PHYSICAL BIOCHEMISTRY

BY

Henry B. Bull, Ph.D.

ASSOCIATE PROFESSOR OF PHYSIOLOGICAL CHEMISTRY

MEDICAL SCHOOL OF NORTHWESTERN UNIVERSITY

NEW YORK

JOHN WILEY & SONS, INC.

Chapman & Hall, Limited

London

THIRD PRINTING, APRIL, 1947

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

G. N. Lewis once defined physical chemistry as a science which includes everything which is interesting and excludes everything which is uninteresting. Evidently, the subject of physical chemistry is a very broad one, and invariably a considerable selection of topics to be presented must be made. I have tried to choose material which aids in the interpretation of living systems. It has long been my feeling that instruction in physical chemistry for pre-medical and for graduate students in the biological sciences is not as well adjusted to the needs of such students as it could be. This book is an attempt to reorient to some extent this type of instruction.

The volume is an outgrowth of a series of about thirty-six lectures which I have given in alternate years to graduate students in biochemistry, physiology, bacteriology, and neurology. The lectures were also attended by medical students. The time is short; the material could be treated more conveniently in about forty-eight lectures. As it is essentially an outline course and must be supplemented by a generous amount of outside reading, literature citations are included for this purpose. To the more serious student of physical chemistry I recommend *Text Book of Physical Chemistry*, by S. Glasstone, published by D. Van Nostrand Company, New York City.

I make no claim for completeness in my presentation of physical biochemistry. Much work has been omitted, not because I did not think it sound or important, but because it did not seem to fit into the story I was telling.

I wish to thank Doctors D. R. Briggs, M. Dole, E. Guth, R. S. Livingston, H. Neurath, and K. Sollner for their critical comments on the sections of the manuscript they were asked to read. They saved me from many a pitfall, and I am duly grateful. They are not, however, to be blamed for any errors which remain. For these and for the method of presentation I alone am responsible.

This book was tested for clarity on John A. Cooper, who served as a highly intelligent guinea pig. My only concern is that he may have misled me because what was clear to him might be obscure to a less talented graduate student. Martin Gutmann also contributed suggestions for the betterment of the manuscript. I am grateful to both these students for their help.

Ross Aiken Gortner, to whom I am most indebted, never saw the manuscript or the book. It is to him, however, that I owe my interest and a large part of my knowledge of physical chemistry. His recent death was a great shock and loss to all of us who knew him. I am proud to be, in a manner of speaking, his scientific son.

This book is dedicated to Fredrica Jane and Fredrica Jean.

HENRY B. BULL

CHICAGO, ILLINOIS
February, 1943

CONTENTS

Chapter I

ATOMS AND MOLECULES

Atoms and molecules being the building blocks from which the chemist makes his science, it is essential that we have clearly in mind some of their more conspicuous characteristics.

For this short discussion of atomic structure we need consider only three elementary bodies, namely, electrons, protons, and neutrons. Electrons carry a unit negative electrical charge, they have a very small mass, and their radii are approximately 2.8×10^{-13} cm. Protons have a unit positive charge, and their masses are 1,836 times those of the electron. Neutrons have masses very nearly equal those of protons, but the net charge on the neutron is zero.

As the tendency in recent years has been for atoms to be treated as abstract mathematical concepts, the physical picture has become somewhat vague and fuzzy. Although a physical picture of the atom has its admitted inadequacies, the mathematical atom is far beyond the scope of this book; accordingly, we must content ourselves with a "horse-and-buggy" atom.

In its simplest terms an atom consists of a central nucleus about which rotate electrons. For example, a hydrogen atom has one electron rotating about a proton. This is the simplest of all the atoms. The next atom as we ascend the scale of complexity is that of helium. The helium atom has two protons and two neutrons in the nucleus about which rotate two electrons. Successive atoms are built up by adding planetary electrons and increasing the charge on the nucleus to match by the addition of protons; neutrons must also be added to have a stable atom. The number of protons in the nucleus of a neutral atom always equals the number of planetary electrons. Isotopes, which we shall discuss presently, result from the addition or subtraction of neutrons from the nucleus.

There is a great deal of open space in atoms. Thus, nuclear radii are all of the order of 10^{-12} to 10^{-13} cm. while the atomic radii are of the order of 10^{-8} cm. In fact, the sum of the volumes of the electrons and the nucleus is about one 10^{-13}th of the effective atomic volume.

As we have noted, the electrons in an atom arrange themselves in a series of groups or shells around the nucleus. The key to this arrange-

TABLE 1

PERIODIC TABLE (rare earths omitted)

Group	0	I	II	III	IV	V	VI	VII	VIII
Sub-group	A B	A B	A B	A B	A B	A B	A B	A B	
Period 1		1 H 1.0080							
2	2 He 4.003	3 Li 6.94	4 Be 9.02	5 B 10.82	6 C 12.01	7 N 14.008	8 O 16	9 F 19	
3	10 Ne 20.183	11 Na 22.997	12 Mg 24.32	13 Al 26.97	14 Si 28.06	15 P 30.98	16 S 32.06	17 Cl 35.46	
4	18 A 39.944	19 K 39.096 29 Cu 63.57	20 Ca 40.08 30 Zn 65.38	21 Sc 45.10 31 Ga 69.72	22 Ti 47.90 32 Ge 72.60	23 V 50.95 33 As 74.91	24 Cr 52.01 34 Se 78.96	25 Mn 54.93 35 Br 79.916	26 Fe 27 Co 28 Ni 55.85 58.94 58.69
5	36 Kr 83.7	37 Rb 85.48 47 Ag 107.88	38 Sr 87.63 48 Cd 112.41	39 Y 88.92 49 In 114.76	40 Zr 91.22 50 Sn 118.70	41 Cb 92.91 51 Sb 121.76	42 Mo 95.95 52 Te 127.61	43 53 I 126.92	44 Ru 45 Rh 46 Pd 101.7 102.91 106.7
6	54 Xe 131.3	55 Cs 132.91 79 Au 197.2	56 Ba 137.36 80 Hg 200.61	57–71 Rare Earths 81 Tl 204.39	72 Hf 178.6 82 Pb 207.21	73 Ta 180.88 83 Bi 209.00	74 W 183.91 84 Po 210	75 Re 186.31 85	76 Os 77 Ir 78 Pt 190.2 193.1 195.23
7	86 Rn 222	87	88 Ra 226.05	89 Ac 227	90 Th 232.12	91 Pa 231	92 U 238.07		

TABLE 2

DISTRIBUTION OF ELECTRONS IN SOME ATOMS

Shell		*K*	*L*		*M*			*N*				*O*
Electron subshell		1*s*	2*s*	2*P*	3*s*	3*P*	3*d*	4*s*	4*P*	4*d*	4*f*	5*s*
Element	Atomic number											
H	1	1										
He	2	2										
Li	3	2	1									
Be	4	2	2									
B	5	2	2	1								
C	6	2	2	2								
N	7	2	2	3								
O	8	2	2	4								
F	9	2	2	5								
Ne	10	2	2	6								
Na	11	2	2	6	1							
Mg	12	2	2	6	2							
Al	13	2	2	6	2	1						
Si	14	2	2	6	2	2						
P	15	2	2	6	2	3						
S	16	2	2	6	2	4						
Cl	17	2	2	6	2	5						
A	18	2	2	6	2	6						
K	19	2	2	6	2	6		1				
Ca	20	2	2	6	2	6		2				
Sc	21	2	2	6	2	6	1	2				
Ti	22	2	2	6	2	6	2	2				
V	23	2	2	6	2	6	3	2				
Cr	24	2	2	6	2	6	5	1				
Mn	25	2	2	6	2	6	5	2				
Fe	26	2	2	6	2	6	6	2				
Co	27	2	2	6	2	6	7	2				
Ni	28	2	2	6	2	6	8	2				
Cu	29	2	2	6	2	6	10	1				
Zn	30	2	2	6	2	6	10	2				
Ga	31	2	2	6	2	6	10	2	1			
Ge	32	2	2	6	2	6	10	2	2			
As	33	2	2	6	2	6	10	2	3			
Se	34	2	2	6	2	6	10	2	4			
Br	35	2	2	6	2	6	10	2	5			
Kr	36	2	2	6	2	6	10	2	6			
Rb	37	2	2	6	2	6	10	2	6			1
Sr	38	2	2	6	2	6	10	2	6			2
Y	39	2	2	6	2	6	10	2	6	1		2
Zr	40	2	2	6	2	6	10	2	6	2		2
Cb	41	2	2	6	2	6	10	2	6	4		1
Mo	42	2	2	6	2	6	10	2	6	5		1
...	43	2	2	6	2	6	10	2	6	6		1
Ru	44	2	2	6	2	6	10	2	6	7		1
Rh	45	2	2	6	2	6	10	2	6	8		1
Pd	46	2	2	6	2	6	10	2	6	10		

ment is obtained from a consideration of the nature of the electronic arrangement of the inert gases, i.e., the zero group in the periodic table (see Table 1). Evidently these rare-gas atoms possess very stable electronic structures; otherwise they would show more chemical activity. This stability is attributed to the possession of completed electronic shells. Since the number of electrons in an atom equals its atomic number, and since the atomic numbers of the zero-group elements are 2, 10, 18, 36, 54, and 86, these figures must represent the completion of successive electronic shells. The number of electrons in each shell should then be obtained by subtracting from the total in any inert-gas atom the number in the preceding inert gas. This gives 2, 8, 8, 18, 18, 32 for the number of electrons in successive complete shells. Starting with potassium, a new outer shell of electrons begins to form while the inner ones are still incomplete. Thus, in no atom does the outer shell have more than eight electrons. The outer shell of electrons is concerned in chemical bonding; we shall consider this important topic presently. Table 2 shows the distribution of electrons in the first 46 atoms.

From the electrons arise the emission spectra of the atoms. As is well known, if the temperature of an element is increased sufficiently, light will be emitted from the element. The wavelength of emitted light will be highly characteristic of the element; in fact, this is the basis for a very delicate test for many elements, i.e., spectroscopic analysis. This type of emission originates from the outer electrons which absorb energy and become excited. They give up the absorbed energy in form of light of a very definite wavelength. If elements are bombarded with high-speed electrons, the electrons in the innermost shells become excited and when they yield up their energy the wavelength of the emitted radiation is very short. Such emitted rays are X-rays. X-rays from the K shells are shorter than those from the L shell, and so on out. The shorter the X-ray, the harder it is said to be. Hard X-rays have a much greater penetrating power than the longer wavelength X-rays. In Chapter XVIII we shall have occasion to discuss the uses to which the biochemist can put X-rays.

Isotopes

The atomic number and not the atomic weight is the fundamental constant of an atom. The atomic weight of an element is simply the average of the atomic masses of all the naturally occurring isotopes of the element. Obviously, it is possible for an element to have a series of atomic masses while retaining its atomic number by the simple expedient of adding or subtracting neutrons from the nucleus. Elements with the same atomic number but different atomic masses are called isotopes.

Isotopes have very nearly identical chemical properties but frequently can be separated from each other by some physical manipulation. Differences in rates of diffusion and of evaporation have been utilized, as well as other techniques.

There are a very large number of isotopes. Tin, for example, has ten isotopes. Many of these isotopes occur in extremely small amounts. A very dramatic and important event was the discovery by Urey and others that hydrogen has an isotope. Ordinary hydrogen has an atomic weight of 1.00813; the heavy isotope has an atomic weight of 2.01472.[1] Since both these hydrogens have an atomic number of 1, the heavy hydrogen must have a nucleus which is made up of one neutron and one proton. The heavy hydrogen has been named deuterium. The coining of a new name for the hydrogen isotope seems to the present author unnecessary and somewhat illogical, since the relation of deuterium to hydrogen is exactly analogous to the relation of the isotopes of other elements to the respective elements. Naturally occurring isotopes of actual or potential interest to the biochemist are deuterium, carbon 13, nitrogen 15, oxygen 18, and sulfur 34. The number after the element is the mass number, which is simply the whole number nearest the isotopic atomic weight. This number is equal to the sum of the neutrons and protons in the nucleus. Table 3 shows the relative abundance of the isotopes of these elements in nature.

TABLE 3

ABUNDANCE OF SOME NATURALLY OCCURRING ISOTOPES

Element	Atomic number	Mass number	Relative abundance
Hydrogen	1	1	99.99
Deuterium	1	2	0.003
Carbon	6	12	99.3
Carbon	6	13	0.7
Nitrogen	7	14	99.86
Nitrogen	7	15	0.14
Oxygen	8	16	99.81
Oxygen	8	18	0.16

Deuterium and nitrogen 15 have been used quite extensively as tracers in physiological reactions. Since the isotopes of a given element have very nearly identical chemical properties, the living organism is unable to distinguish between molecules containing the isotopes of the

[1] J. A. Swartout and M. Dole, *J. Am. Chem. Soc.*, **61**, 2027 (1939).

element. Deuterium is introduced into an organic compound by the proper chemical reaction, and the compound is fed to an animal. After a given time the animal is killed, the deuterium content of the various tissues is determined, and the rate of accumulation of the compound by the various tissues is thus established. More interesting, however, is the tracing of intermediates in physiological reactions by means of isotopes. For this purpose compounds related to the compound fed are isolated and their isotopic content determined. If these chemically related compounds contain deuterium in excess of that naturally present, then they must have been derived from the compound fed. In discussing isotopic concentrations the term atom per cent excess is used, by which is meant the percentage of the isotope in excess of that normally present. In using an isotope for tracing the fate of a compound in the body, care must be taken to place the atom in a position in the compound where it will not be readily exchanged for an isotopic atom incidentally present in the medium. Deuterium atoms in carboxyl, hydroxyl, amino, and other polar groups are extremely labile and exchange with the hydrogen of the aqueous medium very quickly. Deuterium of methyl and methylene groups is, in general, stable and is not removable by treating the compounds even at high temperatures with acid or with alkali. In physiological work with deuterium the tissues or compounds isolated from the tissues are burned to carbon dioxide and water. The deuterium content of the resulting water is measured by refractometric or density methods. In Table 4 are shown some of the physical properties of water and of heavy water.

TABLE 4

PROPERTIES OF ORDINARY AND OF HEAVY WATER

Property	H_2O	D_2O
Melting point	0.00° C.	3.82° C.
Boiling point	100.00° C.	101.42° C.
Relative density at 25° C.	1.0000	1.1076
Refractive index (D line of Na, 20° C.)	1.33300	1.32828
Surface tension 20° C.	72.75 dynes/cm.	67.8 dynes/cm.
Solubility of NaCl at 25° C.	35.917 g./100 g. H_2O	33.042 g./100 g. D_2O

Deuterium can be concentrated by the prolonged electrolysis of water. Hydrogen oxide electrolyzes faster than deuterium oxide, and, accordingly, deuterium oxide accumulates in the electrolysis vessels.[2]

[2] H. C. Urey, F. G. Brickwedde, and G. M. Murphy, *Phys. Rev.*, **39**, 164 (1932); **40**, 1 (1932).

Nitrogen 15 can be concentrated [3] by allowing concentrated ammonium sulfate solutions to flow down a fractionating column into a solution of sodium hydroxide. The ammonia gas set free ascends the column, and an exchange reaction takes place, the heavy nitrogen passing from the ammonium ion into the ammonia molecule. By repeated contact of the ammonia with ammonium sulfate solution, the concentration of the nitrogen 15 in the ammonia gas is increased to more than 2 per cent. Amino acids and other physiologically important nitrogen compounds have been synthesized using nitrogen 15. These compounds have been fed to animals and the fate of the nitrogen 15 in the various organs and tissues of the animal determined. It would take us too far afield to describe the important and interesting physiological experiments with deuterium and nitrogen 15. Schoenheimer and Rittenberg [4] have reviewed the physiological work with deuterium and with nitrogen 15. The amount of nitrogen 15 present is usually determined from the relative intensities of the lines attained with the mass spectrograph. The mass spectrograph depends upon the production of positively charged gaseous ions. These ions are deflected by means of electrostatic and magnetic fields, the extent of deflection being an inverse measure of the mass of the atom while the intensity of the lines on a photographic plate is the measure of the amounts of the element present.

Sulfur 34 and carbon 13 as well as oxygen 18 are rather difficult to obtain in high concentrations, and they have not as yet found extensive application as tracers. Urey and co-workers have described methods for concentrating oxygen 18,[5] carbon 13,[6] and sulfur 34.[7]

Radioactive Isotopes

If atoms of many elements are bombarded with high-speed particles (α-particles, neutrons, protons, or deutrons) the elements will undergo nucleo reactions. The particle will penetrate the nucleus and form a new element (the use of neutrons may lead to the production of an isotope). This clearly is a transmutation of elements. The extent of these nucleo reactions, however, is extremely limited, and only minute amounts of the new element are formed.

