this text. The monograph by Bloor [41] provides an authoritative summary of the present status of knowledge of lipid metabolism. Burr and Barnes [42] have reviewed some of the non-caloric functions of dietary fats.

Fatty Acids Essential in Nutrition. Although the animal organism is capable of manufacturing, from carbohydrates, most of the fatty acids which it requires in its economy, some animals apparently are not capable of synthesizing fatty acids more unsaturated than oleic acid.

In 1929, Burr and Burr [43] reported that the rat cannot synthesize adequate amounts of the polyethenoid fatty acids, and that in the absence of linoleic or linolenic acid from the diet, definite pathologic phenomena occur. The skin becomes rough and scaly, metabolism is abnormal, sexual development is greatly delayed, water consumption is unusually high, reproduction either wholly fails or is abnormal, and lactation is deficient. Usually the animal ceases growth when it has reached about 60–75 per cent of the adult weight, and the life span is reduced to approximately one-third of normal. Very small quantities

[40] A. L. Lehninger, *J. Biol. Chem.*, **157**, 363 (1945); **161**, 437 (1945).

[41] W. R. Bloor, *Biochemistry of the Fatty Acids and Their Compounds, The Lipids*, Reinhold Publishing Corp., New York, 1943.

[42] G. O. Burr and R. H. Barnes, *Physiol. Revs.*, **23**, 256 (1943).

[43] G. O. Burr and M. M. Burr, *J. Biol. Chem.*, **82**, 345 (1929); **86**, 587 (1930).

of purified linoleic, linolenic, or arachidonic acid or the esters of these acids are sufficient to cause normal growth to be resumed, although linolenic acid is less effective than the other two fatty acids in causing the abnormal skin condition to disappear.

The mouse, the rat, the dog, and man have been shown to require a dietary source of essential fatty acid. Hansen and McQuarrie, in the Department of Pediatrics at the University of Minnesota, found that in many cases of infant eczema, the serum lipids are characterized by a lower iodine number than that of normal children. In such instances the low iodine numbers have been increased to more nearly the normal level by the feeding of vegetable oils high in linoleic or linolenic acid, *e.g.*, raw linseed oil, corn oil, or lard. In a number of instances, the inclusion of such oils in the diet has strikingly improved the eczematous condition.

That the essential fatty acids play a significant role in maintaining healthy skin seems evident. In this connection, an interesting interrelationship between the action of the fats and the vitamins is observed in the pyridoxine-sparing action of the essential fatty acids. It has been claimed that the acrodynia produced in vitamin B_6 deficiency is greatly delayed if the level of linoleic acid in the diet is raised.

Burr [44] has prepared an excellent review on the significance of the essential fatty acids, and the reader is referred to this paper for a detailed account of the status of work in this field.

TABLE 66. PER CENT OF GLYCERIDES IN COMMERCIAL EDIBLE FATS *

(Range and Average for 27–60 Samples)

	Lard	Butter	Margarine	100% Hydrogenated †	Blended Hydrogenated ‡
Linoleic	6.5–13.7 (11.7)	1.4– 4.8 (3.3)	1.3–23.4 (10.9)	2.9–22.4 (12.6)	4.7–38.2 (22.0)
Oleic	44.6–57.6 (51.5)	27.9–38.6 (31.5)	17.3–82.3 (60.2)	43.1–65.4 (55.8)	29.5–87.4 (46.6)
Saturated	29.0–44.7 (36.8)	57.9–68.3 (65.1)	10.3–76.0 (28.9)	25.0–39.5 (31.6)	7.9–41.8 (31.4)

* Data adapted from Andrews and Richardson (1943).

† Selective hydrogenation of entire oil to form a plastic fat.

‡ Vegetable oil hydrogenated almost to saturation, then blended with unhydrogenated oil to form a plastic fat.

[44] G. O. Burr, *Federation Proc.*, **1**, 224 (1942).

Because hydrogenation proceeds more readily on the fatty acids with a greater unsaturation than oleic acid (*cf.* Table 60, p. 777), it is evident that the nutritive value of a fat may be adversely affected by hydrogenation. That destruction of the essential fatty acids during manufacture of commercial shortenings is not of serious concern from the standpoint of nutrition has been shown by Andrews and Richardson,[45] who studied the fatty acid fractions in a number of edible fats. Their findings are summarized in Table 66. The essential fatty acids increase in edible fats in the order: butter, margarine, lard, hydrogenated shortening, and blended shortening containing some hydrogenated fat.

Cholesterol Metabolism. Our knowledge of the catabolic changes and the metabolic functions of cholesterol is still extremely meagre. Still less is known about the metabolism of other sterols.

At least three functions have been suggested for cholesterol: (1) it may be an important structural unit in the mosaic of the cell membrane; (2) it may be involved in fatty acid transport; and (3) it may serve as a biological precursor of other compounds necessary for proper functioning of the animal body.

Phospholipid and cholesterol occur in blood and in many tissues in a fairly constant ratio characteristic of the particular tissue. This has been interpreted as reflecting a structural role for these two lipids. In most tissues, the amount of cholesterol esterified with fatty acids is a very small fraction of the total sterol present. In blood, however, the major part of the cholesterol is esterified, the fatty acids being restricted largely to those with a high degree of unsaturation,[46] suggestive of a possible role of the sterol in transport of the unsaturated fatty acids.

As previously indicated, studies with deuterium as a trace element have demonstrated that cholesterol is converted by the body to other biologically active substances, including both bile salts and hormones. As noted by the Columbia group, this implies that the animal can shorten the side chain, introduce hydroxyl groups into both the nucleus and side chain, epimerize preexisting hydroxyl groups, and saturate the double bond in the cholesterol molecule.

The high content of cholesterol in nervous tissue suggests that it may have a significant role the nature of which we are still unaware. Of interest in this connection is the observation of Bloch, *et al.*,[29] that "cholesterol of the central nervous system seems to be the most inert of all tissue constituents which have as yet been studied; it is not regenerated and does not interchange with the dietary cholesterol at appreciable rates."

[45] J. T. R. Andrews and A. S. Richardson, *Oil & Soap*, **20**, 90 (1943).

[46] F. E. Kelsey and H. E. Longenecker, *J. Biol. Chem.*, **139**, 727 (1941).

Cook [47] has reviewed the literature on cholesterol metabolism. His paper will guide the reader to more detailed treatises in this fruitful field for research.

Metabolic Functions of the Phospholipids. The likelihood that phospholipids are important structural materials of the cells has already been mentioned. The phospholipids have lyophilic properties, and they often occur in close association with many proteins in the form of lipoproteins such as the lipovitellin of egg yolk, and tissue fibrinogen or thromboplastin (cf. Chapter 16).

The possible role of the phospholipids in fatty acid transport has been stressed by many workers. Sinclair's work pointed to the functioning of these lipids in the intestinal mucosa as agents for transporting the fatty acids from the epithelium to the lymph system. Studies with radioactive phosphate also indicate that newly formed phospholipids are rapidly distributed to most tissues of the body. The liver appears to be the main site of formation of the plasma phospholipids; inorganic phosphorus is rapidly incorporated into the liver lecithin and cephalin fractions, but not in sphingomyelin.[48] The central nervous system and the testes, however, are very slow in turnover of phospholipid.

Abnormal accumulation of fat in the liver under varying dietary and pathological conditions has also strengthened the belief that fatty acid transport is an important function of the phospholipids. Choline or substances which can furnish active methyl groups for *in vivo* synthesis of choline (*e.g.*, methionine) will rapidly reduce or prevent the accumulation of fat in livers of animals that have been on high fat diets with supplements of thiamine, and to a lesser extent will improve the fatty liver induced by diets high in cholesterol. McHenry [49] notes that this effect, termed *lipotropism*, appears to be explainable on the basis of stimulation of phospholipid formation by choline or related compounds. Such fatty livers would thus be ascribed to a failure of the phospholipid transport mechanism, which prevents the removal of fats from the liver at a sufficiently rapid rate. McHenry states that the lipotropic activity of inositol is also probably due to formation of phospholipid and points out that inositol phosphatides are now known to exist. Choline and inositol do not have similar lipotropic activities for different types of dietary-induced fatty livers. A third lipotropic factor, *lipocaic*, obtained from pancreas,[50] is recognized, but its nature and mode of action are still uncertain.

[47] R. P. Cook, *Nutrition Abstracts & Revs.*, **12**, 1 (1942).

[48] F. E. Hunter, *Proc. Soc. Exptl. Biol. Med.*, **46**, 281 (1941).

[49] E. W. McHenry and J. M. Patterson, *Physiol. Revs.*, **24**, 128 (1944).

[50] L. R. Dragstedt, J. V. Prohaska, and H. P. Harms, *Am. J. Physiol.*, **117**, 175 (1936).

Participation of phospholipids in oxidation-reduction processes associated with cellular respiration has also been suggested. The data bearing on this problem do not offer convincing proof that they do actually participate. That the phospholipids somehow are involved in the active metabolism of tissues is likely, however. Bloor [51] found that the tissues having the highest state of activity also had the highest content of phospholipid. Since then his observation has been repeatedly confirmed and has been extended to include a single tissue at different stages of metabolic activity.[52]

After reviewing the current status of our knowledge of phospholipid metabolism, Sinclair [53] concludes that the functions of the phospholipids remain an unsettled question. Recent recognition of the existence of both serine and inositol in phospholipids may offer a new basis for explaining some of the seemingly anomalous data which bear on this problem.

[51] W. R. Bloor, *J. Biol. Chem.*, **72**, 327 (1927).
[52] W. R. Bloor, R. Okey, and G. W. Corner, *J. Biol. Chem.*, **86**, 291 (1930).
[53] R. G. Sinclair, *Biol. Symposia*, **5**, 82 (1941).

CHAPTER 33

Essential Oils

The essential oils may be defined as *those compounds in plants which are volatile with steam and usually separate as an oily layer in the distillate.* They are present to a small extent in most plants and may be present to a very considerable extent in certain families or groups of plants. Some essential oils are of interest only from the scientific standpoint, involving the identification or the preparation of rare organic compounds. Others such as oils of wintergreen, clove, cinnamon, bergamot, attar of roses, lemon and orange oils, camphor, cedar oil, pine oil, eucalyptus oil, turpentine, etc., are of very considerable commercial importance.

In certain plants essential oils may occur in all the tissues. This is notably true of the conifers. In the rose they occur in appreciable amounts only in the petals, in cinnamon only in the bark and leaves, in the orange in the petals of the flowers and the skin of the fruit, in the nutmeg chiefly in the fruit, and in the camphor tree both in the leaves and throughout the entire woody tissue.

The function of essential oils in the plant kingdom is unknown. Here again, various suggestions have been made, as in the case of the tannins, that they may be waste products, or protection against injury or the invasion of fungi, and that they may provide an odor which will attract insects so as to favor pollination. Very closely allied species may differ greatly in their content of essential oils and in the chemical constituents present in the essential oils.

Miller [1] investigated the essential oils of three species of *Pycnanthemum*, *P. tullia*, *P. incanum*, and *P. lanceolatum*. Miller noted that it is very difficult to separate these three species sharply on morphological grounds. He found, however, that the compounds present in the essential oils were widely different. Thus, the essential oil of *P. tullia* consisted of approximately 50 per cent of cineol; that from *P. incanum* consisted of approximately 90 per cent of pulegone; whereas

[1] E. R. Miller, "A Chemical Investigation of the Volatile Oils of Some Species of the Genus *Pycnanthemum* Michx.," Ph.D. thesis on file in the Library, University of Minnesota, and privately printed, 1918.

the main fraction of the oil from *P. lanceolatum* consisted of cavacrol, only approximately 5 per cent of pulegone being present. Miller accordingly suggested in his discussion that a study of the essential oils might well be undertaken in certain families as at least an assistance in determining botanical classifications.

McNair [2] studied the chemical composition of 398 essential oils from 87 plant families with respect to physical properties and the chemical compounds which were present in the essential oils. He concluded that the nature of the essential oils is intimately related to the position of the plant in the phylogenetic scale, that terpenes and compounds of the fatty series predominate in the volatile oils produced by the plants lowest in the evolutionary scale, and that the plants highest in the evolutionary scale contain more of the volatile oils with aromatic, sulfur, and nitrogenous compounds. He pointed out also that there is almost a straight-line relationship between the refractive index of the volatile oil and the position of the plant on the evolutionary scale, those plants high in the evolutionary scale having volatile oils with a low refractive index. A similar relationship was found to hold for specific gravity; the higher the evolutionary scale of the plant, the greater was the specific gravity of the essential oil. His studies also brought out the fact that there is a climatic influence on the nature of the compounds which are produced. The essential oils of the tropics have lower specific gravities and higher refractive indices than those of temperate climates. He concluded by insisting that the chemical composition of the plant, including particularly the nature of the alkaloids, glycerides, and essential oils, be taken into consideration in taxonomic revision of the various plant groups, since these three groups of chemical compounds appear to form a scale against which the relative degree of evolution of the various groups can be charted.

The field of the chemistry of the essential oils presents many exceedingly difficult problems. Some oils consist almost wholly of a single compound. Other oils are mixtures [3] containing a dozen or more compounds of the most diverse types.

The physical methods for identification of essential oils are somewhat similar to those already noted for the usual fats and oils, such as (1) optical rotation, (2) specific gravity, (3) index of refraction, and (4) boiling point range, *i.e.*, the temperature range over which an oil can be completely distilled.

[2] J. B. McNair, *Am. J. Botany*, **19**, 168 (1932); **21**, 427 (1934); *Bull. Torrey Botan. Club.*, **62**, 515 (1935).

[3] F. B. Power, *J. Ind. Eng. Chem.*, **11**, 344 (1919).

The laboratory methods[4,5] of separating the compounds present in the mixture are in part:

1. Low temperatures, such as freezing point. By this method it is sometimes possible to crystallize out certain of the constituents.
2. Fractional distillation with steam separates an oil into the more volatile and the less volatile fractions.
3. Fractional distillation *in vacuo* of the dried oil again may serve to bring about a partial separation, as may
4. Crystallization from poor solvents of certain oils and fractions of certain oils.
5. The constituents of an essential oil possessing free acidic groups may be removed by shaking the oil with sodium carbonate solution and separating the aqueous solution of the sodium salts so formed from the main bulk of the oil.
6. Similarly, basic compounds may be removed by shaking the oil with a dilute solution of hydrochloric acid and removing the aqueous acid solution containing the bases.
7. Phenols may be removed by shaking the oil with a dilute solution of sodium or potassium hydroxide.
8. Aldehydes and ketones may be removed by shaking the oil with a saturated solution of sodium bisulfite, resulting in the formation of the crystalline bisulfite addition products which are non-oil-soluble.
9. By a determination of the acetyl value, the free hydroxyl groups which are present in the oil may be determined.
10. The acids which were separated by shaking with sodium hydroxide are liberated by the acidification of the solution and are converted into esters or into their silver or barium salts. The saponification number of the esters or the percentage of silver or barium in the salts provides a means of identifying the acids which are present.
11. In many instances, specific compounds which occur frequently in essential oils may be tested for by means of special color reactions.

Undoubtedly the average individual is chiefly interested in the essential oils as the basis of the perfume industry. The perfume industry is very highly developed in France. Grasse alone has used each year[6] 2,200 tons of orange blossoms, 1,650 tons of rose petals, 1,320 tons of jasmine petals, 440 tons of violets, 330 tons of tuberoses, 165 tons of carnations, 110 tons of cassie, 880 tons of mimosa branches, 66 tons of

[4] E. J. Parry, *The Chemistry of Essential Oils and Artificial Perfumes*, Scott, Greenwood and Company, London, 1899.

[5] F. W. Semmler, *Die aetherischen Öle*, 4 vols., Veit and Company, Leipzig, 1906–07.

[6] M. T. Bogert, *J. Ind. Eng. Chem.*, **14**, 359 (1922).

mignonette, and 55 tons of narcissus. One pound of orange blossom oil was obtained from 1,000 pounds of blossoms. One pound of attar of roses required 8 tons of petals.

Three methods of extraction are employed when the oils are to be made into perfumes.[7]

1. *Extraction with Petroleum Ether.* This gives a greater yield, but other substances, extracted along with the true essential oils, result in a lower grade of perfume.

2. *Maceration of the Material with Warm Oil or Fat.* The material is ground in the presence of added oil or fats. The fat containing the essential oil is then expressed by means of a hydraulic press, and the essential oil is recovered by extracting the fat with strong alcohol. The fat (lard) is largely insoluble in the alcohol, and it crystallizes out on cooling, leaving an alcoholic solution of the odoriferous principles. This is a better method than (1) but still yields an inferior grade of perfume for many flowers.

3. *Cold Absorption.* Thin layers of cold fat (40 per cent beef, 60 per cent lard) are spread about $\frac{1}{8}$ inch thick on both sides of glass plates, and a layer of petals about 2 inches deep is placed above these. The layers of fat and petals are racked up one on top of the other as high as the operator can reach. After 24 to 72 hours (depending on the flower being used and the quality of perfume desired), the flowers are removed and new flowers added. Finally the layers of fat are extracted with *cold alcohol*, the fat which dissolves in the alcohol is frozen out, and the alcoholic solution is concentrated or bottled directly. This process yields the finest perfumes.

The essential oils as used in industrial processes or in medical practice may be obtained by three general methods:

1. By some suitable solvent, such as petroleum ether, dichloroethylene, etc., the solvent being of such a nature that it can be readily and completely removed by evaporation at a low temperature.

2. In some instances, by pressure, ground material and a hydraulic press being used.

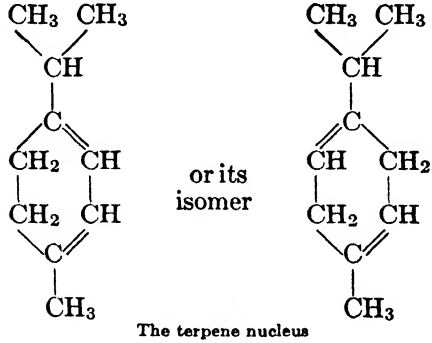
3. By steam distillation—the more general method—steam from a boiler or superheated steam being passed through a mass of raw material which contains the essential oils. This steam passing through the material carries the oil over, the oil separating as a layer in the distillate.

As already noted, various types of chemical compounds, such as hydrocarbons, alcohols, ketones, aldehydes, acids, esters, organic sul-

[7] E. Guenther, *The Essential Oils. History, Origin in Plants, Production, and Analysis*, Vol. I, D. Van Nostrand Company, New York, 1948.

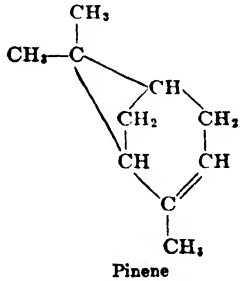
fides, bases, etc., may occur in essential oils. Only a few typical examples will be noted to illustrate the diversity of compounds which may be present.

I. *Terpenes.* These are hydrocarbons of the general formula $C_{10}H_{16}$ and in general are closely related to the parent hydrocarbon. The relationships of these configurations to the cyclic end groups of the carotenoids should be noted.

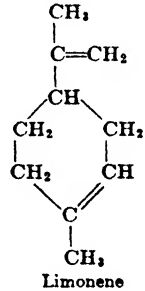


The terpene nucleus

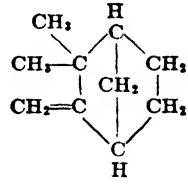
Certain of the typical terpenes are pinene (from conifers), limonene (from lemon-grass oil), and camphene (from the camphor tree).



Pinene

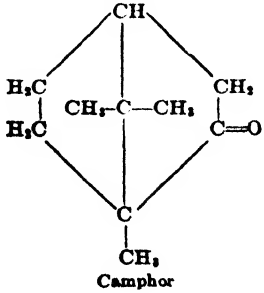


Limonene

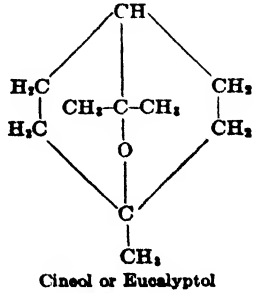


Camphene

II. *Alcohols and Ketones.* These types of compounds are very abundant in essential oils. The camphor series includes camphor, borneol, pulegone, menthol, etc.

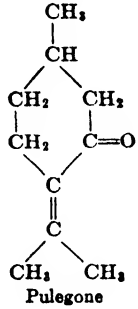


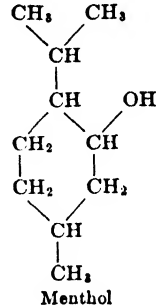
Camphor



Cineol or Eucalyptol

Camphor is obtained from the wood of the camphor tree. It has been synthesized, and the synthetic product is a serious competitor of the natural product. It would appear that the camphor industry faces at the present time the same problems that the natural indigo industry faced a few decades ago.

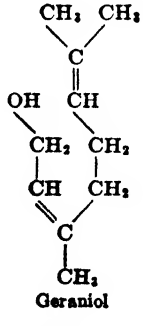

Pulegone

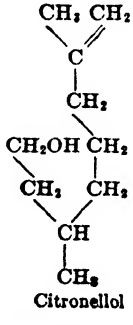

Menthol

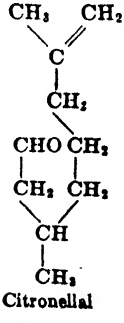
Borneol has the same structural groupings as camphor with the exception that the keto group is replaced by a secondary alcohol group.

Menthol occurs in oil of peppermint to the extent of 65 to 85 per cent, depending on the quality of the peppermint oil. The corresponding ketone, menthone, also occurs in peppermint oil. Cineol or eucalyptol is very widely distributed as a component of many of the essential oils and occurs in a very large amount in oil of eucalyptus.

III. *Geraniol and Citronellol Group* of alcohols, ketones, and aldehydes is another important group. Here we do not have a closed ring, but the formulas may be so written as to indicate their relationship to the closed ring structures, which have already been noted. It will be observed that here there is a marked similarity to the lycopin residue of the carotenoids.

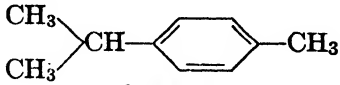

Geraniol


Citronellol

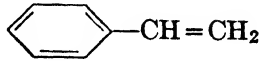

Citronellal

Geraniol occurs in oils of rose, geranium, sassafras, lavender, etc. Citronellal occurs in lemon-skin and in orange-skin oils, eucalyptus oil, lemon-grass oil, etc.

IV. *Benzene Hydrocarbons* (not terpenes) are of common occurrence, *e.g.*, cymene occurs in caraway oil, styrene in styrax balsam.

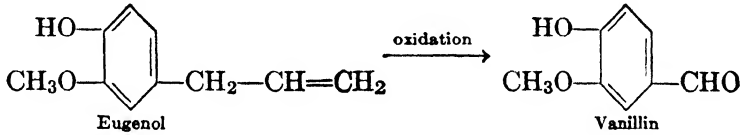


Cymene



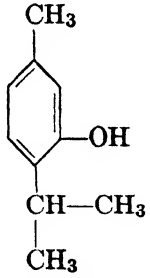
Styrene

V. *Phenols* are widely distributed in the essential oils. Thus, eugenol is the principal constituent of oil of cloves. On oxidation eugenol is converted to the corresponding aldehyde, vanillin, the flavoring principle of the vanilla bean.

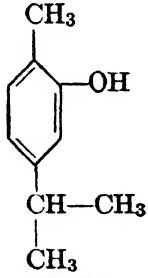


Eugenol Vanillin

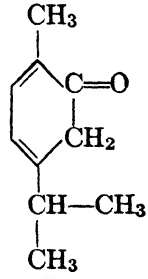
Thymol (isopropyl-*m*-cresol) occurs in a number of essential oils but particularly in the oil of thyme. It is a hydroxy derivative of cymene which is likewise found in oil of thyme and in eucalyptus oil. The corresponding *o*-cresol derivative is carvacrol, occurring in many essential oils, and the corresponding ketone of carvacrol is carvone.



Thymol



Carvacrol



Carvone

VI. *Acids and Esters of Acids* are of common occurrence. These include benzoic acid and its esters, salicylic acid and its esters. The esters of acetic, butyric, and valeric acids are of rather common occurrence. Methyl salicylate, the methyl ester of *o*-hydroxybenzoic acid, is practically the only constituent of pure oil of wintergreen and oil of birch. Ethyl acetate is the chief constituent of oil of banana. Amyl valerianate and ethyl anthranilate (the ethyl ester of *o*-aminobenzoic acid) are the chief constituents of oil of apples. Amyl butyrate occurs in oil of apricots. Ethyl salicylate, ethyl butyrate, and amyl acetate occur in the fruit of the strawberry.

VII. *Aliphatic Alcohols*—a number of these are of common occurrence. Some of these are water-soluble and do not separate with the oil but remain in the aqueous portion of the distillate, from which they must be removed by extraction with a suitable non-miscible solvent or by fractional steam distillation.

VIII. *Aliphatic aldehydes* in small amounts, such as traces of formaldehyde, acetaldehyde, etc., are of rather common occurrence.

IX. *Sulfides* occur in certain oils. Thus, allyl sulfide

$$CH_2{=}CH{-}CH_2{-}S{-}CH_2{-}CH{=}CH_2$$

occurs in oil of garlic, and allyl thiocyanate

$$CH_2{=}CH{-}CH_2{-}N{=}C{=}S$$

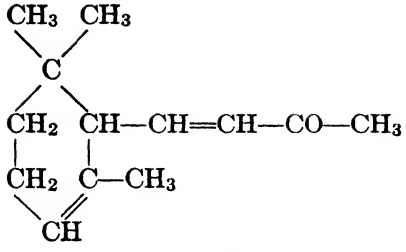
occurs in oil of mustard. Essential oils containing sulfur are especially abundant in the *Cruciferae*. In many instances the sulfur-containing radical is combined with sugars in the form of a glycoside. Thus, allyl thiocyanate occurs in mustard seed as a glycoside which must be hydrolyzed either by enzymatic action or by chemical means before the allyl thiocyanate can be distilled as an essential oil.

X. *Paraffin Hydrocarbons*, C_nH_{2n+2}, are occasionally found in small amounts in essential oils, but their occurrence is rare.

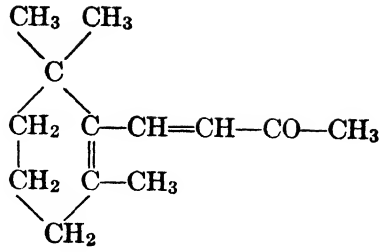
XI. *Organic Bases* occur in small amounts in essential oils. It is still a debatable question whether these are normal constituents of essential oils or have been derived from proteins, lipids, or other nitrogenous constituents of the plant by some decomposition process, either that of the natural autolysis of the plant material or a decomposition brought about by the methods used for preparing the essential oil. It seems probable that they are secondary decomposition products. The one possible exception is indole, which occurs in extremely small amounts in the oil of orange blossoms. Indole, when present in appreciable amount, has a very disagreeable and penetrating odor. The amount which is present in orange-blossom oil is, however, extremely small and in conjunction with the other constituents of the oil produces the characteristic orange-blossom perfume. The synthetic perfume failed to simulate completely the natural perfume until a trace of indole had been added.

Most of the perfumes which are sold to the consumer represent blends which are mixtures of natural essential oils, or blends of compounds which have been synthesized [8] in the organic laboratory. In some instances compounds have been synthesized which very closely resemble

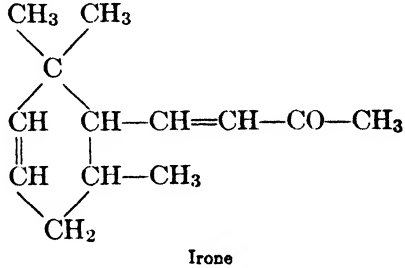
[8] M. T. Bogert, *J. Chem. Education*, **8**, 1311 (1931).



α-Ionone



β-Ionone



Irone

in their chemical structure the constituents occurring in the natural perfumes. Thus, α- and β-*ionone* were synthesized in 1893 in the attempt to synthesize *irone*, the compound responsible for the odor of orris root and probably oil of violets. A mixture of the synthetic α- and β-ionone, which differ from irone only in the position of the double bond,[9] produces a violet perfume so nearly like the natural product as to defy detection by all but the most expert of perfume connoisseurs. The ionones were regarded as wholly synthetic products until in 1929 β-ionone was shown [10] to be present in rather large amount in the essential oil of *Boronia megastigma*. Later Karrer showed the ionone residues to be the characteristic configurations in the carotenoids (p. 886), the ring of α-ionone occurring in α-carotene and that of β-ionone in β-carotene. Irone itself, though it has been known since 1893, was not synthesized [11] until 1935. Although it is isomeric with the ionones, the position of the double bond confers on the molecule a much greater odor value than is possessed by the ionones.

[9] H. Hibbert and L. T. Cannon, *J. Am. Chem. Soc.*, **46**, 119 (1924).
[10] S. Sabetay, *Compt. rend.*, **189**, 808 (1929).
[11] A. Verley, *Bull. soc. chem.*, [5] **2**, 1205 (1935).

V

PLANT PIGMENTS

The whole subject of carotenoids and related pigments is a fascinating one and offers as many unsolved problems as any other phase of experimental biology and biochemistry. . . . The extension of the frontiers of our knowledge regarding these pigments which are so abundantly distributed in so many plants and animals is certain to prove a profitable as well as an interesting undertaking.

LEROY S. PALMER (1922)

Chlorophyll . . . is evidently one of the most remarkable substances in nature, and its activity provides some of the chief theoretical and economic problems which confront the botanist.

EDMUND W. SINNOTT (1946)

CHAPTER 34

Chlorophyll and the Carotenoids [1]

Chlorophyll is the green coloring matter of plants. It occurs in the chloroplasts and is accompanied by one or more of each of the groups of yellow pigments, the carotenes and xanthophylls. The former group consists principally of β-carotene accompanied by varying amounts of α-carotene. The principal xanthophyll in chloroplasts of higher plants is lutein (often called leaf xanthophyll), which is invariably accompanied by very small quantities of zeaxanthin and cryptoxanthin. Table 67, compiled by Strain,[1] summarizes the distribution of these pigments.

Chemists from the time of Berzelius (1839) have struggled with the chemistry of chlorophyll. Willstätter [1] made the first great advance in the determination of its structure. During the past fifteen years, Conant, Hans Fischer, and Stoll, to mention only a few of the workers, have advanced our knowledge of chlorophyll to the point of complete elucidation of its molecular structure.

The state in which chlorophyll exists in the plant cell has been the

[1] For reviews and collected works see R. Willstätter and A. Stoll, *Untersuchungen über Chlorophyll*, Julius Springer, Berlin, 1913; C. J. West, *Biochem. Bull. N. Y.*, **3**, 229 (1914); R. Willstätter, *J. Am. Chem. Soc.*, **37**, 323 (1915); L. S. Palmer, *Carotenoids and Related Pigments*, American Chemical Society Monograph, No. 9, Chemical Catalog Co., New York, 1922; F. M. Schertz and A. R. Merz, Translation of Willstätter and Stoll, with notes, Science Press, Lancaster, Pa., 1928; P. Karrer and H. Wehrli, *Nova Acta Leopoldina*, Neue Folge, **1**, 175 (1933); L. Zechmeister, *Carotinoide*, Julius Springer, Berlin, 1934; H. Willstaedt, *Carotinoide Bakterien- und Pilzfarbstoffe*, Ferdinand Enke, Stuttgart, 1934; F. Mayer, *Chemie der organischen Farbstoffe*, Bd. II, Natürliche organische Farbstoffe, Dritte Umgearbeitete Auflage, Julius Springer, Berlin, 1935; P. Karrer, *Helv. Chim. Acta*, **19** (Fasciculus Extraordinarius), E33 (1936); H. H. Strain, *Leaf Xanthophylls*, Carnegie Institute of Washington Publication No. 490 (1938); H. Fischer and A. Stern, *Die Chemie des Pyrrols*, II Bd., 2 Halfte, Akad. Verlagsgesell., Leipzig, 1940; G. Mackinney, *Ann. Rev. Biochem.*, **9**, 459 (1940); F. P. Zscheile, *Botan. Rev.*, **7**, 587 (1941); H. H. Strain, *Chromatographic Adsorption Analysis*, Interscience Publishers, Inc., New York, 1942; A. H. Cook, translation of F. Mayer, *The Chemistry of Natural Coloring Matters*, American Chemical Society Monograph No. 89, Reinhold Publishing Corp., New York, 1943; L. Zechmeister and L. Cholnoky, *Principles and Practice of Chromatography*, John Wiley & Sons, New York, 1943; L. Zechmeister, *Chem. Revs.*, **34**, 267 (1944); H. H. Strain, *Ann. Rev. Biochem.*, **13**, 591 (1944); H. H. Strain, W. M. Manning, and G. Hardin, *Biol. Bull.*, **86**, 169 (1944).

subject of much investigation.[2] The difficulty of separating "pure" chloroplasts without loss or contamination has left some doubt about their exact composition. However, it is clear that they differ markedly from the surrounding cytoplasm, being much richer in lipids and correspondingly poorer in protein. It is generally believed that the chlorophyll is closely associated with both protein and lipids. The chlorophyll-

TABLE 67.* THE OCCURRENCE OF CHLOROPLAST PIGMENTS IN VARIOUS GROUPS OF PLANTS †

Pigment	Higher Plants	Green Algae	Euglenophyceae	Brown Algae	Diatoms	Dinoflagellates	Yellow-Green Algae	Red Algae	Blue-Green Algae
Chlorophylls									
Chlorophyll *a*	++	++	++	++	++	++	++	++	++
Chlorophyll *b*	++	++	?	−	−	−	−	−	?
Chlorophyll *c*	−	−	−	++	++	++	−	−	−
Chlorophyll *d*	−	−	−	−	−	−	−	++	−
Xanthophylls									
Cryptoxanthin	++	++							
Lutein	++	++	−	−	−	−	−	+	
Zeaxanthin	++	++	−	−	−	−	−		
Violaxanthin	++	++		++	−	−			
Flavoxanthin	++	++		+					
Neoxanthin	++	++			−	−			
Fucoxanthin	−	−	−	++	++	−	−	?	−
Neofucoxanthin *A*	−	−	−	++	++	−	−	?	−
Neofucoxanthin *B*	−	−	−	++	++	−	−	?	−
Diatoxanthin	−	−	−	?	++	−	−		
Diadinoxanthin	−	−	−	?	++	++	−		
Dinoxanthin	−	−	−	−	−	++	−		
Neodinoxanthin	−	−	−	−	−	++	−		
Peridinin (sulcatoxanthin)	−	−	−	−	−	++	−		
Myxoxanthophyll	−	−	−	−	−	−	−	−	+
Unnamed xanthophylls			+	+			+		
Carotenes									
β-Carotene	++	++	++	++	++	++	++	++	++
α-Carotene	+	+	−	−	−	−	−		
ε-Carotene	−	−	−	−	+	−		−	
Flavacin									+
Proteinaceous pigments									
Phycoerythrin	−	−	−	−	−	−	−	+	+
Phycocyanin	−	−	−	−	−	−	−	+	+

* From H. H. Strain, *Ann. Rev. Biochem.*, **13**, 594 (1944).

† ++ Indicates the presence of the pigment in most or in all of the plants examined.
+ Indicates its presence in some of the plants examined.
− Indicates its absence.
? Indicates small quantities that may have come from contamination of the source by other organisms.
A blank space indicates that a thorough search was not made for the pigment.

[2] E. I. Rabinowitch, *Photosynthesis and Related Processes*, Interscience Publishers, Inc., New York, 1945.

protein complex has been given such names as *phyllochlorin*, *chloroplastin*, *photosynthin*, and *chloroglobin*. Wide acceptance of any proposed name probably must await proof of constancy of size and composition of such a compound.

The early view that chloroplasts are heterogeneous, for a time discarded, has been recently revived. The *grana* can be distinguished from the continuous phase (*stroma*) under the microscope. If the chlorophyll is concentrated in these particles it may constitute as much as 10–30 per cent of the dry weight as contrasted with 0.5 to 1.5 per cent of whole leaves. The highest reported concentration of chlorophyll in intact cells is approximately 5 per cent of the dry weight of *Chlorella*.

Chlorophyll may be prepared from either fresh or dried leaves. One kilogram of fresh leaves gives a yield of 0.9 to 2.1 grams; dried leaves yield 5 to 10 grams. The most suitable solvent for extraction is petroleum ether for dried leaves, and pure acetone for fresh leaves, sufficient acetone being added so that, allowing for the moisture in the fresh leaves, the resulting solution is 80 per cent acetone. Chlorophyll can now be isolated as readily as can any alkaloid or any sugar, and within a few hours a kilogram of dried leaves should yield about 6.5 grams of practically pure chlorophyll.[3]

Schertz notes that in the United States alone more than 6,000,000 tons of chlorophyll are produced each year by the corn and small grain crops. He also notes that it possesses some commercial importance and that, in 1924, 3,213 pounds of chlorophyll, valued at $5,799, were imported into the United States to be used in medicines and as coloring for candles, waxes, resins, oils, soaps, foods, etc.

The entire course of chlorophyll utilization by animals is not known. However, many degradation products of chlorophyll have been found in urine, feces, and gallstones. These products include phylloerythrin from bovine gallstones, phylloerythrin and a chlorin (m.p. 221°) in dog feces, a chlorophyll-porphyrin in feces and urine of men. See Fischer and Stern [1] for a more complete list.

Certain of the decomposition products of chlorophyll are extremely resistant to decay. Treibs [4] isolated from oil schists (Alpine Triassic) a compound which he identified as desoxyphylloerythrin which has the same structure as 3-desmethyldesoxyphylloerythrin, except that a $-CH_3$ group replaces the hydrogen on carbon-3. The same product was also identified in bituminous earths, mineral oils, and waxes, and in asphalts, indicating that these had been derived from plant remains.

[3] F. M. Schertz, *Plant Physiol.*, **3**, 211, 323, 487 (1928); *Ind. Eng. Chem.*, **30**, 1073 (1938); **19**, 1152 (1927).

[4] A. Treibs, *Ann.*, **509**, 103 (1934); **510**, 42 (1934).

Properties of Chlorophyll. Prepared according to the method of Schertz, chlorophyll is a bluish black substance with a strong metallic luster, powdering to a greenish or bluish black powder. It has no definite melting point, ranging from 93° to 106°C. for various samples, and is soluble in absolute alcohol, giving a blue-green solution. It shows neither acidic nor basic properties. Acids change its color to olive brown and split off magnesium which is associated with the molecule.

Pure chlorophyll shows the following characteristics:

1. The ash content is 4.5 per cent, the ash being pure magnesium oxide.
2. One-third of the molecule is accounted for in the form of a monoatomic alcohol, phytol, $C_{20}H_{39}OH$.
3. The associated yellow pigments are not a part of the chlorophyll molecule.
4. When chlorophyll is saponified with potassium hydroxide in methyl alcohol, the color changes to a pure brown; impure mixtures give a dirty brown (phase test). Chlorophyll is decomposed by boiling alcoholic potassium hydroxide, and a normal mixture of *phytochlorine e* and *phytorhodin g* is formed.
5. The absorption spectrum of solutions of pure chlorophyll is that of the fresh leaf extract. Chlorophyll *a* shows definite absorption bands at 663, 623, 607, 577, 534, 507, 494, and 432 mμ. Chlorophyll *b* shows absorption bands at 644, 614, 594, 567, 542, 503, 456, and 428 mμ.

Stokes, in 1864, first noticed spectrographically that chlorophyll was a mixture of two components which he called *a* and *b*. He attempted to separate them by means of a differential solubility in alcohol and carbon bisulfide. Willstätter used aqueous methyl alcohol and petroleum ether; chlorophyll *a* goes into the petroleum ether and chlorophyll *b* into the methyl alcohol.

Chlorophyll *a* gives a pure yellow phase in a methyl alcoholic solution of potassium hydroxide and gives only phytochlorine *e* as a decomposition product. Chlorophyll *a* crystallizes in thin lance-like leaflets with a blue-steel luster (m.p. 117°–120°). The solutions in ethyl alcohol are bluish green with a deep red fluorescence. The empirical formula is

$$C_{55}H_{72}O_5N_4Mg \qquad [\alpha]_{720}^{25} = -262°$$

Chlorophyll *b* gives a dark red phase with a methyl alcohol solution of potassium hydroxide and phytorhodin *g* as a decomposition product. The alcoholic solution has a yellow tinge as compared with chlorophyll

a. Chlorophyll *b* is insoluble in cold petroleum ether. It has no sharp melting point. The empirical formula is

$$C_{55}H_{70}O_6N_4Mg \qquad [\alpha]_{720}^{25} = -267°$$

In recent years the Willstätter method of partition between petroleum ether and 90 per cent aqueous methanol has been supplemented by adsorption on Tswett columns. Usually mild adsorbents, like powdered sugar or inulin, are preferred because the chlorophylls tend to undergo changes when being chromatographed.

It must be emphasized, however, that chlorophyll prepared in quantity by the above methods may have undergone changes which markedly affect its spectral quality. Mackinney [5] compared samples of chlorophyll prepared in five different laboratories and found that the absorption coefficients were "so incompatible that their application to spectroscopic assay of the green pigments is useless." In attempting to obtain chlorophyll preparations with constant properties Zscheile and Comar [6] established ratios of absorption coefficients at fixed wave lengths for each of the components. They concluded that, for preparation of samples of reproducible spectral quality after cold acetone extraction, exposure of the pigments to light, heat, drying, or acid must be avoided. When ethyl alcohol is used as an extractive for fresh leaves the enzyme *chlorophyllase* brings about alcoholysis of the chlorophyll, substituting an ethyl group for the phytyl portion of the molecule. The ethyl chlorophyllid is probably the crystalline chlorophyll discovered by the Russian botanist, Borodin, in 1881. On standing, ethyl alcoholic solutions of chlorophyll also readily undergo a change, "allomerization," after which the phase test (see below) is no longer given.

The visible absorption spectra of good preparations of chlorophylls *a* and *b* are given in Fig. 114.

The fluorescent spectrum of chlorophyll *a* has bands at 6685 Å. and 7230 Å.; that of chlorophyll *b* has bands at 6485 Å., 6720 Å., and 7050 Å. (Zscheile [1]).

In general, the ratio of occurrence in higher plants is about 1 molecule of chlorophyll *b* to 3 molecules of chlorophyll *a*. The *a*:*b* ratio is affected by light conditions, averaging 5.5 in alpine plants (maximum value found by Seybold, 9.3) and as low as 1.39 in green algae. There is also a species variation which leads to complete lack of chlorophyll *b* in some forms (Table 67).

In addition to chlorophylls *a* and *b* there are in algae chlorophyll *c* (also called chlorofucine or chlorophyll γ) and chlorophyll *d* (Table 67).

[5] G. Mackinney, *J. Biol. Chem.*, **132**, 91 (1940).

[6] F. P. Zscheile and C. L. Comar, *Botan. Gaz.*, **102**, 463 (1941).

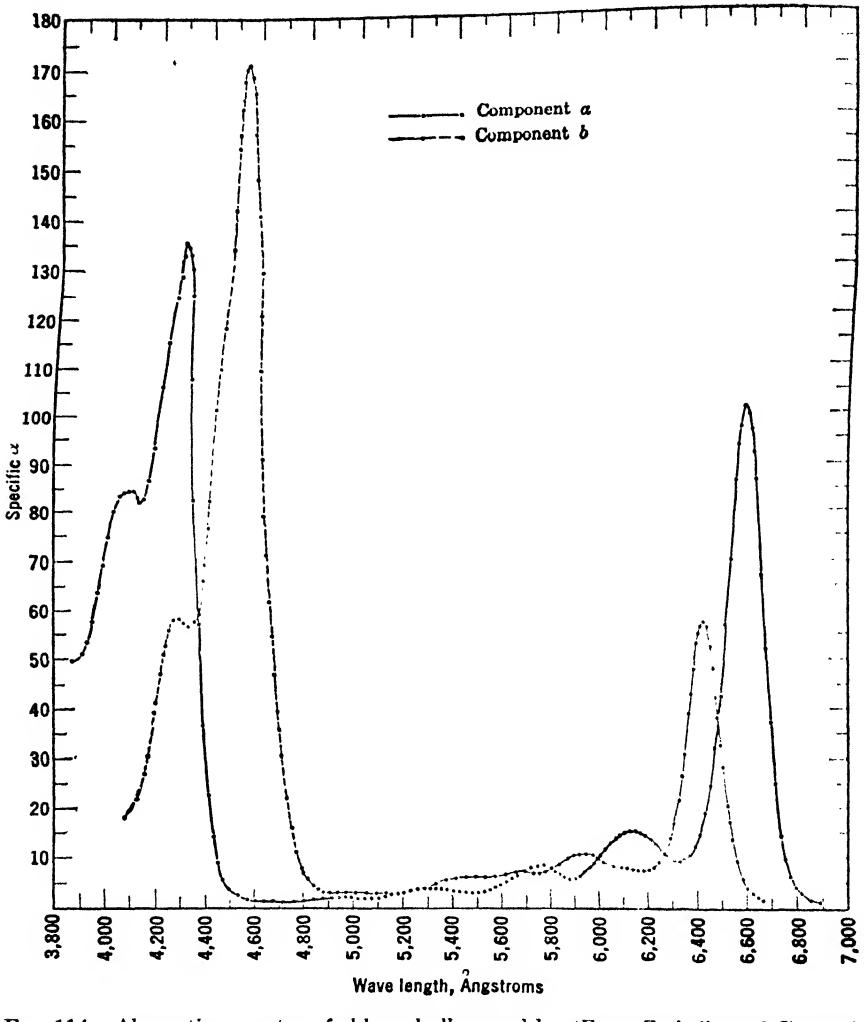


FIG. 114. Absorption spectra of chlorophylls *a* and *b*. (From Zscheile and Comar.) The specific absorption coefficient, α, is used in Beer's law:

$$\alpha = \frac{\log_{10} (I_0/I)}{C \times l}$$

where C = concentration in grams per liter
l = thickness of solution layer in centimeters
I_0 = intensity of light transmitted by solvent
I = intensity of light transmitted by solution.

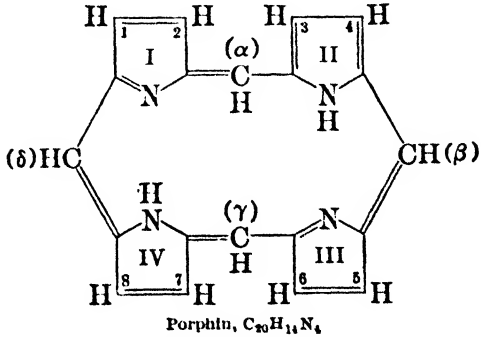
α is equal to the molecular extinction coefficient divided by the molecular weight. Wave length is given in Ångstrom units.

Bacteriochlorophyll is the photosynthetic pigment in all purple bacteria, *Thio-* as well as *Athiorhodaceae.*[7] It is extremely unstable, a fact which may account for the report of more than one form. Bacteriochlorophyll *a* has the empirical formula $C_{55}H_{74}O_6N_4Mg$. It differs from chlorophyll *a* in having an acetyl in place of a vinyl group on carbon-2 and two extra hydrogens on carbon-3 and carbon-4.

Aqueous solutions of the bacteriochlorophyll-protein complex or suspensions of the living bacteria have their most intense absorption band in the infrared. This gives to these organisms the unique ability to do most rapid photosynthesis in wave lengths longer than 800 mμ.[8]

Chlorophyll Degradation Products. Nomenclature of the chlorophyll degradation products requires the use of numerals and subscripts. The subscripts refer to the number of oxygen atoms, *e.g.*, chlorin e_4 contains only the four oxygen atoms in two carboxyl groups. Numerals after a name, *e.g.*, purpurin 7, refer to the percentage strength of hydrochloric acid required to extract two-thirds of the compound from an equal volume of ether. This is the *hydrochloric acid number* of Willstätter and Stoll.

Protochlorophyll is the pigment which gives the faint green color to some seeds and etiolated plants. It has been separated into blue-green and yellow-green components (protochlorophylls *a* and *b*) by adsorption on powdered sucrose and has been identified as an oxidation product of chlorophyll with the two hydrogen atoms removed from positions 7 and 8. It has been described by Fischer and Stern as the magnesium salt of vinyl-pheoporphyrin a_5-phytol ester. Hence, protochlorophyll, chlorophyll, and bacteriochlorophyll have as parent substances porphin, dihydroporphin, and tetrahydroporphin, respectively. The porphin nucleus has the following structure:



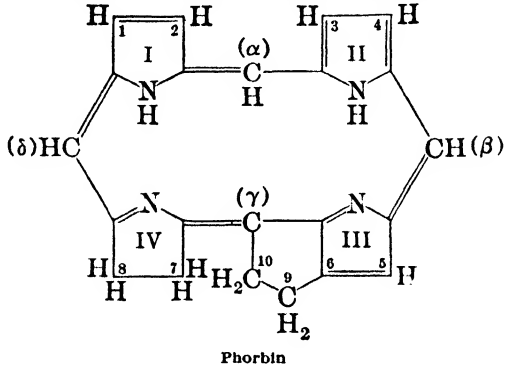
Porphin, $C_{20}H_{14}N_4$

[7] C. B. van Niel, *Advances in Enzymol.*, **1**, 263 (1941).
[8] C. S. French, *J. Gen. Physiol.*, **21**, 71 (1937).

It is impracticable to discuss in detail the many degradation products of the chlorophylls which have been isolated and identified. However, certain of the groupings are of especial interest, and it will suffice to indicate the relationships of these.

7,8-Dihydroporphin is the nucleus of the *chlorin* series of compounds derived from chlorophyll *a*. The *rhodins* are the corresponding derivatives of chlorophyll *b*, *i.e.*, they differ from chlorins in having a formyl group on carbon-3.

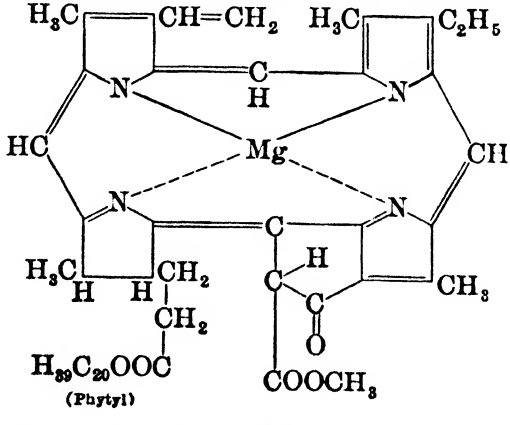
With the introduction of the isocyclic ring between the 6- and γ-positions of the chlorin nucleus *phorbin* is produced.



Phorbin

The prefix *pheo* designates those compounds which have the same substituents found in the chlorophyll molecule. Thus *pheophorbin a* would be the free dicarboxylic acid derived from chlorophyll *a* by removal of magnesium and hydrolysis of the methyl and phytyl esters (see below).

Chlorophylls *a* and *b* are highly substituted phorbins.



Chlorophyll *a*

Magnesium salt of 1,3,5,8-tetramethyl-4-ethyl-2-vinyl-9-oxo-10-carbomethoxy-phorbin-7-propionic acid phytyl ester

Chlorophyll *b* differs from *a* in having a formyl group replace the methyl on carbon-3, giving it two less hydrogens and one more oxygen. This fact has led to many attempts to show that chlorophyll *a* is oxidized to chlorophyll *b* when carbonic acid is reduced to CH_2O by photosynthesis. Such a conversion *in vivo* has not been proved, and the *in vitro* transformation is one of great difficulty.[9]

The bonding of the magnesium is obviously not fixed. It may best be considered to be located at the center of the molecule without primary bonds to any one of the four nitrogens. As resonance within the nucleus takes place its primary bonds could be directed to any two of the unsaturated rings.

Magnesium is readily replaced by hydrogens or other metals (copper, zinc, iron) in the presence of weak acids. Mackinney and Joslyn [10] have shown that the loss of magnesium in 90 per cent acetone with oxalic acid is of first order with respect to acid and probably with respect to chlorophyll. Chlorophyll *a* reacts 7 to 9 times more rapidly than chlorophyll *b*.

Using radioactive magnesium as a tracer, Ruben, Frenkel, and Kamen [11] were unable to detect the interconversion of chlorophylls *a* and *b* during photosynthesis. In some *in vitro* studies they found that unpurified extracts of mixed chlorophylls took up radioactive magnesium and copper, whereas the separated highly purified components of chlorophyll do not exchange magnesium ion in 80 per cent acetone.

Magnesium may be put back into *pheophytin a* by the use of $MgCH_3I$ (Grignard reagent) in ether to give chlorophyll *a*.

Nearly one-third of the molecular weight of chlorophyll is due to the phytyl group which can be reversibly replaced in the molecule by the enzyme *chlorophyllase*. Phytol has been synthesized by Fischer and Löwenberg,[12] who have shown that it is 3,7,11,15-tetramethyl-2-hexadecen-1-ol.

$$CH_3—\underset{\displaystyle CH_3}{\underset{|}{CH}}—(CH_2)_3—\underset{\displaystyle CH_3}{\underset{|}{CH}}—(CH_2)_3—\underset{\displaystyle CH_3}{\underset{|}{CH}}—(CH_2)_3—\underset{\displaystyle CH_3}{\underset{|}{C}}=CH—CH_2OH$$

Earlier workers suggested that phytol might be related to isoprene, the hydrocarbon of rubber and terpenes, four molecules of isoprene condensing with one molecule of water, followed by a reduction:

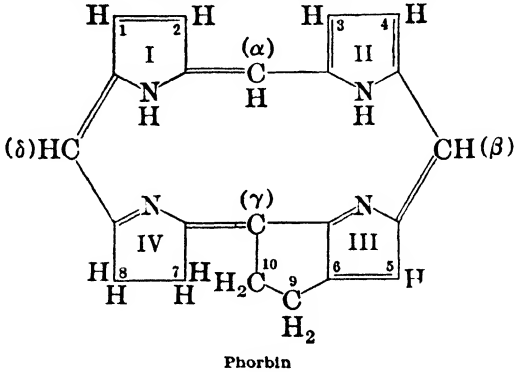
$$4\left(CH_2=\underset{\displaystyle CH_3}{\underset{|}{C}}—CH=CH_2\right) + H_2O + 3H_2 \rightarrow C_{20}H_{39}OH$$

[9] A. Stoll and E. Wiedemann, *Naturwissenschaften*, **20**, 889 (1932).
[10] G. Mackinney and M. A. Joslyn, *J. Am. Chem. Soc.*, **62**, 231 (1940).
[11] S. Ruben, A. W. Frenkel, and M. D. Kamen, *J. Phys. Chem.*, **46**, 710 (1942).
[12] F. G. Fischer and K. Löwenberg, *Ann.*, **464**, 69 (1928); **475**, 183 (1929).

It is impracticable to discuss in detail the many degradation products of the chlorophylls which have been isolated and identified. However, certain of the groupings are of especial interest, and it will suffice to indicate the relationships of these.

7,8-Dihydroporphin is the nucleus of the *chlorin* series of compounds derived from chlorophyll *a*. The *rhodins* are the corresponding derivatives of chlorophyll *b*, *i.e.*, they differ from chlorins in having a formyl group on carbon-3.

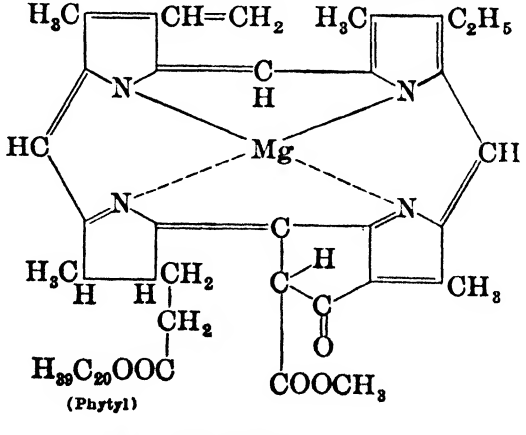
With the introduction of the isocyclic ring between the 6- and γ-positions of the chlorin nucleus *phorbin* is produced.



Phorbin

The prefix *pheo* designates those compounds which have the same substituents found in the chlorophyll molecule. Thus *pheophorbin a* would be the free dicarboxylic acid derived from chlorophyll *a* by removal of magnesium and hydrolysis of the methyl and phytyl esters (see below).

Chlorophylls *a* and *b* are highly substituted phorbins.



Chlorophyll *a*

Magnesium salt of 1,3,5,8-tetramethyl-4-ethyl-2-vinyl-9-oxo-10-carbomethoxy-phorbin-7-propionic acid phytyl ester

Chlorophyll *b* differs from *a* in having a formyl group replace the methyl on carbon-3, giving it two less hydrogens and one more oxygen. This fact has led to many attempts to show that chlorophyll *a* is oxidized to chlorophyll *b* when carbonic acid is reduced to CH_2O by photosynthesis. Such a conversion *in vivo* has not been proved, and the *in vitro* transformation is one of great difficulty.[9]

The bonding of the magnesium is obviously not fixed. It may best be considered to be located at the center of the molecule without primary bonds to any one of the four nitrogens. As resonance within the nucleus takes place its primary bonds could be directed to any two of the unsaturated rings.

Magnesium is readily replaced by hydrogens or other metals (copper, zinc, iron) in the presence of weak acids. Mackinney and Joslyn [10] have shown that the loss of magnesium in 90 per cent acetone with oxalic acid is of first order with respect to acid and probably with respect to chlorophyll. Chlorophyll *a* reacts 7 to 9 times more rapidly than chlorophyll *b*.

Using radioactive magnesium as a tracer, Ruben, Frenkel, and Kamen [11] were unable to detect the interconversion of chlorophylls *a* and *b* during photosynthesis. In some *in vitro* studies they found that unpurified extracts of mixed chlorophylls took up radioactive magnesium and copper, whereas the separated highly purified components of chlorophyll do not exchange magnesium ion in 80 per cent acetone.

Magnesium may be put back into *pheophytin a* by the use of $MgCH_3I$ (Grignard reagent) in ether to give chlorophyll *a*.

Nearly one-third of the molecular weight of chlorophyll is due to the phytyl group which can be reversibly replaced in the molecule by the enzyme *chlorophyllase*. Phytol has been synthesized by Fischer and Löwenberg,[12] who have shown that it is 3,7,11,15-tetramethyl-2-hexadecen-1-ol.

$$CH_3—\underset{\displaystyle CH_3}{\underset{|}{CH}}—(CH_2)_3—\underset{\displaystyle CH_3}{\underset{|}{CH}}—(CH_2)_3—\underset{\displaystyle CH_3}{\underset{|}{CH}}—(CH_2)_3—\underset{\displaystyle CH_3}{\underset{|}{C}}=CH—CH_2OH$$

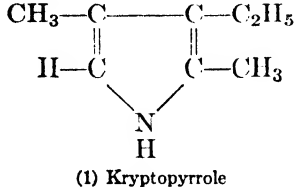
Earlier workers suggested that phytol might be related to isoprene, the hydrocarbon of rubber and terpenes, four molecules of isoprene condensing with one molecule of water, followed by a reduction:

$$4\left(CH_2=\underset{\displaystyle CH_3}{\underset{|}{C}}—CH=CH_2\right) + H_2O + 3H_2 \rightarrow C_{20}H_{39}OH$$

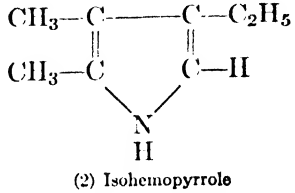
[9] A. Stoll and E. Wiedemann, *Naturwissenschaften*, **20**, 889 (1932).
[10] G. Mackinney and M. A. Joslyn, *J. Am. Chem. Soc.*, **62**, 231 (1940).
[11] S. Ruben, A. W. Frenkel, and M. D. Kamen, *J. Phys. Chem.*, **46**, 710 (1942).
[12] F. G. Fischer and K. Löwenberg, *Ann.*, **464**, 69 (1928); **475**, 183 (1929).

Phytol is readily absorbed [13] from the intestinal tract by the white rat, but its fate in the animal organism is unknown. It has assumed wider biological importance since it was shown to be a part of the vitamin K molecule. It may also be condensed with trimethylhydroquinone to give α-tocopherol (vitamin E).

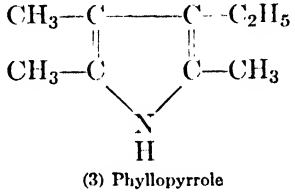
The reduction of chlorophyll yields certain substituted pyrrole rings. The following have been identified:



(1) Kryptopyrrole



(2) Isohemopyrrole



(3) Phyllopyrrole

All three are derived from both hemin (the red pigment of hemoglobin) and chlorophyll.

If one compares the formula of chlorophyll *a* with the formula of hemin (p. 396), it will be observed that there is a very striking similarity in the structure of these two pigments, the one a vital pigment of the autotrophic plants, the other a vital pigment of most representatives of the animal kingdom. The same porphin nucleus occurs in both pigments, and the groups which are attached to the pyrrole rings show a great similarity. There is, however, the fundamental difference in properties due to the fact that hemin is a substituted porphin whereas chlorophyll is a substituted phorbin. It is highly improbable that the close similarity in the structure of these two vital pigments is one of chance. It appears more probable that, in the processes involved in organic evolution, the essential nucleus of the earlier vital pigment, chlorophyll, became modified so as to assume new functions in the developing animal kingdom, the *magnesium* in the chlorophyll being replaced with *iron* in order to care for the new function as an oxygen carrier, the branched-chain aliphatic alcohol, phytol, being similarly replaced by a

[13] H. J. Channon and G. A. Collinson, *Biochem. J.*, **22**, 391 (1928).

protein residue (a histone, globin), possibly because the animal body cannot synthesize such compounds as phytol but can reconstruct a protein molecule from the amino acids which are secured from the food.

Considering *pheophorbin a* the central structure, we have the following relationships:

1. Esterify the 10-carboxylic acid with methanol to form *pheophorbide a*, $C_{32}H_{32}N_4O\begin{pmatrix}COOCH_3\\COOH\end{pmatrix}$.

2. Esterify pheophorbide *a* with phytol to form *pheophytin a*, $C_{32}H_{32}N_4O\begin{pmatrix}COOCH_3\\COOC_{20}H_{39}\end{pmatrix}$.

3. Replace two active hydrogens of pheophytin *a* with magnesium to form chlorophyll *a*. Hence chlorophyll *a* may be given the more descriptive name magnesium-phytyl-methyl-pheophorbin *a*.

4. Hydrolyze phytol from chlorophyll *a* with the enzyme chlorophyllase in aqueous acetone to form *chlorophyllid a*, $MgC_{32}H_{30}N_4O\begin{pmatrix}COOCH_3\\COOH\end{pmatrix}$. If the reaction is carried out in ethyl alcohol ethylchlorophyllid *a* results.

5. Hydrolyze chlorophyllid *a* with cold alkali to form *chlorophyllin a*, $MgC_{32}H_{30}N_4O\begin{pmatrix}COOH\\COOH\end{pmatrix}$. Hot saponification produces isochlorophyllins. Prolonged hot saponification opens the isocyclic ring to give the magnesium-free tricarboxylic acid chlorine e_6 (phytochlorine *e*). The free chlorophyllin is unstable and has been studied largely as the potassium salt.

To summarize, the phyllins, phyllides, and chlorophylls contain magnesium, and the phorbins, phorbides, and phytins are magnesium-free compounds.

The *porphyrins* are substituted porphins, *i.e.*, all four rings are unsaturated and can enter equally into the resonance system. Protochlorophyll is a porphyrin. Fischer's reagent, hydrogen iodide in glacial acetic acid, changes many chlorophyll derivatives to porphyrins. Pheophorbide *a* yields pheoporphyrin a_5-monomethyl ester. Carbon-7 and carbon-8 are dehydrogenated, and the vinyl group is replaced by an ethyl.

Meso compounds have an ethyl group substituted for the vinyl on carbon-2.

Purpurins are chlorines, *i.e.*, their systematic names describe them as substituted chlorine.

Allomerized chlorophyll results from an oxidative reaction in which two equivalents of oxygen are taken up by each chlorophyll molecule. Carbon-10 seems to be a definite point of oxidation.

Photosynthesis. Many hypotheses have been proposed to account for the photosynthesis of carbohydrates through the agency of light and chlorophyll. One of the more recent of these is that of Stoll [14] who postulates a series of five consecutive reactions:

1. The combination of carbonic acid with chlorophyll to form a chlorophyll-carbonic acid complex.
2. The transfer of the carbonic acid to hydrogen acceptors. *This is photochemical reaction I* and is a mechanism similar to a peroxidative rearrangement.
3. The dehydrogenation of the chlorophyll with a stepwise reduction of the carbonic acid.
4. The cleavage of the water molecule attached to the chlorophyll, $H_2O \rightarrow H + OH$, $2OH \rightarrow H_2O_2$, whereby the chlorophyll again becomes saturated with hydrogen. *This is photochemical reaction II.*
5. The removal by leaf catalase of the hydrogen peroxide formed in reaction 4. This is a reaction dependent on temperature, and it results in the liberation of oxygen.

Many experiments have been done during the past fifteen years which either support or deny one or more of the above assumptions. The tendency has been to be less specific in composing theories of photosynthesis.

Absorption of light is the one function that can be definitely assigned to chlorophyll. Eight quanta [15] seem to be the minimum required for the reaction

$$CO_2 + H_2O \xrightarrow{8Nh\nu} CH_2O + O_2$$

in place of the four quanta that had been estimated from earlier experiments. Since the minimum energy requirement for the above equation is about 112,000 cal. the highest efficiency yet attained is about 30 per cent. In full sunlight, outdoor plants are not more than 2–3 per cent efficient. No distinction is now made between the two chlorophylls in this function.

[14] A. Stoll, *Naturwissenschaften*, **20**, 955 (1932).

[15] The energy of the quantum, or photon, is given by $h\nu$, where h is Planck's constant, 6.626×10^{-27} erg-sec., and ν is frequency of the radiation. Multiplying by Avogadro's number, the molar quantum value is $Nh\nu$. This ranges from about 70,000 cal. for violet light to 40,000 cal. for the red.

The mechanism by which the chlorophyll applies absorbed energy to the reduction of CO_2 is entirely unknown. The number of proposed photochemical steps has been increased to eight or more. Molecules of unspecified nature act as hydrogen donors when activated by energy supplied through illuminated chlorophyll. Franck and Gaffron [16] in discussing the eight-step reaction theory of Franck and Herzfeld point out that "four reactions are supposed to belong to the type in which a molecule contained in the complex at the chlorophyll acts as a hydrogen donor and four others are of the type wherein the dehydrogenated donor gets the hydrogen atom back indirectly from the water. The four last-mentioned reactions produce peroxide radicals which act as the source of the oxygen. One immediately sees a close connection to Willstätter's idea, to van Niel's, and others." That the excited chlorophyll molecule itself may act as the hydrogen donor is supported by studies of Franck and Livingston. Since there is evidence that the dehydrogenated chlorophyll (monodehydrochlorophyll) has not changed its visible absorption spectrum, it follows that the bond affected is not a part of the main conjugated system. For a detailed theory of photosynthesis see Franck and Herzfeld,[17] and for a very complete review of the whole problem of photosynthesis see Rabinowitch.[2]

THE CAROTENOIDS

The carotenoids are those light-yellow to deep-red (sometimes even violet or purple) nitrogen-free polyene pigments of the plant and animal kingdoms which can be extracted, by means of fat solvents, from the tissues in which they occur.

In 1930, practically nothing was known of the structural chemistry of the carotenoids. In the last fifteen years, owing primarily to the work of Karrer, Kuhn, Zechmeister, and a number of other workers, the structures of a very considerable number of the carotenoids have been definitely ascertained. Today the number has increased several fold primarily as a result of the extensive use and improvements in Tswett's chromatographic technic. The single recognized leaf xanthophyll of twenty-five years ago has now been shown to be a mixture of twelve or more pigments in some plant species (Strain). Figure 115 shows the striking separation that can be made from leaf extracts. In no field of chemistry has chromatographic analysis been more successfully applied.

[16] J. Franck and H. Gaffron, *Advances in Enzymol.*, **1**, 199 (1941).
[17] J. Franck and K. F. Herzfeld, *J. Phys. Chem.*, **45**, 978 (1941).

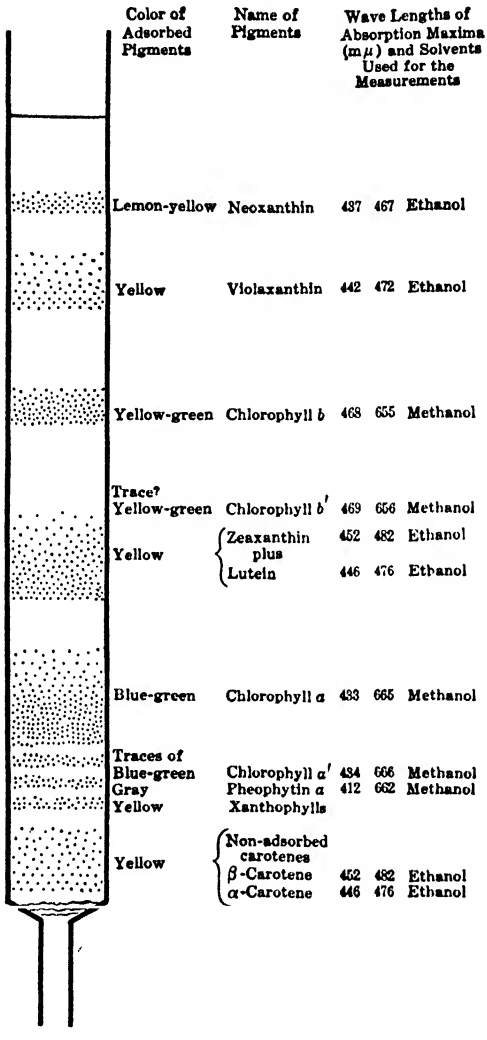
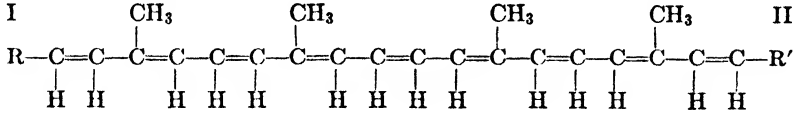


FIG. 115. Pigments from 1 gram of leaves separated by adsorption in a column of powdered sugar. (Courtesy of H. H. Strain.)

Freshly cut, tender leaves were ground with sand and methanol. Pigments in the filtered methanol extract were transferred to petroleum ether, which was then drawn into the column of dry powdered sugar (3.2 by 24 cm.) with the aid of suction (about 0.5 atmosphere). The adsorbed pigments were washed with petroleum ether containing 0.5 per cent propanol to yield the chromatogram represented by the figure. All these operations required only about 1 hour.

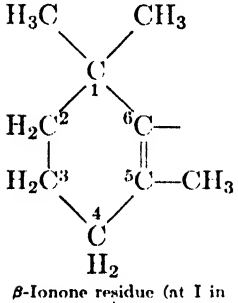
Nonadsorbed carotenes from the column may be separated quickly by readsorption in columns prepared from a mixture of activated magnesia and Celite using petroleum ether plus 4 per cent acetone as solvent. Lutein and zeaxanthin may be separated completely in columns of magnesia plus Celite by use of petroleum ether plus 25 per cent acetone as solvent.

Carotenoids may be hydrocarbons, alcohols, aldehydes, ketones, or acids. Their central structure is a methylated 18-carbon conjugated polyene radical ($C_{22}H_{26}$)

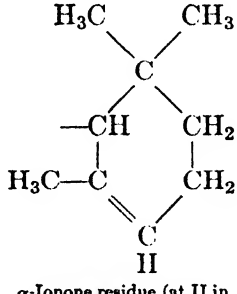


in which the four methyl groups are in the 1,5- or 1,6-position with respect to each other. This middle section never contains oxygen. No uniform system of numbering the carbons has been adopted.

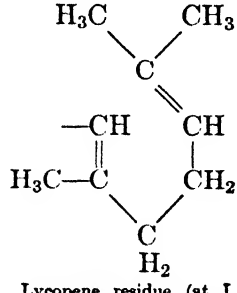
The carotenoids differ according to the nature of the groups R and R′, each of which usually contains nine carbons. Their basic hydrocarbon structures are given below along with the position in which they occur in some well-known carotenes. The marked effect produced by



β-Ionone residue (at I in α-carotene, in γ-carotene, in vitamin A, at I and II in β-carotene, at II in cryptoxanthin)



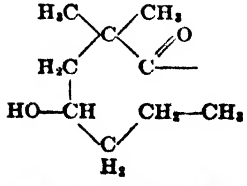
α-Ionone residue (at II in α-carotene and δ-carotene)



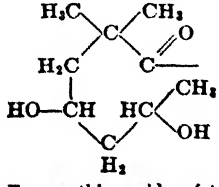
Lycopene residue (at I and II in lycopene, at I in δ-carotene, and at II in γ-carotene and in rubixanthin)

substituting the α-ionone for the β-ionone residue is largely due to the removal of one double bond from the conjugated system. Note that the lycopene residue is not a closed ring. Hence, lycopene, which has this group in both positions I and II, is aliphatic.

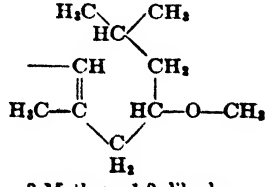
Some of the oxygen derivatives of these three groups are indicated below, and their position in well-known compounds is given.



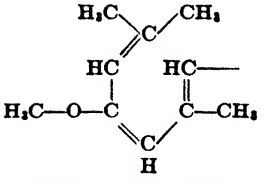
Desoxyfucoxanthin residue (at I and II in capsorubin and at II in capsanthin)



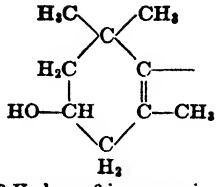
Fucoxanthin residue (at I and II in fucoxanthin)



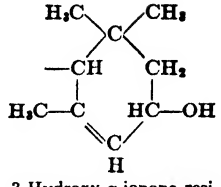
3-Methoxy-1,2-dihydrolycopene residue (at II in rhodoviolascin)



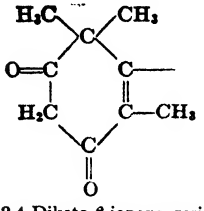
3-Methoxy-Δ-3,4-dehydrolycopene residue (at I in rhodoviolascin)



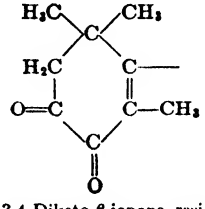
3-Hydroxy-β-ionone residue (at I in capsanthin, rubixanthin, cryptoxanthin, lutein, and helenien, and at I and II in physalien and zeaxanthin)



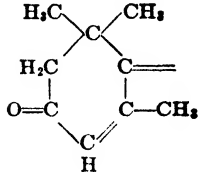
3-Hydroxy-α-ionone residue (at II in lutein and helenien)



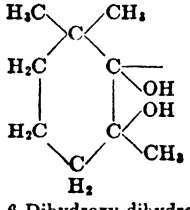
2,4-Diketo-β-ionone residue (at I and II in euglenarhodon)



3,4-Diketo-β-ionone residue (at I and II in astacene)

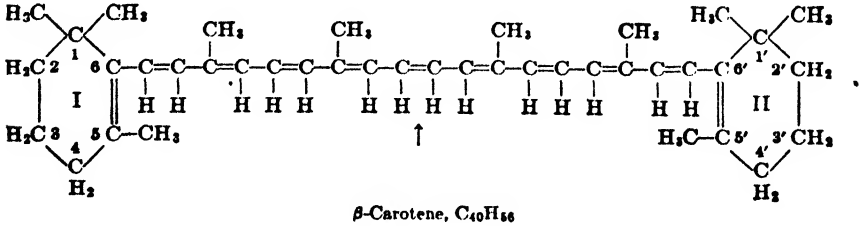


3-Keto-dehydro-β-ionone residue (at I and II in rhodoxanthin)

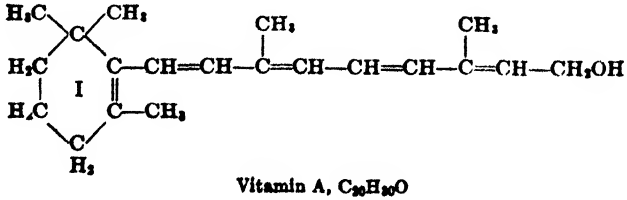


5,6-Dihydroxy-dihydroionone residue (in azafrin)

The relationship of β-carotene to vitamin A is shown in the following formulas:

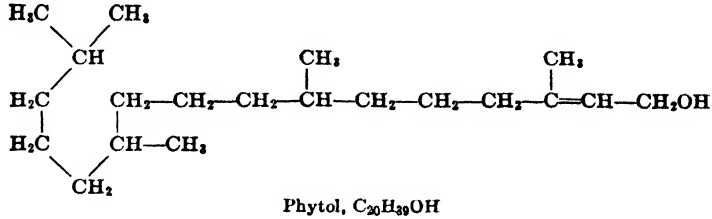


β-Carotene, $C_{40}H_{56}$



Vitamin A, $C_{20}H_{30}O$

When the structural formula of phytol is written with a carbon arrangement similar to the carbon arrangements in the carotenoid molecules, there is a striking resemblance to the carotenoid structure.



Phytol, $C_{20}H_{39}OH$

Nomenclature. No uniform nomenclature for the carotenoids has been adopted. The system proposed by Palmer [18] has not received wide recognition. It has the disadvantage that his α- and β-carotene residues are likely to be confused with α- and β-carotene already so widely implanted in the literature. Strain [1] regards as *carotenes* all hydrocarbons which are essentially aliphatic and which owe their color to a conjugated system of double bonds. Hydroxy, oxo, and hydroxy-oxo derivatives of the carotenes are named *xanthophylls*. Carboxy derivatives of carotenes are *carotenoid acids*. The entire group of pigments and their esters are *carotenoids*. He also follows the practice of naming many of the xanthophylls from some special property or their source, *e.g.*, *cryptoxanthin* (hidden pigment) and *zeaxanthin* (from *Zea mays*).

Mayer [1] lists only the alcohols under xanthophylls but that is not in keeping with long-established usage and would require changing the names of many ketonic xanthophylls. Use of the term *carotenol* to designate the hydroxy derivatives is satisfactory but leaves the ketones and hydroxyketones without designations. Therefore in Table 68 some of the naturally occurring carotenoids are grouped according to their oxygen content.

In addition to the naturally occurring pigments there is now a long list of *neo* compounds produced by isomerization in solution (see below).

There are also colorless fluorescent polyenes widely distributed in plants. Strain, in 1936, adsorbed one or more of these from green leaves and etiolated seedlings on columns of magnesia. He found the entire etiolated seedlings to be strongly fluorescent in ultraviolet light. Recently Zechmeister and Sandoval [19] have listed a large number of plants in which a fluorescent pigment, *phytofluene*, occurs. It is said to differ from the material Strain discovered and is classed as a colorless C_{40} polyene with an isoprenic structure, a molecular weight of about 500,

[18] L. S. Palmer, *Science*, **79**, 488 (1934).

[19] L. Zechmeister and A. Sandoval, *Arch. Biochem.*, **8**, 425 (1945).

and sharp spectral maxima at 367, 348, and 311 mμ (in petroleum ether). Animals fed tomato paste rich in phytofluene deposit a small portion of the total in the liver (rabbits) or in the yolk of eggs (hens). Most of it, however, is rapidly destroyed in the body.[20]

TABLE 68. SOME NATURALLY OCCURRING CAROTENOIDS

Common Name		*Typical Sources*
	HYDROCARBONS, $C_{40}H_{56}$	
α-Carotene	.....	Carrots, palm oil, chestnuts, mountain ash berries
β-Carotene	.....	Carrots, paprika, green leaves, grass, etc.
γ-Carotene	.....	Apricots, fruits of *Convallaria majalis* and *Gonocaryum pyriforme*
δ-Carotene	.....	Fruits of *Gonocaryum pyriforme*
ε-Carotene	.....	Diatoms
ζ-Carotene	.....	Certain strains of tomatoes
Lycopene	.....	Fruits of tomato (*Lycopersicum esculentum*), watermelon (*Cucumis citrullus*), etc.
	ALCOHOLS	
Cryptoxanthin	$C_{40}H_{56}O$	Fruits of *Carica papaya*, yellow *Zea mays*, *Capsicum* sp., etc.
Rubixanthin	$C_{40}H_{56}O$	Rose hips, *Rosa* sp.
Lutein	$C_{40}H_{56}O_2$	Green leaves, many flowers
Zeaxanthin	$C_{40}H_{56}O_2$	Yellow *Zea mays*
Flavoxanthin	$C_{40}H_{56}O_3$	Buttercups (*Ranunculus acer.*)
Violaxanthin	$C_{40}H_{56}O_4$	Yellow pansies (*Viola tricolor*)
	KETONES	
Rhodoxanthin	$C_{40}H_{50}O_2$	Fruit of the yew (*Loxus baccata*), leaves of *Potamogeton*, etc.
Astacin	$C_{40}H_{48}O_4$	Lobster shells
	HYDROXYCARBONYLS	
Capsanthin	$C_{40}H_{58}O_3$	Red peppers (*Capsicum annum*)
Fucoxanthin	$C_{40}H_{54-60}O_6$	Marine alga (*Fucus vesiculosus*)
	CARBOXYLIC ACIDS	
Crocetin	$C_{20}H_{24}O_4$	Saffron (*Crocus sativus*)
Bixin	$C_{25}H_{30}O_4$	Pods of *Bixa orellana*
Azafrin	$C_{27}H_{38}O_4$	Roots of *Escobedia scabrifolia*

The quantity of a given carotenoid in a genus varies widely among species and varieties. The results of an interesting cooperative program

[20] A. Sandoval, E. R. Meserve, H. J. Deuel, Jr., and L. Zechmeister, *Arch. Biochem.*, **11**, 373 (1946).

for breeding tomatoes of higher provitamin A content have been reported recently by Porter and Zscheile.[21] The quantity of lycopene varies from traces in yellow and green fruited varieties to 300–400 γ/g. of fresh red fruit. β-Carotene has been found in concentrations ranging from traces to 130 γ/g. of fresh fruit. α-Carotene is not present in most tomatoes, but in a few experimental selections as much as 15 γ/g. of fruit have been observed. ζ-Carotene, first prepared by Strain from carrots, is peculiar to several selections of tomatoes, the best having 60–70 γ/g. of fresh fruit. This carotenoid is biologically inactive and seems to be related to lycopene.

The Carotenoids of the Tissues of Higher Animals. Palmer and Eckles were largely responsible for demonstrating that the so-called "lypochromes" of animal tissues are in reality carotenoids which originated in the food and which have been deposited in the fat and adipose tissues of the animal. Thus, they showed that the yellow pigment of butterfat is almost wholly carotene and that both carotene and "xanthophyll" occur in the fat of human milk. Carotene admixed with some "xanthophylls" was found by them to be the pigment in the adipose tissue of cattle, whereas they found the pigment of the yolk of hen's eggs to be almost exclusively "xanthophyll."

Palmer, and Palmer and Eckles, have also demonstrated that the carotene content of the cow's tissues, as well as that secreted in the milk fat, is determined by the carotene content of the ration. When foods rich in "xanthophyll," such as yellow maize, were fed, no appreciable pigmentation of the animal tissues or of the milk secretion took place, indicating that the xanthophyll and zeaxanthin in the food were not transported unchanged through the blood stream to the various tissues, in contrast to the behavior of carotene. Similar experiments by Palmer, and Palmer and Kempster, showed that "xanthophyll-rich" rations increased the pigmentation of egg yolk, whereas carotenoid-poor rations or rations in which carotene was the only carotenoid present caused the production of carotenoid-free egg yolks. The "xanthophyll-rich" rations produced a rapid coloration in all parts of the body of the white leghorn fowl, whereas carotene-containing foods had practically no effect upon the coloration of the bird's tissues. Palmer and Kempster were able to rear chickens to maturity on diets which were devoid of the carotenoids that are normally present in the adipose tissue of fowls. The eggs of such chickens possessed carotenoid-free yolks, but this fact did not impair their fertility, nor were the young chickens hatched from such eggs inferior to those hatched from normally pigmented eggs. The success of the experiments of Palmer and Kempster is explained by the

[21] J. W. Porter and F. P. Zscheile, *Arch. Biochem.*, **10**, 537 (1946).

fact that they included an aequate amount of (colorless) vitamin A (pork liver) in their rations.

The yellow pigment of the corpus luteum is almost wholly β-carotene. The human placenta contains both carotene and xanthophyll. The adrenal glands contain carotene both in the medulla and in the cortex, and we have already indicated (*cf.* Chapter 16) that a carotenoid protein constitutes the visual purple of the retina. Recently Wald [22] extracted from human retinas a xanthophyll apparently identical with lutein or leaf xanthophyll itself.

Astacin, the carotenoid of the Crustacea, crystallizes from aqueous pyridine in violet needles which possess a metallic luster. It forms a blood-red solution in pyridine which is orange-red when dilute. It apparently exists in the carapace of the Crustacea in the form of a chromoprotein or perhaps adsorbed upon protein. When the protein is denatured with heat or with alcohol, the red astacin is liberated from the brownish green chromoprotein. This accounts for the change in color when lobsters are boiled.

Identification and Separation of Carotenoids. Tswett, in 1906, introduced a special adsorption technic, or rather adsorption analysis method, which he called chromatographic analysis, and this technic has been used almost universally in the separation of the various carotenoids. It consists essentially of the adsorption of the carotenoids upon calcium carbonate, calcium hydroxide, gypsum, magnesium oxide, aluminum oxide, or other finely divided powders. The powder is packed uniformly in a glass tube, and the carotenoid-containing solution is allowed to percolate through the mass of packed powder. Those carotenoids or other substances which are most strongly adsorbed by the solid powder form a colored band or zone near the top of the powder column. Under proper conditions of concentration of solution, etc., this upper band or zone is followed by a band of the powder free from adsorbed constituents. If only one compound is present in the solution being passed through the powder, only a single uniform adsorption band will form. If the solution being passed through contains several components, the chromatograph will consist of a series of bands somewhat similar to that shown in Fig. 115. After the desired constituents are adsorbed, the column of powder is removed from the tube and separated into the various fractions by cutting the column between the colored bands. The pigment is then eluted, *i.e.*, washed from the powder with alcohol or another strongly polar solvent, and usually readsorbed until it is definitely certain that the eluted fraction consists of a single component. This fraction is then eluted and studied by appropriate technics. In many

[22] G. Wald, *Science*, **101**, 653 (1945).

cases it is essential that the adsorption and elution, as well as the subsequent study of the carotenoid, be carried out in the absence of oxygen, since many of the carotenoids are very subject to oxidation. Zechmeister and Cholnoky and Strain,[1] have outlined in detail the chromatographic technic. The relative positions which organic compounds take on the column depend on the conditions used. Strain [23] has recently published very interesting tables showing the change of adsorption sequences with various combinations of adsorbents and solvents. The pigments used were chlorophyll *a* and *b*, fucoxanthin, violaxanthin, and zeaxanthin.

***Cis-Trans* Isomerization of the Carotenoids.** Zechmeister [1] has discussed in detail the stereochemistry of carotenoids and diphenyl polyenes. The first rearrangement of a carotenoid, bixin, observed by Herzig and Faltis in 1923, was interpreted by Karrer and associates as a *cis-trans* isomerization. In the years 1935–1937 Gillam and associates discovered and studied the new zone on an adsorption column which appeared on elution and readsorption of a sample of β-carotene. This zone was due to a new carotenoid, called "pseudo-α-carotene," and they attributed the change to a rearrangement which took place in the column under the influence of the adsorbent. They showed complete reversibility of the reaction.

$$\beta\text{-Carotene} \rightleftharpoons \text{pseudo-}\alpha\text{-carotene}$$

Zechmeister and Tuzson in 1938 observed that the formation of a new zone from lycopene depended on the time it was kept in solution, rather than on contact with the adsorbent. Strain, in the same year, reported that heat alters xanthophyll pigments in such a way as to produce complex chromatograms. It is now known that these are *cis-trans* rearrangements which take place slowly in solution at room temperature, can be brought to equilibrium in a few hours at elevated temperatures, are catalyzed by iodine and light, and are completely reversible. The compounds formed, called *neo*-carotenoids, are now known for a large number of the natural pigments. The amount of neo-carotenoid found in any extract is proportional to the amount of the stable form. The new adsorption band may be either above or below that of the natural pigment.

A pigment designated by its current name (β-carotene, lycopene, etc.) is usually considered the all-*trans* form in which each double bond possesses *trans*-configuration. The neo-isomers contain *cis*-configurations about some of the double bonds. Zechmeister gives the following formulae for calculating the number of possible isomers: For unsym-

[23] H. H. Strain, *Ind. Eng. Chem., Anal. Ed.*, **18**, 605 (1946).

metrical chain with n effective double bonds the formula is $N = 2^n$. For symmetrical chain and n odd the formula is

$$N = 2^{(n-1)/2} \cdot (2^{(n-1)/2} + 1)$$

but when n is even it becomes

$$N = 2^{(n/2)-1} \cdot (2^{n/2} + 1)$$

However, the methyl groups along the chain introduce steric hindrance which limits the number of likely centers of rearrangement, and it is stated that the number of double bonds available for isomerization in the natural carotenoids is only 4 to 7 and the calculated number of stereoisomers varies between 10 and 128. β-Carotene has 20 isomers instead of 1,056. These include the natural all-*trans*-β-carotene, 3 mono-*cis*-, 6 di-*cis*-, 6 tri-*cis*-, 3 tetra-*cis*-, and 1 all-*cis*-β-carotene. Spatial configuration has a marked effect on biological activity (see Table 69, Chapter 36) and on spectral quality. The latter property is used in conjunction with chromatographic properties to identify the isomers.

Absorption Spectra of the Carotenoids. The chief absorption bands of most of the carotenoids lie between 400 and 500 mμ. There is close agreement among workers on the positions of the band maxima in a given solvent. This is not true of the absolute intensity of the bands whose values vary by more than can be attributed to the instrumental error. The absorption spectra of α- and β-carotene and lycopene are compared in Fig. 116.

Carotenoid studies afford a very striking illustration of how a pure science study, which apparently has no practical importance, may, in an emergency, be found to be of practical use. Before World War I the absorption spectra of the carotenoids and of chlorophyll had been exhaustively studied and mapped. During the war it became necessary for those in charge of the American troops on the western front to devise some means for the detection of enemy troop movements which were being carried out under camouflages so designed as to simulate green foliage. Natural green foliage has a characteristic reflection spectrum not possessed by the green paints used, although, to the eye of an observer in an aeroplane, such camouflaged roads would appear to merge with the green of the surrounding fields.

The Eastman Kodak Company devised a light filter capable of screening out part of the green rays characteristic of chlorophyll but permitting the wide bands in the red at about 700 mμ and a narrow band in the green at about 500 mμ to pass through the filter. When a landscape is viewed through such a filter, the natural vegetation appears more or less bright red or orange-red instead of green, as though it were

on fire. A strip of green paint still appears green when viewed through such a filter. Accordingly, when a landscape was observed from an aeroplane to have the appearance of a prairie fire and a thin green line appeared across such a landscape, it was easy to determine exactly where to drop the bombs.

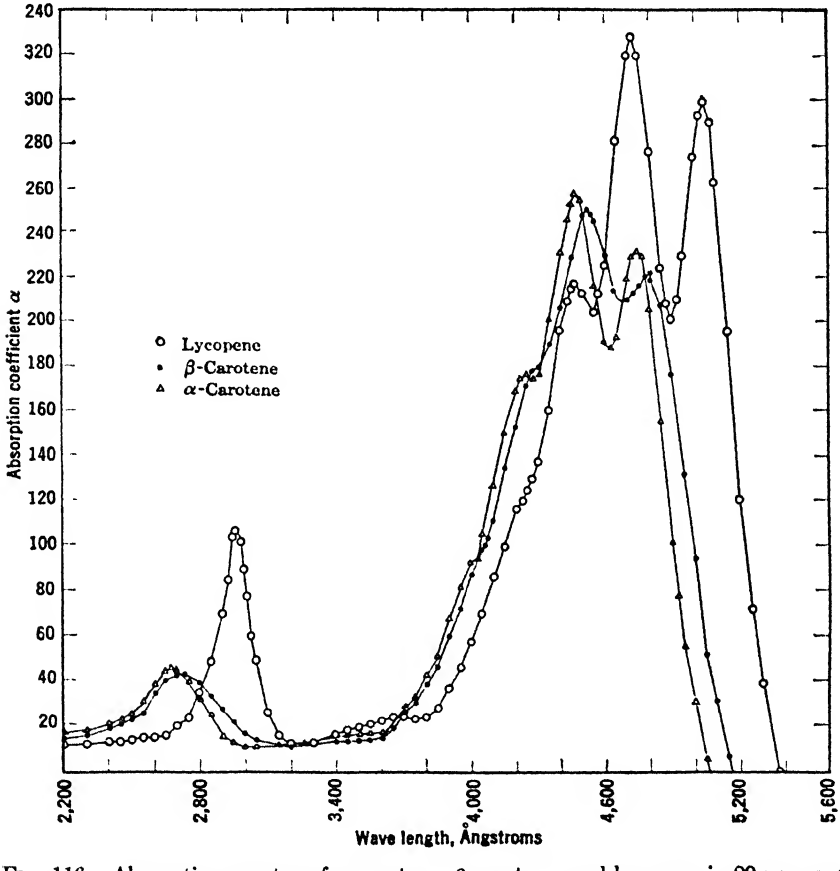


FIG. 116. Absorption spectra of α-carotene, β-carotene, and lycopene in 80 per cent ethanol and 20 per cent diethyl ether. [From E. S. Miller, *Plant Physiol.*, **12**, 667 (1937).]

No one could have predicted that research dealing with the absorption spectra of chlorophyll and the carotenoids would ever be turned to practical use, and it is possible that many of the pure science studies which have been or are being carried out in other fields may have similar or even greater practical importance.

Role of Carotenoids in Plants. There is no general agreement regarding the function of the carotenoids in plants. The fact that they

invariably occur associated with chlorophyll in the chloroplasts permits the assumption that they play some role in the photosynthetic process.

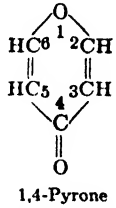
It has recently been found that light absorbed by the carotenoids can be utilized for photosynthesis. This conclusion was reached by careful measurements of quantum yields. This may be their more important function; or it may be that they act as an oxidation-reduction system and are concerned with hydrogen transference or oxygen assimilation. Willstätter and Stoll suggested that one possible function of the carotenoids might be to control an equilibrium between chlorophyll *a* and chlorophyll *b*. However, this general viewpoint was abandoned when it was found difficult or impossible to bring about interchange of carotene and xanthophyll. Other suggestions of physiological function have been made, but as yet there is no proof that they play a major role in any single physiological mechanism.

CHAPTER 35

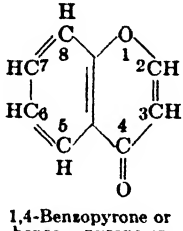
The Flavones, Xanthones, and Anthocyanins

The natural pigments containing heterocyclic oxygen atoms have been given a monographic treatment by Mayer-Cook.[1] In addition to the major groups (flavones, xanthones, and anthocyanins) there is a small group of lichen pigments which contain a five-membered ring. The most widely distributed of these is *usnic acid*, which crystallizes as yellow needles and has the formula $C_{18}H_{16}O_7$. The complete structure is not known but the central nucleus is a *cumaron* group.

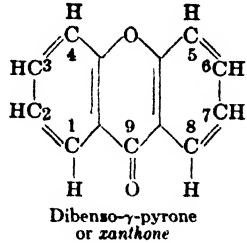
The Flavones (Latin, *flavus* = yellow) **and the Xanthones** (Greek, $\chi\alpha\nu\theta o\sigma$ = yellow). These naturally occurring plant pigments may be regarded as derivatives of the 1,4- or γ-pyrone nucleus.



1,4-Pyrone
or γ-*pyrone*



1,4-Benzopyrone or
benzo-γ-pyrone or
chromone



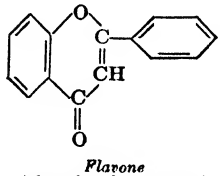
Dibenzo-γ-pyrone
or *xanthone*

The keto group in these compounds does not behave as a typical carbonyl group, *e.g.*, the oxygen is not replaceable by hydroxylamine to form oximes. In addition, the ring oxygen is *basic* and adds acid as though it were *quadrivalent* oxygen, resulting in the formation of oxonium salts. The substitution of a hydroxyl group in position 3 in the chromone nucleus forms *chromonol.*

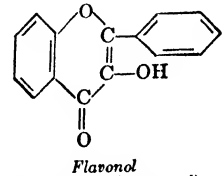
Coloring matters of the chromone and chromonol type have not as

[1] F. Mayer, *The Chemistry of Natural Coloring Matters*, translated and revised by A. H. Cook, American Chemical Society Monograph 89, Reinhold Publishing Corp., New York, 1943.

yet been found in nature, but the benzene derivatives are the *flavones* or *flavonols*. Almost all the derivatives of flavonol are yellow dyes.



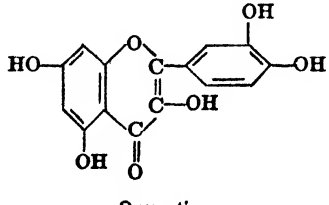
Flavone
(phenylated chromone)



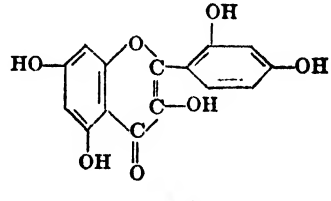
Flavonol
(phenylated chromonol)

Watson [2] has brought together a considerable part of our knowledge in regard to those groups which produce color in an organic compound and those auxiliary groups which convert a colored compound into a pigment or a dyestuff. Those groups which produce color are called *chromophore* groups. Typical chromophore groupings are the nitro group (—NO_2), the azo group (—N=N—), the quinone grouping (*ortho* or *para* O=C_6H_4=O), etc. Such groups do not necessarily produce pigments or dyestuffs. For example, azobenzene is not a dyestuff. If, however, we add one or more *auxochrome* (intensifying or modifying) groups, we obtain a dyestuff. Typical auxochrome groups are amino groups, hydroxyl groups, sulfonic acid groups, etc. In plants the organic compound containing the auxochrome groups but containing the chromophore group in a reduced or otherwise altered form is often spoken of as a chromogen. Thus, indoxyl is a colorless *chromogen* which oxidizes to indigo blue or indigotin. The color of the indigotin can be altered by the introduction of auxochrome groups, *e.g.*, Br, NH_2, NO_2, OH, etc., into the benzene nuclei.

Quercetin, the coloring matter of quercitrin extract, is a tetrahydroxy derivative of flavonol. The *morin* of fustic (Cuba wood) differs from quercetin only in the position of the hydroxyl groups on the benzene nuclei.



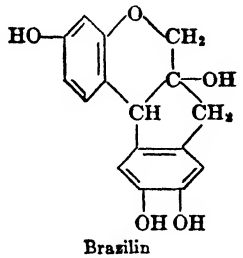
Quercetin



Morin

[2] E. R. Watson, *Colour in Relation to Chemical Constitution*, Longmans, Green and Company, London, 1918.

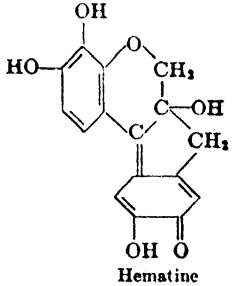
Brazilin, a reduced γ-pyrone derivative, is the coloring principle of hypernic. The constitution was established in 1908 by Sir W. H. Perkin, Jr., who showed it to have the formula



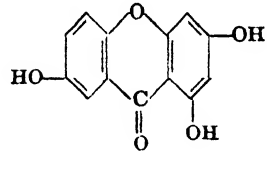
Brazilin

Hematoxylin, from logwood, may be regarded as a molecule of brazilin containing one additional hydroxyl group in position 8. Logwood is the source of an excellent black dye. It "weights" silk to about 200 per cent of the original weight of the silk, and the volume of the silk is enlarged so that the fiber has greater resistance to wear.[3]

By the oxidation of hematoxylin we obtain *hematine*. In the oxidation process an additional chromophore grouping (*p*-benzoquinone) has been introduced into the molecule.



Hematine



Gentisin

The xanthone dyestuffs are not important so far as their natural occurrence is concerned. *Gentisin*, the coloring matter of the gentian root, is a trihydroxyxanthone. Only one or two other natural pigments belong to the xanthone group.

A great deal of our knowledge of the flavones is due to the work of Perkin[4] and those associated with him. The intensity of the color which this group of compounds possess depends markedly on the position of the hydroxyl groups, and as a rule the color is intensified if two

[3] E. S. Chapin, *J. Ind. Eng. Chem.*, **10**, 795 (1918).

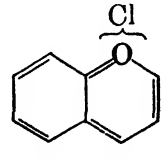
[4] A. G. Perkin and A. E. Everest, *The Natural Organic Colouring Matters*, Longmans, Green and Company, London, 1918.

hydroxyl groups are *ortho* in position to each other. This group of pigments usually occurs in plants as glycosides, one or more hydroxyl groups being combined with a sugar molecule. This has the result that the auxochrome group is rendered inactive so that *in the plant* the flavone glycosides are practically colorless. On hydrolysis of the glycoside, the color develops. Almost all white flowers turn yellow when exposed to the vapors of ammonia, owing to a salt formation of ammonium with the flavones (and possibly in part to hydrolysis of glycosides).

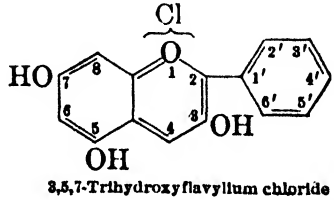
As a rule, the flavone type of pigments does not occur in animal tissues. Thomson,[5] however, isolated a flavone in the study of the pigments of the wings of the butterfly *Melanargia galatea*. The flavone had the properties of quercetin, although it was not positively identified. In a later paper, Thomson [6] noted that the larva of the butterfly feeds upon the grass *Dactylis glomerata*, and that this grass contains a flavone-glycoside, the flavone of which is identical with the pigment isolated from the wings of the butterfly. Apparently, therefore, this pigment of the butterfly wings has a plant and not an animal origin.

The Anthocyanins. The red, violet, and blue pigments present in the blossoms, fruit, and leaves of many plants belong to the group of the anthocyanins. Almost invariably pigments of the above colors belong in this class. All are glycosidic in nature, yielding one or more sugars when boiled with dilute mineral acids. The sugar-free pigments, *aglycones*, are known as *anthocyanidins*. These pigments fall into three major groups, *pelargonidin*, *cyanidin*, and *delphinidin*, which have four, five, and six hydroxyl groups, respectively.

The central nucleus of the anthocyanidins is benzopyrylium chloride (I). 2-Phenyl-benzopyrylium chloride is known as *flavylium* chloride, and with few exceptions all the true anthocyanins are derived from 3,5,7-trihydroxyflavylium chloride (II).



Benzopyrylium chloride
I

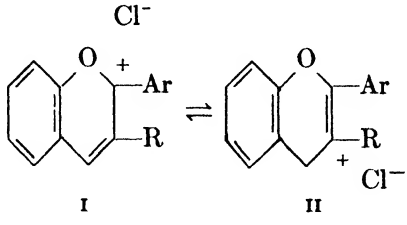


3,5,7-Trihydroxyflavylium chloride
II

Shriner and Moffett [7] present evidence against the above quinoid oxyonium salt structure. They suggest that carbon-2, -3, and -4 of the

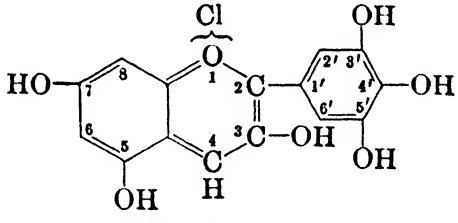
[5] D. L. Thomson, *Biochem. J.*, **20**, 73 (1926).
[6] D. L. Thomson, *Biochem. J.*, **20**, 1026 (1926).
[7] R. L. Shriner and R. B. Moffett, *J. Am. Chem. Soc.*, **63**, 1694 (1941).

heterocyclic ring constitute a mobile allylic system through which the flavylium salts may tautomerize or resonate between structures I and II:



I II

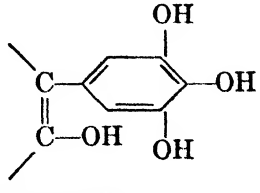
The relation of the anthocyanidins to flavones and flavonols was shown by Willstätter and Mallison, who reduced quercetin with magnesium in alcoholic hydrochloric acid to give pure cyanidin. In a review of the chemistry of the anthocyanins, Robinson [8] states that he and his collaborators have now synthesized all types of anthocyanidins, including the methyl ethers, by convenient methods.



Delphinidin chloride

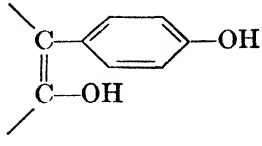
Delphinidin is a violet to blue-black anthocyanidin, obtained from grapes, hollyhocks, petunias, violets, etc. In most instances it occurs as the diglycoside, the sugar residues being apparently attached to the hydroxyl groups on carbon-3 and carbon-5. The violet, *Viola tricolor*, contains this pigment as a rhamnoglycoside. Inasmuch as all the anthocyanins may be regarded as substitution products of the same nucleus, it is unnecessary to repeat all the groupings in order to show the structural formula of the various anthocyanins.

Representing delphinidin as



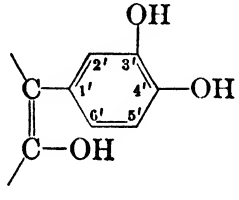
[8] R. Robinson, *Nature*, **135**, 732 (1935).

pelargonidin, the least oxygenated of any of the anthocyanidins becomes



Pelargonidin occurs in many scarlet flowers, scarlet salvia, purple-red asters, etc., as the diglycoside, the residue being attached to the hydroxyl groups on carbon-3 and carbon-5. The diglycoside pelargonin, along with several other related anthocyanin glycosides, has been synthesized by Robinson.[9]

Cyanidin, the anthocyanidin of deep-red dahlias, cornflower, poppies, fruits of cherries, cranberries, currants, mountain ash, etc., contains the grouping



The anthocyanin colors are modified, and the complexity of researches in this field are increased, by the presence of methoxy groups in the molecule. The naturally occurring methyl ethers of the anthocyanidins are *peonidin* (3′-methylcyanidin), *petunidin* (5′-methyl delphinidin), *malvidin* (3′,5′-dimethyl delphinidin), and *hirsutidin* (7,3′,5′-trimethyldelphinidin). Of these, malvidin (also called syringidin and oenidin) is of such frequent occurrence that Robinson suggests that it might be recognized as a fourth fundamental type.

The anthocyanins are very widely distributed throughout the plant kingdom, the majority of the higher plants containing anthocyanin at some stage in their development. It is to anthocyanins that we owe the brilliant colors of our flowers, our foliage plants, and our autumn landscapes. Anthocyanins act as indicators, being red or purple in acid solution and blue or violet in alkali. But in most cases the color of a flower is not a reliable measure of its *p*H. The presence of co-pigments and salts may bring about important color modifications. The coloring of red hydrangea flowers by immersing the stalks in a very dilute solution of ferric chloride is well known.[10]

[9] R. Robinson and A. R. Todd, *J. Chem. Soc.*, 2293, 2299, 2488 (1932).
[10] R. Robinson, *Nature*, **132**, 625 (1933).

A number of our synthetic "coal-tar" dyes, *e.g.*, "rhodamine B" and tetramethylrosamine, are closely allied to the anthocyanin compounds.

Anthocyanin pigments often cause trouble in the canning industry on account of the change in color of the fruit in the canning process or sometimes the actual perforation of the tin cans. In a study of this problem it was found [11] that the color change and the increased corrosion of the metal cans were due to the affinity of the anthocyanins for metals, *e.g.*, tin or iron. The metal salts are essentially insoluble and may precipitate on the inside of the can. The color of the metal salts is "dull" and "muddy," not bright as in the original fruit or juice. The color is shifted toward the violet end of the spectrum. "The reason why fruits of very low acid content but with large amounts of pigment, such as black cherries and blueberries, bring about more extensive corrosion of plain tin or earlier perforation of enameled tin than do more acid, less deeply colored fruits, such as red raspberries or sour cherries, is also clear. In the presence of large amounts of anthocyan, salts of tin with the acids of the fruits can have only momentary existence since they are immediately decomposed with transfer of the tin to combination with the anthocyan," and a liberation of the free acid to attack the metal again. This may continue until all the tin has been removed, the acid being used over and over again. With low anthocyanin and higher acid, the initial attack on the tin may be greater but it soon stops, for the anthocyanin is not present in sufficient amount to act as a reservoir for any large amount of tin.

The function of anthocyanins in the plant is a subject which is still open to speculation. They are apparently rather easily reduced and oxidized, and it has been suggested by many, notably Palladin, that they are respiratory pigments and act as oxygen carriers, assisting in cell oxidations and reductions. This, however, still remains to be definitely proved.

The alpine plants are particularly rich in anthocyanins and lose much of their color when grown at lower levels. One suggestion of a possible function of these pigments is that they may screen out the injurious ultraviolet rays which pass through the thin atmosphere at the higher elevations. The pigment also absorbs strongly the green light transmitted by chlorophyll. Pelargonidin, cyanidin, and delphinidin have their absorption maxima between 500 and 550 mμ, the region of greatest transmission by chlorophyll. Leaves colored with anthocyanins may be as much as 2°C. warmer than other leaves on the same plant which contain only the green pigment and which have the same light exposure.

[11] C. W. Culpepper and J. S. Caldwell, *J. Agr. Research*, **35**, 107 (1927).

Inheritance of Color in Plants. Botanists and geneticists have studied the inheritance of color in plants in order to learn more about the laws governing heredity. Color is easy to see and to follow in the progeny through successive generations. The early workers did not appear to realize that what they were in reality studying was the *inheritance of chemical compounds and the inheritance of chemical reactions* where more than one component was involved. Miss Wheldale [12] (Mrs. Onslow) pointed this out, and later Keeble, Armstrong, and Jones [13] added further details which gave a clear picture of qualitative color inheritance in certain flowers. However, chemical knowledge of these pigments was very limited at that time, and the more quantitative aspects of gene control awaited the researches of Willstätter, Robinson, Karrer, and others to give exact formulae and rapid methods of quantitative analysis.

Mendel's law of inheritance has been found to apply to the inheritance of color in flowers. However, with the anthocyanins we do not always have a simple genetic formula, *i.e.*, a one-factor difference. Studies of the inheritance of color in the sweet pea have shown interesting results. The original wild sweet pea was in all probability a chocolate and purplish blue flower. By breeding, various colors have been selected, and these can be recombined to the original color. Two whites may give purple; this occurs in certain crosses of sweet-pea varieties. Red color is therefore due to two factors, A and B, and the loss of either produces a white flower. A third factor, R, is necessary to produce the blue color which when combined with red produces *purple*, but *R has no color when alone*, only when combined with A and B. Thus, flowers containing only A, B, AR, BR, or R are all colorless. Flowers containing AB are red, and flowers containing ABR are purple. From such data, *deductions* regarding the *chemical* factors which are involved have been drawn.

Miss Wheldale (Mrs. Onslow) proposed a chemical interpretation of the function of factors for flower color in *Antirrhinum*. The factorial composition of the type of this species may be expressed by symbols as follows:

YYIILLTTDDBB

in which the symbols represent the following factors:

Y = a factor representing yellow color in the lips with ivory tube

[12] M. Wheldale, *Biochem. J.*, **5**, 445 (1911); **7**, 87 (1913); M. Wheldale and H. L. Bassett, *ibid.*, **7**, 441 (1913); **8**, 204 (1914); *Proc. Roy. Soc. London*, **B87**, 300 (1914); and M. Wheldale, *J. Genetics*, **4**, 109, 369 (1914 and 1915).

[13] F. Keeble and E. F. Armstrong, *Proc. Roy. Soc. London*, **B85**, 214, 460 (1912); *J. Genetics*, **2**, 277 (1912); F. Keeble, E. F. Armstrong, and W. N. Jones, *Proc. Roy. Soc. London*, **B87**, 113 (1913).

color; the absence of Y precludes the formation of any color in the flower.

I = a factor causing the production of ivory color in the tube and lips and inhibiting the formation of yellow pigment in the lips except on the palate.

L = a factor causing a magenta color in the lips.

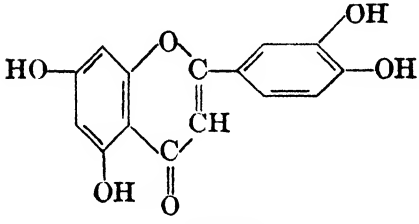
T = a factor causing a magenta color in the tube.

D = a factor causing the production of a deepened or full color; changes a tinged ivory to magenta.

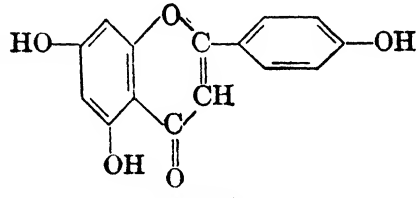
B = a factor converting a red anthocyanin to a magenta anthocyanin.

A study of the chemistry of the pigments involved and the mechanism of color production in *Antirrhinum* makes it possible to interpret these factors in chemical language. Thus, on a chemical basis they are:

Y = a factor causing the production of *luteolin* in the lips and of *apigenin* in the tube.



Luteolin



Apigenin

I = a factor causing the production of *apigenin* in the lips in place of luteolin (note that these two compounds differ only by one hydroxyl group).

L = a factor representing the presence of a reducing (?) or an oxidizing agent, enzyme (?), in the lips, which will produce an anthocyanin from the flavones when they are present.

T = a similar factor representing the presence of a reducing (?) or oxidizing (?) agent in the tube.

D = a factor, emulsin (?), causing the splitting off of the sugars from chains of linked glycosides and the addition or rearrangement of hydroxyl and methoxy groups attached to the rings of the flavone nucleus, thus causing additional anthocyanin to be formed by the hydrolysis of flavone glycosides.