[3] H. G. Thode, J. E. Gorham, and H. C. Urey, *J. Chem. Phys.*, **6**, 296 (1938).

[4] R. Schoenheimer and D. Rittenberg, *Physiol. Revs.*, **20**, 218 (1940).

[5] J. R. Huffmann and H. C. Urey, *Ind. Eng. Chem.*, **29**, 531 (1937).

[6] I. Roberts, H. G. Thode, and H. C. Urey, *J. Chem. Phys.*, **7**, 137 (1939).

[7] H. G. Thode, J. E. Gorham, and H. C. Urey, *J. Chem. Phys.*, **6**, 296 (1938).

Some actual nucleo reactions which have been observed are:

BOMBARDMENT WITH NEUTRONS

$$^{27}_{13}\mathrm{Al} + n \rightarrow {}^{27}_{12}\mathrm{Mg} + {}^{1}_{1}\mathrm{H}$$

$$^{23}_{11}\mathrm{Na} + n \rightarrow {}^{24}_{11}\mathrm{Na}$$

$$^{16}_{8}\mathrm{O} + n \rightarrow {}^{13}_{6}\mathrm{C} + {}^{4}_{2}\mathrm{He}$$

BOMBARDMENT WITH PROTONS

$$^{44}_{20}\mathrm{Ca} + {}^{1}_{1}\mathrm{H} \rightarrow {}^{44}_{21}\mathrm{Sc} + n$$

BOMBARDMENT WITH DEUTRONS

$$^{27}_{13}\mathrm{Al} + {}^{2}_{1}\mathrm{D} \rightarrow {}^{25}_{12}\mathrm{Mg} + {}^{4}_{2}\mathrm{He}$$

In the above nucleo reactions n stands for the neutron; the raised number in front of the symbol of the element denotes the mass number (sum of the number of neutrons and protons in the nucleus); the lowered numbers indicate the atomic number of the element.

All the above transmutations give rise to stable elements. Possibly a more interesting and certainly a more useful development has been the production of radioactive isotopes by bombardment with high-speed particles. For example, if boron is bombarded with high-speed deutrons, it forms the unstable isotope of carbon

$$^{10}_{5}\mathrm{B} + {}^{2}_{1}\mathrm{D} \rightarrow {}^{11}_{6}\mathrm{C} + {}^{1}_{0}n$$

The carbon isotope then disintegrates to yield

$$^{11}_{6}\mathrm{C} \rightarrow {}^{11}_{5}\mathrm{B} + e^{+}$$

The particle e^{+} is known as a positron and has the same mass as an electron but carries a positive charge instead of a negative one. The "half life" of $^{11}_{6}\mathrm{C}$ is 21.5 minutes. By half life is meant the time required for half of the radioactive isotope to disintegrate. Table 5 gives the half life of some of the radioactive isotopes which have been used as tracers in biological reactions.

The bombardment of elements with high-speed particles and the production of isotopes is a highly technical process. There is more than one way of carrying out this bombardment. One of the methods uses the cyclotron. The cyclotron is an apparatus which is so arranged that the ions (protons or deutrons) are made to move in a flat spiral of steadily increasing radius and are given a voltage impulse of about 50,000 volts in each of some two hundred semicircular paths so that a maximum voltage of several million volts can be attained. After being accelerated

through this powerful field, the ions are then focused on the element to be bombarded.

Radioactive isotopes, being much more easily detected and estimated, are far more convenient to use as tracers than the stable isotopes. In carrying out a physiological study, the radioactive isotopes are mixed with the stable element and synthesized into an organic compound. The amount of radioactive isomer is very small and the normal analog is used as a carrier. The compound is then fed. Since the radioactive and normal isotopes have identical chemical properties, they will distribute

TABLE 5

HALF LIFE OF SOME RADIOACTIVE ISOTOPES

Isotope	Half life
$^{3}_{1}H$	30 years
$^{32}_{15}P$	14.5 days
$^{24}_{11}Na$	14.8 hours
$^{42}_{19}K$	12.5 hours
$^{38}_{17}Cl$	37 minutes
$^{58}_{26}Fe$	47 days
$^{128}_{53}I$	26 minutes
$^{11}_{6}C$	21.5 minutes

themselves in the same ratio in all parts of the body to which the normal element has gone. The amount of the radioactive isotope may be quickly estimated by means of a Geiger counter. The Geiger counter consists of a metal cylinder filled with an easily ionizable gas and containing an axial wire charged to a potential of several hundred volts. Upon disintegration the radioactive isotope gives off charged particles which ionize the gas in the Geiger counter. The gas is thus rendered conducting and there is a surge of current from the axial wire to the metal cylinder which is amplified sufficiently to operate a mechanical counter. Each charged particle entering the gas chamber is thus counted.

Hevesy was the first to use radioactive isotopes in physiological work. He has reviewed [8] the use of such isotopes. Perhaps the most significant physiological results obtained with the use of radioactive isotopes have been those of Ruben and co-workers [9] on the process of photosynthesis.

[8] G. Hevesy, *Ann. Rev. Biochem.*, **9**, 641 (1940).

[9] S. Ruben et al., *J. Am. Chem. Soc.*, **61**, 661 (1939); **62**, 3443 (1940); **62**, 3450 (1940); **62**, 3451 (1940).

Photosynthesis is the most fundamental biochemical process of all. Light energy and carbon dioxide are taken into the plant. The carbon dioxide is reduced and carbohydrates are synthesized. Ruben et al., with the use of radioactive carbon, confirmed the suggestion that the initial step is the carboxilation of a compound whose molecular weight is about a thousand. This first step is not of a photochemical nature. The carboxyl group so produced is then reduced. The energy of this reduction comes from the light absorbed by the chlorophyll.

Molecules

A book which all chemists should read is Linus Pauling's *The Nature of the Chemical Bond*, published by Cornell University Press. The author regards this assignment as a "must."

As has been pointed out, the inert-gas type of electronic arrangement is a very stable one, and it is reasonable to assume, therefore, that in chemical combination atoms tend to acquire the inert-gas structure. For example, sodium has one electron in its outer shell (valence shell), and chlorine has seven. The sodium, accordingly, transfers one electron to the chlorine, thereby acquiring the neon structure while chlorine assumes the argon arrangement. This transfer leaves the sodium positively charged while the chlorine becomes negative. These charged particles are known as the sodium and chloride ions, respectively. Since these ions are oppositely charged, there is an electrostatic attraction between them; such an electrostatic bond is called an ionic bond. This is the usual type of chemical bond between the electropositive elements of Group 1 or Group 2 and the electronegative elements of Group 7 of the periodic table. In all these elements the transformation to the inert-gas structure is easily accomplished.

For elements in the other groups of the periodic table some difficulties are encountered. The energy required to remove three electrons from an element in the third group is very large. It is improbable, therefore, that a B^{+++} ion will form in chemical combination. Correspondingly, it is doubtful that nitrogen has sufficient electron affinity to take up three electrons and form the N^{---} ion. The inert-gas structure can be attained, however, in chemical combination by the sharing of electrons in pairs. For example, in the formation of methane, each of the four hydrogen atoms contributes one electron and the carbon atom contributes four electrons

$$4H + C \rightarrow H:\overset{\displaystyle H}{\underset{\displaystyle H}{\overset{\displaystyle \vdots}{\underset{\displaystyle \vdots}{C}}}}:H$$

The carbon in methane thus acquires the neon structure while the hydrogen assumes that of helium. Such chemical bonds as these are called covalent bonds. Bonds of this sort form between elements whose electronegativities do not differ greatly. On the other hand, ionic bonds result when the difference between the electronegativities of the reacting atoms is large. The order of the electronegativity of some elements is

F>O>N>Cl>Br>C>S>P>H>Al>Mg>Li>Ca>Sr>Na>Ba>K>Rb>Cs

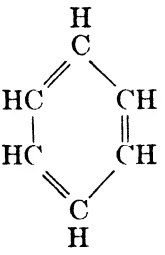
It is evident that there must exist bonds intermediate in nature between the extreme covalent and extreme ionic bond. Consider, for example, hydrogen chloride. This molecule can exist as

$$H^+ \quad :\ddot{\underset{\cdot\cdot}{Cl}}:^-$$

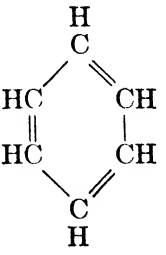
or as

$$H:\ddot{\underset{\cdot\cdot}{Cl}}:$$

When two or more reasonable electronic structures such as the above can be assigned to the molecule, we must take into account what is known as resonance. The hydrogen chloride molecule is said to resonate between the ionic and covalent structure. A resonating structure is more stable than any of the individual structures. Many molecules show resonance between two or more electronic structures, and in a great many cases this extra resonance stability or resonance energy, as it is called, is of crucial importance for the existence of the molecule. A classical example of resonance is shown by the benzene ring, which resonates between the following two structures:



A



B

The benzene ring is thereby stabilized and is more unreactive than would be anticipated from its possession of three double bonds.

Resonance is to be distinguished from tautomerism. A tautomeric mixture actually has in it the different tautomeric compounds, which exist simultaneously. The essential feature of the resonance idea is that a state exists intermediate between the states represented by the

resonating structures. The two or more resonating structures do not exist in a state of equilibrium; the resonating structure is the only structure present.

In hydrogen chloride the bond between the hydrogen and chlorine is said to have a certain amount of ionic character and of covalent character. The actual amount of ionic nature associated with a chemical bond can frequently be calculated with fair accuracy. Hydrogen chloride bond is 17 per cent ionic. In general, the greater the difference between the electronegativity of the reacting atoms the greater will be the ionic character of the resulting bond.

Bond Angles and Distances

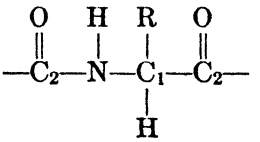
Ionic bonds are without directional effects. Covalent bonds, however, show very pronounced directions, and the angles between the bonds have been in most cases determined. Table 6 gives some common bond angles.

TABLE 6

BOND ANGLES FOUND IN SEVERAL COMPOUNDS

Compound	Group	Bond angle
H_2O	H—O—H	105°
H_2S	H—S—H	92° 20′
NH_3	H—N—H	108°
$N(CH_3)_3$	C—N—C	108°
Propane	C—C—C	111° 30′

There is a small variation of the bond angles of a given atom from compound to compound. Huggins [10] gives the following distances and angles for a peptide chain:



N to C_1, 1.41 Å; C_1 to C_2, 1.52 Å; C_2 to 0, 1.25 Å; C_2 to N, 1.33 Å; C_1 to H, 1.09 Å; the bond angles are: $\angle NC_1C_2$, 112°; $\angle C_1C_2N$, 118°; $\angle C_2NC_1$, 118°.

In many cases fairly accurate values can be obtained for the energy required to break a bond; i.e., the energy required to dissociate a

[10] M. L. Huggins, *Ann. Rev. Biochem.*, **11**, 27 (1942).

compound into its elements. These energies are determined by thermochemical and spectroscopic methods. Some of the energies are given in Table 7.

TABLE 7

ENERGY VALUES FOR SOME SINGLE BONDS

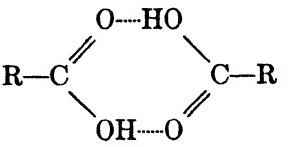
Bond	Energy in kilogram Calories
H—H	103
C—C	58
N—N	20
O—O	35
S—S	64
O—H	110
S—H	88
H—Cl	103
C—N	49
C—O	70

Hydrogen Bonds

One of the most interesting of chemical bonds as far as the biochemist is concerned is the hydrogen bond. The significance of this bond arises not from its strength but from its weakness. Its bond strength is only in the neighborhood of 5 kilocalories. With such a low energy value it is easily possible for this bond to form and to rupture at normal temperatures.

The hydrogen bond is a chemical bond resulting from the attraction of two electronegative atoms for a proton. Accordingly, only the most electronegative atoms are capable of forming hydrogen bonds, and, furthermore, the strength of the bond increases with the electronegativity of the two bonded atoms. It is found that only fluorine, oxygen, nitrogen, and chlorine can form hydrogen bonds, and that the ability to form hydrogen bonds decreases in the above order.

There are numerous examples of hydrogen bonds or hydrogen bridges, as they are sometimes called. The fatty acids have a great tendency to form dimolecules. This structure involves two hydrogen bonds



The hydrogen bond determines the magnitude and nature of the mutual interaction of water molecules and is consequently responsible

for the striking physical properties of this uniquely important substance. The structure of water is discussed in detail in Chapter V.

The hydrogen bond between the nitrogen and oxygen atoms in neighboring peptide chains is in all probability of great importance in determining the arrangement of the peptide chain in protein molecules.[11, 12]

Hydrogen-bond formation can be detected by spectroscopic analysis. The stretching of the valence bond incidental to the formation of a hydrogen bond results in a decrease in frequency of vibration. The shift of the normal absorption spectrum of a compound is, accordingly, taken as evidence for hydrogen-bond formation. Owing to the fact that hydrogen bonds are long-distance bonds, the frequency is of low order and absorption occurs in the infra-red region of the spectrum.

In keeping with all chemical bonds there is an inverse relation between bond strength and bond distance. Hydrogen bonds, in general, being quite weak bonds, their lengths are greater than those of covalent bonds. The energies and distances of some hydrogen bonds are shown in Table 8.

TABLE 8

ENERGIES AND LENGTHS OF SOME HYDROGEN BONDS

Bond	Substance	Energy in kilocalories	Lengths in Å
O—H—O	H_2O	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	1.30	3.38
O—H—N	Peptides		2.85
N—H—F	NH_4F	5.0	2.63

Hydrogen bonds have been ably discussed by Huggins [13] and by Pauling.[14]

[11] A. E. Mirsky and L. Pauling, *Proc. Natl. Acad. Sci. U. S.*, **22**, 439 (1936).
[12] M. L. Huggins, *J. Chem. Phys.*, **8**, 598 (1940).
[13] M. L. Huggins, *J. Org. Chem.*, **1**, 407 (1936).
[14] L. Pauling, *The Nature of the Chemical Bond*, Cornell University Press, 1939.

Chapter II

ENERGETICS

The science of energetics (thermodynamics) deals with energy transformations. Thus, when a person walks upstairs, certain substances undergo chemical reactions in his muscles with the performance of work and liberation of heat. The work energy has increased the gravitational potential energy of the person's body, and the heat has gone into body warmth. This thermodynamic situation has been analyzed with considerable accuracy in the metabolism laboratories.

Although the metabolic studies of the physiologist are of great importance and are appropriate for what they are intended, the biochemist naturally desires a more detailed and intimate study of energy transformations than the physiologist is usually prepared to furnish. The purpose of this chapter is to present a limited exposition of the methods of chemical energetics (thermodynamics).

First, let us inquire into what we mean by the terms heat and work. As is well known, the unit of heat is the calorie, defined as the quantity of heat required to raise the temperature of 1 gram of water from 15° C. to 16° C. A kilogram calorie or large calorie (sometimes written Calorie) is the heat required to raise 1 kilogram of water from 15° C. to 16° C. Physically, heat is a species of molecular motion which gives rise to the physiological sensation of warmth or coldness. Actually, it is extremely difficult to give a rigorous definition of heat. See, for example, Brönsted [1] and MacDougall.[2]

Several temperature scales are in use. The one which is familiar to every scientific worker is the centigrade scale. It has been found, however, that thermodynamic properties which vary with temperature are proportional to the absolute temperature. The absolute temperature is obtained by adding 273.1° to the centigrade scale. Thus, 25° C. is 298.1° absolute.

The unit of work is the erg which is the work done in lifting a gram weight 1 centimeter divided by the acceleration of gravity. It is considered that the term work includes all forms of energy not covered by the term heat.

[1] J. N. Brönsted, *J. Phys. Chem.*, **44**, 699 (1940).

[2] F. H. MacDougall, *J. Phys. Chem.*, **44**, 713 (1940).

It was early discovered that there is an equivalency between work and heat. Thus, if we perform the original experiment of Joule and measure the rise in temperature of a given weight of water as produced by the rotation of paddle wheels which are caused to rotate by a falling weight, we can obtain the equivalency. When this experiment is performed with all its modern refinements, it is found that

$$1 \text{ calorie} = 4.185 \times 10^7 \text{ ergs}$$

Knowing this result, we are in a position to state both heat and work changes in terms of calories.