B = a factor controlling the reaction (hydrogen-ion concentration) of the cell sap, thus shifting the color of a solution of the anthocyanin from red to magenta.

The following examples, which have been worked out for *Antirrhinum* varieties, show how these factors work: [14]

1. yyIILLTTDDBB	"white," lips and tube of corolla are pure white (no *flavone* is produced).
2. YYiilTTDDBB	"yellow," lips yellow, tube ivory [luteolin present but no oxidizing (or reducing) agent].
3. YYiiLLTTddbb	"yellow tinged with bronze," [luteolin present, oxidizing (or reducing) agent present].
4. YYIIlTTDDBB	"ivory," lips and tube ivory [luteolin absent, apigenin present, no oxidizing (or reducing) agent in lips].
5. YYiiLLTTDDbb	"bronze," [luteolin absent, apigenin present in lips, glycoside-splitting enzyme (?) present].
6. YYiiLLTTDDBB	"crimson," lips crimson, tube magenta, luteolin present in lips and tube, glycoside-splitting enzyme present, hydrogen-ion factor present, apigenin absent.
7. YYIILLTTDDBB	"magenta," all factors present producing deepest color.

We thus have hereditary factors which can be expressed by

1. The synthesis of definite chemical substances.
2. The modification of such substances once they are formed.
3. The control of enzyme action.
4. The modification of cell sap reaction.

Anderson [15] found that the hybrid grapes produced by crossing the American and European forms are characterized by the presence of malvidin, the dimethoxy derivative. The tendency to produce a dimethoxy derivative of delphinidin is therefore a dominant in inheritance over the tendency to produce the monomethoxy derivative.

Lawrence and Price [16] and Scott-Moncrieff [17] have reviewed the more recent quantitative work on the inheritance of the three groups of flower

[14] See M. Wheldale, *J. Genetics*, Vol. 4, color plate VII (1914), for an exact reproduction of these varieties.

[15] R. J. Anderson, *J. Biol. Chem.*, **57**, 795 (1923); R. J. Anderson and F. P. Nabenhauer, *ibid.*, **61**, 97, 685 (1924); *J. Am. Chem. Soc.*, **48**, 2997 (1926).

[16] W. J. C. Lawrence and J. R. Price, *Biol. Revs. Cambridge Phil. Soc.*, **15**, 35 (1940).

[17] R. Scott-Moncrieff, *Ergeb. Enzymforsch.*, **8**, 277 (1939).

pigments (anthocyanins, flavones, and carotenoids). They conclude that pigment production is genetically controlled and that the amount of any pigment present can be interpreted in terms of the velocity of the reaction involved in synthesis. The work of Lawrence and Scott-Moncrieff on *Dahlia* flowers established an inverse correlation in production of four pigments which suggests that they are all derived from a common but strictly limited source. Many genes have been identified whose end effects are purely quantitative. In addition there are intensifying and diluting genes whose action is not effective over the whole petal, but is restricted to certain areas, giving rise to patterns.

Modification of chemical structure of the anthocyanins, including the state of oxidation, glycosidal type, and, probably, degree of methylation, are determined by simple gene relationships. The flowers of the majority of "good" species are pigmented by only one anthocyanin. The simplest case of anthocyanin mixture is due to incomplete methylation, *e.g.*, *Geranium psilostemon* contains malvin plus a small amount of petunin. More important are the mixtures arising from different degrees of oxidation.

This discussion may be considered merely a suggestion of the possible aid which biochemistry may furnish in bringing about a solution of the problems of heredity. Only a bare beginning has been made in this important and fruitful field.

Biogenesis of Flavones and Anthocyanins in Plants. When anthocyanins are present, there is, as a rule, an excess of sugars in the leaves. Overton, in 1899, first called attention to the fact that there was more sugar in the red autumn leaves than in the normal green leaves. This has been confirmed by other workers who have also found more glycosides in the leaves containing anthocyanins. It is especially true for plants where the conducting systems have been injured so that the transport of carbohydrates is interfered with. A leaf which is injured may rapidly turn red, while the remainder of the plant retains its full vigor. As Robinson puts it, "low temperature, light, and access to oxygen favor the appearance of the coloring matter; plants which form abundant anthocyanin in the dark are equipped with reserves of foodstuffs, an example being the tulip."

The artificial "feeding" of plants with sugars has resulted in many instances in anthocyanin formation. More intense light and low temperature produce anthocyanin, as has already been noted as a characteristic of alpine plants. Here we have high photosynthetic activity and a low requirement of carbohydrate utilization for energy (low enzyme activity for respiration, due to low temperature), both conducive to the formation of carbohydrates.

Mrs. Onslow [18] believes that chromogens acted on by oxidases produce anthocyanins, and she has produced good *genetical* evidence [19] for her theory. Keeble and Armstrong support her view. Everest, however, in his chemical studies, finds that flavones, when reduced, give colors corresponding to anthocyanins, and he regards the anthocyanins as reduction products of the flavones. Willstätter's studies of the anthocyanins also indicate that they are reduction products of the flavones. Miss Wheldale has questioned the reduction theory, and the genetic data and certain of the biochemical data support her view. Perhaps the observation of Kozlowski [20] may be pertinent to the question. He finds that anthocyanins are reduced by sodium sulfite and sodium hydrosulfite to colorless compounds and that the color is regenerated by oxidation with iodine. He notes that Willstätter's red pigments formed by the *reduction* of flavonols are *not* decolorized by sulfites and that these red pigments of Willstätter are altered in color by elementary iodine. Kozlowski concludes that the hypothesis of Willstätter for the formation of anthocyanins by the reduction of flavonols is not justified and that we must return to the oxidation hypothesis.

The biochemical data are very well summed up by Armstrong,[21] as follows:

> In general the distribution of pigments in flowers coincides exactly with that of oxydases. The oxydases, it is true, are more widely distributed than are the chromogens; but the distribution is in conformity with the oxydase-chromogen hypothesis, as will be illustrated by several typical examples, culled from the many available.
>
> The flowers of *Primula sinensis* and of *Dianthus barbatus* (Sweet William) show most epidermal oxydase in the most deeply coloured varieties, less in the less deeply coloured, and none at all in the white varieties. The white flowers of certain *Primula sinensis*, *Pisum sativum*, and *Lathyrus odoratus* have all been shown to contain oxydases, and the white colour is attributed to the absence of chromogen. In the Mont Blanc Star, the distribution of oxydase again parallels that of pigment. One flower had irregular magenta flaked petals with one exception, this particular petal being of a uniform magenta colour. The latter petal gave a well-marked oxydase reaction, the magenta patches on the others demonstrated a fair reaction, whilst the white portions did not respond to the test.
>
> Similarly, Sweet Williams were grown in full-coloured, white, and almost

[18] M. Wheldale Onslow, *The Anthocyanin Pigments of Plants*, 2nd ed., revised, The University Press, Cambridge, 1925.

[19] M. Wheldale and H. L. Bassett, *J. Genetics*, **4**, 103 (1914); M. Wheldale, *Proc. Cambridge Phil. Soc.*, **15**, 137 (1909); *J. Genetics*, **1**, 133 (1911).

[20] A. Kozlowski, *Science*, **83**, 465 (1936).

[21] E. F. Armstrong, *Chemistry in the Twentieth Century*, E. Benn, Ltd., London, 1924.

white varieties, the latter showing rosy dots or lines. The fully coloured flowers responded definitely to the tests for the presence of oxydases, whilst the white flowers also gave a definite but limited reaction—the white colour being probably due, as explained above, to the absence of chromogen. The white flowers with rosy dots showed oxydases only in the parts of the petals corresponding to the pigmented dots.

Robinson [22] considers as unlikely the synthesis of anthocyanidins by reduction of flavones and favors the oxidase hypothesis of Keeble and Armstrong. The *leuco-anthocyanidins*, much more widely distributed than the colored compounds themselves, are a potential but unproved source of pigments. Parallel formation mechanisms, possibly from a common intermediate, are considered more likely than the flavonol-reduction theory. The C_{15} system (C_6—C_3—C_6) of this group of pigments is regarded as built up from hexoses and triose by means of aldol condensation followed by reduction.

[22] R. Robinson, *Nature*, **137**, 172 (1936).

VI

BIOCHEMICAL REGULATORS

The Vitamins, Hormones, and Enzymes

It would appear to be not an overstatement to say that vitamin research has the capability of revolutionizing the science of physiology and enriching it in every department.

ROGER J. WILLIAMS (1943)

The study of enzyme actions is of importance not only in connection with, and as a part of, chemical science, but also in its biological aspects because of the bearing of such actions upon the chemical changes occurring in life processes. The development of our knowledge of the chemical phenomena underlying living matter is bound up intimately with a knowledge of enzyme actions.

K. GEORGE FALK (1924)

CHAPTER 36

Vitamins

Vitamins may be defined as *very potent organic substances which occur in minute quantities in natural foodstuffs, which must be supplied in the diet of animals or may be synthesized in animals from essential dietary or metabolic precursors, and which exert a hormone-like or enzyme action in the control and coordination of specific chemical reactions in the animal body.* The vitamins are definite nutritive substances, although the quantitative requirements for them are relatively very small, as in the case of some of the inorganic elements now recognized as essential for animal life. The grouping is on a physiological basis and does not represent a series of compounds having similar chemical make-up, such as is found among the nutritionally essential amino acids.

No attempt will be made here to describe the historical steps through which the vitamins were discovered. It will be sufficient to point out that two fields of work eventually combined into what has been referred to by McCollum as the newer knowledge of nutrition. These two fields of work were, first, a study of certain specific diseases, *i.e.*, beriberi and scurvy, and, second, attempts to nourish laboratory animals on synthetic diets composed of the known food components in the purified state. A valuable tool used in more recent work has been the study of the nutritional requirements of microorganisms.

The term "vitamine," from which our present word vitamin is derived, was coined by Funk [1] as the result of an attempt to isolate from food the substance, of as yet unknown nature, that prevented the disease beriberi. When beriberi thus came to be recognized as a "vitamine"-deficiency disease, the hypothesis was advanced that various other "vitamine"-deficiency diseases occur. At about the same time, laboratory studies in nutrition showed that natural foods contain chemical substances necessary for normal nutrition, but not classified under any of the ordinary groups of nutrients, *i.e.*, protein, carbohydrate, fat, and mineral salts. Of special importance was the discovery that these substances were incapable of chemical determination by the methods applicable to the nutrients mentioned. Particularly astonishing was the

[1] C. Funk, *J. State Med.*, **20**, 341 (1912).

fact that the effect exerted by these accessory food ingredients was far greater than that shown by corresponding amounts of known nutrients and apparently quite out of proportion to the amounts evidently present in the food. Because they contributed no significant material substance or energy to the body but acted rather as catalysts of life processes, it was necessary to place vitamins in an entirely different category from the better-known foodstuffs.

The workers in the field of "vitamine"-deficiency diseases and those studying the strictly nutrition problems soon came to recognize that these new substances required for normal nutrition were apparently identical with the "vitamines," a lack of which was primarily concerned with the etiology of the deficiency diseases. It was in this way that these two fields of investigation came together and contributed to the discovery and the extension of our knowledge of vitamins.

Unfortunately, we still have very little knowledge of the relationships between the chemical structure, the biological functions, and the clinical symptoms associated with a deficiency of any of the known vitamins. For a few, such as for vitamins A, D, and K, we see a connection between the biological role and the avitaminosis manifestations. For others, such as for riboflavin, some of the relationships between chemical structure and biological functions are recognized. For none is there a clear picture of the interrelationships between the vitamins of the chemist, the biologist, and the clinician.

FAT-SOLUBLE VITAMINS

Vitamin A. Vitamin A was discovered almost simultaneously by McCollum and Davis [2] and by Osborne and Mendel.[3] The discovery was based on the observations, which have been repeatedly confirmed, that certain fats and oils possess a remarkable power of stimulating the growth of rats, which stimulatory effect is absent in whole or in part from other fats and which is not explainable on the basis of the glycerides contained in these fats.

The chemical nature of vitamin A was established by the brilliant researches of Karrer [4] and associates, later confirmed by Heilbron, Morton, and Webster,[5] and by Karrer [6] himself by means of chemical

[2] E. V. McCollum and M. Davis, *J. Biol. Chem.*, **15**, 167 (1913).

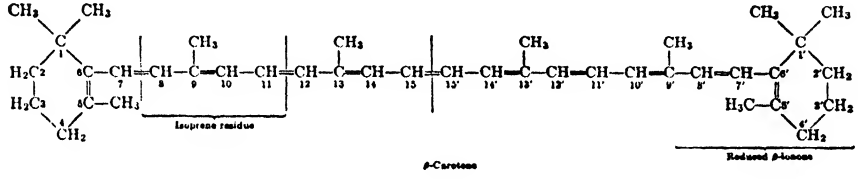
[3] T. B. Osborne and L. B. Mendel, *J. Biol. Chem.*, **15**, 311 (1913).

[4] P. Karrer, A. Helfenstein, H. Wehrli, and A. Wettstein, *Helv. Chim. Acta*, **13**, 1084 (1930); and P. Karrer, R. Morf, and K. Schöpp, *ibid.*, **14**, 1431 (1931).

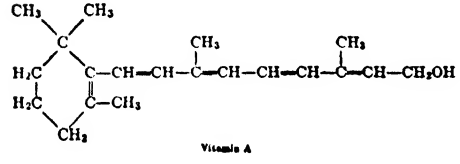
[5] I. M. Heilbron, R. A. Morton, and E. T. Webster, *Biochem. J.*, **26**, 1194 (1932).

[6] P. Karrer, R. Morf, and K. Schöpp, *Helv. Chim. Acta*, **16**, 557 (1933).

synthesis of the hydrogenated derivative of the vitamin. In fact, when Moore [7] established the biochemical basis for vitamin A synthesis in the animal body from the plant pigment carotene, the chemical basis for such a relationship had already been laid by the Karrer school. They had shown β-carotene to be an aliphatic methylated hydrocarbon having two identical, unsaturated, optically inactive terminal rings with the same structure as that occurring in β-ionone. Moreover, the aliphatic polyene central chain could be considered a condensed chain of two pairs of dehydrated isoprene (β-methylbutadiene) residues united in reverse order at the center of the chain (carbons-15, -15′ in the structural formula). Karrer proposed the structure for vitamin A which is now accepted:



β-Carotene



Vitamin A

Heilbron, Morton, and Webster substantiated Karrer's formula for vitamin A by forming from it 1,6-dimethylnaphthalene. Karrer's school completed the proof by an eight-step synthesis of crystalline perhydrovitamin A, beginning with β-ionone, and proving its identity with the completely hydrogenated vitamin from fish liver oil.

More than a decade ago claims were made that the vitamin itself was successfully synthesized,[8,9] in one case through the reduction of the aldehyde condensation product of β-cyclocitrol with two moles of dimethylacrolein, and in the other case through reduction of the same product produced by condensation of β-ionylideneacetaldehyde with β-methylcrotonaldehyde. That vitamin A itself was obtained is disputed by Karrer [10] and numerous others. Apparently the product was a related polyene but not the active vitamin.

[7] R. Moore, *Lancet*, March 9, 1929, p. 499; Aug. 24, 1929, p. 380; *Biochem. J.*, **23**, 803, 1267 (1929); **24**, 672 (1930); **25**, 275 (1931).

[8] R. C. Fuson and R. E. Christ, *Science*, **84**, 294 (1936).

[9] R. Kuhn and C. J. O. R. Morris, *Ber.*, **70B**, 853 (1937).

[10] P. Karrer and A. Ruegger, *Helv. Chim. Acta*, **23**, 284 (1940).

More recently, however, vitamin A activity has been obtained by the synthesis of derivatives of the vitamin. Indeed, Hunter and Williams [11] claim the *chemical* conversion of β-carotene into vitamin A by using hydrogen peroxide in glacial acetic acid to cleave the central double bond, yielding vitamin A aldehyde. This was then reduced to the biologically active alcohol. The yield of vitamin A by this method was very small.

Oroshnik [12] synthesized vitamin A methyl ether (or a stereoisomer), but it is not known whether the product was biologically active. A group in Switzerland [13] also has reported a synthesis of the methyl ether, starting with 1-methoxy-3-methylpentenyne and 4-(trimethylcylohexenyl)-2-methylbutenal. These workers report the synthetic material to have vitamin A activity at least equal to that of β-carotene.

Arens and Van Dorp [14] have synthesized vitamin A acid, starting with β-ionone and γ-bromomethylcrotonate. Dehydration, saponification, and reduction yielded a C_{18} ketone which was converted to vitamin A acid via the Reformatsky reaction. The synthetic vitamin A acid in peanut oil proved to have approximately one-tenth the growth-promoting value of the natural vitamin. However, it must be recognized that in these and other syntheses less potent or inactive stereoisomers may also be present.

At least three modifications of syntheses of ethers and esters of vitamin A have been reported by Milas.[15] The final products and many of the intermediates were biologically active, their potencies being much less than that of crystalline vitamin A, although greater than cod liver oil. One firm has announced the synthesis of vitamin A on a commercial basis, but details of the synthesis have not been released.

The natural vitamin [16] and vitamin A palmitate [17] have been obtained in crystalline form.

The multiple nature of vitamin A now seems well established. In 1937, it was found [18] that the vitamin in the livers of fresh-water fish differed from the vitamin A of commercial fish liver oils from marine fish. The second vitamin, *vitamin* A_2, contains one additional ethylenic

[11] R. F. Hunter and N. E. Williams, *J. Chem. Soc.*, 554 (1945).

[12] W. Oroshnik, *J. Am. Chem. Soc.*, **67**, 1627 (1945).

[13] O. Isler, M. Kofler, W. Huber, and A. Ronco, *Experientia*, **2**, 31 (1946), through *Chem. Abs.*, **40**, 4032 (1946).

[14] J. F. Arens and D. A. Van Dorp, *Nature*, **157**, 190 (1946); *Rec. trav. chim.*, **65**, 338 (1946).

[15] N. A. Milas, *Science*, **103**, 581 (1946).

[16] H. N. Holmes and Ruth E. Corbet, *Science*, **85**, 103 (1937); *J. Am. Chem. Soc.* **59**, 2042 (1937).

[17] J. G. Baxter and C. D. Robeson, *Science*, **92**, 203 (1940).

[18] E. Lederer and V. Rosanova, *Biokhimiya*, **2**, 293 (1937).

linkage. It has a potency of 47,500 U.S.P. units per gram,[19] as compared with 4,300,000 units per gram for vitamin A_1 alcohol.[20] A third form of vitamin A accounting for one-third of the vitamin potency of shark liver oil and other fish liver oils has been called *galol* [21] or *neo-vitamin A*.[22] The compound has been crystallized, and it has a biological potency equal to vitamin A, but it has different physical properties and is more stable to atmospheric oxidation. It forms the same anhydro compound as vitamin A_1. Robeson and Baxter offered evidence for a *cis*-configuration at the double bond between carbon-9 and -10 in both vitamin A_1 and neo-vitamin A, but concluded that a *trans*-configuration at carbon-13 and -14 in vitamin A differentiates this isomer from the *cis*-neo-vitamin A.

In 1943, Embree and Shantz [23] discovered that whale liver oil contained a biologically inactive dihydric alcohol which they termed *kitol*. The compound decomposed upon high-temperature molecular distillation to yield one molecule of active vitamin A. A kitol$_2$ yielding vitamin A_2 also was found in livers of fresh-water fish.

A number of biologically inactive compounds structurally related to vitamin A also occur in nature. *Anhydro* or "cyclized vitamin A," which can be prepared by treating the vitamin with anhydrous hydrogen chloride in absolute alcohol, occurs along with vitamin A in fish liver oils.[24] The conversion of vitamin A to this inactive derivative is apparently one of dehydration rather than of cyclization.[25] Prolonged treatment causes isomerization of the anhydro vitamin A.

Embree and Shantz [26] also have reported on another inactive vitamin A derivative, termed *subvitamin A*, which was found in traces in fish liver oils, and which had properties related to those of vitamin A_1 or A_2.

The ring structure, the unsaturated polyene chain, and the aliphatic primary alcohol group, all are of major importance in the biochemistry of vitamin A. The unsaturated, unoxidized, optically inactive ring evidently plays some as yet undetermined role in the biological activity

[19] J. L. Jensen, E. M. Shantz, N. D. Embree, J. D. Cawley, and P. L. Harris, *J. Biol. Chem.*, **149**, 473 (1943).

[20] J. G. Baxter, C. D. Robeson, and P. L. Harris, cited by K. Hickman, *Ann. Rev. Biochem.*, **12**, 353 (1943).

[21] C. D. Robeson and J. G. Baxter, *Nature*, **155**, 300 (1945).

[22] C. D. Robeson and J. G. Baxter, paper presented at 109th meeting of *American Chemical Society*, Atlantic City, N. J., April 11, 1946.

[23] N. D. Embree and E. M. Shantz, *J. Am. Chem. Soc.*, **65**, 910 (1943).

[24] N. D. Embree, *J. Biol. Chem.*, **128**, 187 (1939).

[25] E. M. Shantz, J. D. Cawley, and N. D. Embree, *J. Am. Chem. Soc.*, **65**, 901 (1943); *cf.* also E. G. E. Hawkins and R. F. Hunter, *Biochem. J.*, **38**, 34 (1944).

[26] N. D. Embree and E. M. Shantz, *J. Am. Chem. Soc.*, **65**, 906 (1943).

of the vitamin, because only those carotenoids having the ionone ring which possesses these characteristics are convertible to vitamin A. Thus one molecule of β-carotene which has two such rings theoretically may form two molecules of vitamin A, whereas α-carotene, γ-carotene, and cryptoxanthin, the only other known carotenoids that may be converted to vitamin A, can yield a maximum of one molecule of the vitamin because they have only one such ring per molecule. Actually, the biological conversion of the carotenoids to vitamin A is usually considerably less efficient than this. Hickman,[27] however, cites data indicating that β-carotene has more than 50 per cent of the biological activity of vitamin A alcohol as proof of *in vivo* fission of the provitamin into two molecules of vitamin A.

The stereochemical configuration of the carotenoids has an important bearing on vitamin A activity. Neo-β-carotene U (the 9,10-*cis* isomer) is only one-third as effective as provitamin A as the natural, all-*trans*-β-carotene.[28] The data of Fraps and Meinke [29] indicate that this decreased potency is not the result of a greater excretion of the isomer. Similarly, *cis*-isomerization of one of the double bonds of α-carotene, leading to the formation of neo-α-carotene U, greatly reduces the provitamin A activity as compared with that of the all-*trans* compound.[30] On the other hand, a naturally occurring stereoisomer of γ-carotene, pro-γ-carotene, having several *cis* double bonds, was found [31] to be a far better provitamin A than γ-carotene itself.

Table 69 brings together some of the data on vitamin A activity of the carotene isomers.

A certain amount of unsaturation in the vitamin A aliphatic chain evidently is required for biological activity because both perhydrovitamin A and perhydro-β-carotene are inactive while α- and β-dihydrocarotene and diiodocarotene are active. Hydrogenation of fats or oils containing vitamin A or its precursors thus destroys this vitamin. The side-chain unsaturation explains for the most part the probable function of vitamin A and the carotenoids as oxidation-reduction catalysts; they should be both hydrogen acceptors and oxidation promoters, the latter especially when actively absorbing oxygen, as they readily do in solution (in fat and fat solvents). Highly oxidized carotenes and vitamin A

[27] K. Hickman, *Ann. Rev. Biochem.*, **12**, 353 (1943).

[28] H. J. Deuel, Jr., C. Johnston, E. Summer, A. Polgar, and L. Zechmeister, *Arch. Biochem.*, **5**, 107 (1944).

[29] G. S. Fraps and W. W. Meinke, *Arch. Biochem.*, **6**, 323 (1945).

[30] H. J. Deuel, Jr., E. Summer, C. Johnston, A. Polgar, and L. Zechmeister, *Arch. Biochem.*, **6**, 157 (1945).

[31] H. J. Deuel, Jr., C. Johnston, E. Summer, A. Polgar, W. A. Schmeder, and L. Zechmeister, *Arch. Biochem.*, **5**, 365 (1944).

are not biologically active, an important fact for vitamin A preservation in foods. Rancid fats are not reliable sources of vitamin A activity. The "carotene oxidase" enzyme found in plant tissues actually is *lipoxidase*, and it does not exert a direct action on the carotene. In the oxidation of the fat, however, carotene is also destroyed.[32]

TABLE 69. BIOLOGICAL ACTIVITY OF CAROTENE STEREOISOMERS *

Type of Carotene	*Relative Biological Activity*
All-*trans*-β-carotene	100
Neo-β-carotene U	38
Neo-β-carotene B	53
All-*trans*-α-carotene	53
Neo-α-carotene U	13
Neo-α-carotene B	16
All-*trans*-γ-carotene	28
Pro-γ-carotene	44
Cryptoxanthin	57

* Compiled from work of Deuel and co-workers by N. Embree, *Oil & Soap*, **23**, 305 (1946).

The alcohol structure in vitamin A explains its ability to form compounds with fat acids, proteins, bile acids, etc., all of which are evidently important in its biochemistry, *i.e.*, the absorption of the vitamin from the intestine as bile acid compound of the free alcohol,[33] its transportation in the blood and lymph as fat acid esters,[34] its storage in the liver [35] in a similar form, the occurrence of a protein-vitamin A [36] compound as precursor of the visual purple of the retina, and the appearance of free vitamin A in non-saponifiable extracts from vitamin A-containing oils and fats. However, the vitamin, being an alcohol, may be separated from its carotenoid precursors by the well-known phase-separation procedure so widely employed in carotenoid chemistry. Vitamin A accompanies the "xanthophylls" in the alcohol layer of a petroleum ether–80–90 per cent methyl alcohol separation.

Vitamin A and each of its carotenoid precursors exhibit characteristic spectral absorption bands, the most important bands of the carotenoids being in the visible blue and violet, that of vitamin A being in the invisible ultraviolet. The wave length of maximum intensity, 328 mμ, of the vitamin A absorption band (300–350 mμ) has proved to be of great

[33] J. B. Sumner and R. J. Sumner, *J. Biol. Chem.*, **134**, 531 (1940).
[34] J. D. Greaves and C. L. A. Schmidt, *Am. J. Physiol.*, **111**, 502 (1935).
[34] J. C. Drummond, M. E. Bell, and E. T. Palmer, *Brit. Med. J.*, **1**, 1208 (1935).
[35] L. Reti, *Compt. rend. soc. biol.*, **120**, 577 (1935).
[36] G. Wald, *J. Gen. Physiol.*, **19**, 351 (1935).

importance for the quantitative estimation of the vitamin. Although not specific for vitamin A, since dehydroergosterol also gives a maximum at 328 mμ, it can be identified because the sterols show three other maxima in the ultraviolet. By using a wave length of maximum intensity of one of the visible bands of the carotenoid precursors of vitamin A, they too can be estimated by spectrophotometric methods, since the absorption intensity depends on the molar concentration of the pigment layer at the point chosen.

The carotenes form an intensely blue compound with antimony trichloride, and vitamin A forms a blue-violet compound with a peak at 620 mμ with this salt under proper conditions. This reaction has been studied extensively and applied both to the detection of vitamin A and its carotenoid precursors in animal tissues and products. It was first developed into a quantitative colorimetric method by Carr and Price.[37] It is useful for preliminary assays of cod liver oil and other vitamin A-rich fish liver oils and for exploratory research in the physiology of the vitamin. The reaction with antimony trichloride, however, is a general one for polyene substances.

Vitamin A exhibits a green, fading fluorescence under ultraviolet light. Popper and others have successfully used fluorescence microscopy in studying vitamin A in tissue sections. A recent review of this subject by Popper [38] describes the technics involved, the factors influencing the use of this method, its specificity, and the qualitative observations that have been made on vitamin A distribution in many organs under both normal and pathological conditions.

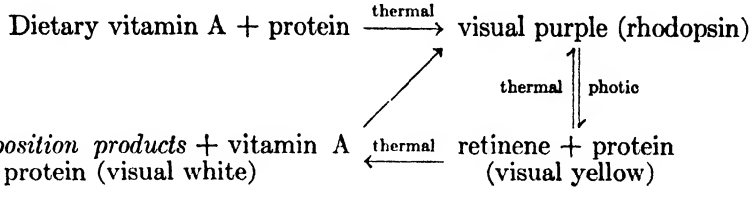
Vitamin A is classified among the growth-promoting factors. It is not clear just how this effect is brought about. There is no evidence as yet that this or any other vitamin actually accelerates the metabolic processes or modifies their efficiency. The probabilities are that the growth-promoting effect of vitamin A is exerted indirectly through its apparent ability to maintain the integrity of the epithelial tissues of the animal body. This is supported by the fact that the primary effect of a deficiency of vitamin A is a keratinization of the epithelial tissues, leading to the diseased condition of the cornea known as *xerophthalmia*.

A more direct relationship between vitamin A and vision than that indicated by the mere prevention of keratinization became evident when Wald [39] showed that the retinal pigment, "visual purple," was a protein carrying vitamin A as a prosthetic group. The sensitivity of the eye to light at low intensities (night vision) is due to elements in the

[37] F. H. Carr and E. A. Price, *Biochem. J.*, **20**, 497 (1926).
[38] H. Popper, *Physiol. Revs.*, **24**, 205 (1944).
[39] G. Wald, *Nature*, **134**, 65 (1934); *J. Gen. Physiol.*, **19**, 351 (1935).

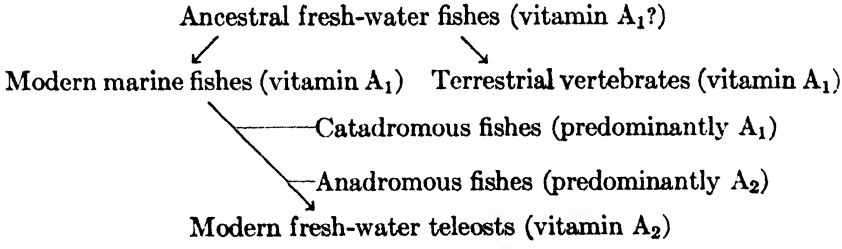
retina called rods. These rods contain "visual purple" or *rhodopsin*, a pigment which is photochemically active and which is necessary for the transformations involved in seeing light. Hecht [40] has pointed out that the absorption spectrum of purified rhodopsin shows a maximum at 510 mμ, coinciding with the region of greatest sensitivity of the eye near the threshold of vision.

Wald suggested the following scheme for transformations in the retina during the visual cycle:



One form of "night blindness" is thus a manifestation of vitamin A deficiency evidenced by poor ability to regenerate rhodopsin. This knowledge has led to use of the so-called dark adaptation test, which uses the biophotometer or the adaptometer as means of clinical diagnosis of mild deficiency of vitamin A.

In fresh-water fish, the retinal pigment is not identical with rhodopsin. Wald [41] has called this substance *porphyropsin*. The pigment differs from rhodopsin in having a different vitamin A as the prosthetic group. This discovery was the first indication that vitamin A may be a group of compounds rather than a single chemical entity. Wald [42] reported that the configuration of vitamin A in the retina is determined genetically rather than by environment. He proposed the following evolutionary scheme:



It is of interest that the feeding of vitamin A_2 in place of vitamin A_1 in the diet of the rat will lead to a slow increase in the former in the

[40] S. Hecht, *Am. Scientist*, **32**, 159 (1944).
[41] G. Wald, *Nature*, **139**, 1017 (1937).
[42] G. Wald, *J. Gen. Physiol.*, **22**, 391 (1939).

retina and other tissues.[43] Vitamin A_2 can successfully provide the vitamin A needed for proper functioning of the mammalian body.

Vitamin A is synthesized by animals only. The high vitamin A activity of green leaves and the association of vitamin A potency with green and yellow seeds and of various fruits and roots with yellow carotenoid pigments is due to the fact that one or more of the carotenoids, and especially β-carotene, are the precursors from which vitamin A is formed. The activity of the provitamin is in part dependent on the extent of its destruction in the gastro-intestinal tract and hence is enhanced by the presence of optimal amounts of the antioxidant tocopherols (see p. 903). The conversion of carotenoids to vitamin A apparently occurs in the liver of animals under the influence of a hydrolytic enzyme which has been called carotenase.[44] It may not be present in all higher animals.

Among the foods of animal origin which are usually reliable sources of vitamin A are whole milk, butter, cheese, egg yolk, and the glandular organs of animals, particularly the liver. The presence of vitamin A in these foods, however, depends on the presence of carotenes and cryptoxanthin in the food of the animals which produce these products.[45] These products may contain both true vitamin A and one or more of its carotenoid precursors, usually chiefly β-carotene, although in egg yolk it is chiefly cryptoxanthin. Standard U.S.P. cod liver oil is a rich source of vitamin A, but halibut and burbot liver oil are many times richer. This fact is of considerable importance when therapeutic doses of the vitamin are necessary. These high concentrations of vitamin A in the liver of fish represent the effect of diets exceptionally rich in carotenes during the fattening season. A similar relationship exists between the diet of cows and hens and the vitamin A content of milk and eggs. The highest concentrations of vitamin A in these foods occur during the season when an abundance of fresh green forage is available. Fortunately the lack of green forage for the winter feeding of dairy cattle and laying fowls may be overcome in part through feeding properly cured, leafy forage, yellow seeds, such as yellow corn, and yellow roots. These facts are valuable to the producers of eggs and milk in helping them control to some extent the vitamin A content of their respective products.

For biological purposes vitamin A and provitamin concentrations in natural foods and medicinal products are not expressed in the usual

[43] E. M. Shantz, N. D. Embree, H. C. Hodge, and J. H. Wills, Jr., *J. Biol. Chem.*, **163**, 455 (1946).

[44] H. S. Olcott and D. C. McCann, *Science*, **74**, 414 (1931); *J. Biol. Chem.*, **94**, 185 (1931).

[45] C. Kennedy and R. A. Dutcher, *J. Biol. Chem.*, **50**, 339 (1922).

quantitative chemical terms but in terms of units equivalent to the units of a standard product. Since 1934 the international standard for vitamin A has been a sample of pure, crystalline β-carotene; one unit of vitamin A activity corresponds to the activity of 0.6 μg. of this International Standard Preparation. In terms of medicinal products and natural foods, U.S.P. cod liver oil must contain at least 600 units of vitamin per gram; the best summer butter (richest in vitamin A) may contain 5,000 units per 100 grams, and winter butter (lowest in vitamin A) averages only 2,400 units per 100 grams.[46] The carotene in carrots may be equivalent to 3,500 to 7,000 units per 100 grams and that in spinach to 9,000 to 36,000 units per 100-gram portion.

The National Research Council has made recommendations for vitamin A allowances in the diet of various animals.[47] Some of these recommendations are shown in Table 70.

TABLE 70. RECOMMENDED DAILY VITAMIN A OR PROVITAMIN A ALLOWANCES IN THE DIET

(National Research Council)

The recommendations for humans are based on an intake of both vitamin A and carotene. The values for farm animals are given in terms of carotene alone.

	Young Growing Animal	*Pregnancy*	*Maintenance of Male Adult*
Humans	3,500 I.U.*	6,000 I.U.	5,000 I.U.
Dairy cattle	25 mg.†	90 mg.‡	60 mg.‡
Beef cattle	25 mg.†	55 mg.‡	55 mg.‡
Swine	4 mg.§	20 mg.	20 mg.
Sheep	3.3 mg.‖	8 mg.¶	8.2 mg.¶
Turkeys	2,500 I.U.**	4,000 I.U.**	
Chickens	1,800 I.U.**	3,300 I.U.**	

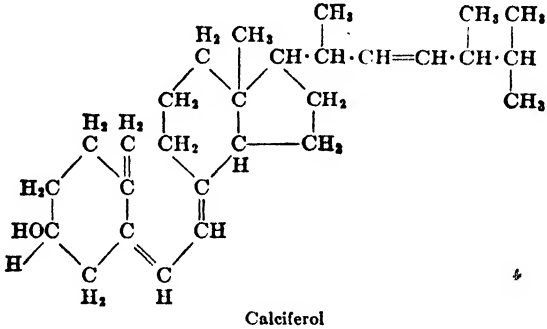
* For a 55-lb. child.
† For a 400-lb. animal.
‡ For a 1,000-lb. animal.
§ For a 100-lb. animal.
‖ For a 60-lb. animal.
¶ For a 150-lb. animal.
** Units of either vitamin A or carotene *per pound of feed.*

[46] Anonymous, *U. S. Dept. Agr. Misc. Pubs.*, 571 (July, 1945).

[47] Food and Nutrition Board, National Research Council, Washington, D. C., Recommended dietary allowances, *Reprint and Circular Series*, No. 122 (August, 1945); Committee on Animal Nutrition, National Research Council, Recommended nutrient allowances for poultry (June, 1944), swine (Aug., 1944), dairy cattle (Aug., 1945), beef cattle (Sept., 1945), sheep (Oct., 1945).

The Vitamins D. Hopkins [48] apparently first recognized that some "accessory" dietary factor is concerned in the etiology of rickets. Funk [49] later classified rickets as an avitaminosis. Mellanby [50] made the first experimental demonstration of rickets by means of a dietary deficiency of fat-soluble vitamin. It remained for McCollum [51] and his co-workers to demonstrate a distinct vitamin related to the etiology of rickets. This vitamin is known as vitamin D, although a considerable number of closely related substances exert a vitamin D effect.

Vitamin D is formed by the irradiation of ergosterol, 7-dehydrocholesterol, and certain other sterols. A number of active materials are now known. The first of the vitamins D for which the structure was determined was vitamin D_2, or *calciferol*. There is no vitamin D_1, since the material to which this name was assigned subsequently proved to contain both vitamin D_2 and lumisterol. Windaus [52] showed that calciferol has the following structure:



Calciferol

Calciferol is a derivative of ergosterol and is the vitamin D in irradiated yeast and in important medicinal forms on the market (viosterol, vigantol, etc.). Formerly regarded as the only vitamin D in nature, it is now known to be absent (essentially) from cod liver oil, although it does occur to some extent in other fish liver oils.[53]

The photochemical conversion of ergosterol to calciferol is effected by

[48] F. G. Hopkins, *Analyst*, **31**, 385 (1906).

[49] C. Funk, *Die Vitamine, ihre Bedeutung für Physiologie und Pathologie mit besonderer Berücksichtigung der Avitaminosen (Beri-beri, Skorbut, Pellagra, Rachitis)*, J. F. Bergmann, Wiesbaden, 1914.

[50] E. Mellanby, *Lancet*, **196**, 407 (1919).

[51] E. V. McCollum, N. Simmonds, J. E. Becker, and F. G. Shipley, *Bull. Johns Hopkins Hosp.*, **33**, 229 (1922).

[52] A. Windaus and W. Thiele, *Ann.*, **521**, 160 (1935).

[53] H. Brockmann and A. Busse, *Z. physiol. Chem.*, **256**, 252 (1938).

both ultraviolet and cathode rays and appears to proceed according to the following series of reactions: [54]

Ergosterol → lumisterol → tachysterol →

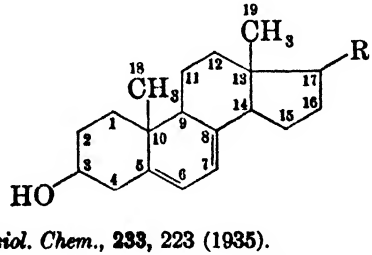
calciferol ↗ isopyro and pyrocalciferol
calciferol ↘ toxisterol → suprasterols I and II

In *lumisterol*, the ring has not yet been opened. *Tachysterol* is an isomer of calciferol, differing from the latter in the position of the double bonds in the opened ring B. *Pyrocalciferol* has one less double bond, since ring closure has regenerated ring B; it differs from ergosterol in the spatial arrangement of the asymmetric centers. In *suprasterol*, the B ring is five-membered.

Vitamin D_3, an irradiation derivative of 7-dehydrocholesterol, is the chief vitamin in cod liver oil and other very active fish liver oils. 7-Dehydrocholesterol appears to be the chief provitamin D in milk and animal fats and in the animal skin in which vitamin D is produced by sunlight. It is this provitamin D, rather than ergosterol, as first believed, which occurs in crude cholesterol; the mistake was due to 7-dehydrocholesterol and ergosterol having similar absorption spectra.

The vitamin D effectiveness of calciferol is increased [55] for chicks by hydrogenation of the double bond in the side chain. This explains in part the greater effectiveness of the cod liver oil vitamins D which appear to be chiefly cholesterol derivatives. The cholesterol side chain is saturated. The irradiation product of 22-dihydroergosterol is *vitamin D_4*, which has been shown to occur in nature.

A number of other materials with vitamin D activity are recognized. All are derived from sterols, the cyclopentanoperhydrophenanthrene or cholane skeleton being the important structure in the provitamin D molecule. The relationship of vitamin D activity to sterol structure has been discussed previously (*cf.* pp. 795 *et seq.*). Of the eleven provitamin compounds listed in Table 63, all but two have the structure



[54] M. Muller, *Z. physiol. Chem.*, **233**, 223 (1935).
[55] F. G. McDonald, *J. Biol. Chem.*, **114**, lxv (1936).

Of the remaining two, 22-dihydrotachysterol is derived by irradiation of a compound having the above structure, and the other provitamin, 7-hydroxycholesterol, will form a similar doubly unsaturated nucleus upon losing one molecule of water. In all forms of vitamin D for which the structure is known, the B ring has been opened to form a methylene group at carbon-18.

Rickets is not entirely a vitamin-deficiency disease. The relative proportions of calcium and phosphorus and their concentration in the diet are important factors in the etiology of rickets. A major function of vitamin D seems to be to maintain the proper concentrations of calcium and phosphorus in the blood, depending on the species of animal, regardless of disproportion between these elements and a deficiency of one or both of them in the diet. When the calcium and phosphorus in the diet are normal the need for vitamin D is apparently at a minimum, although there are definite species differences in this relationship. There are also species differences regarding the influence of vitamin D on the metabolism of calcium and phosphorus.

With regard to the role of vitamin D in the maintenance of normal blood calcium or phosphorus, the naturally occurring rickets in puppies, pigs, colts, and calves is usually preceded by an abnormal drop in calcium, often accompanied by tetany, rather than by a drop in inorganic phosphorus which usually precedes as well as accompanies natural rickets in chicks and infants. Both types of rickets may be produced experimentally in young rats, but the usual one is the low-phosphorus type. With the exception of chicks, and to some extent infants, the dietary deficiency of mineral, either relative or absolute, is that of the element which becomes lowered in the blood. With regard to species differences in the need for vitamin D when the calcium and phosphorus in the diet are normal, we have at the one extreme the rat, whose need for vitamin D is scarcely detectable when these mineral elements are normal, and at the other extreme the chick, which will develop severe rickets in the absence of vitamin D regardless of the calcium and phosphorus content of the diet. Other species fall somewhere in between these extremes, and therefore all have a certain absolute requirement for the vitamin. It is not without biochemical significance that the type of rickets produced in these species when vitamin D is the sole limiting factor is the type natural to the species, referred to above. Other species differences with respect to the response to different forms of vitamin D are also evident.

The decreased concentration of nearly all forms of phosphorus in the blood which accompanies rickets affords an explanation for the general muscular weakness and instability of the nervous system which occur

in this disease, just as the low blood calcium which may also occur explains the tetany sometimes seen in rickets.

Man and other higher animals derive relatively little of the vitamin D which they require from their food, because this vitamin does not occur abundantly in foods. This seems surprising in view of the fact that photochemical activation of provitamins D, which are apparently widely distributed and relatively abundant in plant and animal life, is the only demonstrated origin of vitamin D in nature. Indeed, land animals seem to derive most of their vitamin D by transformation of provitamin or provitamins in their own tissues, brought about through the incidence on their bodies of the proper light rays from the sun.

The establishment of different efficacy ratios for chicks and rats of different forms of vitamin D (see Table 71) has given the biochemist

TABLE 71. ANTIRACHITIC EFFECT OF VARIOUS VITAMINS D

Compound	*Potency, I.U. per gm.*	*Approximate Dose for Chicks Equivalent to 1 I.U. (Rat Unit) of Vitamin D_3*
Vitamin D_2	40,000,000	35
Vitamin D_3	40,000,000	1
Epi-vitamin D_3	4,000,000(?)	.....
Vitamin D_4	20,000,000	20
Dihydrotachysterol	80,000	0.25
A.T. 10 *	30,000	0.25
Irradiated 7-dehydrositosterol	1,000,000(?)	.....
Irradiated 7-dehydrocampesterol	4,000,000(?)	.....

* A preparation containing dihydrotachysterol and having marked antitetany (A.T.) activity in hypoparathyroidism.

an exceptionally useful tool for ferreting out the various vitamins D. However, the final chapter in this story has not been written. By this means Bills and co-workers [56] have been able to differentiate the liver oil of different species of tuna fish, thus introducing a new biochemical basis for taxonomy.

Some forms of vitamin D are more efficient in healing rickets in the chick, some are of equal value for both chicks and rats, and some exhibit greater potency for rats than for chicks. Approximately 35 I.U. of calciferol (vitamin D_2) must be fed to chicks to obtain the anti-

[56] C. E. Bills, *Cold Spring Harbor Symposia Quant. Biol.*, **3**, 328 (1935); *J. Am. Med. Assoc.*, **108**, 13 (1937); C. E. Bills, O. N. Massengale, M. Imboden, and H. Hall, *J. Nutrition*, **13**, 435 (1937).

rachitic effect of 1 I.U. of vitamin D_3, despite the fact that these vitamins have a similar potency (40,000,000 I.U. per gram) for the rat. For rats, vitamin D_4 has only about half the potency of its more unsaturated parent, calciferol, but, when fed to chicks on an I.U. basis, vitamin D_4 proves to be the more effective form, although still far poorer than vitamin D_3. On the other hand, dihydrotachysterol, a very weak antirachitic substance for the rat, is four times as efficient as vitamin D_3 when fed to deficient chicks on an I.U. basis.[57,58] There also is considerable evidence to indicate that other species differences exist, and that even turkeys will respond to the various vitamins D in a manner differing from the chick.

Crystalline calciferol is a very potent drug; as little as 0.025 μg. daily, for several days, produces measurable healing of experimental rickets in a rat. The International Standard Vitamin D is an olive oil solution containing 0.0025 per cent pure crystalline calciferol. The International Unit of Vitamin D is the vitamin D activity of 1 mg. of the standard solution. Thus, the unit is the activity of 0.025 μg. of calciferol, or the activity of calciferol is 40,000 units per milligram. The U. S. Pharmacopoeia, through the Vitamin Division of the U. S. Bureau of Chemistry, has provided a Reference Standard Cod Liver Oil which has been biologically assayed (on rats) against the International Standard.

Biological methods of assay are necessary for the quantitative estimation of vitamin D, but these are subject to the usual errors of such methods. For the D vitamins, in addition, there are species differences in response to the different forms, as already explained. In terms of International Units obtained by rat assay, chicks require many times as much vitamin D in the form of irradiated ergosterol as of cod liver oil to produce normal bone mineralization. Fortunately there seems to be little, if any, practical difference for infants between the different available forms of vitamin D.

Average cod liver oil contains 100 I.U. per gram, although the U.S.P. requirement is 85 I.U. The irradiation of milk to increase the vitamin D content has been promoted by the Wisconsin Alumni Research Foundation. Their process consists of a few seconds' exposure of a very thin film of milk to ultraviolet light at 2,300 to 3,100 Å. Potencies of over 200 I.U. of vitamin D per quart are practical by this means. Fortification of milk with activated vitamin D preparations (viosterol) to a level of 400 I.U. per quart is also common.

[57] J. T. Correll and E. C. Wise, *J. Nutrition*, **23**, 217 (1942).
[58] E. W. McChesney, *J. Nutrition*, **26**, 81 (1943).

Vitamin D allowances for the young, growing animal have been suggested by the National Research Council. The recommended dietary allowances are 400 I.U. for humans, 1,200 I.U. for a 400-lb. dairy calf, 250 I.U. for a 100-lb. swine, 360 A.O.A.C. units (vitamin D_3) per pound of feed for turkey poults, and 180 A.O.A.C. units per pound of feed for chicks.

Vitamin E (α-Tocopherol). Vitamin E was discovered by Evans and Bishop [59] in a study of the estrous cycle of female rats fed for relatively long periods upon synthetic diets with an ample provision for all the previously known vitamins. They found that reproductive failure occurs which is very characteristic in its nature. Although ovulation takes place, and the embryos are implanted and begin to develop normally, the placental function fails before parturition, with the accompanying death and resorption of the implanted embryos. Reproduction in the sense of failure for young to appear may occur as a result of many other dietary deficiencies. It is to be emphasized, therefore, that vitamin E deficiency is at present detected in female mammals chiefly through the special histological technic developed by Evans and Bishop.

Attempts to isolate vitamin E began with Evans and Burr,[60] who prepared a sterol-free fraction from the unsaponifiable ether extract of wheat germ oil, which, after steam distillation at 180°C., 50 mm. pressure, and several distillations *in vacuo*, yielded the most active material at 225–230°C. (0.01 mm.) and represented about 0.25 per cent of the original oil. It cured vitamin E deficiency sterility in female rats in single doses of 5 mg.

In 1936, Evans and his co-workers [61] isolated the vitamin which they called α-tocopherol. Other tocopherols were also identified in natural products.[62] The almost invariable association of vitamin E with antioxidant material in natural foodstuffs [63] was explained when Olcott [64] found that all the forms of the tocopherols were highly antioxygenic, although not equally so.

[59] H. M. Evans and K. S. Bishop, *Science*, **56**, 650 (1922); *J. Metabolic Research*, **1**, 319, 335 (1922); **3**, 201, 233 (1923).

[60] H. M. Evans and G. O. Burr, *Memoirs of the University of California*, Vol 8, p. 129, University of California Press, Berkeley, 1927.

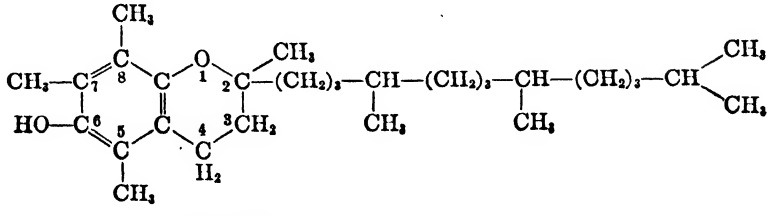
[61] H. M. Evans, O. H. Emerson, and G. A. Emerson, *J. Biol. Chem.*, **113**, 319 (1936); O. H. Emerson, G. A. Emerson, and H. M. Evans, *Science*, **83**, 421 (1936).

[62] O. H. Emerson, G. A. Emerson, A. Mohammed, and H. M. Evans, *J. Biol. Chem.*, **122**, 99 (1937).

[63] M. J. Cummings and H. A. Mattill, *J. Nutrition*, **3**, 421 (1931).

[64] H. S. Olcott and O. H. Emerson, *J. Am. Chem. Soc.*, **59**, 1008 (1937).

In 1938, Fernholz [65] postulated that α-tocopherol was a substituted 6-hydroxychromane with a phytol side chain:



α-Tocopherol

Proof that this formula was correct followed immediately when Smith and his associates at the University of Minnesota [66] reported the synthesis of racemic α-tocopherol with a vitamin potency equal to that of the natural product.[67] Karrer [68] had synthesized α-tocopherol from phytyl bromide and trimethylhydroquinone but believed that a coumaran structure resulted. Smith used three additional and different methods of synthesis, including (*a*) trimethylhydroquinone and γ,γ-dimethylallylbromide, (*b*) trimethylhydroquinone and isoprene, and (*c*) 5,7,8-trimethyl-6-hydroxy-3,4-dihydrocoumarin and methyl magnesium iodide. All methods yielded the same product, definitely indicating a chroman structure as proposed by Fernholz.

An extensive review of the chemistry of vitamin E has been prepared by Smith,[69] who points out that a number of organic compounds appear to have some vitamin E activity. β-Tocopherol is the next lower homolog of α-tocopherol, in which the carbon-7 methyl group has been replaced by a hydrogen atom. In γ-tocopherol, the carbon-5 methyl rather than the carbon-7 methyl group is missing. Recently a δ-tocopherol, with the methyl groups at both positions 5 and 7 missing, has been reported [70] to account for a considerable part of the tocopherols of soybean oil.

Indication that the aromatic ring of the vitamin has biological significance is seen in the fact that β-tocopherol, with a cumohydroquinone ring [71] rather than the duro structure of α-tocopherol, has one-fourth to

[65] E. Fernholz, *J. Am. Chem. Soc.*, **60**, 700 (1938).

[66] L. I. Smith, H. E. Ungunade, and W. W. Prichard, *Science*, **88**, 37 (1938).

[67] H. Evans, G. Emerson, and O. Emerson, *Science*, **88**, 38 (1938).

[68] P. Karrer, H. Fritzsche, B. H. Ringier, and H. Salomon, *Helv. Chim. Acta*, **21**, 520 (1938).

[69] L. I. Smith, *Chem. Revs.*, **27**, 287 (1940).

[70] M. H. Stern, C. D. Robeson, L. Weisler, and J. G. Baxter, Paper presented at 110th meeting of American Chemical Society, Chicago, Ill., Sept. 9–13, 1946.

[71] W. John, *Z. physiol. Chem.*, **250**, 11 (1937).

one-half the vitamin E activity of α-tocopherol. The γ-isomer also has much less biological activity.

The long aliphatic side chain is highly important in determining the vitamin E potency. The phytol moiety is essential for appreciable biological activity of the molecule.

The first oxidation product of α-tocopherol is α-tocoquinone. This compound has no vitamin E activity, but does participate in a reversible oxidation-reduction system with α-tocopherol.

Hathaway and Davis [72] found that artificial drying of alfalfa tended to preserve its vitamin E to a greater degree than did field curing. Vitamin E is stable to heat encountered in ordinary cooking but is readily oxidized. Rancidity or contact with ferric iron leads to loss of activity. The tocopherols serve as antioxidants both *in vitro* and *in vivo*. They are able to protect other fat-soluble vitamins from oxidative destruction in the intestinal tract, and as such have a vitamin-sparing action. The results obtained with the standard U.S.P. diet for vitamin A assay are definitely affected by the addition of tocopherols,[73] which causes an enhanced growth response of the test animals to the U.S.P. vitamin A reference oil. The activity of carotene concentrates is similarly enhanced by simultaneous administration of tocopherols,[74] owing to the protective chemical action of the latter.

Chemical determination of the tocopherols may be based on the use of α,α'-dipyridyl to estimate colorimetrically the ferrous iron formed by the reduction of Fe^{+++} in the presence of tocopherol,[75] or it may involve a photometric determination of the color produced by reacting nitric acid with the vitamin.[76] On the basis of a report by Hume [77] a unit of vitamin E is equivalent to 1 mg. of a standard synthetic *dl*-α-tocopherol acetate.

Reproductive failure after implantation is the most clearly defined symptom associated with vitamin E deficiency. Other manifestations are recognized, however.

A number of favorable reports of the effectiveness of vitamin E in treating threatened and habitual abortion in women have appeared in the medical and scientific literature during the past decade. Vitamin E

[72] I. L. Hathaway and H. P. Davis, *Nebraska Agr. Exp. Sta. Research Bull.* 73 (1934); I. L. Hathaway, H. P. Davis, and R. R. Graves, *Nebraska Agr. Exp. Sta. Research Bull.* 62 (1932).

[73] K. C. D. Hickman, M. W. Kaley, and P. L. Harris, *J. Biol. Chem.*, **152**, 303 (1944).

[74] P. L. Harris, M. W. Kaley, and K. C. D. Hickman, *J. Biol. Chem.*, **152**, 313 (1944); K. C. D. Hickman, M. W. Kaley, and P. L. Harris, *ibid.*, **152**, 321 (1944).

[75] A. Emmerie and C. Engel, *Nature*, **142**, 873 (1938).

[76] M. Furter and R. E. Meyer, *Helv. Chim. Acta*, **22**, 240 (1939).

[77] E. M. Hume, *Nature*, **148**, 472 (1941).

therapy may be of considerable value in such disorders. However, the fact that endocrine as well as nutritional disturbances may be involved in abortion necessarily limits the degree of effectiveness of vitamin E administration.

The physiological effects of vitamin E deficiency are manifested in other ways. A type of male sterility [78] also develops, although much more slowly than the reproductive failures in the female. The effects on the male may be permanent, but this is not likely to be the case for female animals. Evans and Burr [79] studied a form of paralysis in suckling young rats nursed by mothers on certain diets. This condition, which can be cured by changing the diet to natural food or prevented by foods rich in vitamin E or by vitamin-E-rich concentrate prepared from wheat germ oil, has been shown [80] to be a muscular dystrophy of nutritional origin occurring in many species of animals, both mammals and fowl. The effect of cod liver oil in promoting muscular dystrophy under some conditions is associated [81] with the oxidative destruction of the vitamin E in the presence of the highly unsaturated fatty acids of the oil.

Vitamin E activity has been demonstrated in animal tissues and products such as muscle meat, adipose tissue, viscera, and in milk and eggs, probably originating from the food.[82] Synthesis of the vitamin is largely, if not exclusively, a function of plants, certain oil-bearing seeds or parts of seeds being relatively rich in this vitamin, notably wheat germ, cottonseed, palm seed, peanuts, alfalfa seed, and lettuce seed. The oils from rice, barley, oats, corn, and soybeans appear to contain the vitamin but not in such high concentrations as that from wheat. Green leaves, such as alfalfa or lettuce (fresh or dry), are good sources of vitamin E. Most of the vitamin E in legumes occurs as γ-tocopherol, whereas in carrots it all occurs as α-tocopherol.[83]

[78] H. A. Mattill, J. S. Carman, and M. M. Clayton, *J. Biol. Chem.*, **61**, 729 (1924); K. E. Mason, *Proc. Natl. Acad. Sci. U. S.*, **11**, 377 (1925); H. M. Evans, *ibid.*, **11**, 373 (1925); K. E. Mason, *J. Exptl. Zool.*, **45**, 159 (1926); H. A. Mattill, *Am. J. Physiol.*, **79**, 305 (1927); H. M. Evans and G. O. Burr, *Memoirs of the University of California*, Vol. 8 (1927); and F. B. Adamstone and L. E. Card, *J. Morphol.*, **56**, 339 (1934).

[79] H. M. Evans and G. O. Burr, *J. Biol. Chem.*, **76**, 273 (1928).

[80] H. S. Olcott, *J. Nutrition*, **15**, 221 (1938).

[81] C. G. Mackenzie, J. B. Mackenzie, and E. V. McCollum, *J. Nutrition*, **21**, 225 (1941); H. A. Mattill and C. Golumbic, *ibid.*, **23**, 625 (1942).

[82] If all species of higher animals require vitamin E, the opportunity to test its synthesis by animals is afforded in experiments of Wilson, Thomas, and Cannon [*J. Dairy Sci.*, **18**, 431 (1935)], who have reared several generations of goats on rations shown by rat tests to be free from the vitamin.

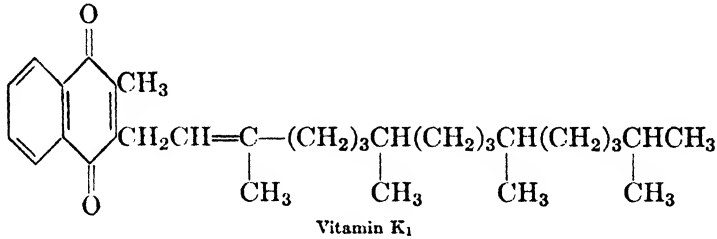
[83] M. L. Quaife and P. L. Harris, Paper presented at 110th meeting of American Chemical Society, Chicago, Ill., Sept. 9–13, 1946.

Although a need for a dietary source of vitamin E has been demonstrated for a number of species of animals, no recommended levels have been generally recognized.

Vitamin K. Dam [84] observed a fatal hemorrhagic disease in young chicks after several weeks of feeding synthetic diets, outstanding symptoms being intestinal, subcutaneous, and intramuscular hemorrhage, inability of the blood to clot, and erosion of the gizzard lining. Dam was not successful in preventing this disease with lemon juice. All the evidence pointed to a deficiency disease, but, since no relation was found to deficiencies in the known vitamins, Dam proposed that the antihemorrhagic factor be called vitamin K (*Koagulations-Vitamin*).

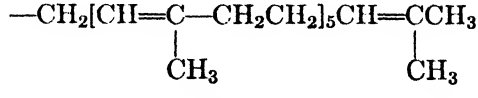
The isolation and much of the work on the chemistry of vitamin K resulted from studies by Dam and his associates [85] and by Almquist.[86]

The chemical nature of the vitamin K of plants (vitamin K_1) was ascertained by synthesis simultaneously in three laboratories.[87] The vitamin proved to be 2-methyl-3-phytyl-1,4-naphthoquinone:



Vitamin K_1

Existence of a somewhat less potent vitamin K_2 in certain microorganisms, in which the phytyl group is replaced by a difarnesyl group,



[84] H. Dam, *Biochem. Z.*, **215**, 475 (1929); **220**, 158 (1930); *Nature*, **133**, 909 (1934); H. Dam and F. Schönheyder, *Biochem. J.*, **28**, 1355 (1934); H. Dam, *Nature*, **135**, 652 (1935); *Biochem. J.*, **29**, 1273 (1935).

[85] H. Dam and F. Schönheyder, *Biochem. J.*, **30**, 897 (1936); H. Dam and L. Lewis, *ibid.*, **31**, 17 (1937); H. Dam, A. Gliger, J. Glavind, P. Karrer, W. Karrer, E. Rothchild, and H. Salomon, *Helv. Chim. Acta*, **22**, 310 (1939).

[86] H. J. Almquist, *J. Biol. Chem.*, **114**, 241 (1936); **117**, 517 (1937); **120**, 635 (1937); A. A. Klose, H. J. Almquist, and E. Mecchi, *ibid.*, **125**, 681 (1938); H. J. Almquist and A. A. Klose, *J. Am. Chem. Soc.*, **61**, 745 (1939).

[87] H. S. Almquist and A. A. Klose, *J. Am. Chem. Soc.*, **61**, 2557 (1939); S. B. Binkley, L. C. Cheney, W. F. Holcomb, R. W. McKee, S. A. Thayer, D. W. MacCorquodale, and E. A. Doisy, *ibid.*, **61**, 2558 (1939); L. F. Fieser, *ibid.*, **61**, 2559 (1939).

was demonstrated by Doisy and his associates.[88] A vitamin K_3 from corn stigma, with a very high antihemorrhagic potency, has been reported by Russian investigators.[89]

Ansbacher and Fernholz[90] showed that considerable vitamin K potency was exhibited by 2-methyl-1,4-naphthoquinone (menadione). This compound is unique in that it represents a simple synthetic material which is more potent than the natural vitamin itself. A number of other derivatives of 1,4-naphthoquinone also show marked antihemorrhagic activity. Phthiocol is very effective.

The proper absorption and utilization of natural (fat-soluble) vitamin K requires the presence of normal bile in the intestinal tract. Oral administration produces more rapid results than intramuscular injections of the vitamin. Of interest in this connection are the observations that the water-soluble 1,4-dihydroxy-2-methyl-naphthalene and 4-amino-2-methyl-1-naphthol[91] and the sodium salt of 2-methyl-1,4-naphthoquinone diphosphoric acid ester[92] are fully as active as vitamin K. They are highly effective therapeutically, since no bile is required for absorption when these compounds are administered parenterally.

Vitamin K_1 is a yellow oil and vitamin K_2 a yellow crystalline solid. The vitamin is labile to alkali and to light.

Vitamin K appears to be necessary for prothromin formation.[93,94] Without prothromin, the process of blood coagulation breaks down owing to lack of a supply of the active coagulating enzyme, thrombin.

Bio-assay remains the principal method of analyzing for vitamin K. Schönheyder[95] developed a curative biological assay procedure for vitamin K based on its ability to restore the clotting of the blood, which is inhibited in severe vitamin K deficiency. One unit of the vitamin is the smallest quantity which will reduce the clotting time to an arbitrary normal when administered during a certain time (several days) to sick chickens having a specified degree of the hemorrhagic disease. Alm-

[88] R. W. McKee, S. B. Binkley, D. W. MacCorquodale, S. A. Thayer, and E. A. Doisy, *J. Am. Chem. Soc.*, **61**, 1295 (1939); S. B. Binkley, R. W. McKee, S. A. Thayer, and E. A. Doisy, *Proc. Am. Soc. Biol. Chem.*, xii (1940).

[89] D. M. Mikhlin, *Biokhimiya*, **8**, 158 (1943), through *Chem. Abs.*, **38**, 1775 (1944); V. V. Babuk, *Compt. rend. acad. sci. U.R.S.S.*, **39**, 277 (1943), through *Chem. Abs.*, **38**, 4014 (1944).

[90] S. Ansbacher and E. Fernholz, *J. Am. Chem. Soc.*, **61**, 1924 (1939).

[91] E. A. Doisy, D. W. MacCorquodale, S. A. Thayer, S. B. Binkley, and R. W. McKee, *Science*, **90**, 407 (1939).

[92] R. H. K. Foster, J. Lee, and U. V. Solmssen, *J. Am. Chem. Soc.*, **62**, 453 (1940).

[93] H. Dam, F. Schönheyder, and E. Tage-Hansen, *Biochem. J.*, **30**, 1075 (1936).

[94] A. J. Quick, *Am. J. Physiol.*, **118**, 260 (1937).

[95] F. Schönheyder, *Nature*, **135**, 653 (1935); *Biochem. J.*, **30**, 890 (1936).

quist and Stokstad [96] developed a method based on prevention of prolonged clotting time and hemorrhage in chicks which had been on the vitamin K-free diet for two weeks after hatching. A revised standard bio-assay has been developed by Almquist.[97] Numerous modifications of these preventive and curative assay technics have been proposed.

In sharp contrast to the chick, vitamin K synthesis in and absorption from the intestinal tract makes the rat normally independent of a dietary source of the vitamin. Although synthesis in the intestine in humans is also appreciable, a deficiency may be brought about as a result of a lowered synthesis (sulfonamide drugs), a lowered supply associated with absence of bacterial synthesis (hemorrhagic disease of newborn infants), or lowered absorption (obstructive jaundice, sprue). In obstructive jaundice, there is a plasma prothrombin deficiency which is associated with insufficient vitamin K, the lowered absorption due to absence of a normal flow of bile to the intestine being a major factor.

The use of the vitamin or an active derivative before and after operations has proved very valuable as an aid to surgery. It is now common practice to administer vitamin K to pregnant women prior to and during labor in order to raise the vitamin level in the infant. Vitamin K is of no value in treating hemophilia, which is not caused by a prothrombin deficiency.

Quick [98] recognized the analogy between vitamin K deficiency and "sweet clover disease" in herbivora and a possible relation to vitamin K of the curative agent for this disease in alfalfa. Link's group [99] has shown that 3,3′-methylene*bis*-(4-hydroxycoumarin) is the agent in spoiled sweet clover hay responsible for hypoprothrombinemia in cattle. It is cured by vitamin K administration.

All green plant material is rich in the vitamin, artificially dried alfalfa being used as the chief source of the vitamin for chemical studies. Hog liver fat is another rich source. Hempseed, as well as certain vegetables, was found by Dam to be a good source, but cereals and other seeds and fruits and vegetables in general are poor sources. Almquist and Stokstad [100] found that the vitamin is produced by microorganisms in fish meal and rice bran when they are allowed to remain in a wet condition for several days at room temperature. Spontaneous cures may occur,

[96] H. J. Almquist and E. L. R. Stokstad, *J. Nutrition*, **14**, 235 (1937).

[97] H. J. Almquist, *J. Assoc. Offic. Agr. Chemists*, **24**, 405 (1941).

[98] A. J. Quick, *Am. J. Physiol.*, **118**, 260 (1937).

[99] M. A. Stahmann, C. F. Huebner, and K. P. Link, *J. Biol. Chem.*, **138**, 513 (1941).

[100] H. J. Almquist and E. L. R. Stokstad, *Nature*, **136**, 31 (1935); *J. Biol. Chem.* **111**, 105 (1935); *J. Nutrition*, **12**, 329 (1936).

possibly explainable by coprophagy. Egg yolk fat and soybean oil were found to contain the vitamin.

Antistiffness Factor. In 1941 it was reported [101] that guinea pigs on a milk diet will develop a nutritional deficiency if the raw milk is replaced by pasteurized whole milk or by skim milk. Symptoms of the deficiency were first a stiffness of the wrist, then emaciation and weakness before death of the animal. Raw cream cured or prevented the condition. Claims that a new and distinct fat-soluble vitamin is involved in this disease have been advanced by van Wagtendonk and Wulzen,[102] who have isolated highly potent, crystalline fractions of the "antistiffness factor" from raw cream, crude molasses, and particularly from crude unheated cane juice.

The antistiffness factor is soluble in ether and other organic solvents, is readily destroyed by heat in the presence of air, and is claimed to be a carbonyl compound with a molecular weight of around 200. The white leaflets from cane juice melted at 81.5–82.0° and had a potency of 500,000,000 guinea pig units per gram. One unit represents that amount which will cure the rigidity of the wrists when administered daily for five consecutive days to a deficient animal.

It is reported that the muscular dystrophy from a deficiency of this factor is not accompanied by a creatinuria commonly found in avitaminosis E. Furthermore, relatively large doses of α-tocopherol failed to relieve the deficiency disease.

THE WATER-SOLUBLE VITAMINS

Ascorbic Acid. Ascorbic acid is the name given the antiscorbutic vitamin, or vitamin C. In American medical literature, it also has been called cevitamic acid.

The classic studies of Holst and Frölich [103] and Fürst [104] furnished the first experimental proof of the existence of an antiscorbutic vitamin, although they were conducted prior to the development of the vitamin hypothesis. However, certain difficulties of technic in producing experimental scurvy and failure to understand the proper relation of the other known vitamins, then called A and B, as well as lack of apprecia-

[101] R. Wulzen and A. M. Bahrs, *Am. J. Physiol.*, **133**, 500 (1941).

[102] W. J. van Wagtendonk and R. Wulzen, *Arch. Biochem.*, **1**, 373 (1943); W. J. van Wagtendonk, V. Schocken, and R. Wulzen, **3**, 305 (1944); W. J. van Wagtendonk and R. Wulzen, *J. Biol. Chem.*, **164**, 597 (1946).

[103] A. Holst, *J. Hyg.*, **7**, 619 (1907); A. Holst and T. Frölich, *ibid.*, **7**, 634 (1907); *Z. Hyg. Infektionskrankh.*, **72**, 1 (1912).

[104] V. Fürst, *Z. Hyg. Infektionskrankh.*, **72**, 121 (1912).

tion of the differences in susceptibility of different species of animals to scurvy, prevented the general acceptance of a vitamin-deficiency etiology of scurvy until about 1919.[105]

The instability of the antiscorbutic vitamin was early recognized as one of its most important properties and was a serious obstacle to its isolation. The demonstration by numerous workers in 1921 that the vitamin is especially susceptible to oxidation, accelerated by alkalinity, was an important advance. Serious attempts to concentrate the antiscorbutic vitamin began with Harden and Zilva.[106]

King and Waugh [107] were the first to report the isolation of the pure crystalline vitamin and the demonstration of its biological activity in 0.5-mg. daily doses to guinea pigs. They recognized the apparent identity of the vitamin with the strongly reducing substance previously isolated by Szent-Györgyi [108] from adrenal cortex, orange, and cabbage juice and called by him hexuronic acid. A few weeks later Svirbely and Szent-Györgyi [109] reported the antiscorbutic activity of the latter's hexuronic acid, which was confirmed by Harris, Mills, and Innes.[110] It was now possible to bring to a swift culmination the determination of the chemical structure of the vitamin and its synthesis. This was greatly aided by the discovery [111] of a rich source of the substance in the ripe fruits of the red pepper and the preparation of large quantities [112] of it, by means of the easily crystallizable monoacetone derivative, discovered by von Vargha.[113] Actually, synthesis of ascorbic acid was accomplished by Reichstein, Grüssner, and Oppenauer [114] before the correct structural formula was established. The first product was the inactive D-series form isolated as the acetone derivative, but later the same workers [115] synthesized the active L-series compound from L-xylosone by cyanhydrin synthesis. The correct structure of the vitamin was suggested by Euler and Martius [116] and was proved correct by the extensive experiments in

[105] J. C. Drummond, *Biochem. J.*, **13**, 77 (1919).

[106] A. Harden and S. S. Zilva, *Biochem. J.*, **12**, 259 (1918).

[107] C. G. King and W. A. Waugh, *Science*, **75**, 357 (1932); W. A. Waugh and C. G. King, *J. Biol. Chem.*, **97**, 325 (1932).

[108] A. Szent-Györgyi, *Biochem. J.*, **22**, 1387 (1928).

[109] J. L. Svirbely and A. Szent-Györgyi, *Nature*, **129**, 576, 690 (1932); *Biochem. J.*, **26**, 865 (1932).

[110] L. J. Harris, I. Mills, and J. R. M. Innes, *Lancet*, **II**, 235 (1932).

[111] J. L. Svirbely and A. Szent-Györgyi, *Biochem. J.*, **27**, 279 (1933).

[112] I. Banga and A. Szent-Györgyi, *Biochem. J.*, **28**, 1625 (1934).

[113] L. von Vargha, *Nature*, **130**, 847 (1932).

[114] T. Reichstein, A. Grüssner, and R. Oppenauer, *Helv. Chim. Acta*, **16**, 561 (1933).

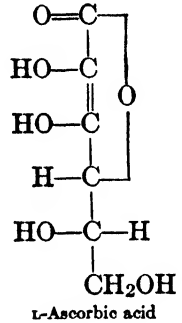
[115] T. Reichstein, A. Grüssner, and R. Oppenauer, *Helv. Chim. Acta*, **16**, 1019 (1933).

[116] H. von Euler and C. Martius, *Arkiv. Kemi, Mineral. Geol.*, [14] **11B** (1933).

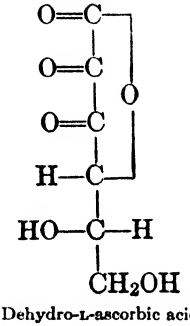
Haworth's [117] laboratory in the University of Birmingham where the synthesis of the vitamin and various isomers was also accomplished.

According to King,[118] a particularly successful method of synthesis accomplished by Reichstein and Grüssner,[119] useful for the commercial synthesis of the vitamin, used these steps: glucose $\xrightarrow{H_2 \text{ and Pt}}$ sorbitol $\xrightarrow{A.\ xylinum}$ sorbose $\xrightarrow{\text{acetone}}$ diacetone sorbose $\xrightarrow{KMnO_4}$ diacetone sorbonic acid $\xrightarrow{HCl}$ L-sorbonic (2-keto-L-gulonic) acid $\xrightarrow{HCl}$ L-ascorbic acid (20–30 per cent yield from sorbose).

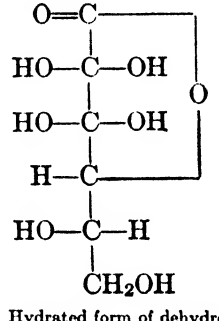
It is now established that the antiscorbutic vitamin is structurally 2,3-dienol-L-gulofuranolactone, which is also active in its dehydro-(2,3-diketo) form. These active lactone forms of the vitamin therefore have the structures



L-Ascorbic acid



Dehydro-L-ascorbic acid



Hydrated form of dehydro-L-ascorbic acid in aqueous solution

The essential aspects of these structures for the biological functions of the compound are not entirely clear but appear to call for D-configuration on carbon-4, a hydroxyl group on carbon-5, and free hydroxyl groups throughout. It is known that the optically opposite D-series ascorbic acid is inactive biologically. Two pentose "ascorbic acids" possess some activity, *i.e.*, the methylated L-lyxose compound L-rhamnoascorbic acid, which has one-fifth the activity of L-ascorbic acid, and D-araboascorbic acid (isoascorbic acid) which has one-twentieth the activity, and L-glucoascorbic acid, which has one-fortieth the activity. The lactone of 2,4-dihydroxyketobutyric acid (hydroxytetronic acid), the simplest substance of the ascorbic acid type, is not antiscorbutically active.[120]

[117] R. G. Ault, D. K. Baird, H. C. Carrington, W. N. Haworth, R. Herbert, E. L. Hirst, E. G. V. Percival, F. Smith, and M. Stacey, *J. Chem. Soc.*, 1419 (1933).

[118] C. G. King, *Physiol. Revs.*, **16**, 238 (1936).

[119] T. Reichstein and A. Grüssner, *Helv. Chim. Acta*, **17**, 311 (1934).

[120] F. Micheel and F. Jung, *Ber.*, **66B**, 1291 (1933).

The chief acidic properties of ascorbic acid are not due to the opening of the lactone, as in the sugars, but are due primarily to the second enolic group on carbon-3. Carter [121] has pointed out that, as expected from Fuson's principle of vinylogy, the carbon-3 hydroxyl of ascorbic acid, which is separated from the carbonyl group by a vinyl group, behaves as a carboxylic hydroxyl group and hence is acidic.

The strong reducing properties of preparations of the vitamin were discovered by Zilva,[122] who later [123] pointed out both the alternate reduction-oxidation character of the reaction toward phenolindophenol and the quantitative reduction of the indicator by both decitrated lemon juice and active vitamin fractions. The latter property was developed by Tillmans and associates [124] as a method for the quantitative estimation of the vitamin.

Dehydro-L-ascorbic acid is a reversible oxidation product of the vitamin, the antiscorbutic activity of which was shown first by Karrer [125] and associates. The structure of the first oxidation product which has lost its activity is not established but is believed to be 2,3-dienol-L-gulonic acid or, more probably, the corresponding diketo compound. There is still much doubt about the role played by the vitamin as a redox agent in the body. The importance of the redox properties of the vitamin in plant life seems much better established, the conditions being much more favorable there. Borsook, *et al.*,[126] have shown that dehydroascorbic acid is rapidly reduced in tissues. When dehydroascorbic acid is injected into the body, part of it is converted into diketogulonic acid and part is reduced to ascorbic acid,[127] according to the scheme

$$\text{Ascorbic acid} \rightleftharpoons \text{dehydroascorbic acid} \rightarrow \text{diketogluonic acid}$$

Ascorbic acid is protected from oxidation in animal fluids by glutathione, proteins, and amino acids, according to Barron, Barron, and Klemperer.[128] These workers show that those fresh vegetables and juices of fresh vegetables and fruits which lose their antiscorbutic value relatively quickly lack the protective mechanisms present in animal

[121] H. E. Carter, *Biol. Symposia*, **5**, 56 (1941).

[122] S. S. Zilva, *Biochem. J.*, **18**, 632 (1924).

[123] S. S. Zilva, *Biochem. J.*, **21**, 689 (1927).

[124] J. Tillmans, *Z. Untersuch. Lebensm.*, **60**, 34 (1930); J. Tillmans, P. Hirsch, and W. Hirsch, *ibid.*, **63**, 1 (1932); J. Tillmans, P. Hirsch, and J. Jacicisch, *ibid.*, **63**, 241 (1932).

[125] P. Karrer, H. Salomon, K. Schöpp, and R. Morf, *Helv. Chim. Acta*, **16**, 181 (1933).

[126] H. Borsook, H. W. Davenport, C. E. P. Jeffreys, and R. C. Warner, *J. Biol. Chem.*, **117**, 237 (1937).

[127] J. R. Penney and S. S. Zilva, *Biochem. J.*, **37**, 403 (1943).

[128] E. S. G. Barron, A. G. Barron, and F. Klemperer, *J. Biol. Chem.*, **116**, 563 (1936).

fluids. Instead they contain "oxidases" which catalyze the reversible aerobic oxidation of the vitamin and later destroy it, as shown first by Szent-Györgyi.[129]

The powerful and specific ascorbic acid oxidase of pumpkins, squash, and cucumbers is a copper protein containing at least 0.24 per cent of copper.[130] The enzyme catalyzes the oxidation of the vitamin to dehydroascorbic acid. However, most plants contain other enzyme systems capable of oxidizing ascorbic acid. The polyphenoloxidases and peroxidases readily oxidize ascorbic acid and are destroyed less easily than ascorbic acid oxidase.

Very small traces of copper will catalyze the rapid autoxidation of ascorbic acid in either acid or alkaline solution.[131]

The existence of a dehydroascorbic acid reductase has been suggested.[132] Bukin has stated that the reductase is destroyed by heating for 3 minutes at 60°, whereas the ascorbic acid oxidase activity is fully preserved.

The practical importance of protecting ascorbic acid from catalytic destruction in vegetables and fruits which are to be canned or otherwise processed has been well established. This is accomplished by a preliminary heat inactivation of the accompanying oxidase systems, by avoiding contact with copper or brass equipment, and by careful exclusion of oxygen during the steps incidental to processing.

The precursors of the antiscorbutic vitamin in plants and in animals which synthesize the vitamin are not known with certainty. The synthesis of ascorbic acid in plants is not dependent on photosynthesis, since it may be formed in the dark and in the absence of chlorophyll, but the importance of this redox substance in subsequent photosynthetic chemical processes seems established. That the ascorbic acid content of plant materials is affected by the intensity of light has been shown for many products. Hamner and his associates have shown experimentally that the ascorbic acid content of tomatoes [133] may be increased by 66 per cent, and of turnip greens [134] by 800 per cent, by moving the plant from a low to a high light intensity a few days before harvesting. Somers, Kelly, and Hamner [135] have used discs cut from leaves of turnips and other plants and floated on a mineral nutrient solution to study the influence of such factors as light intensity, temperature, and carbon dioxide con-

[129] A. Szent-Györgyi, *J. Biol. Chem.*, **90**, 385 (1931).
[130] W. H. Powers, S. Lewis, and C. R. Dawson, *J. Gen. Physiol.*, **27**, 167 (1944).
[131] R. W. Peterson and J. H. Walton, *J. Am. Chem. Soc.*, **65**, 1212 (1943).
[132] E. F. Kohman and N. H. Sanborn, *Ind. Eng. Chem.*, **29**, 189 (1937).
[133] K. C. Hamner, L. Bernstein, and L. A. Maynard, *J. Nutrition*, **29**, 85 (1945).
[134] K. C. Hamner and R. Q. Parks, *J. Am. Soc. Agron.*, **36**, 269 (1944).
[135] G. F. Somers, W. C. Kelly, and K. C. Hamner, *Arch. Biochem.*, **18**, 59 (1948).

centration on the rate of synthesis or destruction of ascorbic acid. They report that the formation of ascorbic acid in these discs requires both light and carbon dioxide and that it is also influenced by temperature in a manner analogous to photosynthetic reactions.

Most species of animals, the guinea pig and man excepted, are not dependent on their food for vitamin C, which is synthesized in the animal itself. The site of ascorbic acid synthesis in animals remains to be established. The intestinal wall has been shown [136] to be a possible site, and the adrenal glands, although normally very rich in the vitamin, are eliminated as the sole site [137] because adrenalectomized dogs do not develop scurvy when supplied with adrenal cortical hormone. The tissues of animals which synthesize their own ascorbic acid are as rich in the vitamin as are those species, human and guinea pig, which require it in their food and are enjoying a liberal intake. Normally the latter excrete the vitamin in the urine, solely in the reduced form, but the excretion continues on a scorbutic diet with the resulting depletion of the ascorbic acid in the tissues. Thus, when a dose of ascorbic acid is not followed promptly by the excretion in the urine, the result is regarded as indicating deficiency of the vitamin.

Scurvy is recognized by a number of clinical manifestations, including hemorrhages ranging from petechiae on the extremities to extensive subcutaneous hemorrhages and excessive bleeding of the gums. Prescorbutic symptoms of vitamin C deficiency also include multiple hemorrhages and less severe degenerative changes in the cartilage and bone matrices. Thus, ascorbic acid is necessary to maintain the integrity of the blood capillaries and appears to play a role in normal bone and tooth development.

Despite our long-standing knowledge of the nature of vitamin C, and despite the availability of simple and accurate analytical methods, surprisingly little is known of the fundamental biological role played by ascorbic acid.

According to King,[138] the two specific roles for ascorbic acid in animal tissues which are likely are: "(*a*) a respiratory function in serving as a hydrogen-transport agent between unidentified metabolytes and other carriers or molecular oxygen, by way of two or more oxidase enzyme systems; and (*b*) regulation of the colloidal condition of intercellular substances as shown by Wolbach and associates." [139] However, there still

[136] E. Harde and J. Wolff, *Compt. rend. soc. biol.*, **116**, 288 (1934).

[137] H. M. Vars and J. J. Pfiffner, *Proc. Soc. Expil. Biol. Med.*, **31**, 839 (1934).

[138] C. G. King, *Physiol. Revs.*, **16**, 238 (1936).

[139] S. B. Wolbach and P. R. Howe, *Arch. Path. Lab. Med.*, **1**, 1 (1926); V. Menkin, S. B. Wolbach, and M. F. Menkin, *Am. J. Path.*, **10**, 569 (1933); S. B. Wolbach, *Am. J. Path.* (Suppl.), **9**, 689 (1933).

is no clear-cut evidence of a specific function for ascorbic acid in cellular oxidation. The role of ascorbic acid in the organism as suggested by its cytology was reviewed by Bourne,[140] who has considered that ascorbic acid is probably intimately bound up in cellular synthesis, particularly in the adrenal and anterior pituitary and perhaps in the corpus luteum and interstitial cells of the gonads.

There is evidence of a protective role played by ascorbic acid against certain infections and of increased requirements resulting from increased eliminations in tuberculosis, rheumatic fever, diphtheria, poliomyelitis, and pneumonia.

The International Unit of antiscorbutic vitamin is the antiscorbutic activity of 0.05 mg. of L-ascorbic acid, which is the International Standard. This is about one-tenth the daily dose necessary to prevent gross scorbutic lesions in a young guinea pig maintained on a scorbutogenic diet.

Certain vegetables and the citrus fruits are outstanding food sources of ascorbic acid. Apples are rather poor sources, usually contributing only 5–8 mg. per 100 grams, but one non-commercial variety contains 37 mg., nearly equivalent to the citrus fruits. Typical values that have been reported per 100-gram serving of material, based mostly on chemical titration methods are: spinach 50–100 mg., oranges and lemons 50 mg., grapefruit 40 mg., raw cabbage 40 mg., tomatoes 15–30 mg., peas 20–30 mg., and strawberries 70 mg. Boiled vegetables, however, may lose up to four-fifths of the ascorbic acid to the cooking water. In contrast to the acid fruits, very considerable losses of the vitamin are encountered in the canning of neutral vegetables.[141] Losses may be due to both leaching and oxidation. Sulfur dioxide inhibits the oxidative destruction of ascorbic acid during processing and has been used extensively during the commercial dehydration of both fruits and vegetables.

The ascorbic acid content of fresh raw cows' milk is 15–25 mg. per liter. Exposure of milk to traces of copper or to daylight through glass in the presence of oxygen changes a large portion of the ascorbic acid to the reversibly oxidized form which does not reduce indophenol. The reversibly oxidized form is readily destroyed by pasteurization but not the reduced form. Human milk is reported to contain as much as five

[140] G. Bourne, *Physiol. Revs.*, **16**, 442 (1936).

[141] N. B. Guerrant, M. G. Vavich, O. B. Fardig, R. A. Dutcher, and R. M. Stern, *J. Nutrition*, **32**, 435 (1946).

times the ascorbic acid occurring in fresh raw cows' milk and appears to fluctuate in ascorbic acid with the diet much more than does cows' milk.

The National Research Council has recommended daily dietary allowances of ascorbic acid on a sliding scale ranging from 30 mg. for a child under 1 year of age to 100 mg. for boys 16–20 years old. A moderately active man should receive 75 mg. In pregnancy, an intake of 100 mg. is recommended, and during lactation, 150 mg. It is recognized that considerably smaller quantities will protect against scurvy.

Citrin (Vitamin P). In 1936 a group of Hungarian workers, including Szent-Györgyi,[142] reported cures of purpura hemorrhagica by administering lemon juice but not when giving ascorbic acid. A flavone glycoside was isolated which cured vascular forms of purpura. This was named citrin (citrus flavone), to indicate its chemistry, and also vitamin P, to indicate its vitamin-like effect on tissue permeability (*i.e., Permeabilität Vitamin*). The citrin also prolonged the life of scorbutic guinea pigs and reduced the extent of capillary hemorrhages.

The Hungarian workers [143] concluded that experimental scurvy is the result of combined ascorbic acid and citrin avitaminoses, pure ascorbic acid deficiency being obtained only when vitamin P is supplied. Bentsáth and Szent-Györgyi [144] assumed that vitamin P requires traces of ascorbic acid for its activity.

That vitamin P activity does not parallel the ascorbic acid content of natural materials has been shown by several workers. Scarborough [145] has reported that grapes have 10–20 times as much P potency as tomatoes, whereas the latter are far richer in ascorbic acid.

Chemical studies [146] of citrin indicated that it is a mixture of glycosides of previously known flavonone derivatives, *i.e.*, hesperitin and eriodictyol. The glycoside *hesperidin* is a methyl derivative of *eriodictin*. Wawra and Webb [147] isolated biologically active hesperidin chalcone

[142] I. Rusznyák and A. Szent-Györgyi, *Nature*, **138**, 27 (1936); L. Armentano, A. Bentsáth, T. Beres, I. Rusznyák, and A. Szent-Györgyi, *Deut. med. Wochschr.*, **62**, 1326 (1936); A. Bentsáth, S. Rusznyák, and A. Szent-Györgyi, *Nature*, **138**, 798 (1936).

[143] A. Bentsáth, I. Rusznyák, and A. Szent-Györgyi, *Nature*, **139**, 326 (1937).

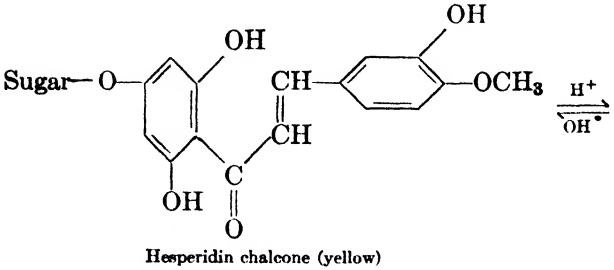
[144] A. Bentsáth and A. Szent-Györgyi, *Nature*, **140**, 426 (1937).

[145] H. Scarborough, *Biochem. J.*, **39**, 271 (1945).

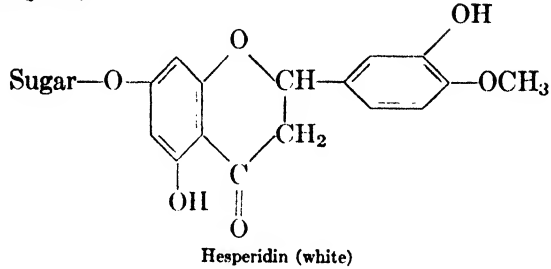
[146] V. Bruckner and A. Szent-Györgyi, *Nature*, **138**, 1057 (1936); S. Lajos and M. Gerendas, *Biochem. Z.*, **291**, 229 (1937); S. S. Zilva, *Biochem. J.*, **31**, 915 (1937).

[147] C. Z. Wawra and J. L. Webb, *Science*, **96**, 302 (1942).

from lemon peel. They considered it the prosthetic group for an enzyme. The chalcone is capable of reversible oxidation and reduction.

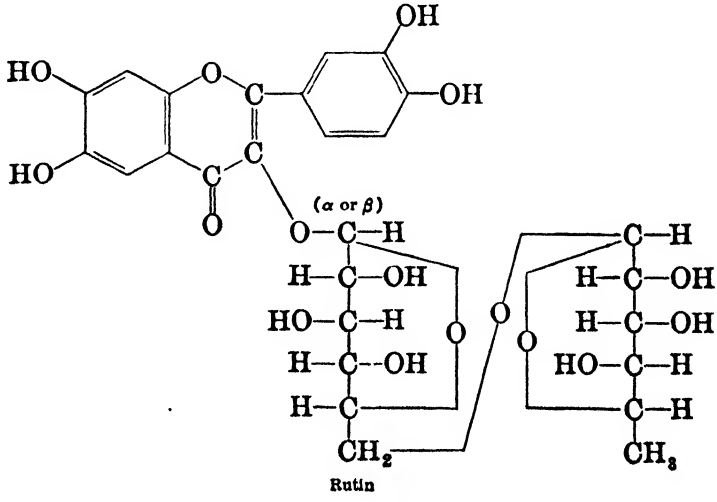


Hesperidin chalcone (yellow)



Hesperidin (white)

Recently, Scarborough [145] has called attention to the probable multiple nature of the capillary fragility factor. He points out that the activity of pure hesperidin is not sufficient to account for the potency of many plant extracts. He considers the active material to be a group of flavonones. Rutin, a rhamnoglucoside of quercetin,[148] has considerable



Rutin

[148] G. Zemplén and A. Gerecs, *Ber.*, **68B**, 1318 (1935).

activity in increasing capillary resistance in man [149] and offers promise in treatment of capillary fragility in retinal hemorrhage and apoplexy.[150] The group at the Eastern Regional Research Laboratory, working on the commercial production of rutin, report [150] that buckwheat is a promising source material. The average rutin content is well above 2 per cent, and occasional samples of buckwheat exceed 6 per cent of rutin.

Considerable controversy has centered on whether citrin should be classed as a vitamin and whether it is specifically associated with prevention of capillary fragility. At the present time, the existence of vitamin P still is not universally accepted, and little is known of its physiological role in animal metabolism. The status of citrin and its therapeutic use has been the subject of a recent review.[151]

The Vitamin B Complex. The "water-soluble B" vitamin has proved to consist of a large number of distinct compounds essential in animal nutrition. At least nine are clearly defined and now available in synthetic form. Each year new factors characterized as possible additions to the family of B vitamins are noted in the literature.

The first subdivision of "water-soluble B" was based on the observation that one fraction (B_1) was heat-labile. The heat-stable fraction (the B_2 group) contains most of the remaining members of the vitamin B complex.

Lepkovsky [152] has reviewed the literature up to 1941 on this group of vitamins.

Thiamine. This is the name suggested by Williams [153] and accepted by the American Institute of Nutrition for the antineuritic or antiberiberi vitamin commonly called B_1. The term aneurin, suggested by Jansen,[154] is used by the British.

So many workers contributed to the early research on experimental and natural beriberi that it is not possible to credit the discovery of the antiberiberi vitamin to one person. Eijkman [155] is credited with first producing beriberi experimentally, and his experiments may now be interpreted as demonstrating the existence of an antiberiberi food factor,

[149] J. Q. Griffith, Jr., J. F. Couch, and M. A. Lindauer, *Proc. Soc. Exptl. Biol. Med.*, **55**, 228 (1944).

[150] J. F. Couch, J. Naghski, and C. F. Krewson, *Science*, **103**, 197 (1946).

[151] *Nutrition Revs.*, **1**, 214 (1943).

[152] S. Lepkovsky, *Nutrition Abstracts & Revs.*, **11**, 363 (1941–42).

[153] *Cf. J. Am. Med. Assoc.*, **109**, 952 (1937).

[154] B. C. P. Jansen, *Nature*, **135**, 267 (1935).

[155] C. Eijkman, *Geneesk. Tijdschr. Nederland Indië*, **30**, 295 (1890); **36**, 214 (1896).

although he did not so interpret it until later.[156] Grijns [157] first suggested a nutritive deficiency theory for experimental polyneuritis and beriberi and made some attempts to determine the nature of the nutritive substance. The first claim of isolation of antiberiberi substance from foods was made by Hulshoff-Pol,[158] who called it X-acid. An intensive search for the substance during the years 1910–1912 reached a milestone in the history of vitamins with the coining of the term "vitamine" by Funk for the substance which he believed to be the specific antineuritic substance. However, it was not until fifteen years later that the actual isolation of the pure crystalline vitamin was accomplished by Jansen and Donath.[159] Rapid progress was now possible in determining the chemistry of the vitamin. The presence of sulfur in the molecule was discovered by Windaus [160] and associates. Methods for obtaining larger yields of pure vitamin were devised by Williams [161] and associates and by Kinnersley, O'Brien, and Peters.[162] These methods made possible the practical employment, for the first time, of the pure antineuritic antiberiberi substance, as well as the determination of its chemical structure. Windaus, Tschesche, and Grewe [163] and Williams [164] demonstrated that the vitamin contains a pyrimidine and a thiazole ring. Sulfite at *p*H 5 quantitatively cleaves thiamine into these two products and destroys the vitamin activity. Makino and Imai [165] presented a number of convincing arguments for the existence of a saturated carbon atom uniting the two cyclic groups. This was confirmed by Williams,[166] who not only gave the first correct structural formula for the vitamin but soon after described [167] its synthesis, thus bringing to a successful

[156] C. Eijkman, *Arch. Hyg.*, **58**, 150 (1906).

[157] G. Grijns, *Geneesk. Tijdschr. Nederland Indië*, **41**, 3 (1901); **49**, 216 (1909); *Mededeelingen Geneesk. Lab.*, 1908 and 1909; *Geneesk. Tijdschr. Nederland Indië*, **51**, 591 (1911). These papers have been published in English, *i.e.*, Professor Dr. G. Grijns' *Researches on Vitamins*, 1900–1911, J. Noorduyn and Son, Gorinchem, Holland, 1935.

[158] D. J. Hulshoff-Pol, *Geneesk. Tijdschr. Nederland Indië*, **47**, 688 (1907).

[159] B. C. P. Jansen and W. F. Donath, *Geneesk. Tijdschr. Nederland Indië*, **66**, 810 (1927).

[160] A. Windaus, R. Tschesche, H. Ruhkopf, F. Laquer, and F. Schultz, *Z. physiol. Chem.*, **204**, 123 (1932).

[161] R. R. Williams, R. E. Waterman, and J. C. Keresztesy, *J. Am. Chem. Soc.*, **56**, 1187 (1934).

[162] H. W. Kinnersley, J. R. O'Brien, and R. A. Peters, *Biochem. J.*, **29**, 701 (1935).

[163] A. Windaus, R. Tschesche, and R. Grewe, *Z. physiol. Chem.*, **237**, 98 (1935).

[164] R. R. Williams, *J. Am. Chem. Soc.*, **57**, 229, 536 (1935).

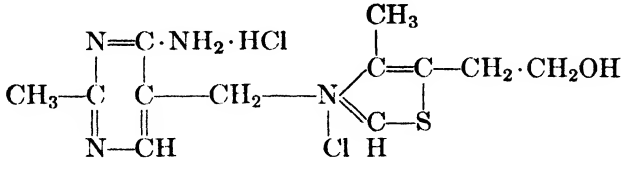
[165] K. Makino and T. Imai, *Z. physiol. Chem.*, **239**, 1 (1936).

[166] R. R. Williams, *J. Am. Chem. Soc.*, **58**, 1063 (1936).

[167] R. R. Williams and J. K. Cline, *J. Am. Chem. Soc.*, **58**, 1504 (1936); J. K. Cline, R. R. Williams, and J. Finkelstein, *J. Am. Chem. Soc.*, **59**, 1052 (1937); *cf.* also R. R. Williams, *Ind. Eng. Chem.*, **29**, 980 (1937); *J. Am. Med. Assoc.*, **110**, 727 (1938).

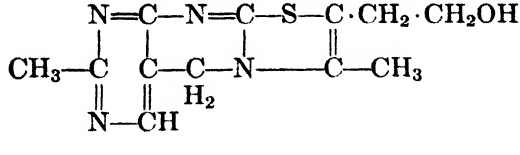
conclusion the studies on the chemical nature of the first vitamin to be definitely postulated.

Thiamine chloride hydrochloride has the structure



It is a colorless, crystalline compound melting at 232–234°C. Below *p*H 7 its aqueous solution shows two characteristic absorption maxima in the ultraviolet at 235 mμ and 267 mμ; its two acid-binding groups have *p*K values of 3.4 and 4.8, the former being that of the thiazole nitrogen, the latter of the free amino group on position-6 of the pyrimidine half of the molecule. When a formaldehyde-treated solution of the compound is added to an alkaline diazotized sulfanilic acid, a characteristic yellow → pink color reaction is shown. This formaldehyde-azo reaction, first described by Kinnersley and Peters,[168] is stable when applied to thiamine solutions which have not been treated with alkali above *p*H 7. It is a reaction given by the thiazole half of the molecule. Strong alkali causes the liberation of hydrogen sulfide (especially when heated), the disappearance of the azo-color reaction, the appearance of a pseudo-acid group with *p*K about 9 (probably the alcohol group), and a marked change in the spectral absorption maximum at 234 mμ, which shows first an increase in intensity and then is replaced by a broad, rapidly fading band at 330–340 mμ. Biological activity is lost on heating, particularly in alkaline medium above *p*H 11.

One of the most interesting properties of thiamine is its conversion on oxidation under suitable conditions to a yellow-blue fluorescent pigment called thiochrome by Kuhn, Wagner-Jauregg, *et al.*[169] This substance has the structure



which shows that its formation from the oxidized vitamin involves loss of water as well as dehydrogenation and the consequent ring formation

[168] H. W. Kinnersley and R. A. Peters, *Biochem. J.*, **28**, 667 (1934).

[169] R. Kuhn, T. Wagner-Jauregg, F. W. van Klaveren, and H. Vetter, *Z. physiol. Chem.*, **234**, 196 (1935).

and shift of double bond. The identity of thiochrome from yeast and the alkaline oxidation products of thiamine was finally proved by Bergel and Todd,[170] who synthesized the pigment. Indeed, the synthesis of thiochrome furnished independent proof of the correct structure of the vitamin.

Thiochrome itself is not biologically active. A chemical method for the estimation of the vitamin, based on the quantitative formation of the fluorescent pigment from the vitamin, has been devised by Jansen.[171] Its modification and application to natural products have given nutritionists a valuable tool for thiamine assay.

Enzymatic conversion of thiamine to inactive degradation products has also been demonstrated. Raw fish meal contains an enzyme capable of destroying thiamine. Foxes fed a diet containing 10 per cent or more of uncooked fish will develop a deficiency disease known as *chastek paralysis*, which is now recognized as an extreme vitamin B_1 deficiency.[172] Large supplements of thiamine or elimination of raw fish from the diet several days a week (and substituting a diet adequate in thiamine) will prevent development of the paralysis. Recently, existence of a thiamine-destroying enzyme in clams (but not in oysters) has also been shown.[173]

That thiamine is concerned with the metabolism of carbohydrates is now well established. In fact, the only specific metabolic function of the vitamin for which there is definite biochemical evidence is its catalytic effect on the removal of pyruvic acid. This was shown first by Meiklejohn, Passmore, and Peters [174] and led to the catatorulin test, an *in vitro* test on polyneuritic pigeon brain tissue devised by Passmore, Peters, and Sinclair [175] and used so extensively by the Oxford University biochemists and others in studies of the antineuritic vitamin. Mathews' [176] suggestion that the vitamin may act as a respiratory catalyst by union with phosphoric acid and a suitable protein was confirmed by

[170] F. Bergel and A. R. Todd, *Nature*, **138**, 406 (1936).

[171] B. C. P. Jansen, *Rec. trav. chim.*, **55**, 1046 (1936).

[172] R. G. Green, W. E. Carlson, and C. A. Evans, *J. Nutrition*, **21**, 243 (1941); **23**, 165 (1942).

[173] D. J. Hennessy and S. Warner, Paper presented at 109th Meeting of American Chemical Society, Atlantic City, N. J., April 8–12, 1946.

[174] A. P. Meiklejohn, R. Passmore, and R. A. Peters, *Biochem. J.*, **26**, 1872 (1932).

[175] R. Passmore, R. A. Peters, and H. M. Sinclair, *Biochem. J.*, **27**, 842 (1933); *cf.* also H. W. Kinnersley, J. R. O'Brien, and R. A. Peters, *Biochem. J.*, **29**, 713 (1935), for modification of the method.

[176] A. P. Mathews, *Principles of Biochemistry*, p. 425, William Wood and Co., Baltimore, 1936.

Lohmann and Schuster,[177] who isolated from yeast a pyrophosphoric acid ester of thiamine which is a specific coenzyme for the carboxylase in the oxidation of pyruvic acid. Animal tissue contains thiamine largely in the ester form, but in plant tissue only a small part of the vitamin is present as the ester *cocarboxylase*. Of interest in this connection is the observation that cocarboxylase is much less stable to heat than is the free vitamin.

Support for the idea of a general relationship between this vitamin and carbohydrate metabolism is found in the series of biological experiments of Evans and co-workers,[178] Salmon and Goodman,[179] and many others on the sparing action of dietary fat on the nutritive requirements for the vitamin. Fat metabolism does not require the presence of cocarboxylase in the tissues. Accordingly the thiamine requirement is related to the *non-fat* caloric intake. McHenry[180] believes that the vitamin catalyzes oxidation of pyruvate in nervous tissue and the synthesis of fat from pyruvate in other tissues and is thus a specific agent in the intermediary metabolism of carbohydrate, the end product of which depends on the tissue involved.

Although the biochemical defect of thiamine deficiency has been demonstrated with only a few tissues, it seems probable that like effects occur also in the peripheral nerves and are responsible for some of the other well-known pathological effects of the vitamin deficiency, such as hypomotility and atony of the gastro-intestinal tract (which must be related to the specific anorexia of thiamine deficiency), a change in the heart beat (bradycardia or decreased heart rate in rats, tachycardia in humans), and other changes for which such a defect in nerve metabolism could be responsible.

The initial biochemical defects in nerve tissue are followed in more advanced stages by serious anatomic lesions, which occur primarily in the peripheral nerves, especially in the distal portions. For a detailed discussion of the pathology of various degrees of thiamine deficiency and beriberi consult Eddy and Dalldorf.[181]

Until recently, the thiamine content of natural foods, special products rich in the vitamin, such as yeast, wheat germ, and rice polish, and of proprietary concentrates was estimated by tedious biological methods,

[177] K. Lohmann and P. Schuster, *Naturwissenschaften*, **25**, 26 (1937); *Z. angew. Chem.*, **50**, 221 (1937).

[178] H. M. Evans, S. Lepkovsky, and E. A. Murphy, *J. Biol. Chem.*, **107**, 439 (1934).

[179] W. D. Salmon and J. G. Goodman, *J. Nutrition*, **13**, 477 (1937).

[180] E. W. McHenry, *Science*, **86**, 200 (1937).

[181] W. H. Eddy and G. Dalldorf, *The Avitaminoses*, Williams and Wilkins Co., Baltimore, 1937.

such as the growth rate of rats, prevention or cure of experimental polyneuritis in rats. Chemical and microbiological assays are now generally employed.

Most of the thiamine assays of foods, etc., are now reported in terms of milligrams or micrograms of the vitamin. The International Standard is a sample of the pure crystalline thiamine chloride hydrochloride, 3 μg. of which represents 1 International Unit.

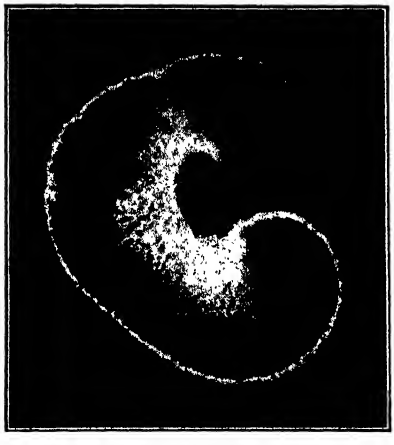


FIG. 117. Cross section of a wheat grain treated with an alkaline ferricyanide solution to convert thiamine to thiochrome, as viewed under ultraviolet light. The aleurone layer of rectangular cells immediately under the bran coats and the cells surrounding the crease of the grain show considerable fluorescence, indicating that the thiamine is located principally in these regions. [From G. F. Somers, M. H. Coolidge, and K. C. Hamner, *Cereal Chem.*, **22**, 333 (1945).]

Microorganisms are an important source of thiamine in nature. Higher plants appear to synthesize it and possibly employ it in their metabolism; they concentrate it in the seeds, especially in the embryo. Animals as widely separated, biologically, as mammals and insects require the vitamin in their food or must live symbiotically with microorganisms capable of forming it. The role of bacteria in supplying the thiamine requirements of ruminants was discovered by Bechdel, *et al.*[182]

Thiamine is fairly abundant in food products normally selected for the human dietary and withstands most cooking processes, although roasting causes quite a significant loss. Processing of foods often lowers the thiamine content, owing either to destruction or to rejection of vitamin-rich portions of the food material. As much as 60–80 per cent of the thiamine in meat may be destroyed during the prolonged heat sterilization. Thiamine is largely present in meat as cocarboxylase, which is more readily destroyed by heat than is the unesterified vitamin.[183] Sulfiting of vegetables or fruits, as for example during the dehydration of cabbage,[184] will destroy the thiamine. The milling of cereals such

[182] S. I. Bechdel, H. E. Honeywell, R. A. Dutcher, and M. H. Knutsen, *J. Biol. Chem.*, **80**, 231 (1928).

[183] R. G. Booth, *Biochem. J.*, **37**, 518 (1943).

[184] M. F. Mallette, C. R. Dawson, W. L. Nelson, and W. A. Gortner, *Ind. Eng. Chem.*, **38**, 437 (1946).

as wheat and rice invariably leads to a refined product of lower thiamine content, since the vitamin is concentrated in the outer layers of the kernel which normally go into the feed by-products (see Fig. 117).

Although beef muscle is relatively poor in thiamine, pork is one of the best sources, containing 0.8–1.5 mg. of vitamin B_1 per 100 grams. When the pigs are on a diet supplemented with this vitamin, the thiamine values for pork may be doubled.[185] Whole wheat contains 0.55 mg. and white flour only 0.07 mg. per 100 grams. Enrichment brings the latter up to 0.44–0.55 mg. per 100 grams (2.0–2.5 mg. per lb.). Fruits in general are rather poor contributors of thiamine, as is milk. Vegetables are not concentrated sources, with the exception of the legumes, which contain up to 0.5 mg. of thiamine per 100-gram serving. Peanuts are rich in the vitamin.

The National Research Council has recommended a daily dietary allowance of 1.0 mg. for a 55-lb. child, 1.8 mg. during pregnancy, and 1.5 mg. for maintenance in a 70-kg. man. For the growing swine, 2.5 mg. of thiamine is suggested, the adult allowance being 3.0 mg. Chicks should have 0.9 mg. of thiamine per pound of feed for optimal nutrition. It is now generally recognized that the requirements for this vitamin are related to the caloric value of the food consumed, and particularly to the carbohydrate intake. Work in Elvehjem's laboratory [186] has demonstrated that, if fat replaces carbohydrate in the diet, the thiamine requirement of growing rats is greatly diminished despite maintenance of the caloric intake.

Riboflavin. As the result of numerous studies culminating with those of Smith and Hendrick,[187] of Goldberger [188] and associates, as well as many other investigators, it became generally recognized that so-called vitamin B as it concerns normal growth of rats consists of some substance in addition to the antineuritic vitamin and that this substance remains after the antineuritic property of yeast and other products is destroyed by autoclaving. This factor came to be known as vitamin G or vitamin B_2 and, because its deficiency seemed to be accompanied by the appearance of pellagra-like skin lesions in rats, became widely accepted as the antipellagra vitamin. When the growth factor was isolated it was found to have neither antidermatitis properties for mammals nor antipellagra properties for humans, but instead marked redox properties, especially

[185] J. W. Pence, R. C. Miller, R. A. Dutcher, and P. T. Ziegler, *J. Animal Sci.*, **4**, 141 (1945).

[186] F. E. Stirn, A. Arnold, and C. A. Elvehjem, *J. Nutrition*, **17**, 485 (1939).

[187] M. I. Smith and E. G. Hendrick, *U. S. Pub. Health Repts.*, **41**, 201 (1926).

[188] J. Goldberger, G. A. Wheeler, R. D. Lillie, and L. M. Rogers, *U. S. Pub. Health Repts.*, **41**, 297 (1926); J. Goldberger and R. D. Lillie, *ibid.*, **41**, 1025 (1926).

as a portion of a ferment molecule. In addition it stimulated growth of young rats and other animals and prevented various pathological conditions. However, the biochemical studies which led to the isolation of the vitamin began in the field of enzymes, not in the field of nutrition.

Banga and Szent-Györgyi [189] described a yellow water-soluble pigment accompanying an oxidation coenzyme in heart muscle. The color of the pigment, named "cytoflav," was destroyed by reducing agents and restored by atmospheric oxygen. Almost simultaneously Warburg and Christian [190] announced their discovery from bottom yeast of the yellow "oxidation" enzyme, the color of whose aqueous solution, like that of "cytoflav," was destroyed on reduction. The suggestion was made that "cytoflav" might be a decomposition product of the yellow enzyme. At the same time the occurrence of such a pigment, characterized by strong green fluorescence, in all extracts containing vitamin B_2 attracted the attention of Kuhn [191] and associates, who soon were able to produce convincing evidence [192] of the growth-promoting effects of the pigment when given to rats placed on so-called vitamin B_2-deficient diets. Also Booher,[193] working in Dr. Sherman's laboratory in Columbia University, produced important evidence that the Bourquin-Sherman [194] units of vitamin B_2, representing a typical biological method for estimating the vitamin, actually measure the growth-promoting flavin content of the test substance.

The widespread distribution of an identical flavin in various plant and animal tissues and products was soon established, and the synthesis of the natural vitamin was accomplished by both Karrer [195] and Kuhn [196] and their associates.

[189] I. Banga and A. Szent-Györgyi, *Biochem. Z.*, **246**, 203 (1932).

[190] O. Warburg and W. Christian, *Biochem. Z.*, **254**, 438 (1932).

[191] R. Kuhn, P. György, and T. Wagner-Jauregg, *Ber.*, **66B**, 317 (1933).

[192] R. Kuhn, P. György, and T. Wagner-Jauregg, *Ber.*, **66B**, 576 (1933); **66B**, 1034 (1933); R. Kuhn and T. Wagner-Jauregg, *ibid.*, **66B**, 1577 (1933); R. Kuhn, H. Rudy, and T. Wagner-Jauregg, *ibid.*, **66B**, 1950 (1933); P. György, R. Kuhn, and T. Wagner-Jauregg, *Z. physiol. Chem.*, **223**, 241 (1934); R. Kuhn and T. Wagner-Jauregg, *Ber.*, **67B**, 1770 (1934).