Dimensions

We shall have occasion to discuss the dimensions of numerous physical quantities in the course of this book. It seems best to anticipate these discussions with a short exposition at this point.

Physical quantities can be conveniently reduced to four fundamental dimensions, length (L), mass (M), time (t), and temperature (T). For example, the dimensions of a volume are length cubed (L^3). Some other physical quantities and their dimensions are:

$$\text{Force} = \text{Acceleration} \times \text{Mass} = \frac{LM}{t^2}$$

$$\text{Tension} = \text{Force per unit length} = \frac{M}{t^2}$$

$$\text{Pressure} = \text{Force per unit area} = \frac{M}{Lt^2}$$

$$\text{Energy} = \text{Work} = \text{Force} \times \text{Distance} = \frac{L^2M}{t^2}$$

Some rules of dimensional analysis are as follows:

1. Pure numbers, such as π, $\frac{1}{2}$, and 50, are dimensionless.
2. Complete exponents are dimensionless, although parts of exponents may have dimensions.
3. Ratios of dimensionally identical quantities have no dimension.
4. Sines, cosines, etc., are dimensionless.
5. Differentials retain the dimensions of the variables.
6. If quantities are additive they must have the same dimensions.

Dimensional analysis is of great aid to one with a limited mathematical equipment in protecting himself against mathematical frauds. If, for example, it is found that the dimensions of a certain equation cannot be made to cancel out, then that equation is incorrect.

Heat Changes

If a chemical reaction is allowed to proceed, it will give off or take up heat; and if the system does no work (reaction carried out in a bomb calorimeter), the heat given off or taken up will always be the same for a given weight of the reacting materials and at a given temperature.

If we burn 1 gram-molecular weight of glucose to gaseous CO_2 and H_2O, we will have 673,000 calories of heat evolved.[3] Since heat has been given off, the system is poorer by 673,000 calories; accordingly, we speak of the negative heat of reaction. We write the reaction to indicate this:

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6H_2O + 6CO_2 \qquad \Delta H_0 = -673{,}000 \text{ calories}$$

If heat had been absorbed from the surroundings instead of being given off, it would have been considered positive.

The heat change of a reaction (carried out so that the reaction does no work) is independent of the nature and number of intermediate steps. It is fortunate that this principle is a valid one. For example, it allows the physiologists to predict with accuracy the heat which will be evolved from the burning of foodstuffs in the body in spite of the fact that, before they are completely burned, they must pass through a complicated series of reactions involving many steps.

Work Changes

If we could build a machine which was completely efficient and which would allow us to utilize all the energy of a chemical reaction to perform work, we would find that upon oxidizing 1 gram-molecular weight of glucose to gaseous CO_2 and H_2O, we could realize as a maximum 676,616 calories of work (at constant temperature and pressure).[4] The term free energy is applied to such work energy; it means that so much energy is free and not bound and can be obtained from the system, provided that we have the proper apparatus for accomplishing it. The same conventions of sign apply to free energy changes as for heat changes. Thus, in the oxidation of 1 gram-molecular weight of glucose at 25° C., free energy is evolved and accordingly its sign is negative.

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6H_2O + 6CO_2 \qquad \Delta F_0 = -676{,}616 \text{ calories}$$

If it had been necessary to do work on the reaction to make it proceed, the free energy change would have been positive.

No actual machine could obtain 676,616 calories of work from the burning of a gram-molecular weight of glucose because no machine is

[3] *International Critical Tables*, Vol. V, p. 166.

[4] G. S. Parks and H. M. Huffman, *Free Energies of Some Organic Compounds*, The Chemical Catalog Co., Inc., New York, 1932.

100 per cent efficient. We could, in fact, calculate the efficiency of our machine by comparing its actual performance with the theoretical. A much more useful purpose, however, to which free energy data can be put is the prediction of the spontaneity of a process. Evidently, if the free energy change for a given reaction is positive, we must do work on the system before it will react and such a reaction will not proceed of its own accord. If, on the other hand, the free energy change is negative, the reaction may proceed spontaneously. We shall see presently how we can use the magnitude of the free energy change to predict the extent of the reaction.

There are certain limitations to the usefulness of predictions based upon free energy changes. In an isolated system we can say definitely that, if the free energy change is positive, the reaction cannot occur. In a living system, however, we are not dealing with a single isolated system but with a multitude of interrelated systems, and the necessary free energy may be supplied by other chemical reactions. Another limitation is the time factor. A reaction with a negative free energy may proceed so slowly as to have no practical significance.

Entropy Changes

We have noted that, in the burning of a gram-molecular weight of glucose, 673,000 calories of heat is evolved whereas the maximum amount of work which can be obtained from the reaction is 676,616 calories. There is an apparent discrepancy of 3,616 calories. The question of the source of these 3,616 calories arises. The answer is that, during the completely efficient oxidation of glucose by the machine, 3,616 calories of heat flowed into the reaction mixture from the outside and were converted into work. That is, if we had conducted our reaction in an isolated container in such a fashion that no heat could flow into the mixture to maintain the temperature constant, we would have found that as the machine worked the temperature dropped. This heat, which flows into or out of a system at constant temperature and pressure while the reaction is allowed to proceed in such a fashion as to obtain its maximum efficiency, is sometimes known as the reversible heat of reaction. It receives this name because a reaction which proceeds in such a fashion as to yield its maximum efficiency must proceed in a reversible manner at every point in its course. It is evident that for the burning of glucose in the manner we have described

$$\Delta H = \Delta F + \text{Reversible heat of reaction} \qquad 1$$

It has been found that equation 1 is true of all reactions and all processes carried out at constant temperature and pressure.

The reversible heat of reaction involves a mysterious function known to the initiated as entropy. At constant temperature and pressure

$$\text{Reversible heat of reaction} = T\Delta S \qquad 2$$

where ΔS is the entropy change and is expressed in calories per degree (absolute) per mole.

Equation 1 now becomes

$$\Delta H = \Delta F + T\Delta S \qquad 3$$

For a reaction to proceed spontaneously, there must be an increase in entropy. The entropy change, therefore, gives us another test for the spontaneity of a reaction.

Entropy is related to the probability of an event occurring. Thus, if P_1 is the probability associated with one situation and P_2 that with a second situation, the entropy change in going from the first to the second situation, provided that there is no energy change, is given by

$$\Delta S = R \ln \frac{P_2}{P_1} \qquad 4$$

where R is a constant into whose nature we shall presently inquire. What we really mean, therefore, by the statement that the entropy must increase in a spontaneous reaction is that the reaction mixture must proceed towards a more probable state. Entropy is a measure of the number of possible configurations of a system having a given energy.

Free Energy Change and the Equilibrium Constant

Our main object in discussing the free energy change or the maximum work obtainable from a process is to be able to predict the occurrence and extent of chemical reactions. It is necessary, therefore, to show that a relation exists between the free energy change of a reaction and its equilibrium constant. The relation between these two quantities is by no means obvious, and it is not an easy one to visualize. Its importance for the continuity of our argument is so great, however, that an attempt will be made to outline the relation between these two factors. This is not intended to be a rigorous derivation.

Suppose that we have a large living cell whose membrane is freely permeable to water. The cell is to be surrounded by an aqueous medium consisting of dilute solutions of substances A and B. The cell has in its interior an enzyme which is capable of converting A into B or B into A so that these two substances are at all times in equilibrium with each other inside the cell. The cell now proceeds to accumulate a certain

volume of the aqueous medium in a vacuole and then concentrates the solution in respect to A, so that its concentration is exactly equal to that in the cytoplasm of the cell. Of course, osmotic work is performed by the cell in carrying out this step. Then the cell, without doing additional work, delivers this small volume of solution of A to the cytoplasm, where it reacts to produce B. The excess of B thus produced is withdrawn by the cell into a second vacuole, where it has the same concentration in respect to B as it has in the cytoplasm. No work is involved in this step.

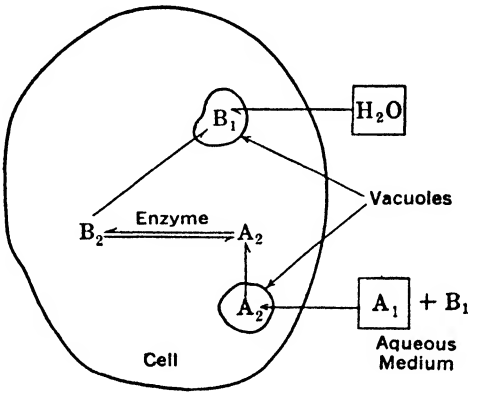


FIG. 1. Diagrammatic representation of living cell doing osmotic work.

The cell is now able to do osmotic work by allowing water from the aqueous medium to diffuse into the vacuole containing B, until the concentration of B in the vacuole is the same as that in the aqueous medium outside the cell. The problem is to calculate the osmotic work done by the cell in taking A from the aqueous medium where its concentration is A_1, into the cell where its concentration is A_2, and the osmotic work gained by the cell in allowing the molecules of B produced from A to be diluted from B_2 to B_1, B_1 being the concentration of B in the outside, aqueous medium. The difference between these osmotic-work terms is the work required to convert A into B.

The osmotic work done in concentrating A is given by

$$\text{Work}_A = RT \ln \frac{A_2}{A_1} \qquad 5$$

while the work gained by the cell in diluting B is

$$\text{Work}_B = RT \ln \frac{B_2}{B_1} \qquad 6$$

The net work is equal to the difference between these terms; accordingly

$$\text{Net work} = RT \ln \frac{A_2}{A_1} - RT \ln \frac{B_2}{B_1} \qquad 7$$

Rearranging terms, and remembering that since the cell is assumed to be completely efficient the free energy change is equal to the net work, we have

$$\Delta F = -RT \ln \frac{B_2}{A_2} + RT \ln \frac{B_1}{A_1} \qquad 8$$

Since A_2 and B_2 are in equilibrium

$$\frac{B_2}{A_2} = K \qquad 9$$

and therefore

$$\Delta F = -RT \ln K + RT \ln \frac{B_1}{A_1} \qquad 10$$

If the concentrations of A and of B are chosen equal to unity, we have

$$\Delta F_0 = -RT \ln K \qquad 11$$

This equation is a completely general one for a system at constant temperature and pressure, and it is valid no matter how many reacting substances are present, provided that the reactants are initially at unit concentration. We shall find later in this chapter that thermodynamic concentrations must be used, in order that this equation may hold for all concentrations.

The constant R in the above equations is the molar gas constant; it is given by the equation

$$R = \frac{PV}{T} \qquad 12$$

At standard temperature (273.1° absolute) and pressure (76.0 cm. of mercury), a gram-molecular weight of an ideal gas occupies 22,412 cc. Accordingly

$$R = \frac{76.0 \times 13.596 \times 980.6 \times 22{,}412}{273.1}$$

$$= 8.315 \times 10^7 \text{ cc. ergs per degree per mole}$$

Since as we have seen

$$1 \text{ calorie} = 4.185 \times 10^7 \text{ ergs}$$

we have

$$R = 1.987 \text{ calories per degree per mole}$$

Converting equation 11 to logarithms to the base 10 and substituting the value of R, we have

$$\Delta F_0 = -4.575T \log K \qquad 13$$

where T is expressed as the absolute temperature.

Equilibrium

This term has a very precise thermodynamic meaning. When a system is in equilibrium, its capacity for doing work is zero; the free energy of the system is at a minimum. Its entropy is at a maximum; it is in its most probable state.

Relation between Heat and Free Energy Changes

As we have seen, for any isothermal process at constant pressure the relation between the free energy change, heat change, and entropy change is given by the equation

$$\Delta H = \Delta F + T\Delta S \qquad 14$$

In other words, it is only when the entropy change is zero that the heat of reaction is equal to the free energy change, and it is, therefore, only under this condition that the heat of reaction is a true measure of the driving force of a process. Under all other conditions the heat change is without direct significance in predicting the extent or spontaneity of a reaction.

By combining in the appropriate manner the necessary mathematical statements of thermodynamics, the relation between the heat and free energy changes at constant pressure is obtained. This relation, known as the Gibbs-Helmholtz equation, is

$$\Delta F - \Delta H = T\left(\frac{d\Delta F}{dT}\right) \qquad 15$$

If the differential $d\Delta F/dT$ is zero, i.e., the free energy change is independent of temperature, ΔF will equal ΔH of the reaction.

If we integrate the Gibbs-Helmholtz equation, assuming ΔH to be constant, there results

$$\frac{\Delta F_2}{T_2} - \frac{\Delta F_1}{T_1} = -\frac{\Delta H(T_2 - T_1)}{T_1 T_2} \qquad 16$$

which allows us to calculate the free energy change as a function of temperature, provided that we know the heat of reaction.

By substituting equation 11 into the Gibbs-Helmholtz equation and carrying out certain mathematical manipulations, we have the van't Hoff equation

$$\frac{d \ln K}{dT} = \frac{\Delta H}{RT^2} \qquad 17$$

If we again assume ΔH to be constant and integrate the van't Hoff equation, at the same time converting the natural logarithm to logarithm to the base 10 and substituting the numerical value for R, we have

$$\log \left(\frac{K_2}{K_1}\right) = -\frac{\Delta H}{4.575}\left(\frac{T_1 - T_2}{T_2 T_1}\right) \qquad 18$$

where K_1 is the equilibrium constant at temperature T_1 and K_2 that at temperature T_2. This equation allows us to calculate ΔH of the reaction if we know K_2 and K_1 or allows us to predict K_2 if we know K_1 and ΔH.

Another form of the integrated van't Hoff equation is

$$\log K = -\frac{\Delta H}{4.575T} + C \qquad 19$$

where C is an integration constant.

Evidently, if log K is plotted against $1/T$, the slope of the straight line will be $-\Delta H/4.575$, from which ΔH may be obtained directly.

TABLE 1

EFFECT OF TEMPERATURE CHANGES ON THE EQUILIBRIUM BETWEEN NATIVE AND DENATURED TRYPSIN IN 0.01 *N* HCl

Temperature °C.	Reciprocal of absolute temperature	Percentage denaturation	K	log K
42	0.0031742	32.8	0.488	−0.3115
43	0.0031635	39.2	0.645	−0 1906
44	0.0031535	50.0	1.000	0
45	0.0031436	57.4	1.347	+0.1294
48	0.0031143	80.4	4.102	+0.6130
50	0.0030950	87.8	7.197	+0.8572

The method by which the heat change of a reaction can be calculated from the dependence of the equilibrium constant on the temperature is well illustrated by the results and calculations of Anson and Mirsky [5] on

[5] M. L. Anson and A. E. Mirsky, *J. Gen. Physiol.*, **17**, 393 (1933–34).

the heat denaturation of trypsin. These workers found that, if a trypsin solution is heated at 60° C. in 0.05 *N* acid, the trypsin is converted into a protein which is completely precipitable by quarter-saturated ammonium sulfate. When the heated trypsin solution is cooled, the trypsin changes back into the original form which is not precipitated by quarter-saturated ammonium sulfate. The equilibrium conditions between the

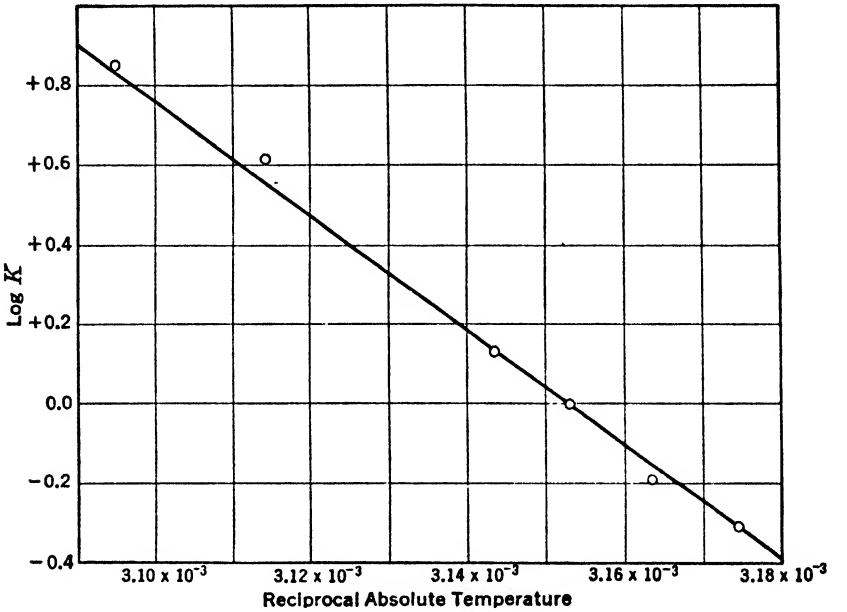


FIG. 2. Plot of log *K* of trypsin denaturation against the reciprocal of the absolute temperature. (Data of Anson and Mirsky.)

native and denatured trypsin were studied over a range of temperatures with the results shown in Table 1.