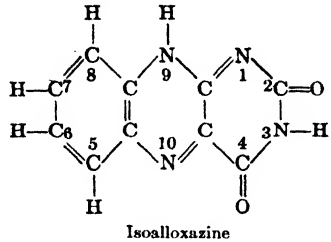
[193] L. E. Booher, *J. Biol. Chem.*, **102**, 39 (1933); **107**, 591 (1934).

[194] A. Bourquin and H. C. Sherman, *J. Am. Chem. Soc.*, **53**, 3501 (1931).

[195] P. Karrer, K. Schöpp, F. Benz, and K. Pfaehler, *Helv. Chim. Acta*, **18**, 69 (1935); P. Karrer, K. Schöpp, and F. Benz, *ibid.*, **18**, 426 (1935); H. von Euler, P. Karrer, M. Malmberg, K. Schöpp, F. Benz, B. Becker, and P. Frei, *ibid.*, **18**, 522 (1935).

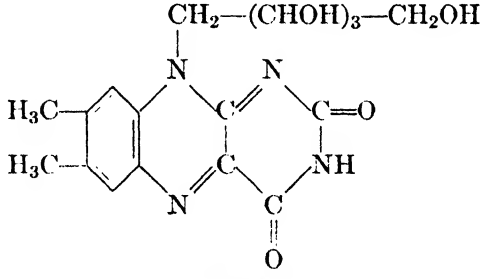
[196] R. Kuhn, K. Reinemund, F. Weygand, and R. Ströbele, *Ber.*, **68B**, 1765 (1935); R. Kuhn, K. Reinemund, H. Kaltschmitt, R. Ströbele, and H. Trischmann, *Naturwissenschaften*, **23**, 260 (1935).

Flavins are derivatives of the heterocyclic isoalloxazine nucleus:



Isoalloxazine

Riboflavin (lactoflavin or ovoflavin or vitamin G) is 6,7-dimethyl-isoalloxazine-9-D-riboside. It is the water-soluble, yellow (with a greenish fluorescence) pigment which occurs in milk whey and in both egg yolk and egg white. Since its identification in these substances, it has been shown to occur in green leaves and in a great variety of plant and animal tissues. In Karrer's synthesis 2-amino-4,5-dimethylcarbethoxy anilide was treated with D-ribose under reducing conditions to form the corresponding amino riboside. The carbethoxy group on the other amino group was then saponified with alkali, and the resulting riboside was condensed with alloxan to form riboflavin:



Riboflavin

It was identical in all respects with the natural product. Riboflavin, when acted on by light under alkaline conditions, gives rise to lumiflavin (6,7,9-trimethylisoalloxazine). Under neutral conditions, in methanol, lumichrome (6,7-dimethylalloxazine) results.

The flavin vitamin thus has the properties of four different classes of organic compounds, those of the sugars through its pentose side chain, of the pyrimidines through its alkali-labile ring, of the azine dyes, and of the benzene derivatives, the ortho-position of the methyl groups on the benzene ring being very unusual. The sugar, attached as non-glycoside, differentiates the vitamin from the nucleosides and is regarded as accounting for its stability toward heat and acids.

Inasmuch as the natural flavin vitamin is widely distributed in nature, its earlier designation as lactoflavin, ovoflavin, hepatoflavin, etc., depending on the source from which it is isolated, has been discontinued in favor of a universal chemical designation, riboflavin (actually D-riboflavin). Karrer has shown that ribose may be replaced by arabinose and that either (but not both) of the benzene methyls may be eliminated or replaced by ethyl without much loss in vitamin activity, but the remarkable result is that the biologically active araboflavin is the one which contains the unnatural optical form of arabinose.

Riboflavin, apparently, occurs chiefly as a monophosphoric ester or as an alloxazine-adenine-dinucleotide and carries out its functions either as such or as part of several enzyme systems concerned with tissue respiration, including xanthine oxidase, D-amino acid oxidase, cytochrome reductase, and Warburg's respiratory yellow enzyme (see Chapter 38). Participation of riboflavin in the oxidation-reduction reactions in tissues is accomplished by the reversible uptake of hydrogen at positions 1 and 10. The phosphorus-free pigment functions as effectively as the ester as a dietary supplement,[197] presumably being phosphorylated in the body cells under the control of the adrenal cortical hormones.

The differentiation of the flavin vitamin from antidermatitis food factors was accomplished by György [198] and confirmed by Harris,[199] by Chick,[200] by Copping,[201] by Bender [202] and associates, and by others. That riboflavin is not concerned with black tongue or with human pellagra was established by Birch and associates,[203] Koehn and Elvehjem,[204] and Sebrell and associates,[205] and other investigators.

Uncomplicated riboflavin deficiency produces alopecia, vascularization of the cornea, and often cataract in rats.[206] Riboflavin is present in appreciable amounts in normal retinal pigment epithelium. In dogs [207] a progressive deficiency disease slowly develops, characterized by loss

[197] P. György, *Proc. Soc. Exptl. Biol. Med.*, **35**, 207 (1936).
[198] P. György, *Biochem. J.*, **29**, 741 (1935).
[199] L. J. Harris, *Biochem. J.*, **29**, 776 (1935).
[200] H. Chick and A. M. Copping, *Biochem. J.*, **29**, 722 (1935).
[201] A. M. Copping, *Biochem. J.*, **30**, 845 (1936).
[202] R. C. Bender, S. Ansbacher, G. E. Flanigan, and G. C. Supplee, *J. Nutrition*, **11**, 391 (1936).
[203] T. W. Birch, P. György, and L. J. Harris, *Biochem. J.*, **29**, 2830 (1935).
[204] C. J. Koehn, Jr., and C. A. Elvehjem, *J. Nutrition*, **11**, 67 (1936).
[205] W. H. Sebrell, D. J. Hunt, and R. H. Onstott, *U. S. Pub. Health Repts.*, **52**, 235 (1937).
[206] P. L. Day, W. J. Darby, and W. C. Langston, *J. Nutrition*, **13**, 389 (1937).
[207] H. M. Zimmerman, G. R. Cowgill, and J. C. Fox, *Arch. Neurol. Psychiat.*, **37**, 286 (1937).

of weight, vomiting and bloody diarrhea, flaccid muscular weakness, incoordination and decrease in the deep reflexes. According to Lepkovsky and Jukes,[208] riboflavin deficiency in chicks results in weakness, emaciation, and diarrhea, but in turkey poults there is an intense dermatitis, accompanied by slow growth and high mortality. All higher animals are assumed to require riboflavin for the synthesis of flavoproteins, which function as intermediary catalysts in vital dehydrogenation processes.

Riboflavin deficiency in humans, as observed by Sebrell and Butler,[209] is evidenced by lip lesions, as well as lesions about the ears, nose, and eyelids. Fatigue, loss of appetite, and other disturbances also were present. Corneal vascularization and photophobia also usually are evident in ariboflavinosis [210] but are not necessarily specifically correlated with a deficiency in the vitamin.

As previously stated, the biological method of assay for riboflavin is the growth method, in which rats are used; it was originally devised by Bourquin and Sherman for vitamin B_2 and later improved and modified by a number of investigators. The marked fluorescence of riboflavin solutions has provided a physical method of assay for this substance. The fluorescent emissions under ultraviolet light of unknowns compared with a standard riboflavin solution make possible the detection of concentrations as low as a fraction of a microgram per milliliter. Microbiological assay methods also are used in analyzing for riboflavin. The most common of these technics is the use of *Lactobacillus casei* as the microorganism whose growth is limited by the riboflavin content of the medium. Titration of the lactic acid produced is a suitable means of assessing growth of the bacteria and, hence, riboflavin in the extract. This method of assay gives values which generally correlate well with those obtained by fluorometric analysis.

The Bourquin-Sherman unit appears to be equivalent to 2.5 μg. of crystalline riboflavin.[211] The concentration of riboflavin in foods now is usually expressed in terms of weight rather than units. Cheddar cheese is very rich in this vitamin, having over 500 μg. per 100-gram portion. Milk contains 200 μg. per 100 grams when protected from light. In sunlight, riboflavin is rapidly destroyed even in opaque material such as

[208] S. Lepkovsky and T. H. Jukes, *J. Nutrition*, **12**, 515 (1936).

[209] W. H. Sebrell and R. E. Butler, *U. S. Pub. Health Repts.*, **53**, 2282 (1938); **54**, 2121 (1939).

[210] O. A. Bessey and S. B. Wolbach, *J. Exptl. Med.*, **69**, 1 (1939); H. D. Kruse, V. P. Sydenstricker, W. H. Sebrell, and H. M. Cleckley, *U. S. Pub. Health Repts.*, **55**, 157 (1940); V. P. Sydenstricker, W. H. Sebrell, H. M. Cleckley, and H. D. Kruse, *J. Am. Med. Assoc.*, **114**, 2437 (1940).

[211] O. A. Bessey, *J. Nutrition*, **15**, 11 (1938).

milk; two-thirds of the riboflavin may be lost in a two-hour exposure at room temperature.[212]

Meat and fish contain riboflavin to the extent of approximately 200 μg. per 100 grams. Organ tissues, such as beef, pork, and lamb liver, are exceedingly rich sources of the vitamin, containing 3–5 mg. per 100-gram serving. Eggs and most green vegetables are good sources, whereas non-leafy vegetables and most fresh fruits are rather poor sources of riboflavin. Even whole wheat flour, which contains approximately 100 μg. per 100 grams, is not a good source of riboflavin. White flour has only about a third of this potency, but when enriched according to government standards it contributes 550–680 μg. per 100 grams (1.2–1.5 mg. per lb.).

Recommended daily dietary allowances of riboflavin range from 0.6 mg. for the infant to 3.0 mg. for a woman in lactation. For adult maintenance, 2.0 mg. of riboflavin is suggested. For growing swine, 3.8 mg. is the recommended allowance. For chicks, the feed should contain 0.9 mg. per lb., and for turkey poults, a level of 2.0 mg. of riboflavin per pound of feed is proposed.

Nicotinic Acid (Niacin) or Nicotinamide (Niacin Amide), the Pellagra-Preventive (P.P.) Factor. Pellagra is the only recognized extreme vitamin deficiency disease that is still widespread in the United States. With the accumulation of man's knowledge of the vitamins, cases of scurvy, xerophthalmia, beriberi, and rickets have become increasingly rarer in this country. But despite our knowledge of the cause and cure of pellagra, this affliction remains as a major medical-nutritional problem in many areas, particularly in the Southern states.[213] The establishment of a distinct vitamin-deficiency etiology for rat acrodynia raised the question of the relation between pellagra dermatitis and pellagra-like symptoms in the rat, dog, and chick. For some time so-called "rat pellagra" and black tongue in dogs [214] had been accepted as analogs of human pellagra, and later [215] similar manifestations in chicks were produced. Evidence was presented by Birch, György, and Harris [216] that the pellagra preventive vitamin was distinct from the vitamin B_6 associated with "rat pellagra." Subsequently, it has also

[212] W. J. Peterson, F. M. Haig, and A. O. Shaw, *J. Am. Chem. Soc.*, **66**, 662 (1944).

[213] See the review on the incidence and mortality of pellagra in *Nutrition Revs.*, **4**, 304 (1946).

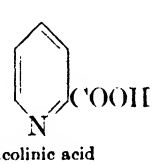
[214] J. Goldberger and G. A. Wheeler, *U. S. Pub. Health Repts.*, **43**, 172 (1928).

[215] A. T. Ringrose, L. C. Norris, and G. F. Heuser, *Poultry Sci.*, **10**, 166 (1931); O. L. Kline, J. A. Keenan, C. A. Elvehjem, and E. B. Hart, *J. Biol. Chem.*, **99**, 295 (1932); C. A. Elvehjem and C. J. Koehn, Jr., *ibid.*, **108**, 709 (1935); S. Ansbacher, G. C. Supplee, and R. C. Bender, *J. Nutrition*, **11**, 529 (1936).

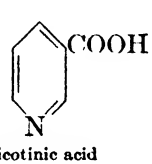
[216] T. W. Birch, P. György, and L. J. Harris, *Biochem. J.*, **29**, 2830 (1935).

been shown that "chick pellagra" is associated with a deficiency of pantothenic acid rather than nicotinic acid. However, Briggs, *et al.*,[217] have demonstrated that chicks also require nicotinic acid and, in its absence from the diet, will develop symptoms analogous to black tongue in dogs. Many independent experimental studies indicate an identical vitamin-deficiency etiology being involved in dog and chick black tongue and human pellagra.

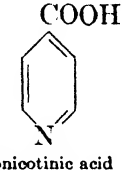
The most successful of the many attempts to concentrate the anti-pellagra vitamin has been made by Elvehjem [218] and associates, who brought the search to a successful culmination in the isolation of nicotinic acid amide from anti-black tongue concentrates from liver extract, this well-known pyridine compound and also the free acid showing marked black-tongue-curing as well as prophylactic properties. The specificity of the structure was evidenced by the failure of picolinic acid or iso-nicotinic acid and many other closely related compounds to prevent canine black tongue.[219]



Picolinic acid



Nicotinic acid



Isonicotinic acid

Nicotinic acid may be prepared by oxidation of nicotine or by oxidation of quinoline to quinolic acid, which upon heating loses carbon dioxide to yield nicotinic acid. The antipellagra vitamin was the first vitamin ever synthesized, since nicotinic acid was known to the chemist two-thirds of a century before its biological significance was recognized. Early in the study of the antineuritic vitamin, nicotinamide was isolated by Funk from natural materials, but its role as a vitamin was not recognized. Indeed, even before its vitamin character was discovered nicotinamide was shown [220] to be a functional group in coenzymes I and II, the diphospho- and triphosphopyridine nucleotides required by a number of dehydrogenases and by the riboflavin-containing "yellow

[217] G. M. Briggs, Jr., R. C. Mills, C. A. Elvehjem, and E. B. Hart, *Proc. Soc. Exptl. Biol. Med.*, **51**, 59 (1942).

[218] C. A. Elvehjem and C. J. Koehn, Jr., *J. Biol. Chem.*, **108**, 709 (1935); C. J. Koehn and C. A. Elvehjem, *ibid.*, **118**, 693 (1937); C. A. Elvehjem, R. J. Madden, F. M. Strong, and D. W. Woolley, *J. Am. Chem. Soc.*, **59**, 1767 (1937).

[219] D. W. Woolley, F. M. Strong, R. J. Madden, and C. A. Elvehjem, *J. Biol. Chem.*, **124**, 715 (1938).

[220] O. Warburg and W. Christian, *Biochem. Z.*, **275**, 464 (1935); H. Euler, H. Albers, and F. Schlenk, *Z. physiol. Chem.*, **237**, 1 (1935); F. Schlenk and H. Euler, *Naturwissenschaften*, **24**, 794 (1936).

enzyme" (see Chapter 38). The nicotinamide moiety is responsible for the oxidation-reduction properties of the coenzymes evidenced in their participation in intracellular oxidation of carbohydrate.

Evidence has also been presented [221] for the existence of another naturally occurring compound containing nicotinic acid. This material accounts for a considerable part of the vitamin in cereal products and is characterized by its extreme susceptibility to alkaline hydrolysis. Work in Elvehjem's laboratory [222] has indicated that the compound is probably not a simple sugar derivative (nucleoside), that it resembles the simple esters of nicotinic acid, and that it is probably biologically active as a source of the vitamin for dogs.

Elvehjem and Teply [223] have prepared an excellent summary of the structure, activity, and estimation of a considerable number of natural products related to nicotinic acid. Chemical assay generally involves the colorimetric determination of the yellow pigment produced by reacting nicotinic acid with cyanogen bromide and an aromatic amine, such as aniline or *p*-aminoacetophenone. The difficulty posed by the presence of other chromogen materials has led to widespread use of the microbiological assay for nicotinic acid, with *Lactobacillus arabinosus* as the test organism.

Pellagra is characterized by a typical rough-red dermatitis accompanied by painful and ulcerated mouth and tongue. Disturbances of the digestive processes and of the nervous system also are present. Shortly after the establishment of nicotinic acid as the anti-black-tongue factor, confirmation was forthcoming as to its value in pellagra therapy. Fouts and associates [224] reported improvement in four pellagra patients after the administration of 0.5–1.0 gram of nicotinic acid. Spies [225] has reported the use of nicotinic acid for the successful and rapid cure of a considerable number of pellagra patients but has again called attention to the multiple nature of the deficiency. Pellagra is generally complicated by accompanying dietary deficiencies in other vitamins of the B complex, including both thiamine and riboflavin.

Although the free acid and the amide are of equal therapeutic effective-

[221] J. S. Andrews, H. M. Boyd, and W. A. Gortner, *Ind. Eng. Chem., Anal. Ed.*, **14**, 663 (1942).

[222] L. J. Teply and C. A. Elvehjem, *Proc. Soc. Exptl. Biol. Med.*, **55**, 72 (1944); W. A. Krehl and F. M. Strong, *J. Biol. Chem.*, **156**, 1 (1944); W. A. Krehl, C. A. Elvehjem, and F. M. Strong, *ibid.*, **156**, 13 (1944).

[223] C. A. Elvehjem and L. J. Teply, *Chem. Revs.*, **33**, 185 (1943).

[224] P. J. Fouts, O. M. Helmer, S. Lepkovsky, and T. H. Jukes, *Proc. Soc. Exptl. Biol. Med.*, **37**, 405 (1937).

[225] T. D. Spies, C. Cooper, and M. A. Bankenhorn, *J. Am. Med. Assoc.*, **110**, 622 (1938); T. D. Spies, W. B. Bean, and R. E. Stone, *ibid.*, **111**, 584 (1938).

ness, nicotinamide has the advantage of not inducing the flushing of the skin—a common after effect of nicotinic acid administration.

A number of investigators have shown that certain diets are associated with a greater incidence of pellagra than other diets with a lower nicotinic acid content. Corn in particular has long been associated with pellagra. Recent findings are beginning to shed light on the bases for these observations. A group working at Wisconsin [226] have shown that corn will inhibit the growth of rats when the other components of the diet are low in tryptophan. Nicotinic acid or tryptophan will restore normal growth. Krehl, *et al.*, present strong evidence for the interrelationship of the dietary requirements for tryptophan and for the antipellagra vitamin and note that pellagra-producing diets tend to be low in both of these nutrients. It is noteworthy that milk, which Goldberger found to be an antipellagra food, proves to be poor in nicotinic acid but an excellent source of tryptophan (and of riboflavin, deficient in many pellagragenic diets). Rosen, Huff, and Perlzweig [227] found an increased excretion of nicotinic acid derivatives in the urine but not in the feces after tryptophan administration, and they consider the amino acid to be a precursor for synthesis of the vitamin in the rat.

Woolley [228] has presented evidence of another explanation for the pellagra-producing role of corn. He has reported that corn contains a weakly basic, water-soluble pellagragenic agent. At a level of 1 mg. of concentrate per 100 grams of ration, this material induced pellagra-like symptoms in mice which could be cured or prevented by nicotinamide. Of interest in this connection is Woolley's earlier observation that 3-acetylpyridine, a structural analog of nicotinic acid, serves as an "antivitamin" and produces a pellagra-like disease in mice which either tryptophan or nicotinic acid can prevent.[229]

Nicotinic acid is excreted as its betaine, trigonelline, as nicotinylglycine or nicotinuric acid, and as an ultraviolet-fluorescing substance F_2,[230] probably identical with N^1-methyl-nicotinamide,[231] which is absent from the urine of pellagrins. Considerable amounts of N^1-

[226] W. A. Krehl, L. J. Teply, and C. A. Elvehjem, *Science*, **101**, 283 (1945); W. A. Krehl, L. J. Teply, P. S. Sarma, and C. A. Elvehjem, *ibid.*, **101**, 489 (1945); W. A. Krehl, P. S. Sarma, L. J. Teply, and C. A. Elvehjem, *J. Nutrition*, **31**, 85 (1946); W. A. Krehl, P. S. Sarma, and C. A. Elvehjem, *J. Biol. Chem.*, **162**, 403 (1946).

[227] F. Rosen, J. W. Huff, and W. A. Perlzweig, *J. Biol. Chem.*, **163**, 343 (1946).

[228] D. W. Woolley, *J. Biol. Chem.*, **163**, 773 (1946).

[229] D. W. Woolley, *J. Biol. Chem.*, **157**, 455 (1945); **162**, 179 (1946).

[230] V. A. Najjar and R. W. Wood, *Proc. Soc. Exptl. Biol. Med.*, **44**, 386 (1940); V. A. Najjar and L. E. Holt, Jr., *Science*, **93**, 20 (1941).

[231] J. W. Huff and W. A. Perlzweig, *Science*, **97**, 538 (1943); *J. Biol. Chem.*, **150**, 395 (1943).

methyl-6-pyridone-3-carboxylamide, an oxidation product of N-methyl-nicotinamide, are excreted along with the latter in normal urine.[232] Transmethylation from a material such as methionine or choline is involved in trigonelline and N-methyl-nicotinamide formation and imposes a drain on the supply of these essential methyl donors in the body.[233] The methylation of nicotinamide apparently takes place in the liver.

Nicotinic acid and its amide are found in only small amounts in milk and eggs (8 μg. per 100 grams) and in only fair amounts in most fruits and vegetables (0.1–0.5 mg. per 100 grams). Most meats are good sources of the vitamin, containing 4–7 mg. per 100-gram portion. Among organ tissues, the liver is especially rich in nicotinic acid, containing 3–4 times more than is present in muscle tissue. Most milled cereals, including corn, contain about 1 mg. of niacin per 100 grams. Although white flour is at about this level, whole wheat flour contains 4–6 mg. per 100 grams. Enrichment of flour brings the nicotinic acid content up to 3.5–4.5 mg. per 100 grams (16–20 mg. per lb.).

Tentative allowances for nicotinic acid have been set up by the National Research Council. They include a dietary allowance of 10 mg. daily for a 55-lb. child, 15 mg. for the adult, and 18 mg. during pregnancy. For swine, an allowance of 12.5 mg. is recommended for a 100-lb. animal. For growing chicks, 8 mg. of nicotinic acid per pound of feed is suggested.

Vitamin B_6 and Related Compounds. When György [234] demonstrated that flavin failed to prevent the dermatitis in rats which had been regarded as one of the specific effects of vitamin B_2 deficiency, he proposed that the "rat-antipellagra" factor be called vitamin B_6. The B_6 deficiency dermatitis was of the florid type, appearing bisymmetrically on the rats and thus resembling human pellagra, although of different etiological origin. It is called rat acrodynia to differentiate it from pellagra.

In 1938, Lepkovsky, Keresztesy, and Stevens, and Kuhn and Wendt, all succeeded in isolating crystalline "vitamin B_6" from natural sources, including yeast and rice polishings. As with most discoveries in the vitamin field in recent years, only a short time elapsed between isolation of the pure compound and assignment of molecular structure. Kuhn and his associates [235] showed vitamin B_6 (adermin) to be 2-methyl-3-

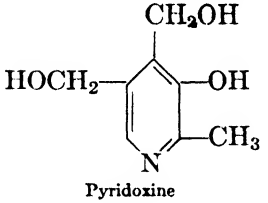
[232] W. E. Knox and W. I. Gussman, *J. Biol. Chem.*, **166**, 391 (1946).

[233] P. Handler and W. J. Dann, *J. Biol. Chem.*, **146**, 357 (1942).

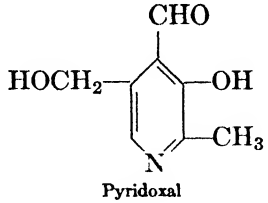
[234] P. György, *Nature*, **133**, 498 (1934).

[235] R. Kuhn, G. Wendt, and K. Westphal, *Ber.*, **72B**, 310 (1939); see also S. A. Harris and K. Folkers, *Science*, **89**, 347 (1939).

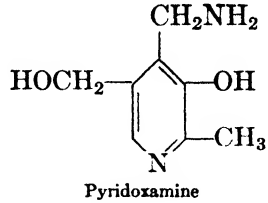
hydroxy-4,5-di-(hydroxymethyl)-pyridine. This compound has been given the name *pyridoxine*. Recently, workers at the University of Texas and at Cornell University have demonstrated that several compounds related to pyridoxine have specific biological activities, indicating the existence of a vitamin B_6 group. The close structural relationships of these various compounds are evident from their formulae:



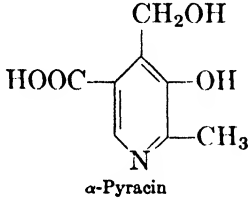
Pyridoxine



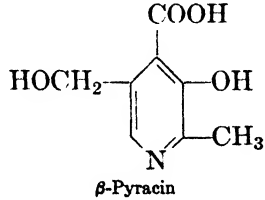
Pyridoxal



Pyridoxamine



α-Pyracin



β-Pyracin

Pyridoxine is stable to heat and to alkali but is light-sensitive, especially to ultraviolet rays and in an alkaline environment. Pyridoxamine and pyridoxal are even more sensitive to light.[236] All three of these compounds are readily destroyed by oxidizing agents.

The multiple nature of vitamin B_6 was first noted by Snell and his associates,[237] who discovered that a "pseudopyridoxine" occurring along with the vitamin in natural materials greatly exceeded pyridoxine in promoting growth of *Streptococcus faecalis* R and *Lactobacillus casei*. "Pseudopyridoxine" was later shown [238] to consist of pyridoxal and pyridoxamine, which are now recognized as active metabolic forms of vitamin B_6. Work of Scott and his associates [239] on factors required by chicks led to the discovery and synthesis of additional B_6 derivatives which they called α- and β-pyracin. Both the free acid and the lactone showed marked biological activity.

[236] E. Cunningham and E. E. Snell, *J. Biol. Chem.*, **158**, 491 (1945).

[237] E. E. Snell, B. M. Guirard, and R. J. Williams, *J. Biol. Chem.*, **143**, 519 (1942).

[238] E. E. Snell, *J. Biol. Chem.*, **154**, 313 (1944); E. E. Snell and A. N. Rannefeld, *ibid.*, **157**, 475 (1945); E. E. Snell, *ibid.*, **157**, 491 (1945).

[239] M. L. Scott, L. C. Norris, G. F. Heuser, W. F. Bruce, H. W. Coover, Jr., W. D. Bellamy, and I. C. Gunsalus, *J. Biol. Chem.*, **154**, 713 (1944); M. L. Scott, L. C. Norris, G. F. Heuser, and W. F. Bruce, *ibid.*, **158**, 291 (1945); M. L. Scott, L. C. Norris, and G. F. Heuser, *Science*, **103**, 303 (1946); M. L. Scott, L. C. Norris, L. W. Charkey, L. J. Daniel, and G. F. Heuser, *J. Biol. Chem.*, **164**, 403 (1946).

Vitamin B_6 deficiency in the rat is characterized by a dermatitis and impaired growth. Dogs, pigs, and chicks develop a hypochromic anemia in addition to a restriction of growth when the diet is inadequate in members of the B_6 group. Synthesis of pyridoxine in the rumen normally makes the cow relatively independent of a dietary source of vitamin B_6, but recent evidence [240] indicates that poikilocytosis in dairy cattle, characterized by abnormal erythrocytes, anorexia, and retarded growth, is due to inadequate pyridoxine in the diet and an inability of the animals to synthesize the vitamin.

The biochemical functioning of these compounds has been elucidated by the brilliant researches of Gunsalus and his associates, who have shown that a phosphorylated derivative of pyridoxal is the coenzyme for decarboxylation of tyrosine [241] and of arginine and glutamic acid,[242] the coenzyme for the transaminase in the glutamate-aspartate system,[243] and for the synthesis of tryptophan from indole [244] by cell-free enzyme preparations. Work of other investigators has extended the list of amino acid decarboxylases for which pyridoxal phosphate is required. The *codecarboxylase* (pyridoxal phosphate) content of bacterial cells and of animal tissues [245] is directly related to the content of pyridoxine or its derivatives in the diet. It thus appears that vitamin B_6 has a major role in protein metabolism, including both decarboxylation and transamination. This phase has been excellently summarized in *Nutrition Reviews*.[246]

The role played by pyracin in hematopoiesis in chicks appears to be an indirect one. The work of Scott and his associates [247] has demonstrated that either α- or β-pyracin (or the pyracin lactones) is needed for the full activity of fermentation *L. casei* factor (folic acid conjugate, see p. 940) in hemorrhagic or certain nutritional anemias. Two lines of evidence, one on the utilization of *L. casei* factor for growth of *S. faecalis* R [247] and the other on the utilization of the conjugate for hematopoiesis in chicks,[248] suggest that pyracin may be an essential com-

[240] J. T. Reid, C. F. Huffman, and C. W. Duncan, *J. Nutrition*, **30**, 413 (1945).

[241] I. C. Gunsalus, W. D. Bellamy, and W. W. Umbreit, *J. Biol. Chem.*, **155**, 685 (1944).

[242] W. W. Umbreit and I. C. Gunsalus, *J. Biol. Chem.*, **159**, 333 (1945).

[243] H. C. Lichstein, I. C. Gunsalus, and W. W. Umbreit, *J. Biol. Chem.*, **161**, 311 (1945).

[244] W. W. Umbreit, W. A. Wood, and I. C. Gunsalus, *J. Biol. Chem.*, **165**, 731 (1946).

[245] W. D. Bellamy, W. W. Umbreit, and I. C. Gunsalus, *J. Biol. Chem.*, **160**, 461 (1945).

[246] *Nutrition Revs.*, **3**, 72, 343 (1945); **4**, 232 (1946).

[247] L. J. Daniel, M. L. Scott, L. C. Norris, and G. F. Heuser, *J. Biol. Chem.*, **160**, 265 (1945).

[248] M. L. Scott, L. C. Norris, L. W. Charkey, L. J. Daniel, and G. F. Heuser, *J. Biol. Chem.*, **164**, 403 (1946).

ponent in the enzyme system which liberates folic acid from its complexes. Free folic acid does not require the presence of pyracin for normal red blood cell maturation.

The metabolic significance of the relationship of the essential fatty acids to vitamin B_6 in rat acrodynia is still uncertain. A number of workers,[249] however, have reported that pyridoxine and linoleic acid supplement each other in curing acrodynia, although it is unlikely that either can effect a complete cure in the absence of the other. Whether pyridoxine is in some way involved in fat metabolism, or whether acrodynia is merely to be considered a multiple-deficiency disease analogous to most cases of pellagra, remains to be established.

Many of the data on vitamin B_6 in natural products deal only with pyridoxine as measured by chemical analysis or by yeast growth. Microbiological and biological technics now are available to measure simultaneously other compounds with B_6 activity. Milk is a very poor source of pyridoxine (6 μg. per 100 ml.) but a relatively good source of the B_6 complex (0.1–0.2 mg. per 100 ml.). Vegetables as a class contain approximately 0.1 mg. per 100-gram portion. Cereals, fish, and meats are good sources of vitamin B_6, containing 0.4–0.5 mg. per 100 grams. White flour contains only half as much of the vitamin as does whole wheat flour.

The Committee on Animal Nutrition of the National Research Council has recommended a daily intake of 3.0 mg. of vitamin B_6 for a 100-lb. swine, and a content of 1.6 mg. per pound of feed for poultry rations. No recommendation has been made for a B_6 allowance for humans.

Biotin. Boas [250] was the first to observe marked dermatitis and other symptoms in rats fed dried egg white as the sole protein in an apparently complete ration for growth. Heat-coagulated egg white did not have this effect. Various substitutions in the original diet or additions to it prevented the syndrome, and a protective factor *X* was postulated. Later this hypothesis was abandoned in favor of a direct toxicity effect caused by a substance produced in the drying and destroyed by heat coagulation.

Various vitamins concerned with normal skin development were called H (from *Haut*) vitamins by György.[251] Later the only "skin" vitamin called vitamin H was the one preventing the so-called egg-white injury.

Extensive studies on this problem were made by Parsons and her associates at the University of Wisconsin. For some years, many workers

[249] F. W. Quackenbush, H. Steenbock, F. A. Kummerow, and B. R. Platz, *J. Nutrition*, **24**, 225 (1942). See also the discussion and citations bearing on this subject in the review article of G. O. Burr, *Federation Proc.*, **1**, 224 (1942).

[250] Margaret A. Boas, *Biochem. J.*, **18**, 422 (1924).

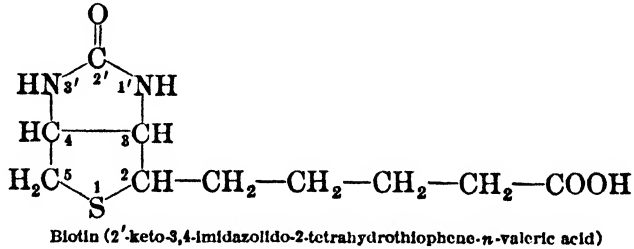
[251] P. György, *Z. ärztl. Fortbild.*, **28**, 377, 417 (1931).

believed that the protective factor in liver and other natural products was an anti-agent against a positive egg-white injury—a toxicity in a broad sense. They admitted the possibility of a vitamin deficiency as advanced by Ringrose and Norris [252] but pointed out that it seemed incompatible with the fact that, the greater the concentration of egg white in the ration, the greater the amount of the protective factor necessary for curing the syndrome. Another interesting point was the demonstration that only slight peptic or acid or tryptic digestion or heat denaturation of the harmful egg white rendered it "non-toxic." Parsons regarded this as a destruction of the toxic principle, Ringrose and Norris as the release (or formation) of the protective vitamin.

We now know that the puzzling syndrome is indeed a deficiency disease in which the vitamin is rendered biologically unavailable by being bound to a protein constituent of egg white.

An important step in this field was the establishment [253] in 1940 of the identity of *vitamin H* [251] (the factor which prevented "injury" due to presence of raw egg white in the diet of animals) with both *coenzyme R* [254] (the growth and respiration factor for the legume nodule organism, *Rhizobium trifolic*) and *biotin* [255] (the growth factor required by yeast and by various other microorganisms).

The chemical nature of this biologically active substance, made available in crystalline form as the result of Kögl's work, was elucidated by studies in a number of laboratories, and especially at the Cornell Medical School. As a result of these studies, the structure of biotin was formulated by duVigneaud and his associates [256] and subsequently was proved to be correct by synthesis in the Merck laboratories.[257]



Biotin (2′-keto-3,4-imidazolido-2-tetrahydrothiophene-*n*-valeric acid)

[252] A. T. Ringrose and L. C. Norris, *J. Nutrition*, **12**, 535 (1936); **12**, 553 (1936).

[253] V. duVigneaud, D. Melville, P. György, and C. Rose, *Science*, **92**, 62 (1940); P. György, C. Rose, K. Hofmann, D. Melville, and V. duVigneaud, *ibid.*, **92**, 609 (1940).

[254] F. E. Allison, S. R. Hoover, and D. Burk, *Science*, **78**, 217 (1933).

[255] F. Kögl and B. Tonnis, *Z. physiol. Chem.*, **242**, 43 (1936).

[256] V. duVigneaud, *Science*, **96**, 455 (1942).

[257] S. A. Harris, D. E. Wolf, R. Mozingo, and K. Folkers, *Science*, **97**, 447 (1943).

Desthiobiotin, in which the sulfur atom is replaced by two hydrogen atoms, has an activity equal to biotin in stimulating growth of *Saccharomyces cerevisiae* but does not stimulate the growth of *Lactobacillus casei* [258] and is not biologically active for animals. Pimelic acid and other degradation products of biotin can also serve as biotin precursor for some microorganisms.[259,260] Biotin assays generally rely on measurement of the growth response of microorganisms (a number of bacteria and yeasts have been used) to biotin in the test material.

Biotin is one of the most potent of the known biologically active substances. Kögl found it to be a definite stimulant for yeast growth in a dilution of 1×10^{-11}. Because of this great activity, and because biotin is present in considerable amounts in many foods and is synthesized by microorganisms in the intestinal tract, deficiencies in this vitamin are not common except on restricted diets. It is interesting to note that chicks are more sensitive to egg-white injury than are rats, apparently owing to the extensive intestinal synthesis (and absorption) of biotin in the rat.

The direct relation of biotin deficiency to egg-white injury in animals has been shown *in vitro* as well as in both mammals and birds, the biotin being rendered unavailable by a firm combination with a constituent in raw egg white.[261] Heating the egg white will release the biotin. The active constituent is a specific protein which has been called *avidin* and is also present in the albumin-secreting portions of the oviduct of the hen as well as in the oviduct and eggs of other fowl and amphibia.[262]

"Egg-white injury" has also been experimentally demonstrated in man.[263] On a diet in which egg white furnished approximately 30 per cent of the caloric intake, the subjects showed deficiency symptoms, including grayish pallor of the skin and mucous membranes, dryness of the skin with fine scaly desquamation, diminution in blood hemoglobin and erythrocyte number and volume, and a greatly reduced biotin excretion. Many of the symptoms of thiamine deficiency were noted, such as muscular pains, anorexia, and extreme lassitude. The deficiency symptoms were rapidly cured by administration of biotin.

[258] D. B. Melville, K. Dittmer, G. B. Brown, and V. duVigneaud, *Science*, **98**, 497 (1943).

[259] R. E. Eakin and E. A. Eakin, *Science*, **96**, 187 (1942).

[260] V. duVigneaud, K. Dittmer, E. Hague, and B. Long, *Science*, **96**, 186 (1942).

[261] R. E. Eakin, E. E. Snell, and R. J. Williams, *J. Biol. Chem.*, **136**, 801 (1940); P. György, C. S. Rose, R. E. Eakin, E. E. Snell, and R. J. Williams, *Science*, **93**, 477 (1941).

[262] R. Hertz and W. H. Sebrell, *Science*, **96**, 257 (1942); R. M. Fraps, R. Hertz, and W. H. Sebrell, *Proc. Soc. Expll. Biol. Med.*, **52**, 140 (1943).

[263] V. P. Sydenstricker, S. A. Singal, A. P. Briggs, M. M. deVaughn, and H. Isbell, *Science*, **95**, 176 (1942).

Hertz [264] recently has prepared a review which summarizes the biological information on biotin and avidin.

Liver, yeast, and egg yolk served as important sources of biotin in the early chemical studies on this vitamin. Other foods containing important amounts of biotin include most fresh vegetables, dairy products, and sea foods. Milk varies considerably in its biotin content but generally contains 2–3 μg. per 100 ml.[265] In most tissues, biotin is present in bound form, from which extraction is difficult without rigorous digestive treatment.

A level of 0.045–0.07 mg. biotin per pound of feed has been suggested by the National Research Council as being adequate for optimal nutrition in chicks. No suggestions have been made for dietary intakes for mammals.

Pantothenic Acid. In 1933, Williams, *et al.*,[266] proposed the name of "pantothenic acid" for the yeast growth factor occurring in many plant and animal tissues. Although its vitamin nature was suspected, it was not until 1939 that pantothenic acid was identified as a nutritive factor for animals. Jukes [267] showed it to be the chick antidermatitis factor, which cures the reddening and swelling of the eyelids and the loss of egg hatchability associated with "chick pellagra." [268] The following year pantothenic acid was identified [269] as part of filtrate factor II required by rats to prevent a dermatitis, graying of the fur, and "blood-caked" whiskers (which arise by the deposition of red coproporphyrin incrustations on the nose and whiskers).[270]

Impure calcium panthothenate was obtained by Williams in 1938, but, even before the pure compound was isolated, synthesis of the vitamin had been effected.[271] The compound proved to consist of β-alanine in a peptide linkage with a dihydroxy acid. Only the *dextrorotatory* form of the acid was biologically active.

[264] R. Hertz, *Physiol. Revs.*, **26**, 479 (1946).

[265] J. M. Lawrence, B. L. Herrington, and L. A. Maynard, *J. Nutrition*, **32**, 73 (1946).

[266] R. J. Williams, C. M. Lyman, G. H. Goodyear, J. H. Truesdail, and D. Holaday, *J. Am. Chem. Soc.*, **55**, 2912 (1933).

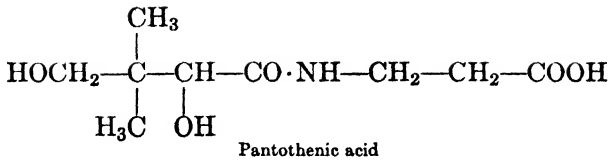
[267] T. H. Jukes, *J. Am. Chem. Soc.*, **61**, 975 (1939).

[268] A. T. Ringrose, L. C. Norris, and G. F. Heuser, *Poultry Sci.*, **10**, 166 (1931); A. T. Ringrose and L. C. Norris, *J. Nutrition*, **12**, 535 (1936).

[269] M. K. Dimick and A. Lepp, *J. Nutrition*, **20**, 413 (1940).

[270] L. W. McElroy, K. Salomon, F. H. J. Figge, and G. R. Cowgill, *Science*, **94**, 467 (1941).

[271] E. T. Stiller, S. A. Harris, J. Finkelstein, J. C. Keresztesy, and K. Folkers, *J. Am. Chem. Soc.*, **62**, 1785 (1940).



Pantothenic acid

As would be expected from its chemical structure, pantothenic acid is hydrolyzed on prolonged autoclaving and is very labile to either acid or alkali.

Both the amino acid and hydroxy acid moieties of pantothenic acid contribute to its specificity. Closely related amino acids, including α-alanine, gave inactive products when coupled with the non-nitrogenous portion of the vitamin.[272] Similarly, slight alterations in the hydroxy acid moiety greatly affect the growth-promoting effects of the compound for microorganisms [273,274] and for animals.[275]

Biological assay for pantothenic acid can be carried on with several species of animals, including the rat, mouse, and chick. More rapid microbiological methods are now available. They indicate that pantothenic acid is very widely distributed in nature. Wheat and other cereals contain somewhat over 1 mg. per 100 grams, but milled flour has only half this amount. Many fruits and vegetables are relatively poor sources of pantothenic acid, although some, such as broccoli, contain approximately a milligram per 100-gram portion. Yeast, liver, and egg yolk are excellent sources, contributing 6 mg. per 100 grams.

For optimal nutrition, swine should have a daily dietary intake of 18.5 mg. of pantothenic acid (100-lb. animal), and poultry rations should contain 5–7 mg. per pound of feed. Ruminants do not appear to require a dietary source of the vitamin, probably because of its synthesis in the rumen.

Folic Acid. This vitamin (including its derivatives) has had a brief but very interesting history under a variety of designations, among which are:

1938. *Vitamin M* (Day and associates [276]), a monkey anti-anemia factor in yeast and liver.

[272] H. H. Weinstock, Jr., E. L. May, A. Arnold, and D. Price, *J. Biol. Chem.*, **135**, 343 (1940).

[273] H. K. Mitchell, H. H. Weinstock, Jr., E. E. Snell, S. R. Stanbery, and R. J. Williams, *J. Am. Chem. Soc.*, **62**, 1776 (1940); H. K. Mitchell, E. E. Snell, and R. J. Williams, *ibid.*, **62**, 1791 (1940).

[274] J. W. Barnett and F. A. Robinson, *Biochem. J.*, **36**, 357 (1942).

[275] E. Zschiesche and H. K. Mitchell, *Proc. Soc. Exptl. Biol. Med.*, **45**, 565 (1940).

[276] W. C. Langston, W. J. Darby, C. F. Shukers, and P. L. Day, *J. Exptl. Med.*, **68**, 923 (1938).

1938. *Factor U* (Stokstad and Manning [277]), a chick growth factor in alfalfa, yeast, and other products.

1940. *Vitamin B_c* (Hogan and Parrott [278]), a chick anti-anemia factor.

1940. *Norit Eluate Factor* (Snell and Peterson [279]), a growth factor for *Lactobacillus casei*.

1940. *Factor R* (Schumacher, *et al.* [280]), present in yeast and required for the prevention of anemia in chicks.

1941. *Folic Acid* (Williams and associates [281]), required for growth of certain bacteria.

1943. *Yeast L. casei Factor* (Stokstad [282]), a new designation for the Norit eluate factor, somewhat less active for *Streptococcus faecalis* R than for *L. casei*.

1943. *Liver L. casei Factor* (Stokstad [282]), equally active for *L. casei* and *S. faecalis* R.

1943. *Streptococcus lactis R (SLR) Factor* (Keresztesy, *et al.*[283]), utilized by *S. lactis* R (*S. faecalis* R) but not by *L. casei*.

1944. *Fermentation L. casei Factor* (Hutchings, *et al.*[284]), active for *L. casei* but not for *S. faecalis* R.

1944. *Vitamin B_c Conjugate* (Binkley, *et al.*[285]), having little activity for either *L. casei* or *S. faecalis* R.

1946. *Pteroylglutamic Acid* (Angier, *et al.*[286]), the synthetic liver *L. casei* factor.

As indicated in this list, the test organism employed, the biological response under investigation, and the source of the material all have affected the results observed. Despite these confusing obstacles, the advances in this field have been very rapid and have led to a reasonably clear picture of the chemistry of the vitamin. We now know that folic acid exists in natural products both in the free form and as one or more "conjugates." A number of these have been isolated in crystalline

[277] E. L. R. Stokstad and P. D. V. Manning, *J. Biol. Chem.*, **125**, 687 (1938).

[278] A. G. Hogan and E. M. Parrott, *J. Biol. Chem.*, **132**, 507 (1940).

[279] E. E. Snell and W. H. Peterson, *J. Bact.*, **39**, 273 (1940).

[280] A. E. Schumacher, G. F. Heuser, and L. C. Norris, *J. Biol. Chem.*, **135** 313 (1940).

[281] H. K. Mitchell, E. E. Snell, and R. J. Williams, *J. Am. Chem. Soc.*, **63**, 2284 (1941).

[282] E. L. R. Stokstad, *J. Biol. Chem.*, **149**, 573 (1943).

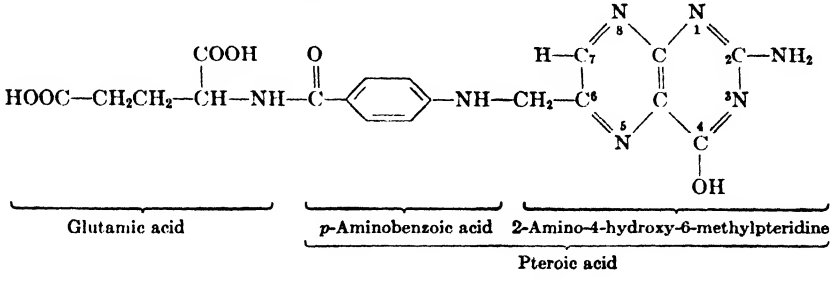
[283] J. C. Keresztesy, E. L. Rickes, and J. L. Stokes, *Science*, **97**, 465 (1943).

[284] B. L. Hutchings, E. L. R. Stokstad, N. Bohonos, and N. H. Slobodkin, *Science*, **99**, 371 (1944). That a fermentation residue was the source of this factor was indicated in later publications. *Cf. Science*, **102**, 227 (1945); **103**, 667 (1946).

[285] S. B. Binkley, O. D. Bird, E. S. Bloom, R. A. Brown, D. G. Calkins, C. J Campbell, A. D. Emmett, and J. J. Pfiffner, *Science*, **100**, 36 (1944).

[286] R. B. Angier, *et al.*, *Science*, **103**, 667 (1946).

form. Folic acid, vitamin B_c, liver *L. casei* factor, and pteroylglutamic acid apparently are identical [282, 286-288] and have the following structure, as shown by synthesis by workers at the Lederle Laboratories and the American Cyanamid Company:



FOLIC ACID

The fermentation *L. casei* factor has full vitamin M activity [289] and appears to be a folic acid conjugate in which two additional glutamic acid residues are present as peptides on the two carboxylic acid groups of folic acid.[286, 290] Pteroic acid (folic acid with glutamic acid removed) is active for *S. faecalis* R but not for *L. casei* and may be the SLR factor; [290] the yeast *L. casei* factor is a more complex glutamic acid conjugate of folic acid, apparently having seven glutamic acid residues. Factor R consists of one or more folic acid conjugates.[291] Even the vitamins B_{10} and B_{11} of the Wisconsin group,[292] required for growth and feathering of chicks, "seem to be related chemically to the various factors with vitamin B_c activity."

It is interesting to note that the peptide of L-glutamic acid with *p*-aminobenzoic acid is a more potent growth factor for *Streptobacterium plantarium* than is *p*-aminobenzoic acid itself, whereas many other peptides of the latter are inactive.[293] It remains to be shown whether the activity of *p*-aminobenzoyl-glutamic acid is related to its possible functioning as a folic acid precursor.

As has been indicated, some microorganisms can utilize folic acid conjugates, some require the simpler form of the vitamin, and others

[287] J. J. Pfiffner, S. B. Binkley, E. S. Bloom, R. A. Brown, O. D. Bird, A. D. Emmett, A. G. Hogan, and B. L. O'Dell, *Science*, **97**, 404 (1943).

[288] B. C. Johnson, *J. Biol. Chem.*, **163**, 255 (1946).

[289] P. L. Day, V. Mims, and J. R. Totter, *J. Biol. Chem.*, **161**, 45 (1945).

[290] E. L. R. Stokstad, *et al.*, Report presented at meeting of the New York Academy of Sciences, May 29, 1946.

[291] L. W. Charkey, Ph.D. thesis, Cornell University, 1945.

[292] G. M. Briggs, Jr., T. D. Luckey, C. A. Elvehjem, and E. B. Hart, *J. Biol. Chem.*, **148**, 163 (1943); **153**, 423 (1944); **158**, 303 (1945).

[293] E. Auhagen, *Z. physiol. Chem.*, **277**, 197 (1943).

can synthesize folic acid from the pteridine residue. Similar differences in activity of various forms of the vitamin are not so evident among animals. The fermentation *L. casei* factor is active for the chick, the rat, and the monkey in preventing a macrocytic anemia and growth retardation, apparently because the liver and other tissues can hydrolyze the conjugate to free folic acid when pyracin (a vitamin B_6 derivative, p. 933) is present along with the enzyme.[294] The enzyme has been termed vitamin B_c conjugase [295] and is widely distributed in both plant and animal tissues. In man, however, pathological conditions exist where enzymic conversion of the conjugates to the active form of the vitamin does not take place. In pernicious anemia, the folic acid conjugates are not of therapeutic value, whereas the free folic acid elicits prompt and significant clinical improvement of the patient.[296]

It has been known for fully a decade that xanthopterin, a pigment in yellow butterfly wings, has some hematopoietic activity for anemic animals. Wright and Welch [297] found that synthetic xanthopterin incubated with rat liver led to *in vitro* synthesis of folic acid, and they suggested that the pterin was part of the vitamin molecule. Ultraviolet absorption curves on the two materials also suggested a chemical relationship.

The activity of xanthopterin as a folic acid precursor is the result of its similarity to the pteridine in the vitamin molecule, the two being identical with the exception that the latter has a methyl group in the 6-position, whereas xanthopterin has a hydroxyl group on carbon-6.[298] It is still not clear why pteroic acid (if this is indeed the SLR factor) will not stimulate folic acid synthesis by rat liver [299] and xanthopterin will.

The use of folic acid in the cure or prevention of various macrocytic anemias, leukopenia, and thrombopenia has written a new and promising chapter in clinical nutrition. Macrocytic anemias encountered in sprue, pregnancy, nutritional anemia, pellagra, pernicious anemia, and after administration of sulfa drugs have responded to folic acid, often in dramatic fashion. Incomplete evidence suggests that folic acid is one

[294] L. J. Daniel, M. L. Scott, L. C. Norris, and G. F. Heuser, *J. Biol. Chem.*, **160**, 265 (1945); M. L. Scott, L. C. Norris, L. W. Charkey, L. J. Daniel, and G. F. Heuser, *ibid.*, **164**, 403 (1946).

[295] O. D. Bird, S. B. Binkley, E. S. Bloom, A. D. Emmett, and J. J. Pfiffner, *J. Biol. Chem.*, **157**, 413 (1945); O. D. Bird, M. Robbins, J. M. Vandenbelt, and J. J. Pfiffner, *ibid.*, **163**, 649 (1946).

[296] F. H. Bethell, M. E. Swendseid, O. D. Bird, M. C. Meyers, G. A. Andrews, and R. A. Brown, *Bull. Univ. Mich. Hospital*, **12**, 42 (May, 1946).

[297] L. D. Wright and A. D. Welch, *Science*, **98**, 179 (1945).

[298] R. Purrmann, *Ann.*, **546**, 98 (1940–41).

[299] J. L. Stokes, J. C. Keresztesy, and J. W. Foster, *Science*, **100**, 522 (1944).

of the materials in liver extract responsible for the effectiveness of the latter in the treatment of both pernicious anemia and sprue. Indeed, Darby, *et al.*,[300] suggest that vitamin M deficiency in the monkey is the experimental analog of sprue in humans.

Spies has been particularly active in experimentation on the clinical use of folic acid. His series of reports during 1945 and 1946, particularly in the *Southern Medical Journal*, should be referred to for a more detailed evaluation of the therapeutic use of folic acid in patients with blood abnormalities. Ruth Woods has prepared a concise review of the clinical use of folic acid.[301]

Analytical values for folic acid in foods have been obtained for the most part by microbiological assays. Many of these data are not reliable in indicating total vitamin activity because of incomplete hydrolysis of the conjugates to the free acid in making extracts for analysis. Free folic acid is the only form which elicits maximum response for all the test organisms that might be employed. The available data indicate that green leafy vegetables (notably spinach) are rich in this vitamin, whereas most cereals and milk are poor sources. Muscle meats contain only moderate amounts of folic acid; liver and kidney have much higher values. Folic acid is quite sensitive to conditions encountered in commercial processing and in cooking foods. Very considerable losses occur during processing. During the cooking of meat and vegetables, losses of one-third to nine-tenths of the vitamin have been reported.[302, 303]

***p*-Aminobenzoic Acid.** During recent years many investigators have classified *p*-aminobenzoic acid as a vitamin. Ansbacher [304] and others have reported that *p*-aminobenzoic acid is effective in preventing a graying of the fur of black or piebald rats on a diet supplemented with other known members of the B-complex. Emerson,[305] however, did not observe this curative effect. It is likely that *p*-aminobenzoic acid is identical with the rat and fox anti-gray-hair factor B_x of Lunde and Kringstad.[306]

That *p*-aminobenzoic acid has a growth-stimulating effect has been noted for chicks and for many bacteria. A number of investigators believe that it acts by stimulating bacterial synthesis of other factors essential to the host. The recent elucidation of the structure of folic acid

[300] W. J. Darby, E. Jones, and H. C. Johnson, *Science*, **103**, 108 (1946).

[301] R. Woods, *Borden's Rev. of Nutrition Research*, **7** (April, 1946).

[302] V. H. Cheldelin, A. M. Woods, and R. J. Williams, *J. Nutrition*, **26**, 477 (1943).

[303] B. S. Schweigert, A. E. Pollard, and C. A. Elvehjem, *Arch. Biochem.*, **10**, 107 (1946).

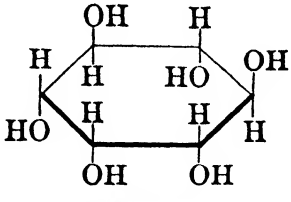
[304] S. Ansbacher, *Science*, **93**, 164 (1941).

[305] G. A. Emerson, *Proc. Soc. Exptl. Biol. Med.*, **47**, 448 (1941).

[306] G. Lunde and H. Kringstad, *Z. physiol. Chem.*, **257**, 201 (1939); *Naturwissenschaften*, **27**, 755 (1939); *J. Nutrition*, **19**, 321 (1940).

suggests that *p*-aminobenzoic acid in some cases may be serving as a folic acid precursor.

Inositol. Mice on a diet deficient in B vitamins develop a condition termed alopecia, characterized by loss of hair and cessation of growth. Woolley reported [307] that inositol was effective in preventing this dietary disorder. More recently he has prepared a review on the significance of this factor in nutrition.[308]



meso-Inositol

Eastcott [309] identified inositol with "bios I," required by yeast cells. Only the meso isomer, in which all the hydroxyl groups except those on carbon-1 and carbon-3 are below the plane of the ring, is nutritionally active.[307]

Little is known of the functions of inositol. It can serve as a lipotropic factor (see the review of McHenry and Patterson [310]) and is a constituent of certain phospholipids occurring in brain tissue, bacteria, and soybeans (see Chapter 31). Foods also contain inositol in the free form and as phytin, the calcium-magnesium salt of the hexaphosphoric ester (phytic acid). The latter is present in cereals and is less completely utilized by the organism.

According to data of Cheldelin and Williams,[311] inositol is present in appreciable amounts in fresh fruits, many vegetables, and in wheat germ, whereas processed cereals and muscle meats are rather poor sources of this factor. Oranges are reported to contain inositol in amounts exceeding 0.2 per cent of the fresh weight.

Vitamin B_{12}. A very recent addition to the list of vitamins is vitamin B_{12}, independently isolated from liver by workers in the United States [312] and in England.[313] This crystalline red compound is a highly potent

[307] D. W. Woolley, *Science*, **92**, 384 (1940); *J. Biol. Chem.*, **139**, 29 (1941); **140**, 461 (1941); *J. Nutrition*, **28**, 315 (1944).

[308] D. W. Woolley, *J. Nutrition*, **28**, 305 (1944).

[309] E. V. Eastcott, *J. Phys. Chem.*, **32**, 1094 (1928).

[310] E. W. McHenry and J. M. Patterson, *Physiol. Revs.*, **24**, 128 (1944).

[311] V. H. Cheldelin and R. J. Williams, *Univ. Texas Pub.*, No. 4237, 105 (1942).

[312] E. L. Rickes, N. G. Brink, F. R. Koniuszy, T. R. Wood, and K. Folkers, *Science*, **107**, 396 (1948).

[313] E. L. Smith, *Nature*, **161**, 638 (1948).

anti-pernicious anemia factor and is of particular interest chemically because it contains cobalt.[314,315] As we have already noted (p. 399) cobalt recently has been shown to be an essential trace element in hemoglobin formation.

Choline. It is debatable whether choline should be classed as one of the vitamins of the B-complex. The choline requirement generally can be met by a suitable intake of methionine, which provides a source of labile methyl groups (see Chapter 19). Choline is a constituent of lecithin (Chapter 31) and is present as acetylcholine in nervous tissue, where it serves in the transmission of nerve impulses (Chapter 21).

One of the principal effects of choline deficiency is the infiltration of fat into the liver. A recent review on lipotropic factors [310] discusses the role of choline in preventing fatty livers.

Antivitamins. The essential nature of *p*-aminobenzoic acid was discovered as the result of studies on bacteriostasis of "sulfa" drugs. The activity of sulfanilamide, the sulfonamide analog of *p*-aminobenzoic acid, was reversed by the latter compound, suggesting that the drug was active by virtue of its ability to prevent bacteria from utilizing *p*-aminobenzoic acid. Sulfanilamide thus was characterized as an antivitamin.

A host of antimetabolites bearing a close structural similarity to essential nutritive factors are now recognized. They have found use both as drugs in chemotherapy and as tools for nutrition research. The sulfonamides in particular have proved very useful in eliminating intestinal synthesis of vitamins, the existence of which has complicated many nutrition studies.

It is not within the scope of this text to discuss in detail this interesting and promising field. Suffice it to say that many of the vitamins listed in this chapter now have one or more antivitamins which are capable of inducing deficiency signs in microorganisms or animals. There is evidence that some of these antivitamins may exist in natural products. Woolley [316] has written an interesting essay on this general development of biochemical antagonism, and Welch [317] has prepared an extensive review of the literature. The reader is referred to these treatises for a more detailed account of this phase of vitamin research.

[314] E. L. Rickes, N. G. Brink, F. R. Koniuszy, T. R. Wood, and K. Folkers, *Science*, **108**, 134 (1948).

[315] E. L. Smith, *Nature*, **162**, 144 (1948).

[316] D. W. Woolley, Chapter 23 in D. E. Green, *Currents in Biochemical Research*, Interscience Publishers, Inc., New York, 1946.

[317] A. D. Welch, *Physiol. Revs.*, **25**, 687 (1945).

CHAPTER 37

The Hormones

The word *hormone* was first applied to this group of *chemical messengers* by Bayliss and Starling [1] in 1902 at the suggestion of W. B. Hardy.[2] The word is derived from the Greek *ορμαω*, meaning, "I arouse to activity." The earlier term, "internal secretions," did not sufficiently indicate their special functions, and the terms "chemical messengers" or "excitants fonctionnels," while indicating physiological properties and special functions, did not seem satisfactory as a permanent nomenclature. The term hormone, as applied by Bayliss and Starling, was intended to designate *those chemical substances secreted by the endocrine glands which, when carried by the blood stream to another organ, profoundly influence the activity of that organ.*

In the highly complex mechanism representative of the bodily activities of the higher mammals, different organs and tissues have taken over very special activities and functions. In order that the entire organism may work in harmony special mechanisms have come into being. One of these mechanisms is the nervous system, which may be likened to an intercommunicating network of telegraph wires centering in the great central switchboards of the brain, spinal cord, and ganglia. In order, however, that life processes may proceed at a *uniform* rate rather than undergo violent fluctuations as the environment changes, a second system of "chemical messengers" was necessitated whereby the increased activity of one organ would be reflected in an altered activity of another organ and thus, through a system of checks and balances, the complex mechanism would work as a unified whole. The hormones provide this "balance wheel." Their presence enables the organism to pass through periods of stress with all forces of the organism mobilized to meet the test and then, when the emergency has passed, to resume the normal coordination of bodily activities.

It should be emphasized that the endocrines are not independent of one another but are at least to a certain extent dependent on one an-

[1] W. M. Bayliss and E. H. Starling, *J. Physiol.*, **28**, 325 (1902).

[2] *Cf.* p. 712, W. M. Bayliss, *Principles of General Physiology*, 4th ed., Longmans, Green and Co., London and New York, 1924.

other's secretions. Because of these interrelationships we can readily see how the entire function of the body may be stimulated, retarded, or perverted by the excess activity, lessened activity, or dysfunction of any one of the endocrine glands. That these changes take place has been abundantly established by endocrinologists. Moreover, certain of the endocrine glands secrete a multiplicity of hormones, and, though this secretion probably arises in each instance from specific cells, nevertheless the unbalance of the organism with respect to a single hormone may account for what was formerly attributed to the dysfunction of the entire gland.

Although the discoveries in the field of the hormones in the past two decades have been all-important, one fact stands out as preeminent, *i.e.*, the apparent master-key control of all the other endocrine glands which resides in certain of the specific hormones elaborated by the anterior lobe of the pituitary. These master-key hormone-controlling hormones of the anterior pituitary appear to be proteins or protein-like in nature.

There is no generally accepted classification of the hormones. For purposes of convenience, however, they may be divided into two groups: (*A*) those hormones having somewhat general physiological and metabolic functions, and (*B*) those having specific physiological or metabolic functions. There follows the listing of thirty-three hormones, for the existence of which there is a reasonable amount of data in the literature. Probably other as yet unknown hormones will be added to the list as the result of further research. In the brief space represented by this chapter it will be impracticable to discuss in detail all the chemical and physiological evidence for the hormones in this list. The interested reader will therefore have to turn to larger treatises [3] for many important details.

A. Hormones having somewhat general physiological and metabolic functions:

1. *Thyroxine*, thyroid
2. *Epinephrine*, adrenal medulla
3. Thymus principle
4. Pineal principle
5. *Pitocin* (or *oxytocin*), posterior pituitary
6. *Pitressin* (or *vasopressin*), posterior pituitary
7. Growth principle, anterior pituitary

[3] A. Grollman, *Essentials of Endocrinology*, J. B. Lippincott Co., Philadelphia, 1941; R. G. Hoskins, *Endocrinology*, W. W. Norton and Co., New York, 1941; C. D. Turner, *General Endocrinology*, W. B. Saunders Co., Philadelphia, 1948; *Ann. Rev. Biochem.*, Vols. I-XVII, Stanford University Press, Stanford University, California, 1932–1948; *Cold Spring Harbor Symposia Quant. Biol.* Vols. 5 and 10 (1937 and 1942); *The Chemistry and Physiology of Hormones*, American Association for the Advancement of Science, Washington, D. C., 1944.

B. Hormones having specific physiological or metabolic functions:

8. *Cortin* (cortical hormones), adrenal cortex
9. *Parathormone*, parathyroids
10. *Insulin*, pancreas
11. *Intermedin*, pars intermedia of pituitary
12. *Prolactin*, anterior pituitary
13. Thyrotropic principle, anterior pituitary
14. Parathyrotropic principle, anterior pituitary
15. Adrenotropic principle, anterior pituitary
16. Diabetogenic principles, anterior pituitary
17. Insulinotropic principle, anterior pituitary
18. Gonadotropic principles (*FSH* and *LH*), anterior pituitary
19. *Estrone*, ovarian follicular fluid
20. *Estradiol*, ovarian follicular fluid
21. *Estriol*, ovarian follicular fluid and placenta
22. *Equilin*, from estrone
23. *Equilenin*, from estrone or equilin
24. *Progesterone*, corpus luteum of ovaries
25. *Pregnanediol*, from progesterone
26. *Testosterone*, testes
27. *Androsterone*, from testosterone
28. *Androstanedione*, testes
29. *Dehydroandrosterone*, testes
30. *Gastrin*, pyloric mucosa of stomach
31. *Secretin*, duodenal mucosa
32. *Cholecystokinin*, duodenal mucosa
33. *Enterogastrone*, duodenal mucosa

The Hormones of the Adrenal Glands. The adrenal or suprarenal glands are small flattened bodies situated in the retroperitoneal tissue at the upper end of each kidney. Each gland consists of an internal medulla and an external cortex, encased in a capsule or sheath of connective tissue. Both the medulla and cortex tissues secrete specific hormones.

The Hormone of the Medulla. The cells of the medulla secrete a hormone known as epinephrine. The secreting cells are stained brownish yellow by potassium chromate and have been designated chromaffine cells. Cells having similar staining reactions are found in ganglia and nerve cells of most invertebrates and contain a substance with physiological properties analogous to if not identical with those of epinephrine. Gaskell [4] suggests that these cells and their secretions in invertebrates are the progenitors of the epinephrine secretory chromaffine cells of the adrenal medulla and of the sympathetic nervous system of the higher vertebrates.

[4] J. F. Gaskell, *J. Gen. Physiol.*, **2**, 73 (1919).

Epinephrine was the first hormone to be isolated and identified as a chemical entity. In 1899, Abel [5] isolated the compound as the benzoate and prepared a number of other derivatives. He also showed that the compound had great physiological activity. Abel gave to the compound the name *epinephrine*. At about the same time von Fürth [6] prepared potent extracts and isolated metallic derivatives of a compound which he called *suprarenine*. Somewhat later Takamine devised essentially the present commercial method for the preparation of epinephrine and called the substance *adrenalin* under which name it was marketed commercially. Epinephrine is β-hydroxy-β(3,4-dihydroxyphenyl)N-methylethylamine. Since it contains an asymmetric carbon atom, it is optically active. The *levo* form occurs in the adrenal gland and is physiologically fifteen to twenty times as potent as the *dextro* form. The function of epinephrine is to maintain smooth muscle in a state of excitability so that it will readily react to nerve impulses transmitted by the sympathetic nerve fibers. Epinephrine contracts the arterioles, thereby exerting a marked effect upon the blood pressure. The entire contractile vascular system is regulated by both the secretion of epinephrine and the sympathetic nervous system acting in conjunction. The hormone is exceedingly potent. The concentration of epinephrine normally present in the circulating blood has been estimated to be of the order of 1:1,000,000,000 to 1:2,000,000,000. Elliott,[7] Tscheboksaroff,[8] and Cannon,[9] and others have supported the view that in periods of stress, such as fear, hunger, pain, rage, etc., the stimulation of the splanchnic nerve causes a discharge of larger amounts than usual of epinephrine into the blood stream, thus causing a rise in blood pressure and a general mobilization of the body forces which enable the organism to rise and meet the emergency. Furthermore, there is definite evidence that epinephrine is involved in carbohydrate metabolism, in both the liver and the muscle. It catalyzes the reaction glycogen $\rightarrow$ glucose in the liver and glycogen $\rightarrow$ lactic acid in muscle tissue.

The secretory activity of the adrenal medulla is intimately associated with the activity of other hormone-secreting organs. There is a close interrelationship between adrenal activity and the secretion of insulin by the pancreas glands. Both epinephrine and insulin influence the blood-sugar level. The activity of the pancreas increases with an in-

[5] J. J. Abel, *Z. physiol. Chem.*, **28**, 318 (1899).

[6] O. von Fürth, *Z. physiol. Chem.*, **24**, 142 (1898).

[7] T. R. Elliott, *J. Physiol.*, **44**, 374 (1912).

[8] M. Tscheboksaroff, *Arch. ges. Physiol.*, **137**, 59 (1911).

[9] W. B. Cannon, *Bodily Changes in Pain, Hunger, Fear, and Rage*, 2nd ed., D. Appleton and Co., New York, 1929.

creased epinephrine output, and *vice versa*, so that the two systems although they supplement each other, may be regarded as mutually antagonistic. The secretory activity of other endocrine glands likewise appears to be influenced by the epinephrine level; in particular, complementary effects are indicated for the activity of the adrenal medulla and the posterior lobe of the pituitary.

In spite of the remarkable and powerful pharmacological and physiological effects which epinephrine exerts on the organism, it is rather surprising to note that the destruction of the medulla of both adrenals is not fatal, and indeed such destruction seems to have no particular effect provided that sufficient functional cortical tissue of the adrenals is left. The removal of both glands in their entirety is fatal. Thus it appears that the medulla is not essential to life and that therefore epinephrine is not an indispensable hormone. In view of this fact it would be anticipated that there are no recognized diseases due to hypomedullary adrenalism. Hyperfunction, however, is to be expected. In all cases of hyperfunction so far described medullary tumors have been involved. The major symptoms of hyperfunction are hypertension, high blood pressure, glycosuria, and other distressing effects.

The Hormones of the Adrenal Cortex. The cortex of the adrenals arises from mesoblast cells which secrete a group of closely related hormones essential to the maintenance of life. In man a pathological condition of the adrenal cortex results in Addison's disease, characterized by anemia, loss of appetite, general languor and debility, feeble heart action, gastrointestinal irritability, and a peculiar and characteristic brownish, splotchy skin pigmentation. In its acute form Addison's disease is invariably fatal, but the symptoms can be rapidly alleviated and complete cure can be obtained by the administration of cortical extract.

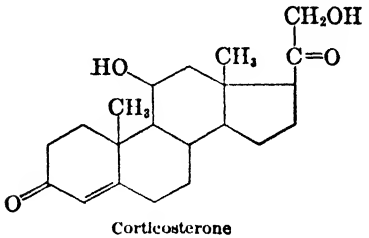
Knowledge of the chemical nature of the adrenal cortex hormones has expanded greatly in recent years. Since 1930, when Hartman, *et al.*,[10] first prepared an active cortical extract, *cortin*, more than twenty different substances have been isolated from cortex tissue. Kendall, Reichstein, Wintersteiner and Pfiffner, together with their co-workers, have been particularly active in this field.[11]

All known adrenal cortex hormones contain the steroid nucleus, relating them structurally to the sex hormones. They have varying de-

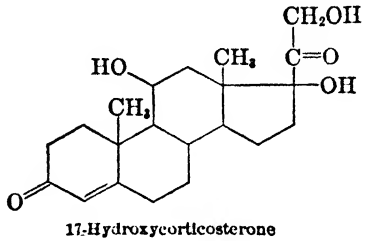
[10] F. A. Hartman, K. A. Brownell, and W. E. Hartman, *Am. J. Physiol.*, **95**, 670 (1930).

[11] For an excellent review of the isolation and chemical characterization of adrenal hormones, the reader is referred to the article by M. H. Kuizenga on p. 57 of *The Chemistry and Physiology of Hormones*, American Association for the Advancement of Science, Washington, D. C., 1944.

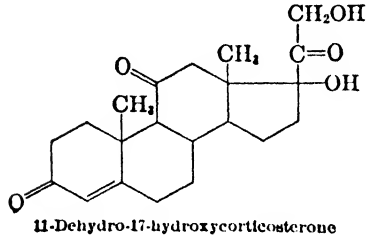
grees of biological activity. The six physiologically most important substances which have been isolated in crystalline form are:



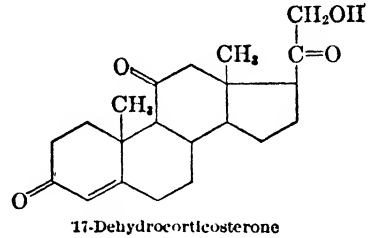
Corticosterone



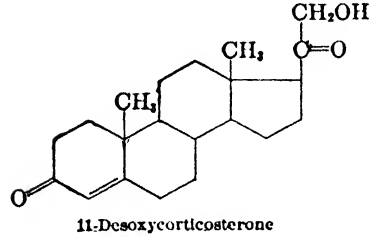
17-Hydroxycorticosterone



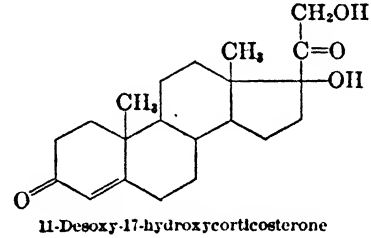
11-Dehydro-17-hydroxycorticosterone



17-Dehydrocorticosterone



11-Desoxycorticosterone



11-Desoxy-17-hydroxycorticosterone

Many metabolic functions have been attributed to the adrenal cortex hormones, but some uncertainty still exists as to whether certain of the effects noted in adrenal insufficiency are direct or secondary. One of the primary functions is to *maintain normal electrolyte and water balances* in the blood and tissues. In the adrenalectomized animal the kidney thresholds for sodium, potassium, and chloride ions are modified so that greater amounts of sodium and chloride are excreted while potassium ion is retained. The resulting ionic changes in the blood cause an increased loss of water from the blood, both through the kidneys and into the tissues. It has been suggested also that the permeability of the capillary walls is directly influenced by the presence or absence of cortical hormones. The administration of large amounts of sodium chloride, for which the adrenalectomized animal has a strong appetite, frequently permits survival for prolonged periods.

Thorn, Engel, and Lewis [12] have compared the effects of various crystalline cortical steroids on the excretion of sodium and chloride ions by the dog. Their observations throw important light on the relationship between chemical structure and biological activity (Fig. 118). Thus

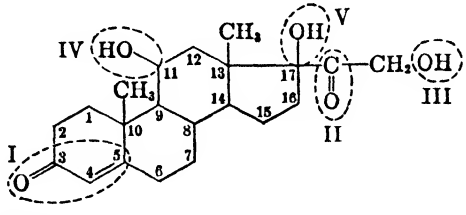
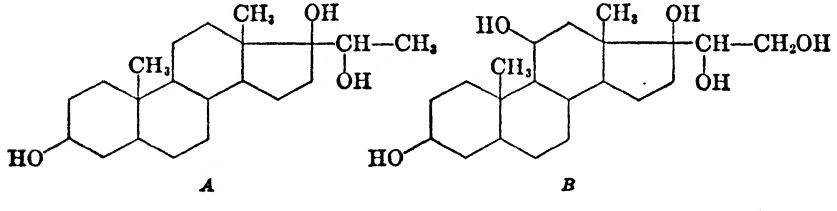


FIG. 118. Relationship of chemical structure to physiological activity of cortical steroids. (After Thorn, Engel, and Lewis.)

I. Essential for all known activity.
II. Essential for all known activity.
III. Enhances sodium retention; necessary for carbohydrate activity.
IV. (Either a hydroxyl or a keto group.) In presence of III, decreases sodium retention and increases carbohydrate activity.
V. In presence of III and possibly IV, induces sodium excretion and increases carbohydrate activity.

11-desoxycorticosterone had a strong effect on sodium and chloride *retention;* corticosterone had a similar but less marked effect. On the other hand, 17-hydroxycorticosterone and 17-hydroxy-11-dehydrocorticosterone decidedly increased the *excretion* of both sodium and chloride ions. Two other cortical steroids, allopregnane-3,17,20-triol (*A*) and allopregnane-3,11,17,20,21-pentol (*B*) were totally inactive.



Another important aspect of adrenal cortical function has to do with the *regulation of carbohydrate metabolism.* In cortical insufficiency there results a depletion of the glucose and glycogen in the blood and tissues. Absorption of glucose from the digestive tract is subnormal, glycogen formation is inhibited, and glucose oxidation is increased. There is evidence that glyconeogenesis, *i.e.*, the conversion of amino acids to carbohydrate in the body, is inhibited by the lack of adrenal cortex principles.

A factor in the cortex which produces effects on the blood pressure similar to epinephrine has been designated *cortipressin.*[13] Its secretion

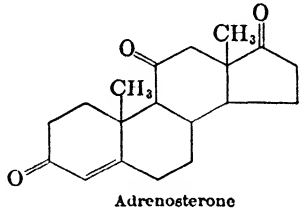
[12] G. W. Thorn, L. L. Engel, and R. A. Lewis, *Science,* **94,** 348 (1941).

[13] J. M. Looney and M. C. Darnell, *J. Biol. Chem.,* **114,** *Proc. Am. Soc. Biol. Chem.,* lxii (1936).

by the cortical tissues probably accounts for the fact that animals survive the complete removal of the medullary adrenal tissue provided that sufficient cortical tissue is left. The hormone, like ephedrine, is effective when taken by mouth. Although concentrated extracts have been prepared, the pure hormone has not been identified.

Another factor, termed *cortilactin*,[14] appears to influence the activity of the mammary glands. Nothing is known concerning its chemical nature. In fact it may be that the effect of the adrenals on mammary function is an indirect effect, the primary effect being on the pituitary and the secretion of prolactin by that organ.

The cortex of the adrenals also elaborates substances which function as sex hormones. Both androgenic (male) and estrogenic (female) substances are produced. One such, *adrenosterone*, has about one-fifth the activity of the testicular hormone androsterone (see p. 973) by the capon comb test. Hyperactivity of the cortex may cause these hormones to

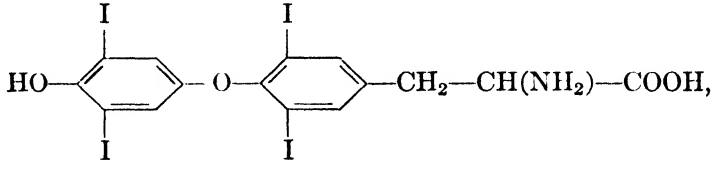


Adrenosterone

be produced in excessive amounts and so affect sex behavior and characteristics. Thus adrenal cortex tumor in young and adolescent boys usually results in precocious growth and maleness. Sometimes the result is merely that of excessive muscular virility. After puberty the result is premature senility and early death. Rarely in adult males is there a definite development of "femaleness." In such instances the breasts enlarge and may even secrete milk. The testes atrophy; there is a loss of sex desire, and a female type of obesity may develop. In the female the predominating change is toward maleness. Before puberty in girls the individual may become fat and muscular with hair appearing early on the pubis and also on the face. The voice becomes coarsened, the uterus and ovaries atrophy, menstruation does not occur, although the external sex organs may show hypertrophy. Similar changes occur when the tumor begins after puberty, with the additional effect of atrophy of the breasts. Removal of tumors in both young girls and adults causes the individual to return to all normal sex characteristics within a few months, although the deep, masculine voice persists for the longest period.

[14] K. A. Brownell, J. E. Lockwood, and F. A. Hartman, *Proc. Soc. Exptl. Biol. Med.*, **30**, 783 (1933).

The Hormones of the Thyroid Gland. The thyroid is a divided endocrine gland situated on each side of the upper end of the trachea. The biochemistry of iodine in higher animals is largely, if not exclusively, associated with the hormone activities of the thyroid gland. The secretions of the thyroid are the primary regulators of the metabolism of the body. They control both the basal and general oxidation levels of all cells concerned with oxidative metabolism. The specific chemical compounds identified with hormonal activity are thyroxine,



discovered by Kendall in 1915 and synthesized by Harington in 1926, and 3,5-diiodotyrosine. Both these compounds have already been briefly considered in our earlier discussions of amino acids and important nitrogen bases. Thyroxine is a much more powerful metabolic stimulant than 3,5-diiodotyrosine, and it may be that the latter compound is simply a precursor or an intermediate in the formation of thyroxine. We have earlier commented on the unusual property of thyroxine, in that a single dose may maintain metabolic activities at an elevated rate for very extended periods of time, sometimes for several weeks.

That the thyroid gland contained iodine was discovered by Baumann [15] in 1895. In 1899, Oswald [16] showed that the iodine was bound in organic linkages and that an iodine-containing *thyroglobulin* could be extracted which possessed the physiological activity characteristic of the entire gland. Accordingly in speaking of the hormones of the thyroid gland it is necessary to include this iodoglobulin as the predominating hormone, although its activity is at least in part due to the presence of the unique amino acid thyroxine in the protein molecule.

Synthetic iodoproteins prepared by iodination of casein and other proteins have been found to contain thyroxine, diiodotyrosine, and monoiodotyrosine. Thyroactive iodocasein has been used to stimulate milk production in cows during the declining phase of their lactation. Reinecke and Turner [17] at Missouri have been particularly active in this field.

[15] E. Baumann, *Z. physiol. Chem.*, **21**, 319 (1895).

[16] A. Oswald, *Z. physiol. Chem.*, **27**, 14 (1899); **32**, 121 (1901).

[17] E. P. Reinecke and C. W. Turner, *Missouri Agr. Exp. Sta. Research Bull.* **355** (1942).

Baker [18] has studied the secretion of iodine by perfused thyroid glands cultivated in the Lindbergh pump. He observed that iodine left the gland in *two* forms, one a globulin (presumably thyroglobulin), the other a low molecular weight substance found in the protein-free filtrate. The latter is probably a degradation product of thyroglobulin formed as result of thyroid gland activity.

Aside from its effects upon oxidative metabolism the hormone exhibits major control over ossification processes and the normal functioning of the central nervous system and a measure of control over the development of the sex organs. Two general types of phenomena are associated with dysfunction of the thyroid: (*a*) in the absence of an adequate iodine supply or in a failure of the synthetic mechanisms within the gland, there is an inadequate secretion of the hormone; and (*b*) the synthetic and secretion mechanisms may become too greatly stimulated and result in an excessive secretion of the hormone and an overstimulation of metabolic activities. Diseases resulting from both hormonal insufficiency and hormonal excess are common.

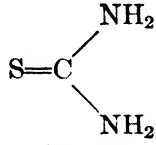
Hormonal insufficiency resulting from a lack of iodine produces both simple goiter and cretinism. In simple goiter there may be a great enlargement of the thyroid gland, apparently a compensatory hypertrophy, more tissue being produced in the attempt to increase hormonal secretion. The body temperature is lowered, nervous symptoms appear, mental processes are dulled, weakness and lethargy become evident, the individual is easily exhausted, and unless the symptoms are relieved the patient progressively becomes worse until death ensues. The cretin is the extreme example of hormonal insufficiency, a congenital idiotic dwarf, usually born of a goiterous mother.

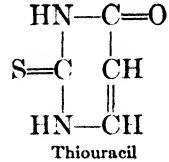
In adults hypothyroidism results in myxedema. Both conditions may be cured by either thyroxine or thyroid. Myxedema may be cured by 3,5-diiodotyrosine, given orally, although over *fifty times* the requisite dose of thyroxine is required, showing that it is not purely an iodine effect. Very definite symptoms of hypothyroidism may exist in adults without myxedema. These are always associated with a low basal metabolic rate. The symptoms are a tired, worn-out feeling, undue fatigability, loss of strength, nervousness, vague pains, sensitiveness to cold and infections, and sterility.

Hyperthyroidism apparently results from two types of pathological goiter, Graves' disease which is frequently (apparently inaccurately) designated exophthalmic goiter, involving diffuse hyperplasia (hypertrophy), and a non-hyperplastic goiter caused by a localized adenoma. The medical profession speaks of both types as toxic goiter. An excess

[18] L. E. Baker, *Science*, **88**, 480 (1938).

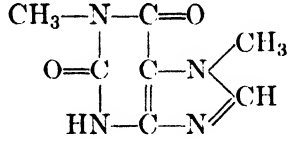
of anterior pituitary thyrotropic factor is believed by some authorities to be involved in the hyperthyroidism. It has been known for some time that various iodine compounds, especially Lugol's solution (iodine in potassium iodide), are effective in the treatment of toxic goiters, particularly pre-operative, but the cause of the beneficial action is not known. More recently thiourea and thiouracil have been found to be effective against hyperthyroidism. Both of these substances function


Thiourea


Thiouracil

by decreasing the production of thyroxine. Thiouracil is the less toxic of the two.

In 1943 Carter, *et al.*,[19] isolated from human urine an antithyroid substance, *paraxanthine*, previously found in whale and ox liver. It functions normally in the body to counterbalance thyroxine; apparently the



control of the basal metabolic rate is dependent on the established balance of thyroid hormone and paraxanthine.

The Hormone of the Parathyroid Glands. The parathyroid glands are small (3–15 mm. by 2–3 mm.) compact glands composed of epithelial cells, usually but not invariably situated adjacent to the dorsal surface of the thyroids. The French physiologist, Gley, in 1891, first demonstrated that the removal of the parathyroid glands resulted in tetany and death. In 1909, MacCallum and Voegtlin [20] found that the removal of the glands was accompanied by a fall in blood calcium to about one-half its normal level and that tetany and death could be prevented by the injection of calcium salts at suitable intervals. It was this finding that led to the theory that the function of the parathyroids is to regulate the calcium metabolism of animals. It has since been shown that sufficient dietary calcium can prevent the results of parathyroidectomy. The rat does not seem to be as dependent as other

[19] G. C. Carter, F. G. Mann, J. Harley-Mason, and G. N. Jenkins, *Nature*, **151**, 728 (1943).

[20] W. G. MacCallum and C. Voegtlin, *J. Expll. Med.*, **11**, 118 (1909).

animals on the parathyroid and accordingly can be used in experimental studies of chronic hypoparathyroidism. Parathyroidectomy in rats results in the development of opaque, brittle, distorted teeth and decalcified bones. In 1924, Hanson [21] and Collip [22] independently prepared active hormone extracts from the glands. However, although we now have a very clear picture of the physiological role played by the parathyroids and although standardized extracts are now available, little is known about the chemistry of the hormone. It is commonly called *parathormone*. It appears to be a protein. It gives the usual protein color tests. It is isoelectric at *p*H 4.8–4.9 and soluble in solutions more acid than *p*H 4.6 or more alkaline than *p*H 5.2. It is precipitated from solution by half-saturation with ammonium sulfate or full saturation with sodium chloride. It does not dialyze through collodion membranes; it is destroyed by boiling with 10 per cent hydrochloric acid or 5 per cent sodium hydroxide for an hour; and its activity is destroyed by either peptic or tryptic digestion. Accordingly the hormone is not effective when given by mouth. The most active preparations contain about 15.5 per cent of nitrogen. It has not been obtained in crystalline form.

Physiologically the hormone seems to exert some direct or indirect action on the labile mineral structures of the skeleton. Either hypo- or hyperparathyroid function results in a decalcification of the skeleton. Clinical hypoparathyroidism occurs most frequently after thyroidectomies where care has not been taken to preserve intact the parathyroid tissue. Parathyroid extracts have proved effective in certain cases of tetany in children and in other cases involving chronic convulsions, irrationalism, acute maniac excitation, etc., so that there is a possibility that hypoparathyroidism may be involved in certain instances in such conditions. Clinical hyperparathyroidism is demonstrated at its extreme in generalized *osteitis fibrosa* (von Recklinghausen's disease of bone) which was first described in 1891 but has only recently been definitely associated with the parathyroids. This disease appears almost invariably to result from a parathyroid tumor, and the removal of the tumor, which was first successfully accomplished in 1926, results in a return to normal. The disease can be simulated experimentally in animals by overdosing with parathyroid extracts. The skeleton is extensively decalcified, the bones no longer throw a dense X-ray shadow, hypercalcemia results, and there is a marked excretion of both calcium and phosphorus. Hyperparathyroidism appears to be invariably the result of a marked hypertrophy of the gland.

[21] A. M. Hanson, *Proc. Soc. Exptl. Biol. Med.*, **22**, 560 (1925).

[22] J. B. Collip, *J. Biol. Chem.*, **63**, 395 (1925); J. B. Collip, E. P. Clark, and J. W. Scott, *ibid.*, **63**, 439 (1925); J. B. Collip and E. P. Clark, *ibid.*, **66**, 133 (1925).

Another disease which is possibly due to chronic hyperparathyroidism is known as the "marble bone disease." In this disease the bones become exceedingly brittle. The effect of the parathyroid hormone on calcium metabolism has led to attempts to associate the action of the hormone with the effects of vitamin D, but no such interrelationship has been established.

Potent parathyroid extracts have been employed to mobilize lead and remove it from the body in cases of chronic lead poisoning. The lead, along with calcium, is largely deposited in the skeleton, probably as the very insoluble tertiary lead phosphate. The administration of the parathyroid hormone, bringing about, as it does, a bone decalcification, also liberates lead so that the amount of lead secreted in the urine is greatly increased.

The Hormone of the Thymus Gland. In most animals the thymus gland is situated in the upper portion of the thorax, although in some it is found in the neck. In the phylogenetic scale it appears first in the fishes. It is lacking in *Amphioxus*. It usually is of a grayish red color and generally consists of two lobes joined across a median plane. Each lobe is divisible into a cortex and a medulla, the latter containing characteristic concentric structures known as "Hassall's corpuscles." In man the gland normally reaches its maximum development in early childhood, during the period of most rapid body growth, and then recedes in size, atrophying after puberty. In castrates it may persist for much longer periods.

The hormone of the thymus gland appears to speed up cell proliferation and cell growth of all tissues and therefore influences the growth rate from birth to puberty. However, it seems to have no effect once normal adult growth has been attained.

The first extract to be made which exerted growth-promoting properties when injected into growing rats was made by Asher[23] in 1930. He called the concentrated purified extract *thymocresin*. The product was stated to be free from protein and lipids. Asher regarded it as a sulfur-containing polypeptide. General growth, growth of the skeleton, and growth of the gonads all appeared to be accelerated by thymocrescin.

In 1930, Hanson prepared a thymus extract by an entirely different method which was not extensively tested until Rowntree[24] began his

[23] L. Asher, *Endokrinologie*, **7**, 321 (1930); *cf.* also L. Asher and P. Stotzer, *Biochem. Z.*, **234**, 1 (1931); V. W. Nowinski, *ibid.*, **249**, 421 (1932); L. Asher and A. Zenklusen, *ibid.*, **252**, 309 (1932); P. Röthlisberger, *ibid.*, **253**, 137 (1932); and Doris Asher, *ibid.*, **257**, 209 (1933).

[24] L. G. Rowntree, J. H. Clark, A. M. Hanson, and A. Steinberg, *J. Am. Med. Assoc.*, **103**, 1425 (1934); *Arch. Internal Med.*, **56**, 1 (1935); L. G. Rowntree, J. H. Clark, A. Steinberg, A. M. Hanson, N. H. Einhorn, and W. A. Shannon, *Ann. Internal Med.*, **9**, 359 (1935); L. G. Rowntree, *J. Am. Med. Assoc.*, **105**, 592 (1935).

series of experiments three years later, using Hanson's original extracts. This extract and similar ones, which have shown such remarkable effects in Rowntree's hands, were so heterogeneous and impure that analyses give no hint of the chemical nature of the hormone. Beginning with young rats, 1 ml. of extract was given daily by intraperitoneal injection until maturity and through gestation and lactation of the females. The young rats of the next generation were treated similarly beginning on the sixteenth to twentieth day of life. This was continued through successive generations. Precocity began to appear in the young born in the later litters of the F_1 generation and continued to be greater in succeeding generations, so that by the F_9 generation the physical, sexual, and psychic development was extraordinarily precocious. Representative data are shown in Table 72.

TABLE 72. APPROXIMATELY AVERAGE EFFECT OF INJECTIONS OF THYMUS HORMONE ON PHYSICAL CHARACTERISTICS OF RAT PROGENY

(Data of Rowntree, *et al.*)

	Normal Control Rats	Effect of Thymus Extracts on Treated Rats		
		F_1 generation	F_8 generation	F_9 generation
Weight at birth	4.6 grams	5.1 grams	6.5 grams	6.0 grams
Ears opened	3 days	$2\frac{1}{2}$ days	Birth	Birth
Teeth eruption	9 "	$8\frac{1}{2}$ "	Birth	Birth
Hair appeared	14 "	11 "	1 day	1 day
Eyes opened	15 "	13 "	2 days	$1\frac{1}{2}$ days
Testes descended	37 "	22 "	$3\frac{1}{2}$ "	$2\frac{1}{2}$ "
Vagina opened	60 "	40 "	17 "	6 "
Pregnant	80 "	70 "	22 "	
First litter	102 "	92 "	43 "	

After the fifth generation the young rats ran about the cage when only 2–3 days of age, being as alert as normal rats at 16–20 days of age. Weaning is possible at 48 hours, after which they can nest for themselves and need no further parental care. They can swim at the third day. These rats do not become larger than normal at maturity, but reach maturity in weight and size in about one-half normal time. The growth-rate acceleration reached its maximum in the F_4 generation. The rats are reported to be more fertile than normal.

Rowntree, *et al.*,[25] have determined the presence of glutathione, cysteine, and ascorbic acid in active thymus extracts. All three substances are capable of accelerating the development of offspring of treated rats, and glutathione also accelerates the rate of growth of the offspring. Whereas all these reducing substances, especially glutathione, simulate the action of thymus extract in certain respects, they are certainly not wholly responsible for the biological effects of the thymus secretion.

There are no known clinical diseases of hypo- or hyperthymus activity. Thymectomy in five successive generations of rats has produced retarded growth, which can be overcome by thymus therapy.

The Hormone of the Pineal Gland. The pineal gland is a small, flattened, pine-cone-shaped body, an outgrowth of the epithalamus, situated in the mid-brain. It contains neurologia and ependymal cells, the latter probably being the secretory cells.

Feeding pineal tissue[26] to tadpoles causes them to become so translucent that the internal organs are rendered visible. The translucency develops about 30 minutes after feeding on the tissue and persists for several hours.

Hanson, at Rowntree's suggestion, prepared an extract of the pineal gland which Rowntree, *et al.*,[27] found to be physiologically potent when injected in a manner similar to that in the earlier thymus studies.

Practically nothing is yet known regarding the chemical nature of the active principle or principles. Inasmuch as Hanson's aqueous extracts (acidified) yielded an apparently active picrate, the chemistry points to a protein-like rather than lipid-like substance.

Rowntree's injection experiments with rats, in which solutions of the picrate were used, indicate a stimulation by it of the gonadotropic hormone of the anterior pituitary. The general effect on the rats in succeeding generations was the production of sexually precocious dwarfs. Growth, in general, was suppressed, the rats being only one-third or less as large as normals at corresponding age in the growth period, but the onset of adolescence and sexual maturity was greatly accelerated. In the F_4 generation the teeth erupted at 4 days, fur appeared at 5 days, the eyes opened at 6 days, the testes descended at 5 days, and the vagina

[25] L. G. Rowntree, A. Steinberg, N. H. Einhorn, and N. K. Schaffer, *Endocrinology*, **23**, 584 (1938).

[26] J. S. Huxley and L. T. Hogben, *Proc. Roy. Soc. London*, **93B**, 36 (1922); and C. P. McCord and F. P. Allen, *J. Exptl. Zool.*, **23**, 207 (1917).

[27] L. G. Rowntree, J. H. Clark, A. Steinberg, A. M. Hanson, N. H. Einhorn, and W. A. Shannon, *Ann. Internal Med.*, **9**, 359 (1935); L. G. Rowntree, J. H. Clark, A. Steinberg, and A. M. Hanson, *J. Am. Med. Assoc.*, **106**, 370 (1936); and L. G. Rowntree, J. H. Clark, A. Steinberg, and A. M. Hanson, *Science*, **83**, 164 (1936).

opened at 24 days of age instead of the normal ages, as shown in Table 72.

Rowntree reports no abnormalities in four successive generations of *pinealectomized* rats. Clear-cut clinical evidence of hypopineal function is lacking. A few cases of possible hyperpineal function, associated with tumorous condition of the gland, have been reported in young children, usually boys. There is abnormal growth, premature genital development, and early death.

The Hormone of the Pancreas Gland. The pancreas is a large, elongated gland located below the stomach between the spleen and duodenum. The direct secretion of the gland, the pancreatic juice containing proteolytic enzymes, passes through the pancreatic duct into the duodenum. The gland contains areas composed of cells smaller than the normal pancreas secretory cells. These areas, known as the *islets of Langerhans*, give rise to the hormone *insulin*, which passes into the blood circulation and not into the pancreatic duct. The islets comprise only $\frac{1}{30}$–$\frac{1}{100}$ of the mass of the pancreas.

The relation of the pancreas to diabetes was shown first by von Mering and Minkowski, in 1889, through experimental pancreëctomy in dogs. It was soon shown, however, by Hédon (1892), Minkowski (1892), and Gley and Thiroloix (1892), working independently, that the relation of the pancreas to the sugar metabolism was through an internal secretion. They succeeded in keeping pancreëctomized dogs in normal condition by grafting the pancreas under the skin, thus removing all possibility of its nerve connections being related to its activity. After this discovery many attempts were made to prepare active extracts of the gland. It remained for Banting, Best, Macleod, and Collip [28] to prepare the first successful extract in 1921. This discovery was recognized by the award of the Nobel prize to Banting and Macleod in 1923. They named the hormone *insulin*. Although insulin does not cure diabetes, by its use diabetic patients are able to have a fairly normal existence. Mathews states that considerably over 1,000,000 people in the United States have diabetes at the present time.

Insulin was isolated as a crystalline protein of low particle weight (35,100) by Abel [29] in 1926. The bulk of the molecule is accounted for [30] by eight amino acids, 30 per cent leucine, 21 per cent glutamic acid, 12

[28] F. G. Banting and C. H. Best, *J. Lab. Clin. Med.*, **7**, 251, 464 (1922); F. G. Banting, C. H. Best, J. Hepburn, J. B. Collip, J. J. R. Macleod, and E. C. Noble, *Trans. Roy. Soc. Canada*, [3] **16** (Sec. V), 27, 31, 35, 39 (1922).

[29] J. J. Abel, *Proc. Natl. Acad. Sci. U. S.*, **12**, 132 (1926).

[30] H. Jensen and E. A. Evans, Jr., *Z. physiol. Chem.*, **209**, 134 (1932); and H. Jensen and O. Wintersteiner, *J. Biol. Chem.*, **98**, 281 (1932).

per cent cystine, 12 per cent tyrosine, 8 per cent histidine, 3 per cent arginine, 2 per cent lysine, and about 10 per cent proline. About 0.5 per cent of all naturally crystalline insulin consists of zinc.[31] The zinc can be replaced by cadmium or cobalt. About three metal ions are associated with each insulin molecule. Zinc seems to have a stabilizing effect, for zinc-free insulin slowly decomposes on incubation at 52°C., but insulin does not decompose when zinc is present or added.

Insulin crystallizes in a rhombohedral cell with hexagonal axes of $a = 74.7$ Å. and $c = 30.2$ Å. In the cell each molecule is surrounded by eight others, two at a distance of 30.2 Å. and six at a distance of 44.3 Å. It is isoelectric at *p*H 5.5–5.6 and is free from phosphorus. It contains 3.2 per cent of sulfur, and its physiological activity appears to be closely associated with the sulfur-containing (cystine) portion of the molecule. Reducing agents destroy the physiological activity, which is not regained on subsequent oxidation. Inactivation parallels the liberation of —SH groups. Acetylation studies have shown that the hydroxyl groups on tyrosine are also vital to insulin activity. Using ketene as an agent, Stern and White [32] showed that acetylation of the free amino groups did not decrease the activity of insulin, but as the hydroxyl groups of tyrosine became slowly acetylated the hormonal activity was progressively reduced.

The protein nature of insulin predicts that it could not be especially effective, if at all, by mouth, inasmuch as it would be digested in the intestinal tract. A major development in insulin therapy involved the preparation of compounds of insulin with *protamines.*[33] Such compounds have already been noted in our consideration of the protamines. Use of these compounds slows down the absorption of injected insulin from the tissues, thereby rendering its action more uniform and more continuous. The use of *globin* (from hemoglobin) and zinc with insulin gives a complex with absorption properties intermediate between insulin and protamine-zinc-insulin.

Insulin profoundly influences carbohydrate metabolism by lowering the blood-sugar level and promoting the reaction glucose → glycogen in liver, heart, and the skeletal muscles. The exact mechanism of insulin action has not been ascertained.

Clinical hypoinsulinism, diabetes mellitus, seems to be associated with degeneration and atrophy of the cells of the islets of Langerhans which

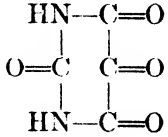
[31] D. A. Scott, *Biochem. J.*, **28**, 1592 (1934); D. A. Scott and A. M. Fisher, *ibid.*, **29**, 1048 (1935); and A. M. Fisher and D. A. Scott, *ibid.*, **29**, 1055 (1935).

[32] K. G. Stern and A. White, *J. Biol. Chem.*, **122**, 371 (1938).

[33] H. C. Hagedorn, B. N. Jensen, N. B. Krarup, and I. Wodstrup, *J. Am. Med Assoc.*, **106**, 177 (1936).

produce insulin. The blood sugar rises to highly abnormal levels, and both carbohydrate and fat metabolism are profoundly altered. The initial cause or causes of the dysfunction are not known.

Until recent years diabetes mellitus could be produced experimentally in animals only by extirpating the pancreas. In 1943 the discovery was made that the injection of *alloxan*, a uric acid derivative, into rabbits



Alloxan

produced typical diabetic symptoms.[34] Subsequently it has been shown to act similarly on rats and dogs. Alloxan causes specific degeneration of the islet tissue of the pancreas; when the dosage is properly regulated this is the only pathological effect of any consequence.

Clinical hyperinsulinism is also known. This is frequently associated with tumor of the pancreas cells which produce insulin. The symptoms are essentially those of overdosage of insulin. In the order of onset they are: (1) fatigue and lassitude; (2) cold perspiration, palpitation, tremor, thirst, fear; (3) clouded senses, pseudo-alcoholic intoxication, hallucinations; (4) convulsions and paralysis, with loss of memory; (5) coma. The blood sugar falls to a negligible value. It is probable that most of the symptoms are traceable to disturbances arising in the central nervous system, an actual brain starvation, since the brain is dependent for its oxidative processes upon the blood sugar and has no store of available carbohydrate which can be drawn upon in an emergency.

The Hormones of the Pituitary Gland. The pituitary is a small, endocrine gland situated beneath the brain in the sella turcica of the sphenoid bone. In man it averages about 0.57 gram in weight. "No other single structure in the body is so doubly protected, so centrally placed, so well hidden." Abundant evidences are accumulating that it exerts a master control over a great variety of bodily functions, including the activities of the other endocrine glands. If one organ of the body can be said to be more important than another, that role must be ascribed to the pituitary.

The pituitary is commonly divided into two lobes, the anterior and the posterior, usually referred to as either the anterior or posterior pituitary, connected by the *pars intermedia*. In whales the *pars intermedia* is absent, and the two lobes are entirely separate. The anterior

[34] J. S. Dunn, H. L. Sheehan, and N. G. B. McLetchie, *Lancet*, **1**, 484 (1943).

lobe or *pars distalis* is composed of both *chromophile* and *chromophobe* cells distributed somewhat indiscriminately throughout the gland. The posterior lobe, on the other hand, is composed of two distinct types of tissue, an inner core or *pars nervosa* and an outer lining of epithelial cells, the *pars intermedia.*

There appears to be no question that the pituitary secretes a multiplicity of hormones. The literature is exceedingly voluminous with respect to the number of hormones, their physiological functions, and their physical properties. All the evidence points toward hormones of protein or polypeptide nature, in many instances of a labile nature so that the activity is lost in attempts to concentrate or purify the extracts. Earlier in this chapter we have listed eight hormones ascribed to the anterior pituitary, one to the pars intermedia, and two to the posterior pituitary. These do not represent all the principles for which claims are made in the literature, nor is there any certainty that still more substances do not remain to be discovered.

Because of the uncertain state of our *chemical* knowledge, only the briefest mention will be made of the individual pituitary hormones.

Posterior Pituitary Hormones.[35] At the present time the evidence points to only two specific principles in the posterior lobe. These are (*a*) *pitocin* (*oxytocin*) and (*b*) *pitressin* (*vasopressin*). Proprietary extracts of the posterior lobe, containing both principles, are available under the name *pituitrin.*

duVigneaud, *et al.*,[36] have shown that the oxytocic and pressor activities of either untreated press-juice of the posterior lobe or of various chemically prepared concentrates have very different electrophoretic migration rates. Therefore, these activities must be associated with different chemical entities. More recently, however, van Dyke and associates [37] have obtained an apparently pure protein which has pressor, diuretic, and oxytocic activities inherent in the molecule.

Pitocin is apparently a cystine-containing polypeptide which is rich in tyrosine. Pitressin seems to be similar, but richer in cystine.[38] The chief effect of pitocin is on smooth muscle. Concentrates have been

[35] E. M. K. Geiling, Chapter X in *Glandular Physiology and Therapy, A Symposium,* American Medical Association, Chicago, 1935; G. W. Irving, Jr., p. 28 in *The Chemistry and Physiology of Hormones,* American Association for the Advancement of Science, Washington, D. C., 1944.

[36] V. duVigneaud, G. W. Irving, Jr., H. M. Dyer, and R. R. Sealock, *J. Biol. Chem.*, **123**, 45 (1936); G. W. Irving, Jr., and V. duVigneaud, *ibid.*, **123**, 485 (1936).

[37] H. B. van Dyke, B. F. Chow, R. O. Greep, and A. Rothen, *J. Pharm. Exp. Therap.*, **74**, 190 (1942).

[38] J. M. Gulland, *Biochem. J.*, **27**, 1218 (1933); J. M. Gulland and T. F. Macrae, *ibid.*, **27**, 1237 (1933); and R. R. Sealock and V. duVigneaud, *J. Pharmacol.*, **54**, 433 (1935).

made which are 1,000–1,250 times as powerful as is histamine. It can be and is employed by obstetricians to stimulate contraction of the uterus and apparently is normally involved in the termination of pregnancy.

Pitressin exerts several effects. One is on the blood capillaries, through which it affects blood pressure. Another is a pronounced antidiuretic effect. This effect seems to be brought about by increasing the reabsorption of water from the kidney tubules.

It would be anticipated that hypo- and hyperfunction of the posterior lobe would affect blood pressure, smooth-muscle contraction, and renal function. Such effects are, however, not common in human experience, nor are they readily produced in experimental animals. The only outstanding abnormal condition which is presumably associated with hypofunction of the posterior lobe is diabetes insipidus. This disease is characterized by the continued excretion of large volumes of a pale urine of low specific gravity, free from sugar. Normal kidney function is restored in many cases by continued injection of pituitrin, the effect being that of the pitressin present, the deficiency of which caused the diuresis. There have been some interesting suggestions that hypersecretion of pitressin may be one cause of gastric and intestinal ulcers, and that abnormal functioning of the pituitary is involved in the migraine-like "pituitary headache."

Pars Intermedia Hormone. This small portion of the pituitary connecting the anterior and posterior lobes secretes a hormone, *intermedin*, which affects the chromatophores of cold-blood animals. When extracts are injected into the minnow, *Phoxinus laevis*, red areas develop around the attachments of the fins. The significance of intermedin in higher animals is not known.

Anterior Pituitary Hormones. This remarkable gland, weighing less than 0.5 gram in the human, produces no less than eight different types of hormones, all but one of which may be regarded as exerting some specific effects. Most of these may be regarded as "master" hormones, in that *they exert specific control either over other endocrine glands or over the production or secretion of other hormones!* Little is known regarding the chemistry of most of the accepted principles, but it seems safe to believe from the facts available that all are proteins or protein-like substances. The "master" hormones of the anterior pituitary are:

1. THYROTROPIC PRINCIPLE. This hormone regulates the secretion of the active principle of the thyroid (thyroxine or thyroglobulin). The absence of this hormone produces all the effects of hypothyroidism; an excess produces all the effects of hyperthyroidism.

2. PARATHYROTROPIC PRINCIPLE. This hormone regulates the proliferation of the parathyroid gland cells.

3. ADRENOTROPIC PRINCIPLE (interrenotropic principle). This hormone controls the normal structure of the adrenal cortex and therefore indirectly controls the production of the adrenal cortical hormones.

4. INSULINOTROPIC PRINCIPLE. This hormone has control over the islets of Langerhans and regulates either insulin production or insulin action.

5. GONADOTROPIC PRINCIPLES. Two such hormones of pituitary origin have been identified. FSH, *the follicle-stimulating hormone*, stimulates male germ cells, ova, and particularly the ovarian follicles, and thus the production of their hormones. The other principle, known as LH or *luteinizing hormone*, stimulates the production of the corpus luteum cells of the ovaries and probably also the interstitial cells of the testes, and thus the production of their hormones.

Other gonadotropins which differ somewhat from those found in the pituitary have been found in the blood, urine, and placental tissue of pregnant women. These principles furnish the basis of the Aschheim-Zondek pregnancy test [39] (development of *corpora lutea* in immature mice by injection of ether-extracted urine). They are *not* found in the pregnancy urine of mice, rats, rabbits, bitches, cats, cows, sows, monkeys, or elephants.

6. DIABETOGENIC PRINCIPLES. The anterior pituitary produces two so-called *diabetogenic principles*, one known as the *blood-sugar-raising principle* and the other the *ketogenic principle*. The blood-sugar-raising principle exerts an antagonistic action toward insulin in the control of blood sugar, possibly acting on nerve centers which control carbohydrate metabolism. There is a strong possibility that the usual diabetes, diabetes mellitus, involves a deficiency of this principle as well as of insulin. The ketogenic principle stimulates the production of the so-called ketone bodies found in the urine in diabetes, especially β-hydroxybutyric acid. Other effects attributed to this principle are (1) the reduction of blood lipids, (2) a depression of the basal metabolic rate, (3) an increase in the specific dynamic effect of proteins.

7. THE GROWTH HORMONE.[40] Probably the anterior pituitary hormone which is capable of producing the most spectacular effects is the

[39] S. Aschheim, Chapter XVII in *Glandular Physiology and Therapy, A Symposium*, American Medical Association, Chicago, 1935.

[40] H. M. Evans, Chapter III in *Glandular Physiology and Therapy, A Symposium*, American Medical Association, Chicago, 1935; W. Marx and H. M. Evans, p. 47 in *The Chemistry and Physiology of Hormones*, American Association for the Advancement of Science, Washington, D. C., 1944.

so-called growth-promoting hormone. Extracts containing this hormone are sold commercially under various trade names, *e.g.*, antuitrin-G (growth), phyone. Only recently, however, have Li and Evans [41] obtained this hormone in a state that appears to be pure on the basis of potency, biological activity, and electrophoretic homogeneity in the Tiselius apparatus over a wide *p*H range.

The growth hormone seems to be produced by the alpha, acidophile, chromophile cells of the anterior pituitary which comprise about 37 per cent of the cell volume. Continued injection of preparations of this hormone into young animals was first shown by Evans, in 1921, to produce giants. His giant rats were essentially normal in proportion in all respects but corresponded in size to a human 10 to 12 feet tall. Similarly hyperproduction of this hormone in childhood produces human giants, several of which have been described in the medical literature.[42] No naturally occurring pituitary giants have been described among other animals, although some no doubt exist. A human hypophysial giant is described [43] as 8 feet 4 inches tall and weighing 390 pounds at 19 years of age. He was normal in size at birth, but began to grow abnormally rapidly at once. He weighed 30 pounds at 6 months of age, 67 pounds at 18 months of age; at 5 years he was 5 feet 4 inches tall; and at 13 years he was 7½ feet tall. Pituitary giants are usually not physically normal in all respects. The joints are generally enlarged and awkwardly formed. The sexual development is usually subnormal and may be expressed by either frigidity or sterility. They usually die before middle age, although there is a record of one attaining the age of 66.

Hyperproduction of growth hormone in adult humans results in the condition *acromegaly*, characterized by the gradual enlargement of the head and features and of the feet and hands, namely, those parts of the skeleton which can enlarge. Pituitary giants gradually develop the acromegalic characteristics. Acromegaly has been simulated experimentally in dogs by injection of hormone extracts.

There is as yet no information as to whether the growth hormone acts directly on the tissues which it stimulates, or indeed any inkling of how it acts. There is increased food consumption and definite indication of increased efficiency of food utilization, particularly of protein, but these effects are hardly sufficient to explain the remarkable results.

Hypofunction of the alpha cells results in dwarfs. Usually there is also hypofunction of the beta, basophile, chromophile cells, which com-

[41] C. H. Li and H. M. Evans, *Science*, **99**, 183 (1944).

[42] H. Cushing, *The Pituitary Body and Its Disorders*, J. B. Lippincott, Philadelphia, 1912.

[43] C. D. Humberd, *J. Am. Med. Assoc.*, **108**, 544 (1937).

prise about 11 per cent of the anterior lobe volume. Inasmuch as these cells apparently produce the sex-controlling hormones of this organ, these hypopituitary dwarfs often remain infantile and become excessively fat (Froelich's syndrome).

Hypofunction of the alpha cells together with hyperfunction of the beta cells produces sexually precocious dwarfs. Pituitary dwarfs develop well mentally and, since the discovery of the growth hormone, have been found to respond well to injection of growth-hormone preparations.