In Fig. 2 the reciprocal of the absolute temperature is plotted against log *K*.

The ΔH of the reaction is found by multiplying the slope of the line in Fig. 2 by 4.575. This turns out to be 67,600 calories. The entropy change of the denaturation of trypsin can also be calculated. At 44° C. the free energy change is zero; therefore, ΔH is equal to $T\Delta S$ (see equation 14). If we divide the heat change by the absolute temperature (317.1), we obtain the entropy change which for this temperature is 213 entropy units. This large value for the entropy change indicates that the transformation from the native trypsin to denatured trypsin involves a change from a highly organized to a much more random structure.

Thermodynamics of Solutions

Since the biochemist deals very largely with solutions instead of with solids or gases, it will be profitable to consider briefly some of the energy relations in solutions.

The analytical chemist expresses concentrations in terms of the number of moles of a given substance in 1 liter of solution, 1 mole of a substance dissolved in 1 liter of solution being a molar solution. The physical chemist, on the other hand, often finds it more convenient to express concentrations in terms of the number of moles per 1,000 grams of solvent. One mole of the substance dissolved in 1,000 grams of the solvent is known as a molal solution. The advantage of this mode of expression is that concentration is independent of temperature, which is not true of molar solutions.

A still more fundamental method of expressing concentrations is by means of mole fractions. The mole fraction of one component is equal to the total number of moles present (solvent and solutes) divided into the number of moles of the constituent in question. That is,

$$N_1 = \frac{n_1}{n_1 + n_2} \qquad 20$$

and

$$N_2 = \frac{n_2}{n_1 + n_2} \qquad 21$$

where N_1 is the mole fraction of one constituent, n_1 is the number of moles of that constituent present, n_2 is the number of moles of the second constituent, and N_2 is the mole fraction of the second constituent. This method of expression is to be preferred because thermodynamic properties which vary with concentration are proportional to the number of particles present and the mole fraction is the ratio of the number of molecules of a substance to the total number of molecules present.

Let us consider a molal solution of sucrose in water and compare these three methods of expressing concentrations. To produce a molal solution of sucrose we dissolve a gram-molecular weight of sucrose (342.17 grams) in 1,000 grams of water. The density of such a solution at 25° C. is 1.2417. Evidently the total volume of this solution is 1,342.17 divided by the density, or 1,081.08 cc. The molar concentration is, accordingly, 1,000 divided by 1,081.08, or it is 0.925 molar. The mole fraction of sucrose is

$$N_1 = \frac{\frac{342.17}{342.17}}{\frac{1,000}{18} + \frac{342.17}{342.17}} = 0.0177$$

It should be noted that the mole fraction of a solute in 1 molal aqueous solution is always 0.0177.

Although, as pointed out above, the mole fraction is a more fundamental unit of concentration than either molarity or molality, it has been the custom of physical chemists to use molality; accordingly, in what is to follow, we shall employ molality as the method of expressing concentrations. It is evident that, in the relatively dilute solutions such as the biochemist is likely to encounter, molality is almost exactly proportional to mole fraction, and we lose very little significance by using molality instead of mole fractions.

Activity

The thermodynamist finds himself compelled to deal with another kind of concentration. It was early discovered that thermodynamic equations, such as 11 previously given, are not exact expressions except in very dilute solutions. As the concentrations are increased, such equations, in general, fail to be exact descriptions of experimental results. The term activity or thermodynamic concentration was introduced to take the place of the molality concentrations. It might be said that the activity of a substance is the concentration the substance apparently has as judged by its chemical effects. More precisely, activity may be defined as being equal to the concentration, as the concentration approaches zero. This is the same thing as saying that the solute approaches ideal behavior as the solution becomes more and more dilute.

The three general causes for the departure of solutions from ideal behavior are: (1) interaction between the solute molecules (such as due to electrical charges); (2) interaction between the solute and solvent (hydration); and (3) interaction between solvent molecules. We shall have occasion to deal with these factors in considerable detail in the course of this book.

The activity coefficient is the expression for the ratio of the activity to the concentration; i.e.,

$$\text{Activity coefficient } (\gamma) = \frac{\text{Activity}}{\text{Concentration}} \qquad 22$$

The correct expression for the free energy change in going from unit activity of the reactants to a condition of equilibrium for a chemical reaction of the type

$$aA + bB \rightleftarrows cC + dD$$

is

$$\Delta F_0 = -RT \ln K = -RT \ln \frac{(\text{Activity } C)^c \times (\text{Activity } D)^d}{(\text{Activity } A)^a \times (\text{Activity } B)^b} \qquad 23$$

Free Energy of Dilution

If a solution behaved in an ideal fashion, the free energy change in going from concentration C_1 to C_2 would be

$$\Delta F = RT \ln \frac{C_2}{C_1} \qquad 24$$

For any actual solution

$$\Delta F = RT \ln \frac{A_2}{A_1} \qquad 25$$

where A_1 and A_2 are the activities.

An important thermodynamic situation arises in respect to a saturated solution of a substance in contact with some of its solid phase. Since the solid is in equilibrium with the dissolved substance, the free energy change in going from the solid to the dissolved state is zero and, accordingly, the molar free energy of the substance in solution is equal to that of the solid. Free energy tables frequently give values for the solid in some definite crystalline state. The problem of finding the free energy of a substance in solution at any given concentration, therefore, consists in adding to the free energy of the solid substance the free energy of dilution experienced in going from the saturated solution to the activity in question.

Activities can be determined in a variety of ways such as by the measurement of vapor pressure, electromotive force of the proper electrical cells, freezing-point lowering, boiling-point raising, and osmotic pressure. We shall have occasion to deal in detail with some of these types of measurements in the course of this book.

Partial Molar Quantities

If a mole of a solute is dissolved in a very large quantity of the solution—so large a volume that the concentration is not significantly changed—the resulting increase in volume is known as the partial molar volume of the added material. In the same way we have partial molar quantities of any extensive property such as the heat content or the entropy. The partial specific volume is the increase in volume resulting from the addition of 1 gram of a solute to a large volume of a solution.

An important partial molar quantity is the partial molar free energy. This quantity was called the chemical potential by Gibbs. It is a necessary condition for equilibrium that the chemical potential of any component must be the same throughout the entire system. For example, if we dissolve iodine in carbon tetrachloride and shake the carbon tetrachloride solution with water, at equilibrium the chemical potential of the iodine in the carbon tetrachloride must equal that in the water.

The chemical potential is closely related to activity. The chemical potential of any component in a liquid solution is given by

$$\mu = \mu_0 + RT \ln A \qquad 26$$

where μ_0 is the chemical potential of the pure solvent and A is the activity of the solute.

Partial molar quantities can be evaluated by graphical methods. For example, to determine the partial molar volume of a component of a solution plot the total volume of the system against the number of moles of this component, keeping the molar number of all other components constant; the slope of the curve will give the partial molar volume of the component at the selected concentrations. (See Chapter 15.)

Free Energy of Formation and Its Use

Four general methods have been used to determine the magnitude of the free energy changes in chemical reactions:

1. The measurement of the equilibrium constant K in a reversible reaction and the calculation of ΔF by the relation $\Delta F_0 = -RT \ln K$.
2. The determination of the reversible electromotive force E of an electrical cell involving the reaction in question, in which case $\Delta F = -NFE$ (N is the number of equivalents, F is a constant, and E is the electromotive force).
3. The determination of heat capacities down to very low temperatures and the utilization of these results together with other thermal data to calculate ΔF by means of the third law of thermodynamics.
4. The combination of suitable chemical equations with known ΔF values to give the reaction in question.

As indicated in a previous section of this chapter, the calculation of the free energy change by the first of the above methods involves the equilibrium constant and, accordingly, for precise work a knowledge of the activities of the reacting molecules. It is also necessary to be certain that the reaction under consideration is a reversible one and that equilibrium is attained within a reasonable time. These conditions limit the applicability of this method of evaluating ΔF.

The second of the above methods will be considered in the chapters on electromotive force cells and on oxidation-reduction.

A consideration of the third method is beyond the scope of this book.

The fourth method, the combining of suitable chemical equations with known ΔF values, is available to everyone, and we shall now attempt to illustrate it. For example, consider the burning of palmitic acid to CO_2 and H_2O. We have

$$C_{16}H_{32}O_2 + 23O_2 \rightarrow 16CO_2 + 16H_2O$$

Looking up in Table 2 the free energy of formation of palmitic acid, CO_2 and H_2O, we have

$$C_{16}H_{32}O_2 \rightarrow 16C + 16H_2 + O_2 \qquad \Delta F_0 = 80{,}000 \text{ calories}$$

$$16C + 16O_2 \rightarrow 16CO_2 \qquad \Delta F_0 = 16(-94{,}450) = -1{,}511{,}200 \text{ calories}$$

$$16H_2 + 8O_2 \rightarrow 16H_2O \qquad \Delta F_0 = 16(-56{,}690) = -907{,}040 \text{ calories}$$

Adding these three chemical equations together, we have the equation for the burning of palmitic acid. Adding also the free energy changes of these reactions, we have the free energy change involved in the combustion of palmitic acid, which turns out to be $-2{,}338{,}240$ calories per mole. Evidently the burning of palmitic acid is thermodynamically a spontaneous process and will proceed to completion with a rich yield of energy.

Usually free energy tables give the free energy of formation of compounds in the solid, liquid, or gaseous states. The free energy of formation means the free energy of the hypothetical reaction whereby the compound in question is synthesized from the elements in their standard state. The free energy of the elements in their standard state is arbitrarily set equal to zero so that, in the reaction

$$16C + 16H_2 + O_2 \rightarrow C_{16}H_{32}O_2 \text{ (palmitic acid)}$$

which evolves 80,000 calories of free energy, we blame the entire energy transformation on the palmitic acid and call it the free energy of formation of palmitic acid.

It is to be remembered that the free energy change of any reaction is equal to the algebraic sum of the free energies of all the reactants involved. The reactants on the left side of the chemical equation have their signs changed previous to summation; those on the right side of the equation remain unchanged.

In order to find the free energy of a compound in solution at unit activity (standard solution state), we must add to the free energy of formation of the compound its free energy of dilution from its saturated solution to unit activity. Frequently, the activities of the substance in solution are not known, and, as a first approximation, the concentrations may be used instead of activities. In dealing with a substance miscible with water in all proportions, a convenient point of departure is the substance in the pure state. Its mole fraction in this state is unity. The free energy change for the transfer of 1 mole of the solute from this state to a very large volume of an aqueous solution in which 1 mole is dissolved in 1,000 grams of water is (for an ideal solution)

$$\Delta F = RT \ln 0.0177 \qquad 27$$

If the substance ionizes in solution, we must also include the free energy of ionization. It is evident that free energy calculations for reactions in solution can be rather involved.

Table 2 is a collection of free energy data from the indicated sources. Those free energies marked (*aq*) refer to a hypothetical 1 molal ideal solution or, in other words, an aqueous solution in which the activity of the solute is unity. Those marked (*g*) refer to a gas at standard conditions; those marked (*s*) refer to solid, crystalline conditions.

TABLE 2

FREE ENERGIES OF FORMATION AT 25° C. AND AT ATMOSPHERIC PRESSURE

Name	ΔF_0	Reference
Acetaldehyde		
(*g*)	−32,000	4
(*l*)	−31,880	4
Acetic acid		
(*g*)	−91,230	4
(*l*)	−94,500	4
(*aq*)	−96,210	4
acetate (*aq*)	−89,720	4
Acetone		
(*g*)	−36,500	4
(*l*)	−37,220	4
d-Alanine		
(*s*)	−88,780	7
(*aq*)	−89,130	7
cation (*aq*)	−92,320	7
anion (*aq*)	−75,910	7
Ammonia		
NH_3 (*g*)	−3,940	6
NH_3 (*aq*)	−6,330	6
NH_4OH (*aq*)	−62,990	6
NH_4^+ (*aq*)	−18,960	6
n-Amyl alcohol (*l*)	−39,100	4
n-Amyl butyrate	−74,900	4

[6] F. H. MacDougall, *Thermodynamics and Chemistry*, 3rd ed., John Wiley and Sons, Inc., New York, 1939.

[7] H. Borsook and H. M. Huffman, *The Chemistry of the Amino Acids and Proteins*, edited by C. L. A. Schmidt, Charles C. Thomas, 1938.

TABLE 2—*Continued*

Name	ΔF_0	Reference
l-Aspartic acid		
(*s*)	−175,440	7
(*aq*)	−172,890	7
cation (*aq*)	−175,480	7
monovalent anion (*aq*)	−167,940	7
divalent anion (*aq*)	−154,990	7
n-Butyric acid (*l*)	−91,500	4
n-Butyl alcohol	−40,400	4
Calcium ion (*aq*)	−132,700	6
Carbon dioxide (*g*)	−94,450	6
Carbon monoxide (*g*)	−33,010	6
Carbonic acid		
(*aq*)	−148,810	7
HCO_3^- (*aq*)	−140,490	6
CO_3^{--} (*aq*)	−126,390	6
Chloride ion (*aq*)	−31,330	6
Creatinine		
(*s*)	−7,550	7
(*aq*)	−7,560	7
cation (*aq*)	−14,160	7
Creatine		
(*s*)	−121,810	7
(*aq*)	−120,560	7
cation (*aq*)	−124,200	7
l-Cysteine		
(*s*)	−82,480	7
(*aq*)	−81,630	7
cation (*aq*)	−83,960	7
anion (*aq*)	−70,270	7
l-Cystine		
(*s*)	−166,630	7
(*aq*)	−162,080	7
monovalent cation (*aq*)	−164,880	7
divalent cation (*aq*)	−166,300	7
monovalent anion (*aq*)	−151,170	7
divalent anion (*aq*)	−137,190	7
Diethyl ether		
(*g*)	−28,090	4
(*l*)	−28,300	4
Dulcitol (*S*)	−223,100	4

TABLE 2—*Continued*

Name	ΔF_0	Reference
Ethyl alcohol		
(*g*)	−38,690	4
(*l*)	−40,200	4
(*aq*)	−41,850	4
Ethyl acetate		
(*g*)	−76,360	4
(*l*)	−77,600	4
Ethyl *n*-butyrate (*l*)	−76,000	4
Formaldehyde		
(*g*)	−26,100	4
(*aq*)	−31,020	4
Formic acid		
(*g*)	−82,520	4
(*l*)	−85,300	4
(*aq*)	−88,110	4
Fumaric acid		
(*s*)	−156,700	4
(*aq*)	−154,800	4
monovalent anion	−150,660	4
divalent anion	−144,620	4
d-α-Glucose (*s*)	−215,800	4
d-β-Glucose (*s*)	−215,400	4
d-Glucose (*aq*)	−217,020	4
d-Glutamic acid		
(*s*)	−174,800	7
(*aq*)	−172,500	7
cation (*aq*)	−175,360	7
monovalent anion (*aq*)	−166,950	7
divalent anion (*aq*)	−154,030	7
Glycerol (*l*)	−113,600	4
Glycine		
(*s*)	−88,920	7
(*aq*)	−89,570	7
cation (*aq*)	−92,720	7
anion (*aq*)	−76,350	7
n-Hexyl alcohol (*l*)	−38,100	4
Hippuric acid		
(*s*)	−90,440	7
(*aq*)	−88,140	7
anion (*aq*)	−82,940	7

TABLE 2—*Continued*

Name	ΔF_0	Reference
Hydrocyanic acid		
(*g*)	28,670	4
(*aq*)	27,280	4
anion (*aq*)	39,140	4
Hydrogen ion (*aq*)	0	
Hydrogen perioxide (*l*)	−28,230	6
Hydroxide ion (*aq*)	−37,585	6
Isopropyl alcohol (*l*)	−44,000	4
d,l-Lactic acid (*l*)	−124,400	4
l-Leucine		
(*s*)	−83,750	7
(*aq*)	−82,800	7
cation (*aq*)	−85,990	7
anion (*aq*)	−69,710	7
Maleic acid (*s*)	−149,400	4
l-Malic acid		
(*s*)	−211,450	4
(*aq*)	−213,570	4
divalent anion (*aq*)	−201,850	4
Mannitol (*s*)	−222,200	4
Methyl alcohol		
(*g*)	−38,890	4
(*l*)	−39,960	4
n-Octyl alcohol (*l*)	−35,100	4
Oxalic acid		
(*s*)	−165,900	4
monovalent anion (*aq*)	−164,380	4
divalent anion (*aq*)	−158,660	4
Palmitic acid		
(*s*)	−80,000	4
(*l*)	−78,560	4
Pentacosane		
(*s*)	13,400	4
(*l*)	15,000	4
Phosphoric acid		
H_3PO_4 (*aq*)	−270,000	6
$H_2PO_4^-$ (*aq*)	−267,100	6
HPO_4^{--} (*aq*)	−257,270	6
PO_4^{---} (*aq*)	−240,970	6

TABLE 2—*Continued*

Name	ΔF_0	Reference
Potassium ion (*aq*)	−67,430	6
Potassium chloride (*s*)	−97,550	6
n-Propyl alcohol (*l*)	−40,900	4
Sodium ion (*aq*)	−62,590	6
Succinic acid		
(*s*)	−178,800	4
(*aq*)	−178,510	4
monovalent anion (*aq*)	−172,780	4
divalent anion (*aq*)	−165,090	4
Sucrose (*s*)	−371,600	4
l-Tyrosine		
(*s*)	−97,640	7
(*aq*)	−94,090	7
cation (*aq*)	−97,130	7
anion (*aq*)	−81,610	7
Urea		
(*s*)	−47,410	7
(*aq*)	−49,010	7
Water		
(*g*)	−54,636	6
(*l*)	−56,690	6

Thermodynamics in Biology

As pointed out in the beginning of the chapter, metabolism studies are of a thermodynamic nature; implicit in them is the assumption of the validity of the law of the conservation of energy.