8. THE LACTOGENIC HORMONES. It is now clear that two distinct sets of hormones are involved in the lactation process. One is required for the growth and development of the mammary gland; the other controls its secretory activity. Experimental work on both phases of this problem began about forty years ago, but the actual solution had to await the isolation of the hormones involved. Early studies on the relation of hormones to secretory activity of the mammary glands included extracts of the posterior pituitary (and later epinephrine itself), the corpus luteum, thymus, placenta, thyroid, testes, and the adrenals. The work eventually centered on the posterior pituitary as the source of the most important milk-secretion hormone. The positive effects which many of the earlier investigators obtained are now explained on the basis that epinephrine, because of its effect on the nervous system and blood pressure, merely causes a more complete discharge of the milk already in the gland and does not actually stimulate milk secretion. Interest was attracted to the anterior pituitary in 1928–1929 by the German workers, Grüter and Stricker,[44] who obtained the first evidence of a true lactation secretion hormone produced in this gland. Further definite evidence was obtained in 1930–1931 by Gardner and Turner,[45] who named it *galactin*. Riddle [46] first isolated the hormone in a high state of purity in 1932 and named it *prolactin*, which name has become more generally adopted. The following facts are fairly well established:

(*a*) The mammary glands develop under the influence of progesterone and the estrogenic hormones. Glands which are caused to develop in virgins, castrates, and males by injection of progesterone and estrone do not secrete milk but do so when prolactin is injected.

(*b*) Prolactin initiates and maintains lactation in the developed gland. In order to demonstrate the presence of this hormone the glands must

[44] F. Grüter and P. Stricker, *Klin. Wochschr.*, [II] **8**, 2322 (1929).

[45] W. U. Gardner and C. W. Turner, *Missouri Agr. Exp. Sta. Res. Bull.* 196 (1933).

[46] O. Riddle, R. W. Bates, and S. W. Dykshorn, *Proc. Soc. Exptl. Biol. Med.*, **29**, 1211 (1932); *Am. J. Physiol.*, **105**, 191 (1933).

first be properly developed through the action of progesterone and estrone. The mammary glands of castrated males may thus be developed and after the injections of prolactin will secrete milk.

(*c*) The ovarian and placental hormones are antagonistic toward prolactin and prevent its secretion during pregnancy. At parturition there is sufficient decline in the female hormone output to permit the secretion of prolactin. These relationships may explain the failure of lactation in certain individuals and suggest a probable clinical use for prolactin.

(*d*) The maintenance of lactation is also influenced by a nervous mechanism.

(*e*) There is some evidence that the adrenal cortex secretes a hormone which has been called *cortilactin* and which is of importance in milk production.

(*f*) Thyroxine has been shown to increase milk secretion in cows, especially at the close of lactation.

(*g*) Prolactin is secreted in the milk; this is true, at least, for women.

(*h*) Prolactin, like the other hormones of the anterior pituitary, is protein [47] in nature. It is completely destroyed by tryptic digestion. When salt-free it can be boiled for considerable periods of time without appreciable loss of activity.

(*i*) Prolactin, when injected into young virgin rats, has been shown by Riddle to stimulate the mothering, cuddling, and nesting instincts.

In recent years a number of reports have appeared which furnish rather convincing proof that prolactin is not the sole lactogenic factor secreted by the anterior pituitary.[48] Thus, purified prolactin preparations fail to stimulate lactation as readily as impure hypophyseal extracts. The other pituitary factors concerned in lactation have not been isolated or characterized.

The Hormones of the Organs of Reproduction. Several hormones already discussed have definite effects on sex characteristics or functions. This is true of thyroxine, the adrenal cortex hormones, the thymus and pineal principles, and certain specific principles of the anterior pituitary. The latter exert a direct effect on either the testes or ovaries, but at present it can only be surmised whether the thyroid, adrenal, thymus, and pineal principles act through the gonads or not. At any rate, we seem to have here the most outstanding example of the

[47] R. W. Bates, O. Riddle, and E. L. Lahr, *Proc. Soc. Exptl. Biol. Med.*, **31**, 1223 (1934); A. White, H. R. Catchpole, and C. N. H. Long, *Science*, **86**, 82 (1937).

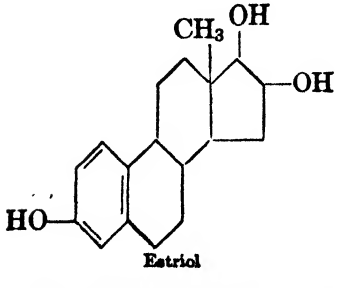
[48] For a comprehensive review of factors involved in lactation see W. E. Petersen, *Physiol. Revs.*, **24**, 340 (1944).

interlocking chain of structures which the glands of internal secretion exhibit.

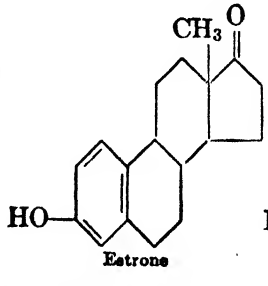
Indirect evidence of hormones produced by the gonads existed prior to the isolation of the first of these in 1923 [49] from the ovarian follicular fluid. Since that time we have seen a most remarkable development in hormone chemistry. As a matter of fact, most of this development has come about within the past ten or fifteen years and was made possible by the discovery (1) that these hormones have the same structural nucleus as cholesterol and its related compounds, (2) that the various gonadal hormones are very closely related chemically, and (3) that, because of the advances in knowledge regarding the chemistry of the sterols in other laboratories, it has been a relatively simple matter to elucidate completely the chemical nature of these gonadal principles. The isolation of these hormones, the elucidation of their structural configuration, and in a number of instances their *synthesis* in the chemical laboratory must be looked upon as one of the major achievements of the organic chemist. Space will not permit an adequate review of this fascinating story. Many workers have contributed their mite. Among the more outstanding are Edgar Allen, A. Butenandt, J. B. Collip, E. C. Dodds, E. A. Doisy, G. F. Marrian, L. Ruzicka, and B. Zondek.

We have earlier commented upon the fact that the sterols, the bile acids, the saponins, the cardiac glycosides, the toad poisons, and the cortical hormones of the adrenal glands all are derivatives of the same fundamental nucleus. To this list we must now add the sex hormones and the carcinogenic hydrocarbons.

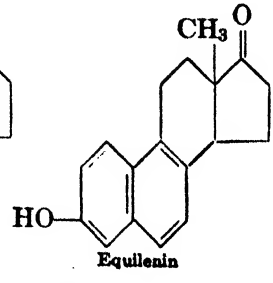
The Estrogenic Hormones. At least five compounds are normally involved in the estrus mechanism. These are (1) estriol, (2) estrone, (3) equilenin, (4) equilin, and (5) estradiol, which possess the respective configurations noted in the formulas



Estriol

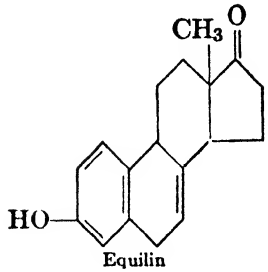


Estrone

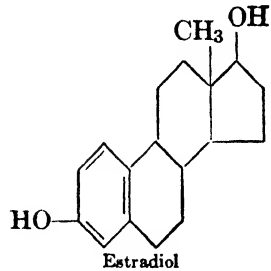


Equilenin

[49] E. Allen and E. A. Doisy, *J. Am. Med. Assoc.*, **81**, 819 (1923).

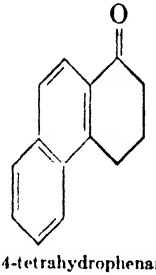


Equilin

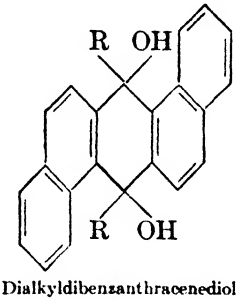


Estradiol

The knowledge of these configurations caused investigators to study the behavior of related compounds, and it was early found, by Cook, *et al.*,[50] that 1-keto-1,2,3,4-tetrahydrophenanthrene possessed estrogenic activity to a high degree, similar to the natural estriol, and estrone.



1-Keto-1,2,3,4-tetrahydrophenanthrene



Dialkyldibenzanthracenediol

These authors prepared an extensive series of related compounds. The series possessing the dialkyldibenzanthracenediol structure has been of outstanding interest. Here the type of alkyl radical in a large measure determines [51] the estrogenic activity of the compound. The dimethyl, di-*n*-amyl, and di-*n*-hexyl compounds are inactive. The diethyl and di-*n*-butyl compounds are active but considerably less so than the di-*n*-propyl compound. The diisopropyl compound was only about one-tenth as active as the *n*-propyl compound, and the isobutyl compound showed only about one-tenth the activity of the *n*-butyl compound. Dodds points out that the phenanthrene nucleus does not appear to be a necessity for estrogenic activity in synthetic compounds, for such compounds as 4,4′dihydroxydiphenyl, diphenyl-α-naphthylcarbinol, 1,2-dihydroxy-1,2-di-α-naphthylacenaphthene, and a number of related compounds are highly estrogenic. He suggests that the diphenyl nucleus or some similar arrangement of the benzene rings may be important.

[50] J. W. Cook, E. C. Dodds, C. L. Hewett, and W. Lawson, *Proc. Roy. Soc. London*, **B114**, 272 (1934).

[51] E. C. Dodds, *Helv. Chim. Acta*, **19** (*Fasciculus Extraordinarius*), E49 (1936).

All the estrogenic hormones arise in the follicular fluid of the ovaries, and apparently all occur together. Their secretion precedes ovulation. Estradiol is the most active and is the parent substance of those which were first isolated. The activity of these substances is progressively lost through either loss or addition of alcohol groups, or through further unsaturation of the ring structure (estradiol > estrone > estriol > equilin > equilenin).

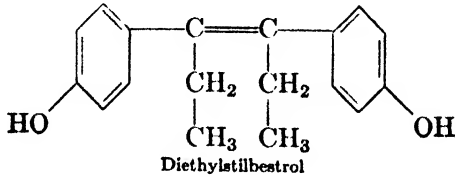
Estradiol, estrone, and estriol control the occurrence of estrus in all species which show this physiological phenomenon and will also produce menstruation and its attending changes in women and primates which have undergone ovariectomy. Estrone, especially, has found clinical use in alleviation of severe subjective symptoms in the menopause in women. These hormones are active when taken by mouth.

The estrogenic hormones have a definite relation to the growth of the mammary glands, especially the ducts and alveoli. They therefore play a specific role in the development of the glands during the attainment of sexual maturity.

Equilin and equilenin may be regarded as derivatives of estrone (theelin). They occur in the urine of pregnant mares and exert only slight estrogenic activity. They perhaps merely represent a mode of inactivation of the estrone which persists during pregnancy and thus account for the failure of the animal to show estrus during pregnancy.

Estradiol, estriol, and estrone combine in the body with glucuronic acid where the acid is esterified on the hydroxyl groups of the hormones, to form *emmenins*. These apparently occur abundantly in the placenta and, as would be expected, in pregnancy urine. The esterification modifies their action so that they have no estrogenic effect on castrates or on mature normal animals. The emmenins do produce estrus in normal sexually immature animals. They are effective by mouth as well as by injection.

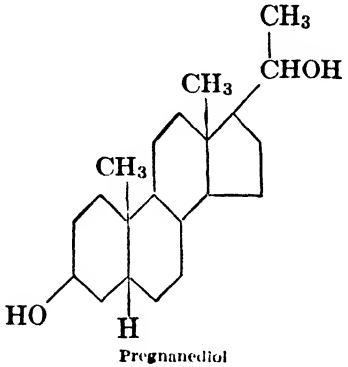
A relatively inexpensive synthetic compound, *diethylstilbestrol*, has come into considerable medical usage. It can be administered orally



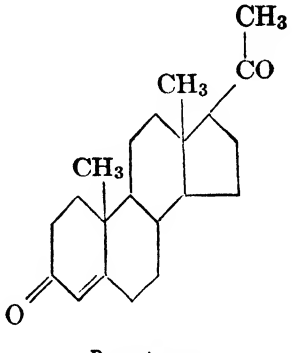
Diethylstilbestrol

and has two to three times the estrogenic potency of estrone. The 4-hydroxyphenyl groups are importantly concerned with the physiological activity of this substance.

The Hormones of the Corpus Luteum and Pregnancy. Two compounds, *progesterone* and its metabolic product *pregnanediol*, closely allied in structure to the other sex hormones, occur abundantly in the corpus



Pregnanediol



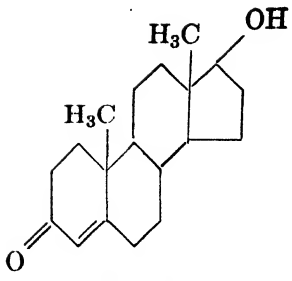
Progesterone

luteum and in pregnancy urine. Some endocrinologists believe that they also are produced in the placenta. The active agent of this pair, progesterone (also called progestin, corporin, luteosterone, lutin), has the special functions of (1) inhibiting ovulation, (2) stimulating the growth of the uterus, and (3) sensitizing the uterus for the implantation of the embryo (or embryos). The absence of this hormone through the removal of the corpus luteum during the earlier part of pregnancy causes failure of implantation of the embryo, or, if the embryo is already implanted, it causes destruction of the pregnancy. So far as is known the corpus luteum has no useful action except in pregnancy.

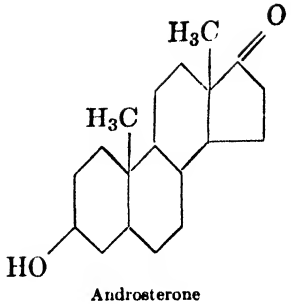
Progesterone is not effective by mouth, but must be injected, usually subcutaneously, in oil. Progesterone, along with the estrogenic hormones, causes the development of the mammary gland in pregnancy. It appears to be the chief agent concerned with the preparation of these glands for lactation.

The Male Sex Hormones. The male sex hormone, *testosterone*, and its relatively feeble derivatives, *androsterone*, *dehydroandrosterone*, and *androstanedione*, are remarkably closely related to the ovarian estrogenic hormones.

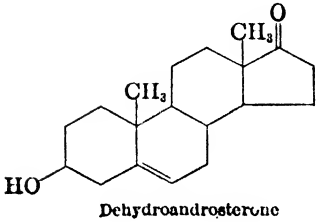
Testosterone differs from estradiol only in having an additional —CH_3 group and a ketone group in place of one of the —OH groups. Because of these differences there is, of course, less unsaturation in the ring nucleus. These slight differences suggest the possibility of interconversion. In fact, interconversion does occur, and their closely related structure explains the apparent anomalous demonstration of Zondek that the materials richest in estrogenic substances are stallions' testes



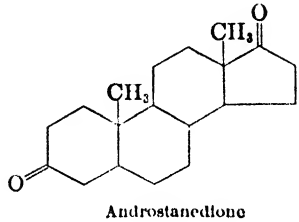
Testosterone



Androsterone



Dehydroandrosterone



Androstanedione

and urine, where both estradiol and estrone are found. The reverse conversion occurs in females, although not so extensively. Mathews [52] believes that the formation of relative excess of testosterone or estradiol, respectively, in early embryonic life determines the sex of the embryo. In the human this would have to occur before the third week, at which time the sex may be distinguished. Mathews emphasizes that in many particulars every individual not only starts as a hermaphrodite but remains one at all stages of its life. This appears to the true so far as the ovarian and testicular hormones are concerned. His idea is that a slight overbalance of male or female hormone starts the individual in the direction of male or female, which is kept going in that direction by the formation of the glands which produce the male or female hormone predominantly. In support of this he cites the experiment of Koch and Willier, who produced an ovary out of a left protestis in the developing chick embryo by injecting estrone into the incubating egg. In fowls the left protestis develops into the functional ovary of the hen, both sides developing testes in males. Thus, according to Mathews' reasoning, the so-called genital ridge of the mesoblastic tissue of the chick embryo from which testes and ovaries develop is hermaphroditic, not neutral, and needs only the predominance of male or female hormone to develop into male or female tissue. Mathews also speculates as to the possibility

[52] A. P. Mathews, *Principles of Biochemistry*, William Wood and Co., Baltimore, 1936.

that the chromosome theory of sex determination is likewise explainable on the basis of predominance of male and female hormone.

The testicular hormone is very similar in structure to the corpus luteum hormones, pregnanediol and progesterone. There is definite evidence that testosterone is changed to progesterone in the stallion because the latter occurs in stallion urine. Another striking example of either specific function of male hormone in the opposite sex or its change to the female hormone is seen in the use of the female bittering as a test animal for the male hormone. The addition of testosterone or androsterone to the aquarium water in which the female bittering is to be placed causes growth of its ovipositor. This is not caused by addition of estriol or estrone.

Testosterone and the other androgens are produced in the interstitial tissue of the testes. The functions of these hormones are (*a*) the determination and control of all the secondary sex characteristics of males, and (*b*) the maintenance of the normal functional condition of the male accessory sex organs, especially the prostate and seminal vesicles. The control exercised over the secondary sex characteristics is manifested by ability to promote comb growth and male plumage in cockerels and capons, the former being employed for biological assay. The male hormones have been shown to maintain the normal viability and mobility of spermatozoa and the normal chemical and biochemical properties of the semen. The prostate degenerates in the absence of these hormones and is restored by their injection. This is also true of seminal vesicles. It should be remembered, however, that the production of these hormones is under the control of principle LH of the anterior pituitary. Simple prostate hypertrophy appears to be caused by unbalance between this gonadotropic hormone of the pituitary and the testicular hormones. When the testicular hormones are deficient, the hypophysis produces an excess of LH which in turn causes an excessive stirring of the interstitial cells to produce testosterone. The excess testosterone causes the prostate to enlarge.

An odd effect of the male sex hormone (testosterone propionate) has been recorded by Hamilton and Hubert.[53] They noted that surgically castrated or hypogonadal males tanned but little, if at all, when exposed to the sun or ultraviolet light. When, however, the hormone was supplied some weeks (or even up to five months) later, without further exposure a decided tan developed except in the areas previously protected by a bathing suit. They likened the effect to a photographic process, where the "exposure" to sunlight lays down a rather colorless material in the

[53] J. B. Hamilton and G. Hubert, *Science*, **88**, 481 (1938).

skin which is subsequently "developed" into pigment by the male sex hormone. Furthermore, the tan so brought out on patients with insufficient testicular secretion would gradually fade when hormonal treatment was discontinued and reappear when therapy was reinstituted.

Gastro-Intestinal Hormones.[54] Bayliss and Starling[1] applied the term "hormone" to a secretion of the duodenal mucosa which induced a flow of pancreatic juice. Pavlov had earlier observed that acids in the duodenum caused such a flow, but Bayliss and Starling showed that the action was not a nerve response and were able to prepare from duodenal mucosa an extract which, when injected into the blood stream, activated the pancreas and caused a copious flow of pancreatic juice. They named this hormone *secretin* and showed that it was stable to heat but was destroyed by oxidizing agents or by proteolytic enzymes. Secretin is insoluble in the usual organic solvents, it dialyzes through collodion or parchment membranes, it contains sulfur, it is precipitated by picric, tannic, phosphotungstic, and trichloroacetic acids, and the weight of evidence is that it is a polypeptide or a proteose. It has been isolated as the crystalline picrolonate.[55] An enzyme which destroys secretin has been shown to be present in blood serum.[56] This substance, *secretinase*, probably accounts for the subsiding of pancreatic secretion after a meal and the absence of secretin in urine.

Two other hormones have been separated from the mucosa of the upper intestine. *Cholecystokinin* stimulates contraction and emptying of the gall bladder; *enterogastrone* apparently inhibits the secretion of hydrochloric acid by the parietal cells of the gastric glands and, accordingly, is effective in offsetting experimental ulcer. A substance, *urogastrone*, similar in biological activity to enterogastrone, has been found in urine; presumably it is an excretory product of enterogastrone. The exact chemical nature of these hormones is not yet known.

For many years the presence of one or more substances in the stomach (pyloric) mucosa which stimulate gastric secretion has been recognized. Edkins,[57] in 1906, first put forward evidence that a "gastric secretin," as he termed it, existed in pyloric extracts. Since that time the active principle, *gastrin*, has been widely investigated, but some controversy regarding its chemical nature still exists. Histamine is known to be

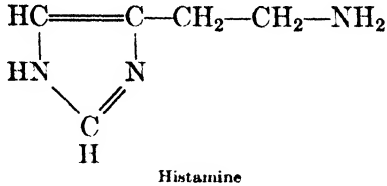
[54] For excellent reviews, see A. C. Ivy, Chapter XXX in *Glandular Physiology and Therapy*, American Medical Association, Chicago, 1942; and H. Greengard, p. 174 in *The Chemistry and Physiology of Hormones*, American Association for the Advancement of Science, Washington, D. C., 1944.

[55] E. Hammarsten, G. Agren, A. Hammarsten, and O. Wilander, *Biochem. Z.*, **264**, 275 (1933); H. Greengard and A. C. Ivy, *Am. J. Physiol.*, **124**, 427 (1938).

[56] H. Greengard, I. F. Stein, Jr., and A. C. Ivy, *Am. J. Physiol.*, **133**, 121 (1941).

[57] J. S. Edkins, *J. Physiol.*, **34**, 133 (1906).

present in most "gastrin" extracts [58] in amounts sufficient to account for all or nearly all the vasodepressor and secretory-stimulating properties of such extracts. On the other hand, Komarov [59] has obtained a

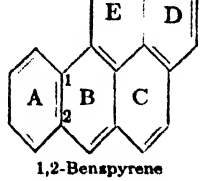


Histamine

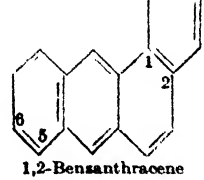
histamine-free preparation which does not affect the blood pressure but still causes gastric (and pancreatic) secretion. He has fractionated this preparation into two active principles which he refers to as gastrin and gastric secretin, the first eliciting the flow of gastric juice. the second functioning on the pancreas like secretin.

THE CARCINOGENIC HYDROCARBONS

It has been known for a number of years that repeated application of coal tar to the skin will eventually induce cancerous growths. Cook [60] and collaborators at the Research Institute of the Cancer Hospital, London, have been especially active in the investigation of this phenomenon. In 1933, they [61] reported the isolation and identification of the active constituent in coal tar. This proved to be 1,2-benzpyrene. 1,2-Benzanthracene was also isolated from the tar, and, although it is not itself markedly carcinogenic, a number of its derivatives have been shown by Cook, *et al.*, to have great carcinogenic activity. It can be regarded as the parent substance of a host of carcinogenic compounds. Substitutions in the 5- and 6-positions are especially potent; substitu-



1,2-Benzpyrene



1,2-Benzanthracene

[58] J. Sacks, A. C. Ivy, J. P. Burgess, and J. Vandolah, *Am. J. Physiol.*, **101**, 331 (1932).

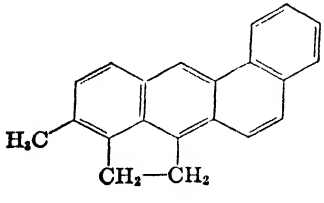
[59] S. A. Komarov, *Proc. Soc. Exptl. Biol. Med.*, **38**, 514 (1938); *Am. J. Physiol.*, **126**, 558 (1939); *Rev. can. biol.*, **1**, 191, 377 (1942).

[60] J. W. Cook, *et al.*; see numerous papers in *Proc. Roy. Soc. London, Series B*, and *J. Chem. Soc.*, 1930 to date.

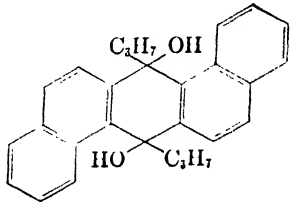
[61] J. W. Cook, C. L. Hewett, and I. Hieger, *J. Chem. Soc.*, 395 (1933).

tions in position-5 have in general somewhat greater carcinogenic activity than similar substitutions in position-6. Substitutions in other positions do not appear to affect carcinogenic activity greatly.

Among the derivatives having powerful carcinogenic power is methylcholanthrene (*A*), which can be considered a 5,6-dialkyl-1,2-benzanthracene and which bears close structural relationships to the sterols, the bile acids, and the sex hormones. It has been synthesized from desoxycholic acid, a normal bile constituent, and shows some estrogenic activity in addition to carcinogenic properties. Another synthetic substance (*B*) which has about the estrogenic potency of natural estriol, yet is distinctly carcinogenic, is a derivative of dibenzanthracene. Al-



A



B

though the thesis has not been proved, it seems probable that methylcholanthrene or certain of its potent derivatives may arise in the body from the sterols, the bile acids, or the sex hormones and thus be the normal carcinogenic agents.

The reader who wishes to pursue this interesting and important series of studies, with particular reference to the space relationships of the organic chemistry involved, is referred to the excellent summaries of Fieser.[62]

THE PLANT HORMONES

Whereas the term "hormone" was originally intended to designate only specific principles elaborated by the animal body and in particular by the endocrine glands, it has been adopted by plant physiologists to indicate those special factors which greatly stimulate plant growth.[63]

[62] L. F. Fieser, p. 81 in *The Chemistry of Natural Products Related to Phenanthrene*, Reinhold Publishing Corp., New York, 1936; p. 1 in *Cause and Growth of Cancer*, University of Pennsylvania Press, Philadelphia, 1941.

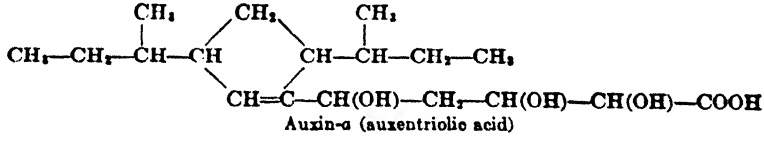
[63] For literature see P. Boysen-Jensen, *Growth Hormones in Plants*, translated by G. S. Avery and P. R. Burkholder, McGraw-Hill Book Co., New York, 1936; K. V. Thimann and J. Bonner, *Physiol. Revs.*, **18**, 524 (1938); F. W. Went, *Ann. Rev. Biochem.*, **8**, 521 (1939); J. van Overbeek, *ibid.*, **15**, 631 (1944); F. Skoog, *ibid.*, **16**, 529 (1947); P. W. Zimmerman and A. E. Hitchcock, *ibid.*, **17**, 601 (1948).

The work in this field really dates back to Charles Darwin, who, in 1881, in his book, *The Power of Movement in Plants*, suggested that, "When seedlings are freely exposed to a lateral light, some influence is transmitted from the upper to the lower part, causing the latter to bend." This hypothesis was first proved experimentally by Boysen-Jensen, in 1910–1911, who blocked off the path of transmission with mica plates and thus demonstrated that something which promoted cell growth was being transmitted from the growing coleoptile tip toward the base of the coleoptile. These observations were followed by a host of investigations which finally culminated in the isolation and identification of the chemical structure of at least certain of the more important chemical compounds involved. In this series of studies the contributions from the laboratories of Kögl and Went, Boysen-Jensen, and Thimann should be especially recognized.

One method of testing for the presence of the growth-stimulating substances is either (*a*) to apply the material unilaterally directly to decapitated coleoptiles of *Avena sativa*, or (*b*) to make up the material to be tested in a 3 per cent agar gel or in lanolin and apply these preparations, under standardized conditions, unilaterally to decapitated *Avena* coleoptiles. The curvature of the treated coleoptile after a standard period of time—110 minutes (Thimann), 120 minutes (Went), 3 hours (Boysen-Jensen)—depending on the exact procedure being followed, is then measured. The curvature is proportional to the *concentration* of the growth substance in the material applied to the decapitated coleoptile.

Kögl named these growth substances *auxins*, from the Greek $\alpha\upsilon\xi\epsilon\iota\nu$, to increase. Three compounds that show these remarkable growth-promoting properties have been isolated from natural sources, *i.e.*, *auxin-a*, *auxin-b*, and *heteroauxin*. All three are monobasic acids.

Auxin-*a* or *auxentriolic acid*, $C_{18}H_{32}O_5$, is a monocyclic trihydroxy-monocarboxylic acid with one double bond; it has the formula

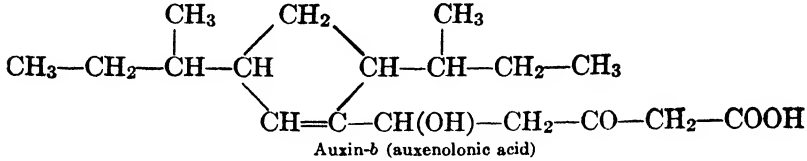


Auxin-a (auxentriolic acid)

It is thermostable, not decomposed by light, but undergoes slow spontaneous isomerization so that it becomes physiologically inactive after a few months, even when kept in the dark in a vacuum. It is relatively stable to acids but rather sensitive to alkalies.

Auxin-*b* or *auxenolonic acid* is closely related in structure to auxin-*a* but has only one hydroxyl group, the other hydroxyl groups of auxin-*a*

being replaced by a carbonyl oxygen and hydrogen respectively. It likewise is heat- and light-stable but is readily oxidized and becomes inactive in a few months, presumably through isomerization. *Heteroauxin* has proved to be *β-indoleacetic acid.*



Auxin-*b* (auxenolonic acid)

Both auxin-*a* and auxin-*b* have been obtained from malt and the oils of maize, mustard, sunflower, and flax. Maize oil is, in general, richer in auxins than are the other vegetable oils which have been examined. Kögl,[64] *et al.*, investigated thirty-six maize oils and found that oils from different geographical areas (and presumably maize sorts) differed widely in auxin content, the range being from 0–540 μg. per kg. with an average of 103 μg. Twenty-two samples of unmalted barley showed a range of 0–43 μg. of auxin with an average of 5 μg. per kg. Malting increased the auxin content up to the fifth day of malting when a maximum content of 420 μg. per kg. was observed. From one 16-kg. lot of maize oil, Kögl isolated 15 mg. of crude crystals of auxin-*a* and 15 mg. of crude crystalline auxin-*b*. On purification 6 mg. of pure auxin-*b* was obtained. One sample of maize oil was found which contained 10–13 mg. of auxin per kilogram. From 160 kg. of this oil, Kögl isolated 238 mg. of auxin-*b* and 7 mg. of auxin-*a*. These data are given to indicate the enormous dilutions in which these growth substances occur in nature, and the difficulties which attended their isolation and identification.

Human urine was found by Kögl [65] to be a rich source of auxin-*a* and auxin-*b*, having an auxin content of approximately 2 mg. per liter. Urine also contains heteroauxin.[66] The relative growth-promoting activities of these three compounds are 50,000,000 AE [67] for both auxin-*a* and auxin-*b*, and 25,000,000 AE for heteroauxin (β-indoleacetic acid). Various synthetic compounds, including many derivatives of indole, benzene, naphthalene and anthracene, and even such substances as ethylene gas, have been tested for growth-promoting properties. Although a

[64] F. Kögl, H. Erxleben, and A. J. Haagen-Smit, *Z. physiol. Chem.*, **225**, 215 (1934).

[65] F. Kögl, A. J. Haagen-Smit, and H. Erxleben, *Z. physiol. Chem.*, **214**, 241 (1933); **220**, 137 (1933).

[66] F. Kögl, A. J. Haagen-Smit, and H. Erxleben, *Z. physiol. Chem.*, **228**, 90 (1934).

[67] AE = *Avena Einheit, i.e.*, that amount of growth substance which, when present in a block of 3 per cent agar, 2 by 2 by 0.5 mm. in size, will in 2 hours cause a curvature of 10° in the decapitated coleoptile held at 22–23°C. in a relative humidity of 92 per cent.

considerable number of these synthetic agents show "hormonal" activity, most of them do not approach the natural auxins in efficiency. Thus β-(α-methylindole)-acetic acid has an activity of only 125,000 AE. β-Indolepropionic acid, β-indolecarboxylic acid, β-2,3-dihydroindoleacetic acid, and β-(α-ethylindole)-acetic acid are all inactive in this respect.

It is generally believed that the auxins are responsible for the growth of plant cells. Their function seems to be limited to cell enlargement, not to cell proliferation, although some workers have claimed that they speed up and facilitate cell division. They appear to be formed in the rapidly growing tips of the coleoptile and to be transported downward. They are also formed in root tips, buds, and apparently in the extremities of all rapidly growing plant parts. Heteroauxin has been isolated from various fungi, including yeasts, *Aspergillus niger*, and *Rhizopus* sp., and appears to be the principal, if not the sole, growth-promoting principle in these lower plant forms. Heteroauxin has already found a commercial application in inducing the formation of roots on cuttings, thereby making possible the commercial propagation of certain plants which normally could not be propagated in this manner. Two synthetic substances, β-indolebutyric acid and α-naphthaleneacetic acid, likewise are highly effective root-forming agents.

In 1939 a "wound hormone," so called because it elicited renewed growth in mature parenchymatous cells, was isolated in crystalline form from green string beans.[68] It was named *traumatic acid* and shown to be 1-decene-1,10-dicarboxylic acid.

$$HOOC—CH{=}CH—(CH_2)_8—COOH$$

In the preface to the first edition of this text it was emphasized that in the last analysis the chemistry of the cell is essentially the same for both plants and animals. That this is true is becoming more and more evident as research progresses. Numerous workers are finding that vitamins which hitherto have supposedly affected only animal processes, and even certain of the sex hormones, likewise play important roles in the development of plants. Bonner[69] and Robbins and Kavanagh[70] have reviewed some of the work in this field, and their papers may be consulted for literature sources.

[68] J. English, Jr., J. Bonner, and A. J. Haagen-Smit, *Science*, **90**, 329 (1939); *J. Am. Chem. Soc.*, **61**, 3434 (1939).

[69] J. Bonner, *Botan. Rev.*, **3**, 616 (1937).

[70] W. J. Robbins and V. Kavanagh, *Botan. Rev.*, **8**, 411 (1942).

CHAPTER 38

Enzymes

Catalysis. The phenomenon of catalysis is widely recognized. A catalyst is a substance which alters the speed of a reaction without appearing as part of the final products; stated in another way, the catalyst influences the speed of a reaction without altering the energy contents of the reactants and the resultants. Many of the changes which occur in living matter are regulated by certain biocatalysts called *enzymes.*

The criteria for an ideal catalyst have been accepted since the time of Ostwald: (*1*) The substance should be returned unchanged chemically after a reaction, although it may be altered physically; thus a platinum gauze used in fixing nitrogen as ammonia becomes pitted and gray with use although its efficiency is thereby increased. (*2*) The catalyst should function in small quantities, and the amount of change should be proportional to the amount added, other conditions being constant. (*3*) The catalyst should not change the equilibrium point of the reaction. It follows as a corollary that the same catalyst should function whenever the equilibrium point is approached from both sides; this has been verified in many instances. At times the catalyst appears to alter the equilibrium point, but such observations can generally be ascribed to a side reaction. Kendall and Booge [1] showed cryoscopically that the organic acids employed to catalyze the hydrolysis of ethyl acetate formed a complex with the ester and that the apparent shift in the equilibrium point was satisfactorily accounted for on this basis.

It would also follow from (3) that a catalyst cannot initiate a reaction. This postulate has been seriously challenged by such men as Schönbein, Armstrong, and J. J. Thomson. Experimentally this is difficult to test because many reactions proceed too slowly to give measurable changes. Furthermore, it is difficult to devise conditions under which the presence of a catalyst may be ruled out; it is known that traces of moisture and dust particles, the solvent, and even the walls of the container may greatly influence the nature of a reaction. Even the

[1] J. Kendall and J. E. Booge, *J. Am. Chem. Soc.*, **38**, 1712 (1916).

method of preparing a catalyst, and hence its surface condition, may determine the course of events, as was shown by Adkins and Nissen [2] in the case of aluminum oxide. Boswell and Dilworth [3] examined the same catalyst and concluded that variations in the method of preparation determined the nature of the catalytic surface and that an oriented adsorption layer of water was the real seat of the catalysis. They continue, "The catalyst from this point of view does not accelerate a reaction already in progress . . . but actually initiates the change . . ." Also, there is abundant proof that in many instances the catalyst does enter into the reaction—at least to the extent of forming labile compounds with the reactive molecules of the substrate. These labile compounds then are rearranged or altered while associated with the catalyst and break down with the formation of new products and the regeneration of the catalyst.

Two earlier hypotheses to account for catalytic action have more recently been reconciled in a general explanation. The first postulate was due to Bayliss,[4] who argued that capillary adsorption played the major role in that it led to a localized high concentration of the reactants, which in turn would give rise to an increase in the rate of reaction. There are objections to this postulate. First, it does not account for the specificity of catalysts; thus, isobutyl alcohol when passed over copper at 300°C. yields the aldehyde and hydrogen, whereas with aluminum oxide at the same temperature the products are isobutylene and water. Second, there does not appear to be any quantitative relationship between the amount of adsorption and the extent of catalysis.

The second explanation involved the formation of an unstable intermediate compound from the reactants and the catalyst, which in turn decomposed to give the products of the reaction. The formation of an intermediate complex would depend on the selective action of the catalyst and hence would be in conformity with the recognized specificity of catalysts. Several intermediates have been described; for example, nitrosylsulfuric acid has been demonstrated as an intermediate in one process for making sulfuric acid.

The view most widely held today is a concept of activated adsorption involving the formation of intermediate complexes through a graded series of linkages varying from those of the van der Waal type of binding to that approaching the usual stable chemical type of binding. The former is that involved in the adsorption (capillary condensation) de-

[2] H. Adkins and R. H. Nissen, *J. Am. Chem. Soc.*, **46**, 130 (1924).

[3] M. C. Boswell and H. M. Dilworth, *J. Physical Chem.*, **29**, 1489 (1925).

[4] W. M. Bayliss, *The Nature of Enzyme Action*, 5th ed., Longmans, Green and Co., New York, 1925.

scribed by the Freundlich adsorption isotherm (see p. 178), whereas the latter requires higher heats of adsorption and an activation energy for its achievement.[5] The complexes formed by activated adsorption are often spoken of as compounds because of the nature of the linkages involved even though these "compounds" are not stoichiometrical in composition. Later we shall refer to Michaelis' computation of the extent of combination between an enzyme and its substrate.

Enzymes as Catalysts. The enzymes may be defined as *thermolabile catalysts of a definite organic nature, elaborated by living tissue but capable of action outside these tissues.* By this definition auxins and hormones are excluded, if they are at all catalysts, because they carry on their physiological roles only in living tissues; the vitamins probably are not catalysts although several function as components of enzyme systems. The phrase "of a definite organic nature" does anticipate what appears to be a generalization but is added here to exclude strictly surface phenomena.

Many enzymatic processes, such as the making of wine and cheese and the tanning of leather, were familiar arts since the dawn of history. Words like putrefaction and fermentation, in their Latin forms, were in general use, although nothing was known of the chemical changes taking place. In 1811 Kirchhoff observed that starch could be transformed into soluble sugars by boiling with dilute sulfuric acid and that no acid was consumed in the process. In 1833 Payen and Persoz discovered that an extract from germinating seeds would produce the same transformation on starch. Gay-Lussac had shown that blood possessed a component which could decompose hydrogen peroxide in a manner similar to certain metals.

At the time of Pasteur it was believed that, for certain processes in nature, the intact cell or tissue was necessary, *e.g.*, yeast in the alcoholic fermentation of sugars; such enzymes were called "organized ferments" in contrast to others, like the diastase of the salivary gland, which were designated "unorganized ferments," since no organism (or tissue) was necessary for their actions. In 1878 Kühne suggested the term *enzyme* for the latter group. In a famous experiment Büchner[6] carefully ground yeast cells and prepared an aqueous extract which was free from every trace of living cells or their fragments. He was able to demonstrate that this solution could ferment glucose to alcohol, although at a slower rate than did the intact cells. His experiment overthrew the vitalistic concept which had definitely related the actions of the organized ferments

[5] G. M. Schwab, *Catalysis from the Standpoint of Chemical Kinetics*, translated by H. S. Taylor and R. Spence, D. Van Nostrand Co., New York, 1937.

[6] E. Büchner, *Ber.*, **30**, 117, 1110 (1897).

to the life process of the organisms, and from that time the word enzyme has carried its present more inclusive meaning.

The systematic nomenclature of the enzymes involves the suffix *ase* added to the stem name of the substrate. Thus, amylase is the generic term for the enzymes hydrolyzing starch (amylose), to which may be added the source of the enzyme, *i.e.*, malt amylase. There are certain exceptions to the rule; certain enzymes were known before the system was adopted and we still retain names like pepsin, trypsin, and rennin. Certain enzymes are elaborated as the *zymogen* or *proenzyme*, which is inactive until it has undergone some slight change. Thus, the pepsinogen from the pyloric end of the stomach is rapidly converted to active pepsin by the hydrochloric acid from the fundus region. Particularly in systems which catalyze oxidation-reduction reactions, more than one component constitutes the system. The enzyme proper is called the *apoenzyme* and the smaller, dialyzable component is the *coenzyme*; the complete system is designated the *holoenzyme*. As an example, carboxylase is a protein which of itself has no action but will decarboxylate pyruvic acid when thiamine pyrophosphate is added.

Occurrence of Enzymes. All forms of life produce enzymes. Certain of these are secreted by tissues and are known as *extra-cellular* enzymes; others appear to be associated with the protoplasm and to be non-diffusible through the cell membrane, so are designated *intra-cellular* enzymes. As examples, wood-rotting fungi secrete a cellulase which hydrolyzes cellulose to soluble sugars, in which form the organism can utilize the carbohydrate; on the other hand, yeast can take up sucrose which is acted upon by enzymes within the cell. In general enzymes may be found wherever an organism produces some change upon its nutrients. Thus one observes that certain glands in higher organisms secrete digestive enzymes into the alimentary tract and that the ripening seed lays down enzymes which can be mobilized to digest the insoluble storage materials when germination takes place. One curious exception occurs in the seed of the jack bean, which is notably rich in urease although urea does not appear to be involved in the nitrogen cycle of this plant.

Preparation of Enzymes. An enzyme is not always employed in the pure form. The preparations which have been used for study range in composition from tissues containing relatively low concentrations of the active principle to those of the crystalline product. Often an aqueous extract is made from defatted tissue by employing saline, acid, alkaline, or glycerol solutions. Such a preparation is suitable for immediate use but generally is inactivated upon standing at room temperature; stable preparations may be made by precipitating the desired fraction with

alcohol or with a soluble salt like ammonium sulfate. This is essentially the degree of concentration found in many commercial preparations. The enzymes appear to be *protein* in composition; consequently the methods of protein isolation have been most useful in the preparation of enzyme concentrates. Repeated fractional precipitation, alternated with dialysis and with selective adsorption, have usually resulted in a highly active preparation. In certain cases such technics have yielded a crystalline product constant in composition and in activity.

Composition of Enzymes. The first enzyme to be obtained in the crystalline state was *urease* (Fig. 119), obtained by Sumner [7] as a globulin with a molecular weight of 483,000. Crystalline pepsin was first prepared by Northrop,[8] trypsin by Northrop and Kunitz,[9] and chymotrypsin by the same workers.[10] Like urease, these enzymes and the corresponding zymogens, which they also isolated, are *simple proteins* as is rennin, first crystallized by Hankinson.[11]

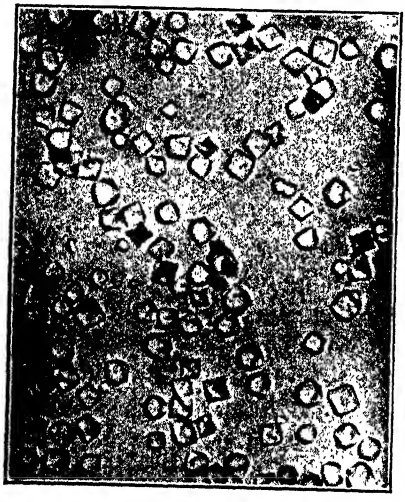


FIG. 119. Crystalline urease. (Courtesy of Prof. J. B. Sumner.)

However, certain enzymes are *conjugated* proteins; the first of this type to be prepared in crystalline form was the "old yellow enzyme" of Warburg, isolated by Theorell,[12] and shown to be a chromoprotein in which the apoenzyme is conjugated with riboflavin pyrophosphate. Crystalline tyrosinase [13] and ascorbic acid oxidase are copper proteins, whereas carbonic anhydrase [14] is a zinc protein; in no case is the linkage of the metal known. Crystalline catalase [15] has been obtained from beef liver and erythrocytes; both preparations are iron porphyrin proteins. These

[7] J. B. Sumner, *J. Biol. Chem.*, **69**, 435 (1926).
[8] J. H. Northrop, *J. Gen. Physiol.*, **13**, 739 (1930).
[9] J. H. Northrop and M. Kunitz, *J. Gen. Physiol.*, **16**, 267, 295 (1932).
[10] M. Kunitz and J. H. Northrop, *J. Gen. Physiol.*, **18**, 433 (1935).
[11] C. L. Hankinson, *J. Dairy Sci.*, **26**, 53 (1943).
[12] H. Theorell, *Biochem. Z.*, **272**, 155 (1934).
[13] H. R. Dalton and J. M. Nelson, *J. Am. Chem. Soc.*, **61**, 2946 (1939).
[14] D. A. Scott and A. M. Fisher, *J. Biol. Chem.*, **144**, 371 (1942).
[15] J. B. Sumner and A. L. Dounce, *J. Biol. Chem.*, **121**, 417 (1937); also M. Laskowski and J. B. Sumner, *Science*, **94**, 615 (1941).

are examples of some thirty enzymes which have been isolated in crystalline form.

Many enzymes have been prepared essentially free of concomitant materials as judged by constant composition and activity but have not yet been isolated in crystal form. All are protein in nature; the occasional statements to the contrary have generally been shown to be due to the failure of certain tests to detect minute quantities of proteins. In the instance of holoenzymes, the coenzymes are bound in varying degrees to the apoenzymes, and the resulting products are in reality conjugated proteins.

In 1922 Willstätter [16] proposed his *"Träger" theory,* which was accepted by Oppenheimer,[17] Waldschmidt-Leitz,[18] and most of the European workers. This viewpoint holds that enzymes contain a smaller reactive group which possesses a particular affinity for definite groupings in the substrate and this accounts for the specificity of enzyme behavior. The reactive group, or enzyme proper, is considered attached to a colloidal carrier, and the enzyme action is determined in part by the affinity of the active group for the substrate, and in part by the colloidality of the entire aggregate. When the colloidal properties of the aggregate are destroyed, the activity of the enzyme disappears. Willstätter in his earlier statements indicated that a single colloidal carrier did not appear to be essential but that any "suitable" colloidal carrier could act as a protective colloid for the active group. This viewpoint has been more or less modified so that certain of the proponents of the theory now suggest that the specificity of enzyme action is determined not alone by the nature of the prosthetic group but likewise by a colloidal carrier which shows specific affinities for the substrate.

Fodor [19] has emphasized even more strongly the importance of colloidality in the determination of enzyme action. He believes that enzymes are essentially commonly known substances such as proteins and carbohydrates in a peculiar colloidal state, the activities being due solely to the energies characteristic of colloidal systems. It is difficult to see, however, how enzyme specificity can be accounted for on the basis of Fodor's view.

The American workers have, in general, maintained that the *entire protein molecule* is required for enzymatic action and that the arrange-

[16] R. Willstätter, *Ber.*, **55**, 3601 (1922): *cf.* also R. Willstätter, J. Graser, and R Kuhn, *Z. physiol. Chem.*, **123**, 1 (1922) (note particularly pp. 45–60).

[17] C. Oppenheimer, *Die Fermente und ihre Wirkungen*, Vol. I, 5th ed., Georg Thieme, Leipzig, 1925.

[18] E. Waldschmidt-Leitz, *Enzyme Actions and Properties*, translated and extended by R. P. Walton, John Wiley & Sons, New York, 1929.

[19] A. Fodor, *Das Fermentproblem*, Theodor Steinkopff, Dresden and Leipzig, 1922.

ment of the groupings in the protein molecule is not only responsible for the enzyme activity but also determines the specificity. Thus Sumner, Kirk, and Howell [20] demonstrated that the loss of activity of urease paralleled the extent of splitting of this enzyme by papain or by pepsin. Northrop [21] has adduced similar evidence to show that the same is true for his crystalline preparations of pepsin and trypsin. It will be indicated later that many of the desmolases consist of an apoenzyme which requires a coenzyme (prosthetic group) for enzymatic function. One would ask whether the apoenzyme, the coenzyme, or the holoenzyme is responsible for specificity; it appears to be universally true that the apoenzyme is the seat of this specificity. As an example, Warburg and Christian [22] separated the apoenzyme of their "old yellow enzyme" from its coenzyme which is a flavin mononucleotide. The apoenzyme was then caused to combine with a flavin-adenylic acid dinucleotide, and the result was a synthetic yellow enzyme with catalytic properties identical with those of the "old yellow enzyme." From the considerations indicated it would appear that *the protein, rather than an indifferent "Träger," is the enzyme proper.* The fact that certain units or groupings in the protein molecule must be free for enzyme action does not weaken the argument because any alteration in groupings results in a derived or denatured protein; since it is no longer the original protein one need not be surprised to find that it has lost its enzymatic function.

It will be recalled that all proteins possess colloidal properties. Their molecular weights range from 20,000, through values like 37,000 for pepsin and 248,000 for catalase, up to the value of 483,000 reported for urease. These have been calculated from osmotic pressures, rates of diffusion, and from sedimentation velocities. The enzymes diffuse very slowly and do not pass through ordinary dialyzers; on the other hand, the coenzymes are much smaller in size and are often removed upon dialysis. Like all proteins, the enzymes are amphoteric and exhibit characteristic isoelectric points. Their ionic properties make them sensitive to changes in electrolyte concentrations, particularly *p*H changes, and their protein natures render enzymes strongly thermolabile.

Classification of Enzymes. Two large classes of enzymes are differentiated on the basis of the reactions which they catalyze. The *hydrolases*, or hydrolyzing enzymes, produce a cleavage in the substrate by the addition of water; as such they come into action whenever lipids carbohydrates, and proteins are to be hydrolyzed *in vivo*. The second

[20] J. B. Sumner, J. S. Kirk, and S. F. Howell, *J. Biol. Chem.*, **98**, 543 (1932).

[21] J. H. Northrop, *Crystalline Enzymes*, Columbia University Press, New York, 1939.

[22] O. Warburg and W. Christian, *Biochem. Z.*, **298**, 368 (1938).

group of enzymes consists of the *desmolases*, which catalyze the rupture of linkages that are not hydrolyzable, such as the bonds in the carbon chain of glucose or in the oxygen molecule. The enzymes involved in oxidation-reduction reactions and in fermentations are of this type. The effects produced by coagulating enzymes, like thrombin and rennin, are not clearly understood, but these enzymes are usually classed with the hydrolases.

A classification of the more commonly recognized hydrolyzing enzymes is shown below. It is not possible to present in this scheme the various limitations as to substrate and extent of action pertaining to each enzyme, or to consider whether the indicated enzyme consists of more than one component. For example, it is known that pancreatic "trypsin" contains trypsin, several chymotrypsins. heterotrypsin, and carboxypeptidase.

CLASSIFICATION OF THE HYDROLASES

Enzyme	*Typical Substrate*	*End Products*
A. Esterases		
1. Glyceridases		
a. Lipase	Fats and oils	Glycerol + fatty acids
b. Lecithinase A	Lecithin	Lysolecithin + unsaturated fatty acid
c. Lecithinase B	Lysolecithin	Choline-glycerophosphate + fatty acid
2. Specific esterases		
a. Choline esterase	Acetylcholine	Choline + acetic acid
b. Cholesterase	Cholesterol esters	Cholesterol + fatty acid
c. Chlorophyllase	Chlorophyll-a	Phytol + chlorophyllide a
d. Pectase	Pectin	Pectic acid + methyl alcohol
e. Tannase	Tannin	Glucose + digallic acid
f. Sulfatase	Phenyl sulfate	Phenol + H_2SO_4
g. Butyrase	Ethyl butyrate	Butyric acid + ethyl alcohol
3. Phosphatases		
a. Phosphomonoesterases (several)	Glycerol-phosphate	Glycerol + H_3PO_4
b. Nucleotidase	Nucleotide	Nucleoside + H_3PO_4
c. Phosphodiesterase	Diphenyl phosphate	Phenyl phosphate + phenol
d. Pyrophosphatase	$Na_2H_2P_2O_7$	NaH_2PO_4
e. Metaphosphatase	$NaPO_3$	NaH_2PO_4
f. Phosphoamidase	H_2PO_3NHR	$RNH_2 + H_3PO_4$
g. Phytase	Phytin	Inositol + H_3PO_4
4. Phosphorylases		
a. α-Glucosan phosphorylase	Glycogen (starch) + H_3PO_4	Glucose-1-phosphate

CLASSIFICATION OF THE HYDROLASES (*Continued*)

Enzyme	*Typical Substrate*	*End Products*
B. Carbohydrases		
1. β-h-Fructosidases (β-fructofuranosidases)		
a. Invertase	Sucrose	Glucose + fructose
2. α-Glucosidases		
a. Maltase	Maltose	Glucose
3. β-Glucosidases		
a. Emulsin	β-Glucosides	Glucose + non-sugar residue
b. Cellobiase	Cellobiose	Glucose
c. Myrosinase	S-β-glucosides	Glucose + sulfur derivative
4. α-Galactosidases		
a. Melibiase	Raffinose	Sucrose + galactose
5. β-Galactosidase		
a. Lactase	Lactose	Glucose + galactose
6. Polysaccharidases		
a. Amylases		
1. α-Amylase	Soluble starch	Dextrin + maltose
2. β-Amylase	Soluble starch	Maltose
b. Cellulase	Cellulose	Cellobiose
c. Hemicellulase (cytase)	Hemicellulose	Simple sugars
d. Inulase	Inulin	Fructose
e. Lichenase	Lichenin	Cellobiose
f. Hyaluronidase	Hyaluronic acid	Glucosamine + uronic acid
g. Protopectinase	Protopectin	Pectin
h. Pectinase	Pectic acid	Galactose + uronic acid
C. Enzymes hydrolyzing nitrogen compounds		
1. Proteases		
a. Rennin	Casein	Paracasein
b. Thrombin	Fibrinogen	Fibrin
c. Pepsin	Proteins	Peptones
d. Trypsin	Proteins	Polypeptides + amino acids
e. Cathepsin	Proteins	Polypeptides + amino acids
f. Ficin	Proteins	Polypeptides + amino acids
g. Papain	Proteins	Polypeptides + amino acids
h. Bromelin	Proteins	Polypeptides + amino acids
i. Lysozyme	Glycoproteins	Sugars + proteins(?)
j. Renin	α_2-Globulin	Hypertensin + residue

CLASSIFICATION OF THE HYDROLASES (*Continued*)

Enzyme	*Typical Substrate*	*End Products*
k. Keratinase	Keratins	Polypeptides
l. Protaminase	Protamines	Arginine + residue
m. Aminopeptidase	Polypeptides with free amino group	Amino acids
n. Carboxypeptidase	Polypeptides with free COOH group	Amino acids
o. Prolinase	Prolylpeptides	Proline + residue
p. Prolidase	N-substituted proline peptides	Proline + residue
q. Dipeptidase	Dipeptides	Amino acids
r. Leucylpeptidase	Leucyl peptides	Amino acids
2. Amidases		
a. Urease	Urea	Ammonia + carbon dioxide
b. Asparaginase	Asparagine	Aspartic acid + ammonia
c. Arginase	Arginine	Urea + ornithine
d. Histozyme	Hippuric acid	Glycine + benzoic acid
3. Deaminases		
a. Guanosine deaminase	Guanosine	Xanthosine + ammonia
b. Adenosine deaminase	Adenosine	Hypoxanthosine + ammonia
c. Guanase	Guanine	Xanthine + ammonia
d. Adenase	Adenine	Hypoxanthine + ammonia
4. Nucleases		
a. Depolymerases		
1. Ribonuclease	Ribopolynucleic acid	Nucleotides(?)
2. Thymonuclease	Polynucleic acids (all)	Nucleotides(?)
b. Nucleosidase	Nucleoside	Base + sugar

It is more difficult to present in schematic form the actions of the desmolases. This term has been applied to all non-hydrolytic enzymes, but it is also used in a narrower meaning to include only those enzymes concerned in the rupture of a carbon chain. As will be seen, we have chosen to employ the term in both senses. Because certain of these enzymes catalyze the approach to an equilibrium equally well from either side, it is pointless to speak of their substrates as distinct from the corresponding end products. In these cases the equilibrium reactions are indicated. For example, the last step in muscle glycolysis utilizes an enzyme called lactic dehydrogenase which catalyzes both phases of the reaction pyruvic acid $\rightleftharpoons$ lactic acid, although *in vivo* the important reaction is that proceeding to the right.