The application of some of the considerations of thermodynamics to biological systems has certain important limitations. One must be very careful about calculations which assume a state of equilibrium to exist. A point in question is the application of the Donnan equilibrium equations to the general problem of the accumulation of electrolytes by living cells. As shown by Stewart [8] and others, the accumulation of electrolytes by many cells involves the expenditure of energy by the cell and is in no sense an equilibrium condition; and the use of equations based upon the assumption of an equilibrium condition is manifestly unjustified. There is no doubt that, in some respects, living cells may

[8] F. C. Stewart, *Trans. Faraday Soc.*, **33**, 1006 (1937).

be in equilibrium with their surroundings. For example, many cells appear to attain true osmotic equilibrium in a remarkably short time.[9]

It must not be imagined that, even though the free energy change for a reaction which is postulated to occur in a living system is found to be positive, the reaction cannot take place. The positive free energy change may mean that the reaction under consideration is only a part of the total process. For example, we know that ammonia is converted into urea in the body. It has been postulated by Krebs and Henseleit [10] that the synthesis of urea from ammonia takes place through the ornithine, citrulline, arginine, ornithine cycle. The net reaction of this cycle is

$$2NH_3 + CO_2 \rightarrow CO(NH_2)_2 + H_2O$$

The concentrations of the reactants of the above reaction in the body being taken into consideration, it is found that the free energy change for this reaction is plus 13,800 calories. From this calculation we might assume that urea would be converted into CO_2 and NH_3 in the body. This assumption is not in accord with the experimental observations, and, accordingly, we are driven to the conclusion that the Krebs-Henseleit cycle is only part of the total reaction which takes place when urea is formed from ammonia and carbon dioxide.

The question sometimes arises whether or not a living system is able of its own accord to proceed away from a condition of equilibrium instead of towards it, as is demanded by thermodynamics. Suppose, as a matter of argument, we were to assume that a living cell membrane could show one-way permeability to some chemical compound, say glucose; i.e., the glucose could enter the cell by simple diffusion and the cell could, by doing no work, block the exit of glucose molecules. Evidently in time the cell would build up a considerable concentration of glucose within its interior and it could use this excess concentration to do osmotic work although the cell itself had done no work in the process of accumulating the glucose. Such a situation would completely violate thermodynamic calculations made on the cell. Thermodynamics states that a membrane can never show one-way permeability without doing the required osmotic work.

Actually, in those systems which have been investigated, the work done by living cells is only a small fraction of the total work available as estimated from free energy calculations. In this connection Wilson and Peterson [11] have reviewed the energetics of heterotropic bacteria and Baas-Becking and Parks [12] those of autotropic bacteria.

[9] B. Lucke and M. McCutcheon, *Physiol. Revs.*, **12**, 68 (1932).
[10] H. A. Krebs and K. Henseleit, *Z. physiol. Chem.*, **210**, 33 (1932).
[11] P. W. Wilson and W. H. Peterson, *Chem. Revs*, **8**, 427 (1931).
[12] L. G. M. Baas-Becking and G. S. Parks, *Physiol. Revs.*, **6**, 85 (1927).

Wilson and Peterson state that part of the energy of metabolism of heterotropic bacteria is immediately converted into heat because of the irreversible nature of the processes carried out by the bacteria. A portion of the free energy available for work is used for cell growth and synthesis. It appears that most of the free energy (1) is transformed into surface energy, (2) performs osmotic work, (3) displaces equilibrium conditions within the cell, and (4) possibly maintains oxidation-reduction potentials. Eventually, all the energy appears as heat.

Chapter III

REACTION KINETICS

In many respects chemical kinetics is of more importance to the biochemist than thermodynamics. Biochemical processes can, in general, occur in a variety of ways; i.e., the free energy relations permit of a number of possible reactions arising from a given situation. The relative speeds of the various reactions determine which reaction will be the predominating one.

The speed of a reaction will evidently depend upon a number of factors, such as the specific nature of the reacting molecules, the concentrations, the temperature, the presence or absence of appropriate enzymes, and the nature and condition of the medium in which the reaction takes place.

Since reaction kinetics deals with the speed of chemical reactions, it is essential that we understand clearly what we mean by the speed or velocity of a chemical reaction. We must also know in what manner the reaction velocity is to be measured and expressed.

Reaction Velocity

The velocity of a chemical reaction is the instantaneous rate of disappearance of a particular molecular species at any given time. It is possible to express this concept graphically. If the concentration is plotted as ordinate against the time as abscissa, then the slope of the line at any given time is equal to the velocity of the reaction at that time.

According to the mass-action law, the velocity of a chemical reaction is proportional to the product of the concentrations of the reacting molecules. For example, in the reaction

$$A \rightarrow \text{one or more products}$$

we have

$$\text{Velocity of reaction} = \text{Constant} \times \text{Concentration of } A \qquad 1$$

Such a reaction is known as a monomolecular reaction because only one molecular species A is undergoing a change. If two molecules are involved, whether they are two molecules of the same compound or

molecules of two different compounds, the reaction is known as a bimolecular one and may be represented as

$$A + B \rightarrow \text{one or more products}$$

and the velocity of the reaction is

$$\text{Velocity} = \text{Constant} \times \text{Concentration } A \times \text{Concentration } B \qquad 2$$

A knowledge of the reaction velocity as a function of concentration of the reacting molecules frequently gives us valuable information concerning the number of molecules involved in the reaction (order of reaction). We also wish to be in the position to judge the influence of varying temperatures, enzyme concentrations, hydrogen-ion concentration, etc., on the velocity of reactions. The determination and the expression of reaction velocities correctly are, therefore, of fundamental importance.

Reaction velocities may be measured in several different ways. Three of these are:[1]

(*a*) The evaluation of the velocity constants (constants in equations 1 and 2).

(*b*) The determination of the amount of reactant changed in a given time during the first part of the reaction.

(*c*) Comparison of the reciprocals of the times required to effect a given change.

From equations 1 and 2, we see that for a monomolecular reaction

$$K = \frac{\text{Velocity}}{\text{Concentration } A} \qquad 3$$

and for a bimolecular reaction

$$K = \frac{\text{Velocity}}{\text{Concentration } A \times \text{Concentration } B} \qquad 4$$

Evidently, in a monomolecular reaction the velocity constant is equal to the velocity of the reaction when the concentration is unity, whereas for a bimolecular reaction the velocity constant is equal to the velocity of the reaction when the product of the concentrations of the reacting molecules is unity (1 mole per liter). For this reason these velocity constants are called the specific-velocity constants. They may also be expressed as the actual number of molecules reacting per unit time when the concentration is unity (1 mole per liter). Whichever way they are expressed, they may be considered a true measure of the reaction velocity, and it is entirely appropriate to use them in comparing the effect of various factors on the reaction velocity.

[1] O. Bodansky, *J. Biol. Chem.*, **120**, 555 (1937).

It is not always possible, however, to formulate the kinetics of a given reaction. In short, it is not always possible to find a velocity constant which is truly constant and independent of time and concentration. We must distinguish two kinds of deviations from constancy: (1) that due to unavoidable experimental error, and (2) that due to incorrect formulation of the reaction kinetics. If the deviations are haphazard and exhibit no evident relation to time and concentration, it may be assumed that the deviations are due to experimental error. On the other hand, if the velocity "constant" shows a regular variation with time and concentration, the deviations may be due to incorrect formulation of the kinetics of the reaction or to the presence of a constant error. If the deviations are due to experimental errors, the constants may still be used to compare reaction velocities with the realization that they are subject to an experimental error, the magnitude of which can be estimated from a statistical study.[2]

Order of Reaction

It is important to realize that, before we can express reaction velocities in terms of velocity constants, we must first be able to formulate the kinetics of the particular reaction. This formulation is commonly accomplished by trial-and-error method. If the differential equation 1 is integrated and converted to logarithm to the base 10, we have for a monomolecular reaction that

$$\log \frac{C_0}{C} = \frac{Kt}{2.3} \qquad 5$$

where C_0 is the initial concentration, and C is the concentration at any time t. If the reaction in question follows the kinetics of a monomolecular reaction, the plot of $\log C_0/C$ against the time t will yield a straight line whose slope is $K/2.3$.

It should be noted from equation 5 that the magnitude of K is independent of the manner in which the concentration is expressed. K has the dimensions of reciprocal time and is usually expressed in reciprocal seconds. Actually, most, if not all, biochemical reactions which follow the kinetics of a monomolecular reaction are really bimolecular. If, in a bimolecular reaction, one of the molecular species is in great excess, so that a decrease in its concentration is relatively slight, as compared to that of the other molecular species which is at a much lower concentration, the velocity of the reaction will be principally dependent upon the concentration of the molecular species at the lower concentration, and,

[2] L. S. Kassel, *The Kinetics of Homogeneous Gas Reactions*, The Chemical Catalog Co., Inc., New York, 1932.

accordingly, the reaction will appear to be a monomolecular one. The acid hydrolysis of sucrose is such a reaction. In spite of the fact that the reaction is really bimolecular, it shows the kinetics of a monomolecular one. The velocity constant calculated on the basis of a monomolecular reaction is, however, a true measure of its velocity. All reactions which follow the kinetics of a monomolecular reaction are known as first-order reactions regardless of how many species of reacting molecules are involved.

If equation 2, which is a differential equation expressing the velocity of a bimolecular reaction, is integrated, we have

$$t = \frac{2.3}{K(a-b)} \log \frac{b(a-x)}{a(b-x)} \qquad 6$$

where a is the initial concentration of A, and b that of B, and x is the amount of substance decomposed in time t. If the reaction follows the kinetics of a bimolecular reaction (second-order reaction), the plot of $\log (a-x)/(b-x)$ against t will yield a straight line. If a is equal to b or if two molecules of the same substance combine, then

$$\frac{1}{A} = Kt + D \qquad 7$$

where A is the concentration of A at any time t, and D is the integration constant. Under these circumstances the plot of $1/A$ against t will yield a straight line. It must be pointed out, however, that the plot of the reciprocal of the concentration of a molecular species against time is not a general test for second-order reactions. It is a proper test only if the given molecular species is reacting with itself or if the initial concentrations of two molecular species are equal.

Examination of equations 6 and 7 shows that, in second-order reactions, the value of K is not independent of the manner in which the concentrations are expressed. K for such reactions has the dimensions of L^3/t and is usually expressed in liters per mole per second.

Enzymatic reactions as well as surface reactions frequently exhibit, over at least a part of their course, the kinetics of what is known as a zero-order reaction. In such a reaction the velocity is independent of the time and of the concentration, and

$$\text{Velocity} = -K \qquad 8$$

This behavior is usually due to a saturation of the surface (or enzyme) at which the reaction is taking place by the reactants; the further increase of the concentration of the reactants cannot increase the number

of molecules which are active in the chemical reaction, and hence the velocity is independent of concentration.

Although the order of a reaction can frequently be assigned by a simple inspection of the graphs of the proper functions of the concentrations of the reactants against time as described above, at times a more quantitative formulation is desired. The most probable order of a reaction can be determined by means of statistics as Lauffer and Price [3] have done in their study of the kinetics of the thermal denaturation of tobacco mosaic virus protein. Their method is as follows: The best straight line is calculated by the method of least squares for the cases in which C, $\ln C$, $\ln \frac{(a-x)}{(b-x)}$, and $\frac{1}{C}$ are plotted against time. It will be recalled that the method of least squares involves the evaluation of the constants a and b in an equation of the type

$$Y = aX + b \qquad 9$$

These constants are calculated by means of the following two statistical equations:

$$a = \frac{\Sigma(X)\Sigma(Y) - N\Sigma(XY)}{[\Sigma(X)]^2 - N\Sigma(X^2)} \qquad 10$$

and

$$b = \frac{\Sigma(X)\Sigma(XY) - \Sigma(X^2)\Sigma(Y)}{[\Sigma(X)]^2 - N\Sigma(X^2)} \qquad 11$$

where N is the number of experimental points and the symbol Σ means that all the values of the indicated functions are to be added together. The standard error of estimate is then calculated for each assumed order of the reaction by the equation

$$\text{Standard error} = \frac{\Sigma(Y - Y')^2}{(N-2)\Sigma(\bar{X} - X)^2} \qquad 12$$

where Y is the calculated value of C, $\ln C$, $\ln \frac{(a-x)}{(b-x)}$, or $\frac{1}{C}$ as the case may be; Y' is the observed value; N as before is the number of observations; $\bar{X}$ is the mean time; and X is the observed time. The slopes of the best straight lines are divided by their standard errors, and the greater the magnitude of this statistic the better the fit of the data to the straight line in question. A choice of the most probable order of the reaction can then be made immediately.

[3] M. A. Lauffer and W. C. Price, *J. Biol. Chem.*, **133**, 1 (1940).

As suggested before, the formulation of the kinetics of a reaction is not always practical, and we wish, therefore, to examine the two other ways mentioned above for expressing reaction velocities.

Some workers take the amount of reactant changed in a given time during the first part of the reaction as a measure of reaction velocity. Provided that the time interval is short, i.e., the amount of decomposition is small, this method is acceptable. In a first-order reaction or a bimolecular reaction which involves two different molecules whose initial concentrations are unequal, the ratio of the two velocity constants under different experimental conditions is equal not to the ratio of the two concentrations after a given time but to the ratio of a logarithmic function of their concentrations. If, therefore, the extent of reaction is appreciable, a comparison of the velocities in these cases and by this method will lead to a significant error.

The third method which has been used as a measure of reaction velocity is the determination of the reciprocal of the time required to effect a given amount of chemical change. If the amount of reactant changed is kept constant, then, no matter how complicated the reaction kinetics may be, we will have

$$K = \frac{M}{t} \qquad 13$$

where M is the term involving the concentration changes and which is by definition constant. Therefore, under different experimental conditions

$$\frac{K_2}{K_1} = \frac{1/t_2}{1/t_1} \qquad 14$$

where K_1 and K_2 are the velocity constants of a reaction under two different conditions and t_1 and t_2 are the times required to bring about a given chemical change. The reciprocals of the times required for a given change are, therefore, a true measure of the respective rates of reaction, and no knowledge of the reaction kinetics is necessary. The only assumption implicit in the argument is that, whatever the form of the true (but unknown) kinetic relation, it must be the same under the several conditions of the experiment; i.e., this method fails if the reaction changes its order from one condition to another.

The preceding discussion of reaction kinetics is a highly streamlined and simplified description. Many complicating factors have not been dealt with, such as opposing reaction, side reactions, and consecutive reactions. To enter into an exposition of all such complications would involve more time and space than we have at our disposal.

Having shown how the velocity of a reaction might be expected to vary with concentration of the reactants and having outlined the methods of measuring reaction velocity, we now wish to consider the effect of temperature on a chemical reaction.

Effect of Temperature

Increase of temperature generally increases the velocity of a chemical reaction, and the temperature variation is usually expressed in terms of what is known as the Q_{10}. The Q_{10} is the ratio of the speed of reaction at one temperature to that at a temperature 10° lower. As a rule, the Q_{10} is in the neighborhood of 2. It is, however, not universally true that increase of the temperature increases the speed of a reaction. For example, the speed of the urea denaturation of egg albumin actually decreases with increasing temperature.[4] This is also true of the surface coagulation of egg albumin.[5] The speed of all reactions catalyzed by enzymes decreases above a sufficiently high temperature. This is due to the destruction of the enzyme at higher temperatures.

It was noted by Arrhenius [6] that, if the logarithm of the velocity constant of a reaction is plotted against the reciprocal of the absolute temperature, a straight line is obtained. The equation for such a relation is

$$\log K = -\frac{G}{T} + D \qquad 15$$

where G and D are constants.

At this point we wish to point out a certain confusion in the use of the term "temperature coefficient of a reaction."

If the above equation of Arrhenius is differentiated and rearranged we have

$$\frac{1}{K}\frac{dK}{dt} = \frac{G}{T^2} \qquad 16$$

$\frac{1}{K}\frac{dK}{dt}$ is the correct expression for the temperature coefficient of a reaction, i.e., the increase of velocity per unit velocity per degree rise in temperature. It is evident from equation 16 that, since G is a constant, the temperature coefficient of a reaction will diminish with increasing temperature and the rate at which it will diminish will depend upon the value of G. It is, however, the custom among chemists to call the ratio of the velocities at two temperatures 10° apart the temperature coefficient (Q_{10}).

[4] G. F. Hopkins, *Nature*, **126**, 383 (1930).
[5] H. B. Bull, *Cold Spring Harbor Symposia Quant. Biol.*, **VI**, 140 (1938).
[6] S. Arrhenius, *Z. physik. Chem.*, **4**, 226 (1889).

These two ways of expressing the temperature dependence are by no means equivalent. For example, with a reaction in which G is 4,500 at 30° C., the true temperature coefficient $\left(\frac{1}{K}\frac{dK}{dt}\right)$ is 4.9×10^{-2}, while the ratio of the velocity of the reaction at 35° C. to that at 25° C. is 3.1. The true temperature coefficient can have positive or negative values, depending on whether the speed of reaction increases or decreases with the temperature, whereas the value of the Q_{10} may be less than 1 but never less than zero.

The constant G in the Arrhenius equation has been given a definite physical meaning. It is related to the energy of activation. Equation 15 can be written

$$\log K = -\frac{E}{2.3RT} + B \qquad 17$$

and between the limits of two temperatures T_1 and T_2

$$\log \frac{K_2}{K_1} = \frac{E}{2.3R}\left(\frac{T_2 - T_1}{T_2 T_1}\right) \qquad 18$$

The constant E, called the energy of activation, represents the minimum energy which the reacting molecules must have before they will react.

A consideration of the energy of activation helps us to understand the rather remarkable observation that a 10° rise in temperature can double and sometimes more than double the velocity of the reaction. The number of molecules with an energy content greater than the energy of activation can be calculated from the integrated Maxwell-Boltzmann distribution equation, which is

$$\frac{n_1}{n_0} = e^{-E/RT} \qquad 19$$

where n_1 is the number of molecules with an energy in excess of E, and n_0 is the total number of reactant molecules present. If the energy of activation is assumed to be 12,000 calories and the temperature to be 300° K. (under these conditions the velocity of the reaction would be exactly doubled in going from 22° C. to 32° C.) and we substitute these data in equation 19, we have

$$\frac{n_1}{n_0} = e^{-12{,}000/(2\times 300)} = 2.06 \times 10^{-9}$$

while at 310° K.,

$$\frac{n_1}{n_0} = e^{-12{,}000/(2\times 310)} = 3.93 \times 10^{-9}$$

Thus, whereas the average kinetic energy of the molecules has been increased only 3 per cent (10/300) by the 10° increase in temperature, the fraction of molecules with energy greater than 12,000 calories has practically doubled.

Collision Theory and Activation

On the basis of the collision theory of chemical reactions, the absolute velocity of a chemical reaction is given by the number of molecules reacting per second, which, in turn, equals the number of molecules colliding per second multiplied by the chance that the colliding particles have sufficient energy to react.

The Arrhenius equation 17 may be expressed in the following form:

$$\ln K = -\frac{E}{RT} + \ln Z \qquad 20$$

where $\ln Z$ is equal to the integration constant B. Rearranging equation 20 we have

$$\ln \frac{K}{Z} = -\frac{E}{RT} \qquad 21$$

and converting this equation to the exponential form

$$K = Ze^{-E/RT} \qquad 22$$

From what has been said, $e^{-E/RT}$ is evidently the chance that the colliding particles will have sufficient energy for a reaction to occur ($e^{-E/RT}$ is the ratio of the activated molecules to the total number of reactant molecules present), Z is then the number of molecules colliding per second in unit volume, and K is now the actual number of molecules reacting per second per unit volume. In a monomolecular reaction, Z is to be regarded as the frequency of vibration of the activated bonds in the reacting molecules and is approximately the same for all monomolecular reactions (of the order of 10^{13}).

Sometimes it is possible to obtain Z for a bimolecular reaction by an independent calculation from molecular kinetic theory. This value can then be compared with the Z value found from the velocity constant of the reaction and the energy of activation. In a number of instances the two methods yield results of the same order of magnitude, which indicates a basic reality to this interpretation of Z. The agreement of the two values of Z so calculated means, of course, that every collision of the activated molecules is fruitful in producing a chemical reaction. In reactions in which the two methods yield divergent results, steric factors are considered responsible for the lack of agreement. For example,

it is believed that the activated molecules must strike each other in some definite position, and, accordingly, only a fraction of the activated collisions are effective in producing a reaction. Then equation 18 is written

$$K = PZe^{-E/RT} \qquad 23$$

where P is now a steric factor which may have any value from unity to zero, depending upon the particular reaction under consideration.

Consider the specific example of the hydrolysis of sucrose by saccharase.[7] In this particular experiment the concentration of sucrose was 5.0 grams per 100 cc. and that of the saccharase was 2.5×10^{-5} gram of pure saccharase per 100 cc. The molecular weight of saccharase (from diffusion experiments) is about 50,000. The temperature was 310° absolute. This information gives us the opportunity of calculating the number of collisions between the sucrose and saccharase molecules occurring in one cubic centimeter of solution per second. We can substitute this information in the gas collision formula

$$Z = n_1 n_2 \left(\frac{\sigma_1 + \sigma_2}{2}\right)\left[8\pi RT\left(\frac{1}{M_1} + \frac{1}{M_2}\right)\right]^{1/2} \qquad 24$$

where n_1 is the number of molecules of sucrose per cubic centimeter (8.86×10^{19}), n_2 is the number of molecules of saccharase per cubic centimeter (3.03×10^{12}), σ_1 is the diameter of the sucrose molecule (9.0×10^{-8} cm.), σ_2 is the diameter of the saccharase molecule (5.5×10^{-7} cm.), R is the gas constant (8.31×10^7 ergs per degree), T is the absolute temperature (310°), M_1 is the molecular weight of sucrose (342), and M_2 is the molecular weight of saccharase (50,000). From this calculation Z is found to be equal to 1.21×10^{24} collisions per cubic centimeter per second. The number of effective collisions is obtained by multiplying Z by $e^{-E/RT}$. E has a value of 8,700 calories, and, therefore, the fraction of the collisions in which the energy is greater than 8,700 calories is $e^{-8{,}700/(2\times 310)}$ or 7.0×10^{-7}. This figure multiplied by Z gives 8.47×10^{17} as the total number of activated collisions per cubic centimeter per second. The observed value of K (specific reaction rate) is 2.1×10^{-4} per second. The number of molecules decomposing per cubic centimeter per second is, therefore,

$$\frac{0.05 \times 6.06 \times 10^{23} \times 2.1 \times 10^{-4}}{342} = 1.86 \times 10^{16}$$

Comparison of this value with the number of activated collisions (8.47×10^{17}) shows that only 1 in 46 activated collisions results in a

[7] P. W. Wilson, *Respiratory Enzymes*, Burgess Publishing Co., Minneapolis, 1939.

reaction. The interpretation of this result is that only $\frac{1}{46}$ of the surface area of the saccharase is active.

The electrostatic charges on the activated molecules are also believed to cause a departure of P from unity, and equations have been derived which enable one to calculate P as a function of the valence and radii of the reacting ions.[8]

Transition State Theory

There is, beside the collision theory outlined above, another approach to the whole problem of kinetics which involves a consideration of an activated complex. Eyring and Stearn [9] and collaborators have been most active along this line. The two methods of approach do not appear mutually exclusive; both have their advantages. The collision theory has its grounding in ideas that are simple and familiar to most scientific workers, and it involves calculations that are easy to make. The transition-state theory, on the other hand, involves basic ideas that are foreign to the experience of most biochemists, although the general picture which is presented is not a difficult one to grasp. One advantage of this type of approach is that all kinds of reactions are dealt with in a unified manner, whereas, on the basis of the collision theory, bimolecular and monomolecular reactions are treated differently.

The picture which Eyring and co-workers have given for reaction kinetics is about as follows: Before molecules can react they must pass through a configuration known as the activated state which has an energy content greater than that of the normal reactants. The average of this energy increment is the activation energy. The plot of the potential energy for various configurations shows that this activated state corresponds to a saddle point or pass between two hills. The pass leads to the decomposition of the activated configuration or, as it is called, the activated complex. Figure 1 shows a diagrammatic representation of a cross section of such a pass B leading from one energy valley A, which represents the normal state of the reactants, into a second energy valley C, which is the energy level of the products of the reaction.

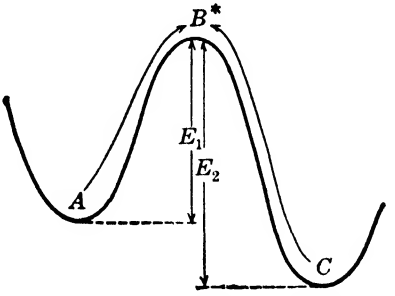


FIG. 1. Energy pass of a reaction.

[8] E. A. Moelwyn-Hughes, *Proc. Roy. Soc.* (*London*), (A)**155**, 308 (1936).

[9] A. E. Stearn, *Ergeb. Enzymforsch.*, VII, 1 (1938).
H. Eyring and A. E. Stearn, *Chem. Revs.*, **24**, 253 (1939).

In Fig. 1, E_1 represents the activation energy of the forward reaction and E_2 that of the opposing reaction.

The molecules pass from the normal state A to the activated state B and back again to the normal state A. Therefore, between the normal reactants and the activated complex, an equilibrium is established, which, however, is being continually disturbed by the activated complex "spilling" over into C.

To calculate the reaction velocity constant, the concentration of the activated complex in equilibrium with unit concentration of the reactants is multiplied by the rate at which the activated complex is crossing the pass B and spilling down into C. The velocity constant K is then

$$K = K^* \frac{kT}{h} \qquad 25$$

where K^* is the equilibrium constant between the normal reactants and the activated complex; kT/h is a frequency which represents the rate of passage of the activated complex through the pass and into C; k is the Boltzmann constant, equal to 1.37×10^{-16} erg per degree; h is Planck's constant, equal to 6.54×10^{-27} erg seconds; and T is the absolute temperature. K^*, the equilibrium constant, is related to the heat change, free energy change, and entropy change of activation in much the same manner as the equilibrium constant of any reaction is related to these thermodynamic quantities. We have, for example,

$$RT \ln K^* = T\Delta S^* - \Delta H^* = -\Delta F^* \qquad 26$$

and, accordingly,

$$RT \ln \frac{Kh}{kT} = T\Delta S^* - \Delta H^* = -\Delta F^* \qquad 27$$

ΔS^*, ΔF^*, and ΔH^* are the standard entropy, standard free energy, and heat change for the reaction in which the activated complex is formed from the reactants.

It turns out that ΔH^* is related to the energy of activation E as calculated by Arrhenius's equation in the following way

$$\Delta H^* = E - RT \qquad 28$$

It is evident that, for the usual values of the energy of activation and at lower temperatures (physiological temperatures), the energy of activation of Arrhenius is practically equal to the heat change of the activation reaction.

It is interesting to compare the collision treatment of the energy of activation with that of the transition-state treatment. In the exponential form equation 27 can be written

$$K = \frac{kT}{h} e^{\Delta S^*/R} \times e^{-\Delta H^*/RT} \qquad 29$$

At lower temperatures and appreciable energies of activation ΔH^* and E are practically equal and we can write equation 29 (approximately) as

$$K = \frac{kT}{h} e^{\Delta S^*/R} e^{-E/RT} \qquad 30$$

Comparing equation 23 with equation 30, we see that

$$\frac{kT}{h} e^{-\Delta S^*/R} = PZ \qquad 31$$

since Z and kT/h have the dimension of frequency and $e^{\Delta S^*/R}$ and P are dimensionless,

$$\frac{kT}{h} = Z \qquad 32$$

and

$$P = e^{-\Delta S^*/R} \qquad 33$$

The so-called steric factors of the collision theory are, therefore, related to the entropy change of the activation reaction.

From equation 27 we see that the velocity of a reaction is dependent upon the free energy change of the activation reaction and that neither the heat change alone (energy of activation of Arrhenius—approximately) nor the entropy change of the activation reaction alone will give a fair measure of the velocity of reaction.

Eyring and Stearn have applied the quantum-statistical mechanical treatment to protein denaturation with interesting results. As is well known, if solutions of many proteins are heated or treated with a variety of reagents (urea, alcohol, strong acid, strong base, etc.), the protein undergoes an intramolecular change known as protein denaturation, the most characteristic results of this reaction being (1) decreased solubility, (2) appearance of sulfhydryl groups, and (3) increased viscosity of the solution of the protein molecules. In general, heat denaturation of proteins exhibits enormous temperature coefficients which yield correspondingly large values for the energy of activation as Table 1 shows.

TABLE 1

ENERGIES OF ACTIVATION OF PROTEIN DENATURATION BY HEAT

(Eyring and Stearn)

Protein	Temperature, ° C.	K sec^{-1}	E
Insulin	80	1×10^{-4}	35,600
Pepsin	25	5.47×10^{-4}	63,500
Egg albumin (hens)	65	2.54×10^{-4}	132,100
Hemoglobin	60	4.3×10^{-4}	75,600

Eyring and Stearn [9] have calculated the ΔF^*, ΔH^*, and ΔS^* changes associated with the denaturation reactions of a number of proteins. In order to do this, they made use of equation 27.

It would take us too far afield and involve a discussion of protein structure to describe in detail their most interesting results and conclusions. It can be said, however, that ΔF^* gives a direct measure of the speed of reaction, while ΔS^* gives a measure of the degree of randomness involved in the activation. ΔH^* can be roughly interpreted in terms of the number of bonds which have to be broken in order to form the activated complex.

The kinetics of pepsin denaturation was investigated by Steinhardt [10] in considerable detail. He concludes from his experimental data that, before denaturation of pepsin can occur, 5 protons (hydrogen ions) must be ionized off the pepsin molecule and that a large fraction of the apparent energy of activation is really due to the heat required to dissociate off the 5 protons. This heat of dissociation for pepsin amounts to 45,000 calories, which leaves only 18,300 calories as the energy required to break the intramolecular bonds in the activation of pepsin.

Q_{10} *in Biology*

The activity of life is greatly influenced by the temperature of the environment; in general, life processes increase their speed with increasing temperature. Attempts have been made to analyze this temperature dependence in terms of the Arrhenius equation 17. It is believed by Crozier and his associates that the speed of a given process (such as the rate of beat of insects' wings, for example) is controlled by a master reaction (slowest reaction) and that the change in physiological activity

[10] J. Steinhardt, The Stability of Crystalline Pepsin, *Kgl. Danske Videnskab. Selskab. Math.-fys. Medd.*, XIV, II (1937).
V. K. La Mer, *Science*, 86, 614 (1937).

with temperature is a reflection of the change in the velocity of the master reaction. Not infrequently the energy of activation of a living process is not constant but varies with temperature. The explanation advanced by Crozier is that the living process is made up of a series of reactions and as the temperature is changed a different reaction of the series becomes the slowest, which results in an altered energy of activation.

Crozier has found that the energies of activation for life processes have a tendency to cluster about certain values, i.e., 8,000, 12,000, and 18,000.