THE DESMOLYZING ENZYMES

Enzyme	*Substrate*	*End Products*
A. Catalase	H_2O_2	$H_2O + O_2$
B. Peroxidases		
1. Plant and verdoperoxidase	H_2O_2 + reduced compounds	Oxidized compound
2. Cytochrome-*c* peroxidase	H_2O_2 + reduced cytochrome *c*	Oxidized cytochrome *c*
C. Carbonic anhydrase	H_2CO_3	$CO_2 + H_2O$
D. Oxidases which use O_2 of the air		
1. Enzymes containing iron		
a. Cytochrome oxidase	Reduced cytochrome *c*	Oxidized cytochrome *c*
2. Enzymes containing copper		
a. Tyrosinase (phenol oxidase)	Tyrosine	Hallochrome
b. Laccase	Phenols	*o*- and *p*-Quinones
c. Ascorbic acid oxidase	Ascorbic acid	Dehydroascorbic acid
E. Dehydrogenases		
1. Enzymes which transfer electrons to cytochrome		
a. Succinic dehydrogenase	Succinic acid ⇌	fumaric acid
b. Lactic dehydrogenase (yeast)	Lactic acid ⇌	pyruvic acid
c. α-Glycerophosphate dehydrogenase	α-Glycerophosphate ⇌	phosphoglyceric aldehyde
2. Enzymes which pass hydrogen to the yellow enzymes		
a. Xanthine oxidase (Schardinger)	Xanthine	Uric acid
b. Diaphorase	Reduced coenzyme I	Oxidized form
	Reduced coenzyme II	Oxidized form
c. "Old yellow enzyme"	D-Amino acids	α-Keto acid
d. D-Amino acid oxidase	L-Amino acids	α-Keto acid
e. L-Amino acid oxidase	Glucose	Gluconic acid
f. Glucose oxidase (molds)		
3. Enzymes which pass hydrogen to coenzymes I and II		
a. Alcohol dehydrogenase	Ethanol ⇌	acetaldehyde
b. Glucose dehydrogenase	Glucose	Gluconic acid
c. Lactic dehydrogenase	Lactic acid ⇌	pyruvic acid
d. Robison ester dehydrogenase	Hexose-6-phosphate	Phosphohexonic acid
e. α-Glycerophosphate dehydrogenase	Glycerol-3-phosphate	3-Phosphoglyceric aldehyde
f. Diphosphoglyceric aldehyde dehydrogenase	1,3-Diphosphoglyceric aldehyde	1,3-Diphosphoglyceric acid

THE DESMOLYZING ENZYMES (*Continued*)

Enzyme	*Substrate*	*End Products*
4. Enzymes which involve pyridoxine (or its derivatives)		
a. Transaminases	Pyruvic acid + glutamic acid ⇌	alanine + ketoglutaric acid
b. Amino acid decarboxylases (bacterial)	Amino acid	CO_2 + amine
F. Miscellaneous oxidases or dehydrogenases		
1. Monoamine (tyramine) oxidase	Amine	NH_3 + aldehyde
2. Diamine oxidase (histaminase)	Amine	NH_3 + aldehyde
3. Uricase	Uric acid	Allantoin + CO_2
4. Luciferase	Luciferin	Oxyluciferin
5. Lipoxidase	Unsaturated fatty acid	Peroxide derivative
G. Desmolases (in the narrower sense)		
1. Aldolase (zymohexase)	Glucose-1,6-diphosphate	Dihydroxyacetone-phosphate + phosphoglyceric aldehyde
2. Amino acid decarboxylases	Amino acids	Amine + CO_2
3. Carboxylase	Pyruvic acid	Acetaldehyde + CO_2
4. Carboligase	Two aldehydes	Benzoin-type product
H. Hydrases and mutases		
1. Glyoxylase	Methylglyoxal	Lactic acid
2. Aldehyde mutase	Aldehyde	Acid + alcohol
3. Fumarase	Fumaric acid	L-Malic acid
4. Enolase	2-Phosphoglyceric acid ⇌	phosphopyruvic acid
5. Phosphotriose isomerase	Dihydroxyacetone phosphate ⇌	phosphoglyceric aldehyde
I. Transphosphorylases		
1. Hexokinase	Glucose + ATP	Adenylic acid + glucose-6-phosphate
2. Phosphoglucomutase	Glucose-1-phosphate ⇌	glucose-6-phosphate
3. Phosphohexose isomerase	Glucose-6-phosphate ⇌	fructose-6-phosphate
4. Phosphocarboxyl transphosphorylase	Acetyl phosphate + adenylic acid	Acetic acid + ATP

Coenzymes. Earlier it has been indicated that many enzymes consist of an apoenzyme which requires some coenzyme in order to function. This is particularly true of the dehydrogenases for which the coenzyme serves as a mediator whose function is to accept hydrogen or an electron and, in turn, to pass this on until the oxygen of the air can act as the ultimate acceptor. We may define a coenzyme as a *natural, thermostable, and often readily removable fraction of an enzyme system which is necessary to the action of that system.* The term is usually restricted to organic compounds, although such metals as zinc and manganese are rather easily removed from carbonic anhydrase and from arginase, respectively, with resulting inactivation. The coenzymes as prosthetic groupings of conjugated proteins are separated from their appropriate apoenzymes with varying degrees of ease; for example, the iron porphyrin group of catalase is very firmly attached to its protein, whereas the riboflavin phosphate nucleus of the flavin enzymes is readily removed by dialysis against dilute acids at low temperatures.[23] Like many other systems, the yellow enzyme can be regenerated by adding the coenzyme to the specific protein, as shown in Fig. 120.

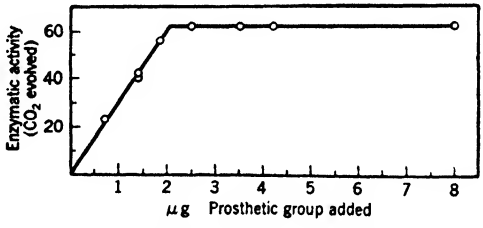


FIG. 120. The effect of adding varying amounts of the prosthetic group to the specific protein in resynthesizing the yellow enzyme. (After Theorell.)

The first coenzyme preparation was that of Harden and Young,[24] who observed that a porcelain filter impregnated with gelatin would retain the colloidal apoenzyme from the crystalloid coenzyme. Neither fraction was active but the ability to ferment sugar was restored upon mixing the two systems.

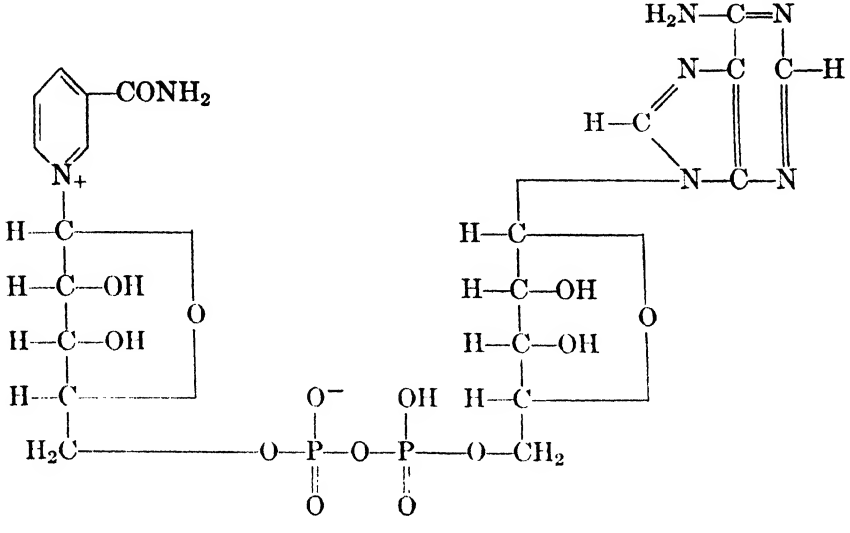
The following are the more commonly recognized coenzymes. Several have been indicated in the classification of the dehydrogenases as the first acceptor of hydrogen or an electron. The roles of others are not understood, but it is known that their presence is required.

Coenzyme I (often abbreviated Co I and also called diphosphopyridine

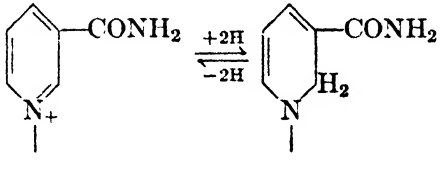
[23] H. Theorell, *Biochem. Z.*, **278**, 263 (1935).

[24] A. Harden and W. J. Young, *Proc. Roy. Soc. London*, **B78**, 369 (1906).

nucelotide, DPN) is the dinucleotide of nicotinamide and of adenine, as shown by Euler and Schlenk.[25]



It is widely found in nature and is involved in over thirty different enzymatic reactions. When it acts as a hydrogen acceptor, dihydrocoenzyme I is formed by the addition to one of the double bonds of the pyridine nucleus:



The reaction is reversible, for the reduced form can be converted to the oxidized state by diaphorase and other flavoproteins which remove the hydrogen and are in turn reduced.

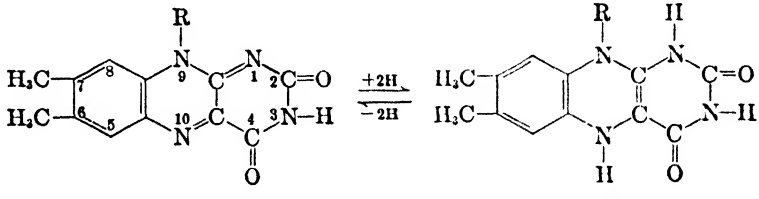
Coenzyme II (triphosphopyridine nucleotide, abbreviated Co II or TPN) is required for a limited number of reactions. It has been shown by Warburg, Christian, and Griese [26] to be similar in structure to coenzyme I except that it contains three phosphate units. Most workers consider that the third phosphate is tied into the chain of the other two units, but from its behavior it seems more probable that the third phosphate is linked to the adenylic acid nucleus as suggested by Schlenk.[27]

[25] H. von Euler and F. Schlenk, *Z. physiol. Chem.*, **246**, 64 (1937).

[26] O. Warburg, W. Christian, and A. Griese, *Biochem. Z.*, **282**, 157 (1935).

[27] F. Schlenk, *Symposium on Respiratory Enzymes*, Univ. Wisconsin Press, Madison, 1942.

Another important group of coenzymes are derivatives of riboflavin phosphate; when combined with the appropriate apoenzyme these are referred to as *flavoproteins* or the *yellow enzymes*. At least twelve enzymes of this type are well characterized and differ in function because of the specificity of the apoenzyme. However, the prosthetic groupings are of two types. The "old yellow enzyme" of Warburg and Christian and cytochrome-*c* reductase are known to contain riboflavinphosphate as the coenzyme; this structure is considered a nucleotide since it contains a base (isoalloxazine), a sugar, and phosphoric acid. The coenzyme for the other two flavin systems is a dinucleotide since it contains isoalloxazine, ribose, phosphate, phosphate, ribose, and adenine, in the order mentioned. Like coenzymes I and II the yellow enzymes are hydrogen mediators and are capable of existing in both the oxidized and the reduced forms as shown (R indicates the residual portion of the molecule):

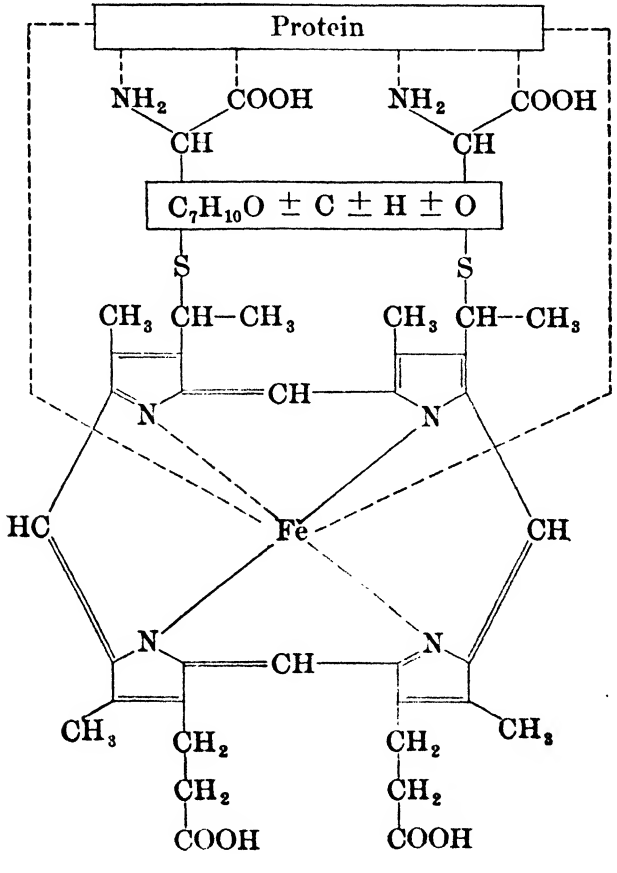


Oxidized state Reduced state

It will be noted that the addition of two hydrogen atoms to the oxidized form results in a 1,4-addition across the conjugated system represented by positions 1 and 10 of the isoalloxazine nucleus.

Cytochrome oxidase, cytochrome peroxidase, catalase, peroxidase, and the cytochromes have in common some *iron porphyrin* as the prosthetic group. In all cases, when reduced with sodium hydrosulfite essentially the same absorption bands are observed as with hemoglobin the differences observed are shifts due to the varying protein apoenzymes. In the natural state, however, the iron of hemoglobin and of oxyhemoglobin is in the ferrous state, whereas in catalase and peroxidase the most active preparations contain *ferric* iron. Because cytochrome oxidase and the cytochromes are involved in oxidation-reduction reactions, the iron should be capable of existing in both states. When oxidized cytochrome accepts an electron from a hydrogen atom the iron is reduced to the ferrous state, and a hydronium ion is liberated into the surrounding medium; this reaction is reversible under suitable conditions. Eventually neutral oxygen of the air becomes the acceptor for the electrons and the hydronium ions.

Cytochrome *c* is much more stable than are cytochromes *a* and *b*. Theorell [28] has proposed the following structure for cytochrome *c*:



Cytochrome *c*

He notes that the iron porphyrin is attached to the protein portion through a thio-ether bridge and postulates that the iron is further co-ordinated with two imidazole rings in the protein. These are in keeping with the observations that the prosthetic group in cytochrome *c* is firmly bound to the protein and that the product is not auto-oxidizable under ordinary biological conditions.

Thiamine (see Chapter 36) is, *in vivo*, readily esterified to the pyrophosphate, *cocarboxylase* (*i.e.*, the coenzyme for carboxylase) and thus is involved in the conversion of pyruvic acid to carbon dioxide and acetal-

[28] H. Theorell, *Biochem. Z.*, **298**, 242 (1938); *J. Am. Chem. Soc.*, **63**, 1820 (1941).

dehyde. This does not involve a change in the state of oxidation, as the term is generally understood; however, cocarboxylase may act as a hydrogen acceptor in certain dismutation reactions of pyruvic acid.

Glutathione is necessary to the conversion of methyl glyoxal to lactic acid; this appears to involve the sulfhydryl group of reduced glutathione or related compounds as shown by Jowett and Quastel [29] and by Behrens.[30] *Pyridoxal monophosphate* has been shown to be the coenzyme for transaminase and for amino acid decarboxylase [31] as well as for the specific glutamic acid decarboxylase of higher plants.[32]

Transphosphorylation is an important phase of many reactions, as has been indicated in the discussion of intermediary carbohydrate metabolism. In both muscle glycolysis and alcoholic fermentation the sugar is converted to a phosphate ester, and the subsequent phosphate derivatives take part in all steps of the reaction up to the point where pyruvic acid results. *Free adenylic acid and its related phosphate derivatives* are the most common acceptors and donators of the phosphate ion. It will be recalled that muscle contains free adenylic acid which consists of 5-phosphoribose linked to position-9 of adenine. Adenylic acid may also be considered its nucleoside, adenosine, to which is tied one phosphate nucleus. Hence adenylic acid is often called adenosine monophosphate to distinguish it from its derivatives of higher phosphate content, which are adenosine diphosphate (or pyrophosphate) and adenosine triphosphate; the two latter are often abbreviated to ADP and ATP, respectively. The roles played by these compounds have been illustrated in the chapter dealing with carbohydrate metabolism.

Kinases. Certain enzymes are elaborated as the zymogen which is converted to the active enzyme as needed. In the case of pepsinogen this is readily accomplished by the acidity of the stomach. Trypsinogen can be converted to trypsin by ammonium or magnesium salts or by active trypsin; however, a natural kinase, *enterokinase*, secreted by the intestinal mucosa is the principal agent to bring about the change *in vivo*. A kinase may be defined as *a naturally occurring organic substance which can convert a zymogen to the active form*. Apparently the kinase has no function after the conversion has taken place. There has been considerable controversy as to the nature of kinase action, but Kunitz,[33] from a study of the kinase elaborated by a *Penicillium*, has adduced evidence that the change is enzymatic in nature.

[29] M. Jowett and J. H. Quastel, *Biochem. J.*, **27**, 486 (1933).
[30] O. K. Behrens, *J. Biol. Chem.*, **141**, 503 (1941).
[31] I. C. Gunsalus, W. W. Umbreit, W. D. Bellamy, and C. E. Frost, *J. Biol. Chem.* **161**, 743 (1945).
[32] O. Schales, V. Mims, and S. G. Schales, *Arch. Biochem.*, **10**, 455 (1946).
[33] M. Kunitz, *J. Gen. Physiol.*, **21**, 601 (1938).

Activators. The rates of reaction for many enzymes may be increased from a few to several hundred per cent by the presence of certain compounds or ions which are variously known as activators, stimulators, or protectors. It would seem desirable to reserve the term coenzyme for those substances essential to any action and to employ the term activator for those which *merely enhance the rate of the reaction;* however, this distinction is not always possible because many enzyme systems have not been purified to the stage where the absolute need can be established. In many instances one particular compound appears to be the natural activator although related compounds may be employed. As an example, the chloride ion, *in vivo*, activates pancreatic amylase although bromides, iodides, nitrates, chlorates, thiocyanates, and fluorides are less effective, in the order mentioned.[34] To list other examples, pancreatic lipase is activated by bile salts, albumin, and by calcium oleate, whereas papain and bromelin are activated by hydrocyanic acid and by hydrogen sulfide.

Several hypotheses may be listed to account for these phenomena, some of which may apply in one case and not in others. (1) The activator may act as an *emulsifier* and thus increase the surface of contact. This appears to explain why the activators for lipase listed above are effective, since the enzyme is in the aqueous phase distinct from the substrate oil phase. (2) Salt solutions may serve better to *peptize the enzyme* with the result that additional functional groups can be brought into action. (3) The activator may *make the substrate more reactive to the enzyme.* Linderström-Lang and Duspiva [35] demonstrated that the clothes moth secretes both an enzyme and a natural activator which together will produce hydrolysis of wool and that the natural activator can be replaced by thioglycolic acid. Without the activator the enzyme cannot attack wool but will hydrolyze casein. It is also known that keratins are not attacked by the ordinary proteinases until the substrate has been altered by certain compounds, most of which contain sulfhydryl groups. (4) It has been pointed out that many of the activators are reagents which *stabilize or protect the sulfhydryl and other groups of the enzyme* necessary to its action, as pointed out by Singer [36] and by Hellerman.[37] Since the compounds which favor the stability of the sulfhydryl group are essentially reducing agents, one might expect the effect to be one of poising the redox potential. Fishgold [38] has shown, however, that the rate of

[34] H. C. Sherman, M. L. Caldwell, and M. Adams, *J. Am. Chem. Soc.*, **50**, 2538 (1928).

[35] K. Linderstrøm-Lang and F. Duspiva, *Z. physiol. Chem.*, **237**, 131 (1935).

[36] T. P. Singer, *Brewers Digest*, **20**, 85, 104, 109 (1945).

[37] L. Hellerman, *Cold Spring Harbor Symposia Quant. Biol.*, **7**, 165 (1939).

[38] H. Fishgold, *Biochem. J.*, **28**, 406 (1934).

urease action is, over a large range, independent of the redox potential. Scott and Sandstrom [39] found that several mercaptans had a greater stimulating effect upon papain than did hydrogen sulfide or cysteine, but again such effects were not related to any poised potential. Although no clear-cut effect of the redox potential has been demonstrated in these cases, it is recognized that in the instance of the hydrogen mediators, which act with the dehydrogenases, the level of reduction potential is an important factor. (5) Many of the effects of activators can be traced to a protector effect; in other words, the activator *inhibits the action of some enzyme poisons.* Sumner and Hand [40] demonstrated that proteins, amino acids, gums, hydrogen sulfide, and hydrocyanic acid offset the harmful action of traces of heavy metals in distilled water upon urease, which, like so many enzymes, requires for its activity free sulfhydryl groups in its molecule. In further support of this hypothesis it is often observed that the activators produce less effect as the enzyme preparation is progressively purified. Then, too, enzymes are protein in nature, and hence tend to be denatured upon shaking or prolonged standing; this effect can at times be avoided by the addition of a lyophilic colloid like gum arabic.

Inhibitors. By contrast to the activators, certain substances are inhibitors or poisoners of enzymes. Singer [36] has recently compiled an extensive list of enzyme inhibitors and discussed their possible roles. Owing to the chemical nature of enzymes, those compounds which either *denature* or *precipitate proteins* would inhibit enzymatic function. Among these are the soluble salts of mercury, silver, gold, and copper which are toxic even in very dilute solutions; also organic compounds like alcohol, ether, and formaldehyde have a denaturing effect. The halogens, hydrogen peroxide, and other mild oxidizing agents are injurious owing to their action on groupings like the sulfhydryl radical. In general, the common salts are injurious in higher concentrations because they tend to precipitate proteins. Most antiseptics and the preservatives for biological materials are protein denaturants and should be avoided in enzyme preparations; washed toluene is less harmful and is often used to avoid the growth of microorganisms.

A second group of enzyme poisons *act upon the prosthetic grouping* in enzymes. For example, hydrogen sulfide and hydrocyanic acid inactivate the oxidases because they immobilize the iron or copper in these enzymes; similarly, carbon monoxide and sodium azide act upon the iron-containing enzymes. At times it may be desirable to prevent enzymatic action with specific poisons; thus the oxalate or citrate ion is generally

[39] E. M. Scott and W. M. Sandstrom, *Arch. Biochem.*, **1**, 103 (1942).

[40] J. B. Sumner and D. B. Hand, *J. Biol. Chem.*, **76**, 149 (1928).

employed to prevent the clotting of blood by removing the calcium ion which is essential to the process.

Other poisons act by *competition with the substrate for the enzyme.* These are known as *competitive inhibitors;* the literature on this subject has increased rapidly during the past few years.[41] The antagonism between the sulfonamides and *p*-aminobenzoic acid has features similar to the competitive inhibition. Quastel [42] found that succinic dehydrogenase can be inactivated by malonic acid (and related compounds) because this acid is sufficiently similar in structure to succinic acid to combine with the enzyme. Because the enzyme cannot dehydrogenate malonic acid, the latter remains attached to the "active center" of the enzyme and thus prevents it from acting upon the normal substrate.

Antienzymes. Several naturally occurring organic compounds inhibit enzyme action. It has been known for a long time that the parasitic roundworm, *Ascaris,* could persist in the intestines of hogs because it contained an antipepsin and an antitrypsin.[43] The inhibitor is not destroyed by boiling in neutral or weakly acid solutions but is hydrolyzed by alkalies; it is dialyzable and has the properties of a polypeptide. Robbins and Lamson [44] observed that the *Ascaris,* however, could not withstand the action of ficin. Kunitz and Northrop [45] have isolated from the pancreas a crystalline antitrypsin with a molecular weight of 40,000 which inhibits tryptic action in the *p*H range below 7.0. An antitrypsin [46] has also been prepared from the white of egg. From solutions of pepsinogen Herriott [47] has isolated an antipepsin, which is a polypeptide with a molecular weight of 5,000 and is highly specific against pepsin. Clinically, the mucins from gastric mucosa and from okra have been used to prevent the pepsin and the acid of the stomach from attacking an ulcer area. These substances resist hydrolysis and in that limited sense are antipepsins.

This phenomenon is not restricted to the proteinases, for amylase inhibitors have been found in wheat,[48] which appears to be protein in nature; an antiamylase of a different composition has also been reported from certain sorghums.[49] The early literature contains statements to the effect that normal blood contains inhibitors for several enzymes. Many

[41] M. G. Sevag, *Advances in Enzymol.,* **6,** 33 (1946).
[42] J. H. Quastel, *Biochem. J.,* **20,** 166 (1926).
[43] B. K. Harned and T. P. Nash, *J. Biol. Chem.,* **97,** 443 (1932).
[44] B. H. Robbins and P. D. Lamson, *J. Biol. Chem.,* **106,** 725 (1934).
[45] M. Kunitz and J. H. Northrop, *J. Gen. Physiol.,* **19,** 991 (1936).
[46] A. K. Balls and T. L. Swenson, *J. Biol. Chem.,* **106,** 409 (1934).
[47] R. M. Herriott, *J. Gen. Physiol.,* **24,** 325 (1941).
[48] E. Kneen and R. M. Sandstedt, *Arch. Biochem.,* **9,** 235 (1946).
[49] B. S. Miller and E. Kneen, *Arch. Biochem.,* **15,** 251 (1947).

of these observations were made before the importance of *p*H controls was recognized; others appear to be due to a partial adsorption upon blood proteins of the enzyme tested for.

The term antienzyme has a second meaning, that of immunochemistry. Since the enzymes are proteins they would be expected to show antigenic properties and hence give rise to *antibodies*, in this case antienzymes. The early data were examined by Wells,[50] who doubted that the findings were typical of antienzymes but thought they might largely be accounted for as the reactions due to concomitant materials. Later trials with crystalline enzymes have demonstrated that the enzymes do give rise to antienzymes. This has been shown for luciferase by Harvey and Dietrick,[51] for pepsin by Northrop,[52] and by Kirk and Sumner [53] with urease. The injection of urease into rabbits stimulated the production of sufficient antiurease to enable the animal to withstand 1,000 times the lethal dose of urease. From an *in vitro* precipitate of the urease and antiurease complex, the workers were able to obtain the antiurease fraction in a purified form; it appears to be a glycoprotein.

Enzyme-Substrate Complex. Several observations point to the conclusion that enzymes combine with their substrates as the first step in their action. It is known that many enzymes may be adsorbed from solution upon a solid substrate; thus stearin will selectively remove lipase from pancreatic juice, whereas casein will adsorb most of the proteolytic enzymes. If invertase is adsorbed upon colloidal iron oxide it cannot readily be removed by washing with water or with solutions of fructose or lactose, but it is eluted by solutions of sucrose or raffinose. In other words, solutions of the latter sugars are substrates for invertase, and they form complexes which are water-soluble and at the same time form stronger linkages with the enzyme than does the iron oxide.

Stern [54] has given definite proof of an intermediate compound between catalase and monoethyl hydrogen peroxide. The system was studied spectrophotometrically.

(1) The purified catalase enzyme yields a brown solution with characteristic absorption bands at 650, 646–620, and 610 mμ.

(2) When ethyl hydrogen peroxide is added to the catalase solution, there is a short period when the brown color characteristic of the enzyme disappears and a pale green color develops. During this stage the ab-

[50] H. G. Wells, *The Chemical Aspects of Immunity*, Chemical Catalog Co., New York, 1925.

[51] E. N. Harvey and J. E. Dietrick, *J. Immunol.*, **18**, 65 (1930).

[52] J. H. Northrop, *J. Gen. Physiol.*, **13**, 739 (1929).

[53] J. S. Kirk and J. B. Sumner, *J. Biol. Chem.*, **94**, 21 (1931).

[54] K. G. Stern, *J. Biol. Chem.*, **114**, 473 (1936).

sorption bands of catalase disappear and no new characteristic absorption bands develop.

(3) After the initial period there develops a red color of the catalase-substrate compound. This intermediate compound has very pronounced absorption bands at 576–564 and 540–529 mμ.

(4) As the red catalase-substrate compound breaks down, there coexists in the solution the absorption spectrum of both the free enzyme and the enzyme-substrate compound.

(5) When all the enzyme-substrate compound has been broken down, the brownish color characteristic of the enzyme and the characteristic enzyme absorption bands alone remain.

It seems probable that similar enzyme-substrate combinations take place as intermediary compounds in most enzyme reactions. The difficulty has been to demonstrate the presence of specific compounds inasmuch as they may exist for only short periods of time.

Michaelis and Menten [55] have developed an expression for the extent of complex formation. Their value, K_m, is the equilibrium constant

$$K_m = \frac{(E)(S)}{(ES)} \tag{168}$$

for the reaction

$$E + S \leftrightharpoons ES$$

where E, S, and ES are the concentrations of the enzyme, the substrate, and the complex, respectively. From theoretical considerations they write the equation

$$v = V \frac{S}{S + K_m} \tag{169}$$

in which v is the initial short-time velocity of the reaction at the concentration of substrate S; and V is the maximum velocity attainable at some high concentration of S where all of the enzyme is combined with the substrate. By plotting v as a function of S a sigmoid curve is obtained. It will be seen from equation (169) that, when v is one-half of V, K_m is numerically equal to S.

At times it is difficult to attain the maximum reaction rate by varying the concentrations of the substrate. Lineweaver and Burk [56] suggest that equation (169) be expressed in the form of its reciprocals:

$$\frac{1}{v} = \frac{1}{V} + \frac{K_m}{VS} \tag{170}$$

[55] L. Michaelis and M. L. Menten, *Biochem. Z.*, **49**, 333 (1913).

[56] H. Lineweaver and D. Burk, *J. Am. Chem. Soc.*, **56**, 658 (1934).

If $1/v$ is plotted as a function of $1/S$ a straight line will result. The intercept on the Y-axis represents $1/V$, and the slope of the line gives the value for K_m/V, from which K_m may be calculated.

Hanes [57] prefers to multiply equation (170) by S to give:

$$\frac{S}{v} = \frac{S}{V} + \frac{K_m}{V} \tag{171}$$

He plots S/v as a function of S and obtains a straight line. In this case $1/V$ is the slope of the line, and K_m/V becomes the intercept on the Y-axis.

By considering the *velocity* constants involved in the equations

$$\text{Enzyme} + \text{substrate} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \text{complex} \xrightarrow{k_3} \text{enzyme} + \text{products}$$

Briggs and Haldane [58] conclude that

$$K_m = \frac{k_2 + k_3}{k_1} \tag{172}$$

Chance [59] has confirmed this by a careful study of peroxidase, in which he was able to measure k_1, k_3, and k_2/k_1. An independent approach to the mechanism of urease action was made by Van Slyke and Cullen,[60] who concluded that a complex of the enzyme and urea does exist.

Haldane [61] has compiled the values of K_m obtained for a large number of enzymes. In the desmolases the values range from 10^{-3} to 10^{-8} M concentrations and for the hydrolases the figures range from 0.0005 M to solutions of 5 per cent concentration; the latter was obtained for trypsin on casein.

The Michaelis constant appears to have validity as a property serving to characterize a particular enzyme. In their original paper Michaelis and Menten found that K_m was independent of the concentrations of invertase employed. For the same enzyme Josephson [62] showed that K_m was constant over a pH range of 4 to 8. Kuhn [63] demonstrated that the degree of purification had no effect on K_m; however, the invertases from different yeasts did yield various values. The latter was not un-

[57] C. S. Hanes, *Biochem. J.*, **26**, 1406 (1932).
[58] G. E. Briggs and J. B. S. Haldane, *Biochem. J.*, **19**, 338 (1925).
[59] B. Chance, *J. Biol. Chem.*, **151**, 553 (1943).
[60] D. D. Van Slyke and G. E. Cullen, *J. Biol. Chem.*, **19**, 141 (1914).
[61] J. B. S. Haldane, *Enzymes*, Longmans, Green and Co., New York, 1930.
[62] K. Josephson, *Z. physiol. Chem.*, **134**, 50 (1924).
[63] R. Kuhn, *Z. physiol. Chem.*, **125**, 28 (1923).

expected since the catalysts would be expected to vary as do other properties within different yeasts.

The expression of K_m is arbitrarily fixed so that the concentration of the complex appears in the denominator. Some writers prefer to consider the reciprocal value, which is a measure of the affinity of an enzyme for its substrate. This concept has been extended to calculate the affinity of a coenzyme for its apoenzyme.

Factors Influencing the Rate of Enzymatic Reactions. The external factors such as the temperature and the acidity of the medium, as well as the time duration of a run, have considerable effect upon the amount of change produced upon substrates by enzymes. To determine the extent of these effects it is customary to conduct runs in which all factors are kept constant except the one being studied. This has the limitation that factors like time, temperature, and *p*H show optimum effects which shift with changes in the other conditions. It is, therefore, necessary either to determine the effects of several factors simultaneously to arrive at optimum conditions or to recognize that the factor determined gives optimum activity under the conditions arbitrarily selected.

Effect of pH. Because the enzymes are lyophilic colloids, their activities are greatly influenced by the *p*H of the medium. Constant acidity of the medium is usually insured by the use of buffers. Figure 121 illustrates for three different diastases the great difference in activity due to changes in *p*H; [64] the curves are typical of enzymes in that some show a narrow region and others a broader zone of optimum activity. Pepsin acts best at a *p*H in the neighborhood of 1.5 to 2; in general, the enzymes from fungi and plants are most active in the range from *p*H 4 to 6.5, and those from the higher animals function best from *p*H 6.5 to 8. Bone phosphatase, arginase, and many of the oxidases have optima above *p*H 8.

The optimum *p*H for an enzyme preparation may be shifted as concomitant materials are removed. Willstätter and others [65] found that several purification procedures on canine gastric lipase caused the optimum to move from *p*H 5.5–6.3 to 7.1–7.9.

Especially in the proteases, the optimum will vary with properties of the substrate. Papain [66] and other similar enzymes appear to act upon the substrate at its isoelectric point; thus the optimum action of papain and fibrin is at *p*H 7.2, which is the isoelectric point of that protein, whereas the action on gelatin and albumin peptone (isoelectric point of

[64] H. C. Sherman, A. W. Thomas, and M. E. Baldwin, *J. Am. Chem. Soc.*, **41**, 231 (1919).

[65] R. Willstätter, F. Haurowitz, and F. Memmen, *Z. physiol. Chem.*, **140**, 202 (1924).

[66] R. Willstätter, W. Grassmann, and O. Ambros, *Z. physiol. Chem.*, **151**, 307 (1926).

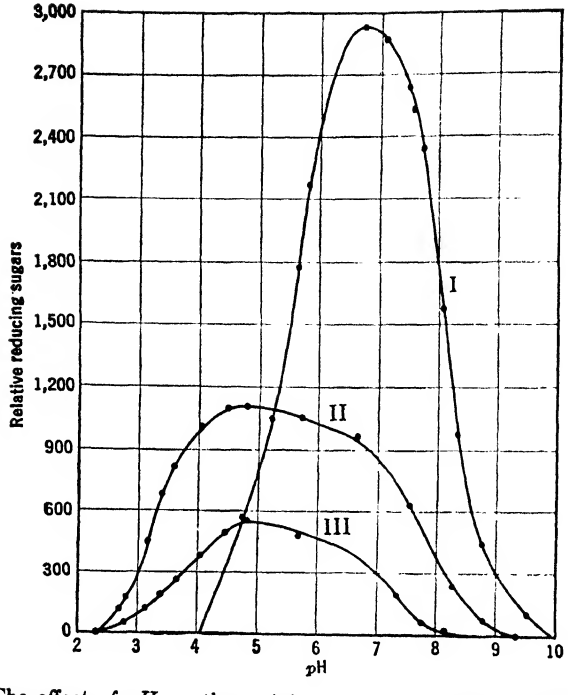


FIG. 121. The effect of *p*H on the activity of (I) pancreatic amylase, (II) malt amylase, and (III) taka-amylase.

these proteins is at 4.8) is most rapid at a *p*H near 5.0. Northrop [67] found that the optima for pepsin and for trypsin are also influenced by the protein substrate as shown in Table 73. The isoelectric point of pepsin is 2.7 and that of trypsin is at 7.0; hence the cation of pepsin acts

TABLE 73. OPTIMUM *p*H FOR PEPTIC AND TRYPTIC ACTION IN RELATION TO THE ISOELECTRIC POINT OF THE SUBSTRATE

Protein	Isoelectric Point of Protein	*p*H for Optimum Action by:	
		Pepsin	Trypsin
Casein	4.6	1.8	8.2
Hemoglobin	6.8	2.2	8.7
Gelatin	4.8	2.0	7.4

[67] J. H. Northrop, *J. Gen. Physiol.*, **5**, 263 (1922).

upon the cation of the substrate, and the anion of trypsin acts upon the protein anion. This rules out the formation of a coacervate as an explanation of this phenomenon.

The composition of the buffer mixture may alter the optimum *p*H, as found for urea by Howell and Sumner [68] and shown in Table 74. The variations may be due, in part, to differences in ionic strengths since that factor was not kept constant. Sherman and his co-workers [69] demonstrated that the optimum activity of malt amylase was at *p*H 4.5–4.8 in an acetate buffer of 0.00005 *M* and that it increased to the range 5.0–5.4 in a 0.1 *M* acetate buffer. Similarly, the optimum shifted from *p*H 4.5 to 4.9 with phosphate buffers ranging from 0.00005 to 0.1 *M*.

TABLE 74. OPTIMUM *p*H FOR UREASE ACTION OF VARIED SUBSTRATE CONCENTRATIONS AND IN VARIOUS BUFFERS

Buffer	*0.1% Urea as Substrate*	*2.5% Urea as Substrate*
Acetate	6.7	6.4
Citrate	6.7	6.5
Phosphate	7.6	6.9

Table 74 also indicates that, in this case, the concentration of the substrate influences the optimum *p*H. This may mean that urease functions better in the more acid medium which is at the same time more destructive to the enzyme unless the latter is "protected" in the form of its complex with urea, which condition the greater concentration of urea would insure. The effect of varying substrate concentration to change the optimum *p*H has been observed in only a few cases.

It is not generally known what phase of the reaction is influenced by a change in the acidity of the medium. The effect may be on the formation of an unstable complex of the enzyme and its substrate, on the rate of decomposition of this complex, on the stability of the enzyme, or on the substrate by production of some slight change such as a tautomeric shift. It can be shown for some enzymes that the downward slope of the activity-*p*H curve on the alkaline side of the optimum hydrogen-ion concentration parallels the denaturation curve of the enzyme in the same range when no substrate is present. In some enzymes the optimum activity is not at the point of greatest stability. Trypsin is most stable in the range from *p*H 5 to 6, whereas its optimum activity is always in a more alkaline range; similarly, pepsin is most stable at *p*H 4, but its

[68] S. F. Howell and J. B. Sumner, *J. Biol. Chem.*, **104**, 619 (1934).

[69] H. C. Sherman, M. L. Caldwell, and H. H. Boynton, *J. Am. Chem. Soc.*, **52**, 1669 (1930).

activity is most pronounced at a lower level. By contrast, yeast invertase is most stable at its isoelectric point, where it also displays the greatest catalytic effect.

Effect of Temperature. The generalization of van't Hoff states that a rise of ten degrees will double the speed of a chemical reaction. This is often expressed as Q_{10}, which is the ratio of the rates of reaction at two temperatures ten degrees apart. Certain values obtained upon enzyme systems are illustrated in Table 75. For most enzymatic re-

TABLE 75. THE Q_{10} FOR CERTAIN ENZYME SYSTEMS

Enzyme	*Temperature*	Q_{10}
Steapsin	0–10	1.50
	10–20	1.34
	20–30	1.26
Pepsin	0–10	2.60
	10–20	2.00
	30–40	1.60
	40–50	1.40
Trypsin	21–31	5.30
	31–39	5.30
Invertase	25–35	1.61
Emulsin	20–30	2.62
Pancreatic amylase	30–40	2.00

actions Q_{10} is around 2 at lower temperatures but gradually drops off until the ratio is 1 and finally a fractional value at higher temperatures. Figure 122 illustrates a typical experiment. Between 0° and 10°, Q_{10} is 2, whereas from 30° to 40° it is 1.12. If the value of Q_{10} had remained constant over the region we would expect 16 times as much product produced at 40° as at 0°, whereas experimentally the result was 6 times that at 0°. The favorable effect of temperature on the rate of the reaction is counterbalanced by the denaturation of the enzyme which also increases with temperature until the two factors are equal just below 50°. In this particular run 62.5 per cent of the activity shown at 0° appears to be lost at 40°. It will be noted that the right-hand branch of the curve is the steeper, and we find that the temperature coefficient of denaturation is greater than the effect on the reaction velocity. In this discussion we have considered Q_{10} as the expression of the temperature coefficient; this is generally done by chemists although it is not strictly correct, as pointed out by Bull.[70]

[70] H. B. Bull, *Physical Biochemistry*, John Wiley & Sons, New York, 1943.

Arrhenius proposed the equation

$$\log \frac{k_2}{k_1} = \frac{E}{2.3R}\left(\frac{T_2 - T_1}{T_2 T_1}\right) \tag{173}$$

where k_1 and k_2 are the velocity constants at temperatures T_1 and T_2; E is the energy of activation and represents the energy per mole necessary to convert the molecules in the normal state to a higher energy level at which they become reactive. The value for E may be calculated by substituting the appropriate values in equation (173) or by plotting log k against $1/T$; the slope of the line obtained represents $4.58E$.

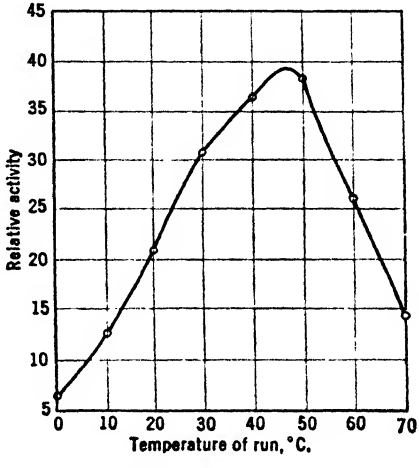


FIG. 122. Effect of temperature on the action of malt amylase on starch.

Sizer [71] has given a critical survey of the effect of temperature on enzyme kinetics. His paper lists the values of E for those enzymes which obey the Arrhenius law over a wide temperature range; we discover that most of the values ie between 8,000 and 16,000 calories per mole. One can calculate that E would be approximately 12,000 calories between 23° and 33° for an enzyme with a Q_{10} equal to 2. Another table presents the activation energies for the denaturation of enzymes by heat as determined out of contact with the substrate. These energies are from 40,000 to 100,000 calories per mole. From an inspection of these values one would predict a very slow denaturation of enzymes at room temperature. Such is not the case, as Fig. 122 shows. Sizer has calculated the entropy changes involved and suggests that "the very high entropy increases in the order of magnitude of 100 cal./degree, which occur on enzyme activation, counterbalance the large activation energies . . ." at the lower temperatures.

Several generalizations have grown out of the studies cited above. Many workers have employed temperatures of 37° for their studies, probably because this is body temperature. For short time runs most enzymes of animal origin show an optimum above 40°, whereas most enzymes from plants and microorganisms are most active near 50°.

[71] I. W. Sizer, *Advances in Enzymol.*, **3**, 35 (1943).

We shall later see that other variants influence the optimum temperature. It is generally found that flash heating at 100° will inactivate most enzymes.

By contrast, it was earlier assumed that temperatures below 0° would inhibit enzymatic function. Recently Sizer [72] and Lineweaver [73] have shown that enzymes do function below 0°, but that there is a sharp

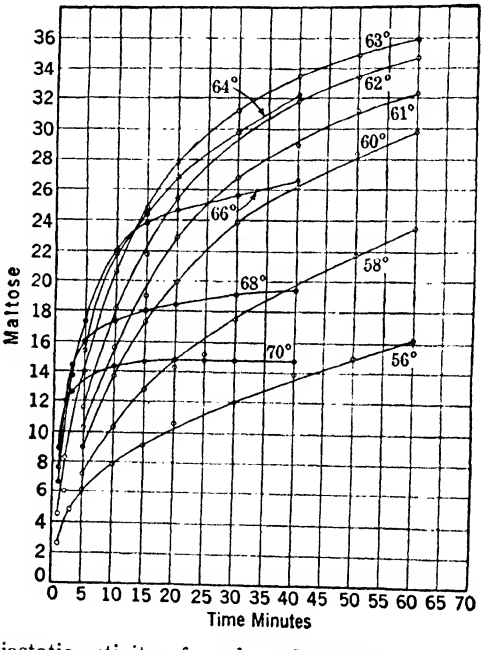


FIG. 123. The diastatic activity of a wheat flour with time and temperature as variables.

break in the rates of reaction and that the activation energies are several times greater at these lowered temperatures, as measured upon lipase, pepsin, and trypsin.

Time Factor. The amount of change produced by an arbitrary amount of enzyme should be proportional to the time duration of the run, provided that no change has taken place in the potency of the enzyme and that the concentration of the substrate is not a limiting factor. We have seen, however, that the acidity and the temperature of the environment play a role in determining the rate of inactivation. Consequently it can be stated that time, temperature, and *p*H are interrelated in de-

[72] I. W. Sizer and E. S. Josephson, *Food Research*, **7**, 201 (1942).
[73] H. Lineweaver, *J. Am. Chem. Soc.*, **61**, 409 (1939).

termining the optimum conditions. In general, the shorter time runs of a few minutes may show an optimum at higher temperatures and at acidities which could not be employed for experiments designed to run for hours or even days. Figures 123 [74] and 124 [75] illustrate the interrelationship of these factors upon diastases.

Concentration of Enzyme. In general for short time runs the amount of change produced is proportional to the concentration of enzyme employed, which is one of the criteria of catalysis. Schütz early proposed a rule that the amount of change produced was proportional to the square root of the concentration of enzyme employed. This was first announced after a study of the action of crude pepsin on egg albumin. Northrop [76] re-examined the problem using his crystalline pepsin and trypsin; he found that the observation could be ascribed to the inhibiting effects of the products of the reactions upon the enzymes.

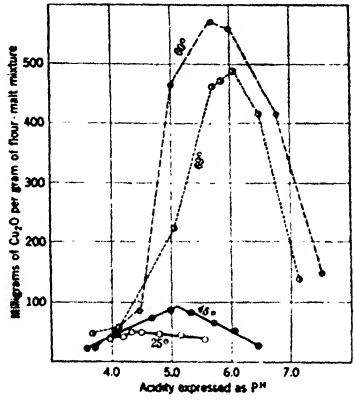


FIG. 124. The influence of temperature on the optimum hydrogen-ion concentration for the saccharogenic activity of malt.

Kinetics of Enzyme Action. The law of mass action states that the speed of any chemical reaction is proportional to the active concentration of the reacting materials. This is true whether or not a catalyst is employed, although the latter will influence the magnitude of the velocity constant, which is the proportionality factor that makes it possible to equate the time of a run with the concentrations of the reactants or the end products. We shall see that enzymatic reactions where the course can be described mathematically are of zero order or the first order.

For many enzymatic reactions it is found that the amount of product x is a linear function of the time interval t; in other words x is constant during successive units of time interval. This is described as a reaction of *zero order*. Since x is proportional to t we may write:

$$x = kt \tag{174}$$

This law holds whenever the concentration of the substrate a is high compared with the amount of enzyme employed. Since a is decreasing

[74] E. Munz and C. H Bailey, *Cereal Chem.*, **14**, 85 (1937).
[75] A. G. Olsen and M. S. Fine, *Cereal Chem.*, **1**, 215 (1924).
[76] J. H. Northrop, *J. Gen. Physiol.*, **2**, 471 (1920); **4**, 245 (1922).

in value the speed must be determined by some component which remains constant in concentration. From a consideration of the Michaelis constant, K_m, it follows that a high value for a will insure that the enzyme-substrate complex is maintained at constant value. It is assumed that the speed of a zero-order reaction is determined by this constant concentration of the complex and not by the value of a. Zero-order reactions appear to be limited to catalyzed reactions; Hinshelwood and Burk [77] found that the thermal decomposition of ammonia on tungsten was of this type.

Reactions of the *first order* are characterized by a gradual slowing-up of the rate of formation of x because its rate of production is a function of the concentration of the unreacted substrate, $a - x$, which is diminishing as x increases. The velocity constant k for a first-order reaction is given by the expression

$$k = \frac{1}{t} \log \frac{a}{a - x} \tag{175}$$

For purposes of calculation the equation is more useful in the form

$$k = \frac{1}{t_2 - t_1} \log \frac{C_1}{C_2} \tag{176}$$

where C_1 and C_2 represent the substrate concentrations at t_1 and t_2. Most hydrolases are permitted to function in dilute solutions so that the concentration of the second reactant, water, is essentially constant. An aqueous solution of sucrose which contains 34.2 grams per liter would consist of 0.1 mole of sucrose plus over 55.5 moles of water; even when the sucrose is completely hydrolyzed the water concentration would be in excess of 55.4 M.

A *second-order* reaction is one in which the product of the varying concentrations of two reactants determines the rate of reaction. The writer has never found an enzymatic reaction which was clearly of this order. Nelson and Schubert [78] showed that, for the hydrolysis with invertase of sucrose above 10 per cent, water was a limiting factor; however, no "constant" for a second order can be obtained from a calculation of their data.

It is possible for the reactions of the same system to be either of zero order or the first order, depending on the relative concentrations of the enzyme and the substrate. Kuhn [79] demonstrated this in studies in

[77] C. N. Hinshelwood and R. E. Burk, *J. Chem. Soc.*, **127**, 1105 (1925).
[78] J. M. Nelson and M. P. Schubert, *J. Am. Chem. Soc.*, **50**, 2188 (1928).
[79] R. Kuhn, *Z. physiol. Chem.*, **125**, 28 (1923).

which the enzyme concentration was fixed in a series of runs employing varying concentrations of sucrose; a typical experiment is shown in Fig. 125. In concentrations above 0.05 *M* the graph is a straight line, and the reaction is of zero order. In concentrations below 0.05 *M*

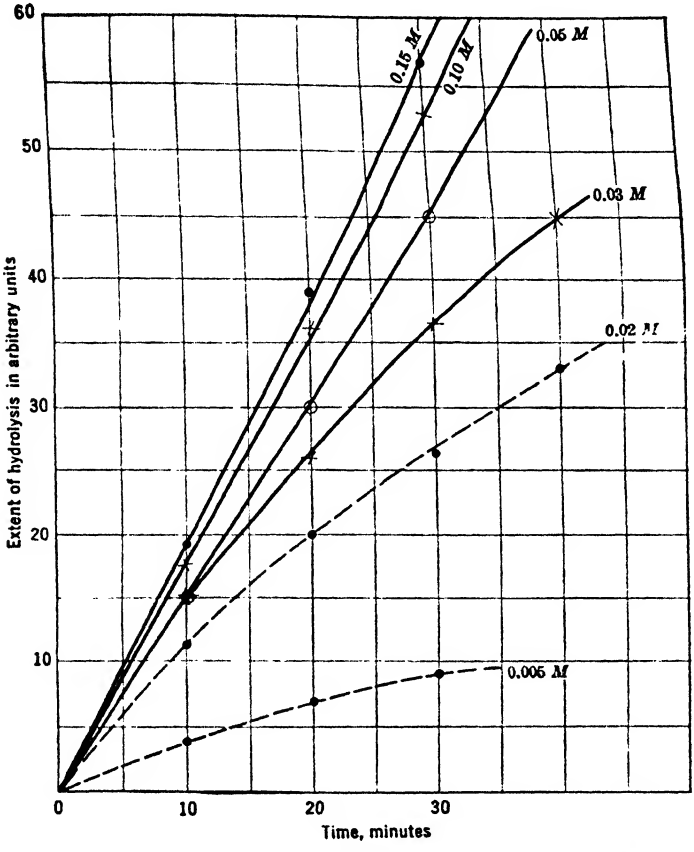


FIG. 125. The course of hydrolysis by invertase on solutions of sucrose of different concentrations.

the reaction is of the first order. Van Slyke,[80] in a survey article on the kinetics of hydrolytic reactions, points out that both phases of this phenomenon may be observed in one solution provided the initial concentration of the substrate is high enough to insure a zero-order reaction in the first phase; after a time the substrate can no longer saturate the enzyme, and the conditions for a first-order reaction obtain. He writes

[80] D. D. Van Slyke, *Advances in Enzymol.*, **2**, 33 (1942).

an overall equation to describe the entire course of the reaction:

$$t = \frac{1}{K_c} \log \frac{a}{a - x} + \frac{x}{K_d} \tag{177}$$

K_c is the velocity constant for the formation of the enzyme-substrate complex, and K_d the velocity constant for the decomposition of this complex into the resulting products with the liberation of the enzyme. The other terms have the meanings indicated for equations (174) to (176). It is obvious that when x is so small that x/K_d is negligible, the equation simplifies to the first-order expression.

Whenever a plot of the amount of change against time gives a curve which looks like a "die-away" curve characteristic of a first-order reaction, it is desirable to calculate the values for k at various time intervals by equation (176) to make certain that the first-order law prevails. A simpler method is to plot the log of $(a - x)$ as a function of time; this should yield a straight line. If a deviation from the law is detected, several factors may be operating. First, the enzyme may have been denatured during the course of the run, and the result will be a progressive lowering of the value of k. A second factor should be considered, *i.e.*, that the reaction may proceed in more than one step with the result that only the "overall" effect is being measured. Unless one of the steps goes much faster than the others no definite order of reaction is expected. This situation might obtain in the hydrolysis of a fat made up essentially of mixed glycerides, or in the hydrolysis of a protein, the linkages of which are probably not equally readily hydrolyzed. A third complicating factor might arise from a slight reversal of the reaction, since many enzymes have been shown to catalyze both the hydrolysis and the synthesis of biological compounds. To obviate these difficulties it is customary to conduct kinetic studies under conditions favorable to the life of the enzyme and during the initial stages of the reaction.

Enzymatic Synthesis. The accepted theory of catalysis postulates that the catalyst should not change the equilibrium point of a reaction. Enzymes are no exception to this generalization; consequently they would, under appropriate conditions, be capable either of decomposing or of synthesizing compounds. Most of the actions of the desmolases give rise to equilibrium reactions with measurable quantities of both phases, but in the use of the hydrolases the reactions take place in such dilute solutions that the equilibrium is far to the side of hydrolysis.

Several products have been synthesized by enzymes; these include ethyl butyrate, methyl oleate, triolein, amygdalin, and other β-glucosides. Whenever a protein hydrolysate has been concentrated to re-

move most of the water and treated with one or several proteinases, new peptide linkages are produced, but the products do not have many of the properties of proteins; the products are generally known as *plasteins*. When working with pepsin, Wasteneys and Borsook [81] found that compounds like benzene, xylene, and benzaldehyde could replace that enzyme.

Sym [82] was able to isolate 0.31 mole of butyl benzoate from a solution of 0.45 mole of benzoic acid and 0.5 mole of butyl alcohol by the action of steapsin activated with sodium glycocholate for 24 hours. Small quantities of water inhibited the reaction to a greater degree than would be demanded by the equilibrium. Sym got his best results when the system was suspended in carbon tetrachloride.

Other enzyme syntheses have led to the predicted end-point. Mack and Villars [83] added urease to solutions which were 10 *M* with respect to ammonium carbonate and carbamate. With 1.0 and 0.1 per cent of urease, equilibrium was attained after 10 and 98 hours, respectively; they calculated that 600 days would be required in the run containing no enzyme.

Recently both animal and plant phosphorylases have been used to synthesize polysaccharides. By the use of muscle phosphorylase Cori and Cori [84] were able to synthesize glycogen from glucose-1-phosphate, provided a trace of glycogen was added. A product resembling the amylose fraction of starch has been synthesized by phosphorylases from various sources by Hanes [85] and by Hassid and McCready.[86] By contrast, Haworth, Peat, and Bourne [87] obtained a product more similar to the amylopectin of starch by the action of an enzyme system found in the potato. It will be noted that the phosphorylases can catalyze both the formation and the hydrolysis of starch and glycogen, whereas the diastases have to date been shown only to catalyze the hydrolysis of these compounds.

Specific Enzyme Reactions. We have already noted that the literature dealing with specific enzymatic reactions is exceedingly voluminous, and from time to time in the preceding chapters we have discussed certain interrelationships between enzymes and substrates. In view of these facts, it seems wise to close the discussion of enzymes at this point,

[81] H. Wasteneys and H. Borsook, *Physiol. Revs.*, **10**, 110 (1930).
[82] E. A. Sym, *Enzymologia*, **1**, 156 (1936).
[83] E. Mack and D. S. Villars, *J. Am. Chem. Soc.*, **45**, 501 (1923).
[84] G. T. Cori and C. F. Cori, *J. Biol. Chem.*, **131**, 397 (1939).
[85] C. S. Hanes, *Proc. Roy. Soc. London*, **B129**, 174 (1940).
[86] W. Z. Hassid and R. M. McCready, *J. Am. Chem. Soc.*, **63**, 2171 (1941).
[87] W. N. Haworth, S. Peat, and E. J. Bourne, *Nature*, **154**, 236 (1944).

having emphasized only some of the broader principles involved. To extend the discussion to make this text adequate in the field of enzymes would be beyond the purposes for which the volume is intended. For detailed discussion of the various enzyme systems the reader is referred to the earlier extensive monographs of Oppenheimer [88] and of Euler.[89] For shorter yet excellent presentations, the more recent texts by Tauber [90] and by Sumner and Somers [91] are suggested; the latter authors have given a most adequate picture of the current trends in enzyme chemistry.

[88] C. Oppenheimer, *Die Fermente und ihre Wirkungen*, 5th ed., 4 vols., Georg Thieme, Leipzig, 1925–1929; Supplement Bd., W. Junk, The Hague, 1936–1937.

[89] H. von Euler, *Chemie der Enzyme*, 3rd ed., 2 vols., J. F. Bergmann, Munich 1925–1934.

[90] H. Tauber, *The Chemistry and Technology of Enzymes*, John Wiley & Sons, New York, 1949.

[91] J. B. Sumner and G. F. Somers, *Chemistry and Methods of Enzymes*, 2nd ed., Academic Press, New York, 1947.

AUTHOR INDEX

SUBJECT INDEX

SEVEN
DAY
BOOK