The interpretation of Crozier has been severely criticized by certain workers.[11, 12] For example, Ponder and Yeager state that (1) "If there are several reactions involved in bringing about the total response, it is mathematically demonstrable that the slope of the line (or curve) is not determined by the temperature coefficient of any one of them alone; an activation value of 10,000, say, derived from this line, does not indicate that any one of the individual reactions has this temperature coefficient, or even that the average energy of activation of the reaction is 10,000. (2) The fact that the entire system appears to follow the Arrhenius equation does not even necessarily indicate that any one of the underlying reactions does so; further, if a series of values are obtained for various parts of the temperature range, there is no reason to suppose (*a*) that each new slope corresponds to a new reaction or (*b*) that the values obtained for various slopes have any direct relation to the temperature coefficients of the various underlying reactions in the system under consideration."

It would seem best, in physiological work dealing with the effects of temperature on various processes, to express the results in terms of Q_{10} rather than as an energy of activation. This would avoid any implication concerning the physical meaning of such experimental work.

Enzymatic Activation

Enzymes are a group of organic compounds of a highly specific nature whose function is to direct biochemical reactions into useful channels. A number of them have been isolated and purified to a sufficient extent to obtain crystals. All enzymes so far prepared in a pure state are proteins, either simple or conjugated.

In common with all true catalysts, enzymes do not and cannot influence the equilibrium point of a reaction; they serve only to decrease the

[11] L. V. Heilbrunn, *An Outline of General Physiology*, W. B. Saunders Co., Philadelphia, Pa.

[12] E. Ponder and J. F. Yeager, *J. Exptl. Biol.*, **7**, 390 (1930).

time required to attain equilibrium. They serve to "lubricate" the reaction.

Since an enzyme increases the speed of reaction in the forward direction, and since the equilibrium constant is unchanged, it follows that speed of the reverse reaction also must be increased. This is clear from the simple relation

$$\frac{K_f}{K_r} = K_e \qquad 34$$

where K_f and K_r are the velocity constants in the forward and in the reverse direction respectively, and K_e is the equilibrium constant of the reaction which is unchanged by the presence of the enzyme. We can say, therefore, that enzymes which have a hydrolyzing action on fats, proteins, etc., must likewise have a synthesizing action if the enzyme is placed in a medium which contains the hydrolyzed products. Some success has been attained in the reversal of enzymatic reactions. The most successful syntheses have been accomplished in the case of alcohols and carboxylic acids under the influence of lipases and esterases.

The kinetics of enzymatic reactions is frequently complicated, and the order of the reaction may change during the course of the reaction. At the beginning of the reaction and at high substrate concentration the reaction may exhibit the kinetics of a zero-order reaction. This stage corresponds to a saturation of the enzyme surface by the substrate. As the reaction proceeds and the substrate concentration decreases, the reaction usually reverts to the kinetics of a first-order reaction. It cannot be too strongly urged that, in comparing the influence of various factors on the velocity of enzymatic reactions, a proper measure of the velocity be chosen. This question has been dealt with in the first part of this chapter.

Figure 2 shows the velocity of conversion of urea to ammonia and carbon dioxide in the presence of ureas as a function of time.[13]

A number of factors influence the velocity of an enzymatic reaction, some of them having profound effects. Among them may be mentioned the substrate concentration, the enzyme concentration, the hydrogen-ion concentration, electrolyte concentration, the presence of activators or inhibitors, and temperature.

The influence of the substrate concentration has been noted above. With increasing concentration of enzyme the reaction rate increases. At low enzyme concentration the increase is directly proportional to the enzyme concentration. As the amount of enzyme is increased above a certain amount, the direct proportionality is lost and the rate increases

[13] D. D. Van Slyke and G. E. Cullen, *J. Biol. Chem.*, **19**, 141 (1914).

more slowly than the enzyme concentration. Let us consider a specific example. The standard method for the estimation of the activity of a proteolytic enzyme is that of Anson.[14] It depends upon the determination of the amount of the amino acid tyrosine liberated from a standard

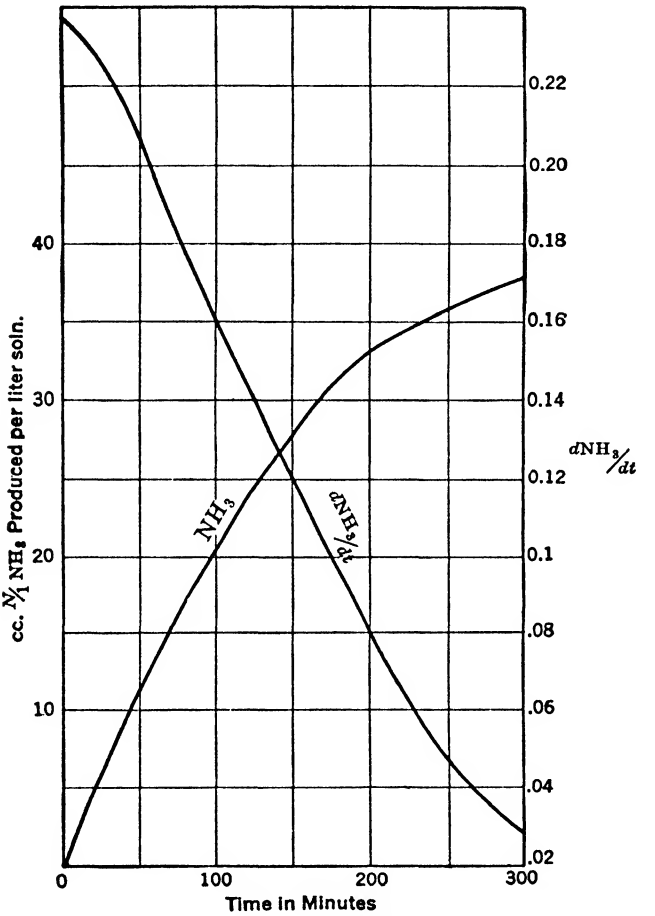


FIG. 2. Velocity of decomposition of urea in the presence of ureas as a function of time in minutes. (Van Slyke and Cullen.)

hemoglobin solution under standard conditions. Shown in Fig. 3 is the amount of tyrosine liberated from hemoglobin by the action of bromelin, the proteolytic enzyme from pineapples as a function of the cubic centimeters of a solution of a given enzyme preparation at 37.5° C. and at *p*H 7.5 in the presence of a veronal buffer.[15]

[14] M. L. Anson, *J. Gen. Physiol.*, **22**, 79 (1938).

[15] J. A. Cooper, unpublished results.

The standard activity curve of bromelin can now be constructed from Fig. 3. A tangent is drawn to the curve from the origin. The tangent is extrapolated to the amount of enzyme which gives 0.001 milliequivalent of tyrosine. This amount by definition contains 1 unit of proteolytic enzyme. The curve in Fig. 3 can then be replotted. The ordinate should

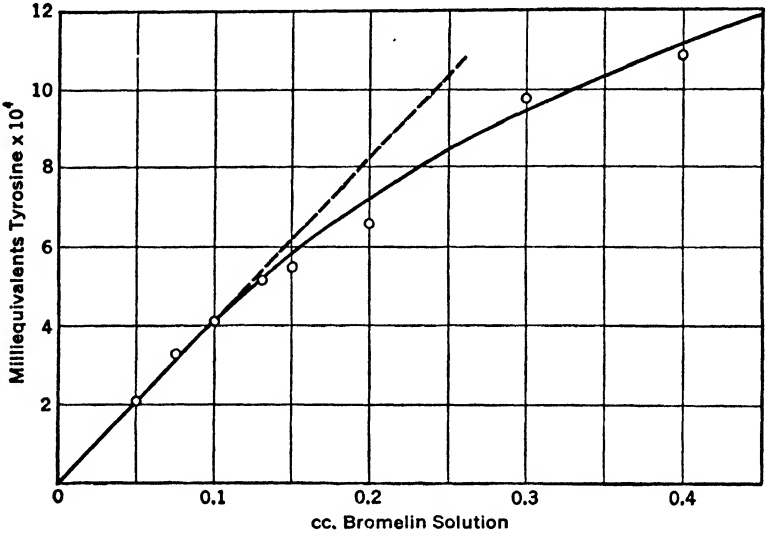


FIG. 3. Digestion of hemoglobin by bromelin under the standard conditions defined by Anson. (Cooper.)

show the milliequivalents of tyrosine and the abscissas the units of bromelin. For those planning work on proteolytic enzymes, Anson's paper should be consulted.

Since enzymes are proteins and have an ampholytic character, it would be anticipated that the hydrogen-ion concentration would have considerable effect on enzyme activity, and this is found to be the case. Each enzyme has its optimum *p*H for maximum activity. With some enzymes this is a very narrow range; with others it is broader. The optimum *p*H may depend upon the type of buffer used, i.e., phosphate, acetate, citrate, etc.

Michaelis and Menten [16] proposed that an enzymatic reaction proceeds by way of a substrate-enzyme intermediate compound. The rate of reaction would, therefore, be limited by (1) rate of complex formation and (2) rate of complex decomposition. This theory may be expressed by

$$\text{Enzyme} + \text{Substrate} \leftrightarrows \text{Complex} \rightarrow \text{Resultants} + \text{Enzyme}$$

[16] L. Michaelis and M. L. Menten, *Biochem. Z.*, **49**, 333 (1913).

The kinetics of such a reaction have been worked out, and the velocity of an enzymatic reaction is given as

$$v = \frac{VS}{K_s + S} \qquad 35$$

where K_s is the dissociation constant of the substrate enzyme complex, S is the substrate concentration, and V is the maximum velocity, i.e., the velocity of the enzymatic reaction when the concentration of the substrate is so high that all the enzyme will be tied up in the intermediate complex.

If the reciprocals of both sides of the Michaelis-Menten equation are taken and the equation rearranged, we have

$$\frac{1}{v} = \frac{K_s}{V}\left(\frac{1}{S}\right) + \frac{1}{V} \qquad 36$$

If $1/v$ is plotted against $1/S$ a straight line is obtained with a slope equal to K_s/V and an intercept equal to $1/V$. This is the method used in evaluating K_s and V from the experimental data.

Lineweaver and Burk [17] have considered various aspects of the Michaelis-Menten theory and have investigated seven complex cases in which this theory is used. They consider, among other things, competitive and non-competitive inhibitors. In competitive inhibition of an enzymatic reaction the inhibitor competes for the enzymes with the substrate, whereas in non-competitive inhibition the inactivation of the enzyme is independent of the substrate concentration. As pointed out above, in the general enzymatic reaction, if $1/v$ is plotted against $1/S$ a straight line is obtained. In competitive inhibition, the slope of the straight line is increased, but the intercept is unchanged; with a non-competitive inhibition the intercept as well as the slope of the line is increased.

The general function of enzymes is to reduce the energy of activation for a given reaction. The enzyme thus "tunnels" under the energy pass between the reactants and the resultants of the reaction. In general, therefore, enzymatic reactions have a lower temperature coefficient than corresponding non-enzymatic reactions. Lineweaver [18] gives the energies of activation of some of the non-catalyzed and catalyzed reactions. (See Table 2.)

It must be pointed out and emphasized that the ΔF^* of the activation reaction controls the speed of reaction. The speed of reaction can,

[17] H. Lineweaver and D. Burk, *J. Am. Chem. Soc.*, **56**, 658 (1934).
[18] H. Lineweaver, *J. Am. Chem. Soc.*, **61**, 403 (1939).

TABLE 2

ENERGIES OF ACTIVATION OF SOME CATALYZED AND NON-CATALYZED REACTIONS (Lineweaver)

Reaction	Catalyst	E
H_2O_2 decomposition	None	18,000
	Colloidal platinum	11,700
	Liver catalase	5,500
Sucrose inversion	H ions	26,000
	Yeast invertase	11,500
Casein hydrolysis	HCl	20,600
	Trypsin	12,000
Ethyl butyrate	H ions	13,200
	Pancreatic lapase	4,200

however, be analyzed in terms of ΔH^* and ΔS^* of the activation reaction. The balance of these two functions, of course, determines the ΔF^* of the reaction in accord with the thermodynamic relation

$$\Delta F^* = \Delta H^* - T\Delta S^* \qquad 37$$

According to Stearn et al.,[19] if one deals with a system involving reactants, catalyst (enzyme), activated complex, and reaction products, two general cases arise:

1. The enzyme is more strongly bound to the reactant than to the activated complex. Here the formation of the activated complex involves the breaking of the enzyme-reactant bond with consequent increase in ΔH^*. The release of the enzyme, however, will cause a compensating increase in ΔS^*.

2. The enzyme is more strongly bound to the activated complex than to the reactant. Here the formation of enzyme-activated complex will lower ΔH^* but the tying up of the enzyme by the activated complex will involve a compensating decrease in ΔS^*. When these two compensating effects balance each other, there will be no change in ΔF^* and, accordingly, no catalytic effect. In case 1, when the effect of ΔH^* predominates, a negative catalyst results. When ΔS^* is predominately affected, positive catalyst follows. In case 2, just the opposite results.

One of the outstanding and interesting features about reactions catalyzed by enzymes is the specificity of the enzyme. In some instances this specificity is absolute, the enzyme acting upon only one particular linkage in one compound. In others the action is more general. In

[19] A. E. Stearn, H. P. Johnston, and C. R. Clark, *J. Chem. Phys.*, **7**, 970 (1939).

order that the enzyme may act there must be a union of some kind between the substrate and the enzyme. This requires that the substrate have a very definite structure and that the active groups of the enzyme, those groups which bring about reaction, be adjacent to the reactive groups of the substrate. The exact action of the enzyme is, no doubt, manifold. For example, the oxidation-reduction enzymes (see Chapter VII) have the property of accepting or donating electrons. It is the feeling of the author that hydrolytic enzymes, such as proteases, lipases, carbohydrases, act by accepting or donating protons, and that the groups in the enzyme which are responsible for the action of the enzyme are probably proton-accepting or -donating groups, such as NH_3^+ and COO^-.

Chain Reactions

In a chain reaction a molecule is activated by contact with the sides of a container, by an enzyme, by absorbed radiation, or by solvent molecules. This "hot" or activated molecule is able to activate another molecule by collision, after which it decomposes; the newly activated molecule is able to repeat this performance. This activating series forms what is known as a chain which may have several thousand links in it before it is broken. If the chains become branched and are not inhibited, an explosion results. A chain reaction bears the following six characteristics: [20]

1. At least one of the stages in the sequence of the reaction must be exothermic (give off heat); accordingly, it is able to activate other molecules.

2. The velocity of a chain reaction is a function of the number of active centers produced per unit time in unit volume, and of the average length of the chain. Therefore, there will be no simple relation between the concentration of the reactants and the instantaneous rate of reaction.

3. Velocity of the reaction may be profoundly altered by the presence of an added foreign substance which breaks the chain.

4. The reaction may involve an induction period during which the chains are forming.

5. If the chains are initiated or terminated on the walls of the vessel, the velocity of the reaction will depend on the size or shape or on the nature of the walls of the vessel.

6. The chains may branch, and unchecked acceleration leads to an explosion.

It is doubtful that living systems make use of such a poorly controlled process as a chain reaction; certainly the kinetics of reactions catalyzed

[20] E. A. Moelwyn-Hughes, *Ergeb. Enzymforsch.*, VI, 23 (1937).

by proteases, carbohydrases, and lipases do not suggest a chain reaction. It is possible that the oxidative enzymes, since they involve an exothermic step, may pass through a chain reaction, although to the author this seems improbable. It is true, however, that the catalase system and the oxidation of luciferin, catalyzed by luciferase, are adequately described by the kinetics of a chain mechanism.[21]

It seems very probable that a chain mechanism is involved in the drying of oils, as well as in oil rancidity.[22] Polymerization of the oil is also thought to be a chain reaction. In fact, the two reactions are probably closely connected. The activation energy received from oxidation may serve to activate molecules for the polymerization reaction. In rancid oils the chains can be broken by small amounts of anti-oxidants, hydroquinone being an example of such an anti-oxidant.

[21] E. N. Harvey, *J. Gen. Physiol.*, **10**, 875 (1927).
[22] J. A. Christiansen, *Trans. Faraday Soc.*, **24**, 714 (1928).

Chapter IV

ELECTROSTATICS AND DIELECTRICS

We shall have occasion to deal with electrostatic phenomena in a number of topics to follow. It is appropriate, therefore, to discuss at this point the first principles of electrostatics together with the methods of measuring, meaning, and results of dielectric-constant studies.

Elementary Electrostatics

It is an old axiom of physics that like electrostatic charges repel and unlike charges attract each other. For example, we have all noticed how the hairs on a cat's back stand apart after they have been stroked on a dry winter day. The hairs repel one another because they carry electrostatic charges of the same sign. This is one example of a physiological response which is due to electrostatic charges. We shall discover that electrostatic interactions are of extreme importance to physiology and biochemistry—far more indeed than keeping cats' hairs apart.

Unit electrostatic charge is defined as that quantity of electricity which will repel, in a vacuum, a like quantity of electricity of the same sign at a distance of 1 centimeter with a force of 1 dyne, the two charges being "point charges." In general, the force between two point charges is given by Coulomb's law, which is

$$\text{Force} = \frac{Q_1 \times Q_2}{Dd^2} \qquad 1$$

where Q_1 and Q_2 are the magnitudes of the two point charges and d is the distance between them. D is the dielectric constant of the medium separating the point charges, whose properties we shall discuss presently. If the charges Q_1 and Q_2 are of unlike sign, the force will be one of attraction rather than repulsion.

The difference in potential between two points A and B will be the work in ergs required to move a unit charge from A to B. This is true irrespective of the nature of the medium or how the process is carried out. If the charges at A and B have opposite signs, work can be obtained from the system. The potential difference between A and B is exactly equal to the free energy required to move the charge from A to B. If

A and B bear the same sign of charge, the free energy change is positive; if opposite, the free energy change is negative.

The general relation between charge and potential is given by

$$Q = C\Phi \qquad 2$$

where Φ is the potential difference, Q is the charge, and C is a constant known as capacity. The magnitude of the capacity depends upon the geometry of the system.

The fundamental relation between the potential and the charge is expressed in a more explicit manner by Poisson's equation, which involves the second differential of the potential with respect to the X, Y, and Z axes. This equation is

$$\frac{d^2\Phi}{dx^2} + \frac{d^2\Phi}{dy^2} + \frac{d^2\Phi}{dz^2} = -\frac{4\pi\rho}{D} \qquad 3$$

where ρ is the net charge density, i.e., the net number of unit charges per cubic centimeter.

As it would be extremely difficult to explain the full meaning of the Poisson equation in the space at our disposal it is better not to make the attempt. To find the relation between the charge and potential in any specific case we integrate the Poisson equation under the specified conditions. Carrying out this integration for two parallel plates, we have

$$\Phi = \frac{4\pi dQ}{DA} \qquad 4$$

where A is the area of the plates and d is the distance between them. Combining equations 2 and 4 we find that the capacity of two parallel conducting plates is

$$C = \frac{AD}{4\pi d} \qquad 5$$

If we integrate Poisson's equation for two conducting concentric spheres, we have

$$\Phi = \frac{Q}{Dr}\left(\frac{d}{r+d}\right) \qquad 6$$

where r is the radius of the inner sphere and d is the distance between the inner and outer spheres.

For two conducting concentric cylinders, we have

$$\Phi = \frac{2Q}{D}\ln\left(1 + \frac{d}{r}\right) \qquad 7$$

where r is the radius of the inner cylinder, d is the distance between the inner and outer cylinder, and Q is the charge per unit length. The relation between charge and potential may be complicated and not always easy to predict.

The force between charged bodies likewise depends on the geometry of the system. As we have seen, for two point charges

$$\text{Force} = \frac{Q_1 \times Q_2}{Dd^2} \qquad 8$$

and between two parallel conducting plates

$$\text{Force} = \frac{2\pi\sigma^2 A}{D} = \frac{\Phi^2 AD}{8\pi d^2} \qquad 9$$

where σ is the charge density per unit area.

Dimensions of Electrostatic Quantities

It is evident that, if the dimensions of the dielectric constant are known, the dimensions of potential and charge can be obtained from equation 9. As a fundamental constant relating potential and charge, the dimensions of the dielectric constant are unknown. If, however, it is regarded simply as the ratio of the force between two charges in a vacuum to that in a given medium, the dielectric constant is, of course, dimensionless. We are thus faced with the anomalous situation that the electrostatic potential, while numerically equal to the free energy of the charging process, may not have the dimensions of energy. For a fuller discussion of the dimensions of electrostatic quantities, see Guggenheim.[1]

The three systems of electrical units in use are the electrostatic, as defined above, the electromagnetic, and the so-called practical system. The relations between the electrostatic units and ordinary units are:

Electrostatic (e.s.u.)	Practical
1 volt	299.8 volts
1 unit charge	3.33×10^{-10} coulomb
1 capacity unit	1.113×10^{-6} microfarad
1 resistance unit	8.988×10^{11} ohm-centimeters

To convert ordinary (practical volts) to e.s.u. volts, the practical volts are divided by 299.8.

[1] E. A. Guggenheim, *Nature*, **148**, 751 (1941).

Dielectric Constant

As we have seen, the dielectric constant can be defined as the ratio of the electrostatic potential existing between electrostatic charges in a vacuum to that of the medium in question. For a vacuum D is thus set equal to unity. For all other media D will be greater than unity; for example, the dielectric constant of water at 25° C. is 78.54. To illustrate the meaning of this value, suppose that we charge two plates of a condenser with a certain quantity of electricity of opposite sign, the charging process to take place in a vacuum. We will then find that a certain potential difference exists between the two plates; now, if pure water is placed between the plates, the potential difference between the plates will be reduced to 1/78.54 of that previously found in a vacuum. It has been assumed that no conduction has taken place between the plates.

The measurement of the dielectric constant with any degree of precision requires careful work and good apparatus and is not to be lightly undertaken by the uninitiated. The various ways of measuring this important constant may be summarized as follows: [2]

1. The bridge method [3] is considered the best for lower frequencies. Fundamentally, it is the same circuit as is used for conductivity measurements except that in this case the capacity of the cell is a virtue instead

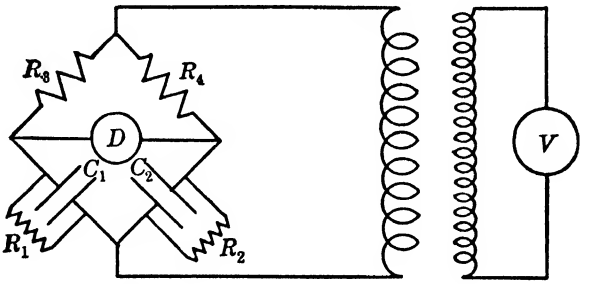


FIG. 1. Bridge circuit for dielectric constant measurements. D is telephone, V is source of alternating current, C_1 is unknown capacity (containing solution whose dielectric constant is to be measured), C_2 is a variable condenser, and R_3 and R_4 are known resistances.

of a nuisance. The resistance and capacity of a cell containing the solution under investigation are compared with those of a standard resistance and capacity. The circuit is diagrammed in Fig. 1.

[2] J. L. Oncley, *Chem. Revs.*, **30**, 433 (1942).

[3] F. Daniels, J. H. Mathews, and J. W. Williams, *Experimental Physical Chemistry*, 2nd Ed., McGraw-Hill Book Co., Inc., New York, 1934.

P. S. Albright, *J. Am. Chem. Soc.*, **59**, 2098 (1937).

When the bridge is balanced (minimum noise in the phone D), we have

$$\frac{R_3}{R_4} = \frac{C_2}{C_1} \qquad 10$$

Knowing R_3, R_4, and C_2, we can calculate C_1. Knowing the size of the plates of the condenser and the distance between the plates, we can calculate the dielectric constant from the value of C_1.

2. The resonance method [4] finds its greatest usefulness in the range of higher frequencies. The resonance of a circuit is determined, and in this circuit is included a cell containing the solution whose dielectric constant is to be measured. The resonance frequency of an electrical circuit depends upon the capacity of the circuit. The capacity of the circuit, in turn, depends upon the dielectric constant of the solution in the cell. Most investigations on dipolar ions have been carried on by this method.

3. The force method [5] makes use of the deflection produced by an applied electric field upon a conducting ellipsoid suspended in the solution under investigation. It is evident that this method could be effective only at quite low frequencies.

4. The calorimetric method [6] involves the measurement of the expansion of a solution resulting from the heat generated by an applied field of

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
i-Propanol	18.62	20
n-Butanol	17.0	20
Petroleum oil	2.12	20
Benzene	2.28	20

[4] J. Wyman, *Phys. Rev.*, **35**, 623 (1930).
M. A. Elliott and J. W. Williams, *J. Am. Chem. Soc.*, **61**, 718 (1939).
[5] J. W. Dunning and J. W. Shutt, *Trans. Faraday Soc.*, **34**, 467, 479 (1938).
[6] P. Debye, *Trans. Faraday Soc.*, **30**, 679 (1934).

high frequency. Incidentally, the medical profession makes use of heat generated by high-frequency currents for an entirely different purpose. It is used to produce artificial fevers in patients in the control of syphilis and in other pathological conditions. This technique goes under the name of diathermy. Table 1 shows the dielectric constant of some common substances.

Dielectric Constant of Cell Membranes

The dielectric constant of biological surfaces such as cell membranes is of especial interest. There is considerable reason to believe that living cell membranes have, in general, a much lower dielectric constant than water. Various values have been given. Admittedly, these are only intelligent guesses, but it is agreed that the dielectric constant in such membranes may well be as little as 15 or lower. This means that these membranes will have a "non-aqueous" character. In such a medium the common electrolytes, such as NaCl, will behave as if they were scarcely ionized at all. This has important consequences for the question of membrane penetration by electrolytes.

In an oriented layer of molecules the dielectric constant is increased above that in bulk. For example, Kallmann and Kreidl [7] found that the dielectric constant of a thin layer of palmitic acid inclosed between parallel metal plates rose with decreasing thickness of the layer of palmitic acid, attaining a limiting value when the thickness was 10 μ to 20 μ.

There is always a serious ambiguity when one talks about the dielectric constant of films and membranes. What is measured is the total effect of all the molecules. The dielectric constant in the space between giant molecules may, however, have quite a different value, and it is this value which determines the behavior of small ions, such as Na^+ and Cl^-.

Molecular Structure and the Dielectric Constant

In general, dielectric-constant studies can, under suitable conditions, yield two kinds of information about molecules, namely a knowledge of (1) the electrical asymmetry (dipole moment) of molecules, and (2) the size and shape of molecules. We shall see how it is possible to obtain this information from dielectric-constant measurements. We must, however, first inquire why there is a reduction in the potential between two charged plates when a medium of high dielectric constant is substituted for one with a low constant.

The answer to this question is that the medium with a higher dielectric constant is polarized to a greater extent than a medium with a

[7] H. Kallmann and W. Kreidl, *Z. physik. Chem.*, **A159**, 322 (1932).

lower constant. Polarization manifests itself as a separation or displacement of charges in a medium. This displacement requires that work be done on the system and results in a decreased potential across the plates of a condenser.

The three ways in which a substance may become polarized are:

1. The electrons within the atoms are displaced relative to the positive atomic nuclei.

2. The molecules are stretched, twisted, or bent by displacement of atoms within the molecules.

3. The dipolar molecules are oriented in the electrical field, the positive end of the molecule being directed toward the negative condenser plate, and the negative end towards the positive plate.

These three factors are known as the electronic, the atomic, and the molecular contributions to the dielectric constant. The magnitude of the three factors can be determined individually by measuring the dielectric constant of the substance as a function of the frequency of the alternations of the sign of the charge on the condenser plates. The electrons have a small mass and, hence, can respond to frequencies even higher than visible light. In fact, the dielectric constant at these high frequencies is intimately related to the index of refraction of the substance, i.e.,

$$\text{Electronic dielectric constant} = n^2 \qquad 11$$

where n is the index of refraction of the medium to visible light.

Atoms respond to infra-red light but not to alternations more rapid than those of infra-red light. The square of the index of refraction of a medium to infra-red light is, therefore, equal to the sum of the electronic and atomic contributions to the dielectric constant.

The time required for the orientation of molecules, on the other hand, is relatively long, being of the order of a billionth of a second. The larger the molecule, the greater is the time required to orient it.

The question arises as to actual technique of the measurement of polarization in any particular case. It should be clearly realized that, although the polarization is related to the dielectric constant and, in fact, is calculable from it, polarization is an entirely separate physical quantity.

Polarization can be calculated from the dielectric constant by means of the Clausius-Mosotti relation, which is

$$P = \frac{M}{\rho}\left(\frac{D-1}{D+2}\right) \qquad 12$$

where ρ is the density of the liquid or gas and M is its molecular weight.

The derivation of the Clausius-Mosotti equation is rather involved. It consists in resolving the electrical field acting on a molecule into three components and in calculating each component separately and then adding them up. There is considerable uncertainty regarding one of these components, and the Clausius-Mosotti relation is not on as sound a basis as could be desired.

Not very much information can be gained from the determination of the polarization of a pure polar liquid such as water or alcohol, because there is too much interaction between the molecules. The procedure in the study of polar molecules is as follows. A small amount of the particular substance is dissolved in a non-polar solvent, i.e., a solvent whose molecular polarization is negligible, such as carbon tetrachloride, benzene, or para-dioxane. The dielectric constant of the dilute solution is then measured as a function of the concentration of the solute. The polarization of the solute and solvent is assumed to be additive under these conditions so that

$$P_{12} = N_1 P_1 + N_2 P_2 \qquad 13$$

where P_{12} is the polarization of the solution, N_1 is the mole fraction of the solvent, and N_2 the mole fraction of the solute. P_1 is the polarization of the solvent and P_2 that of the solute. Equation 12 then becomes

$$P_{12} = \frac{M_1 N_1 + M_2 N_2}{\rho} \left(\frac{D - 1}{D + 2} \right) \qquad 14$$

where M_1 is the molecular weight of the solvent and M_2 that of the solute, and ρ is the density of the solution. The polarization of the solution, P_{12}, being known, the polarization of the solute can be calculated from equation 13. The value of P_1 is determined in the absence of the solute. The polarization calculated in this manner, known as the molar polarization, has the dimensions of volume and, in fact, is expressed in cubic centimeters. The molar polarization having been obtained, the dipole moment of the solute molecules can be calculated by proceeding in one of two ways.

The dipole moment which we wish to calculate is a measure of the electrical asymmetry of a molecule and is equal to the distance in centimeters separating two charges in a molecule multiplied by the electronic charge. That is

$$\text{Dipole moment } (\mu) = 4.77 \times 10^{-10} \times d \qquad 15$$

The dipole moment is always of the order of magnitude of 10^{-18}. Such units are called Debye units. Symmetrical molecules such as carbon tetrachloride, methane, benzene, and para-dioxane have very small or

zero dipole moments, whereas unsymmetrical molecules such as chloroform, nitrobenzene, and fatty acids have fairly large dipole moments. The importance of dipole-moment measurements for structural organic chemistry is evident.

Let us return to the two methods of calculating the dipole moment from polarization. As we have indicated, polarization is made up of three factors: the electronic, the atomic, and the molecular orientation. The total polarization is the sum of these factors,

$$P = P_0 + \frac{4\pi N\mu^2}{9kT} \qquad 16$$

where P_0 is the part due to the combined effect of the electrons and the atoms. The term $\frac{4\pi N\mu^2}{9kT}$ is the part due to molecular orientation. N is Avagadro's number, 6.023×10^{23}; k, the Boltzmann constant, is

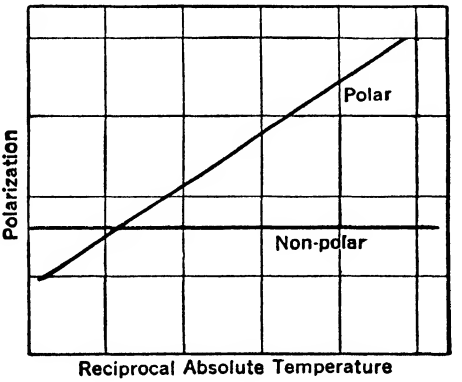


FIG. 2. Diagrammatic plot of molar polarization P against the reciprocal of the absolute temperature.

1.3805×10^{-16} erg per degree; μ is the dipole moment; and T is the absolute temperature. Introducing the numerical values for the constants in equation 16 and rearranging, we have

$$\mu = 0.0127 \times 10^{-18}\sqrt{(P - P_0)T} \qquad 17$$

We can determine P by experiment, and, if we knew P_0, the dipole moment could be calculated directly. P_0 could be calculated from the index of refraction of infra-red light, but, as such measurements are difficult, the next best thing is done and P_0 is assumed to be calculable from the index of refraction of visible light. This introduces a small error. It must be remembered, however, that for a polar molecule P_0

is quite small in comparison with the polarization due to molecular orientation, and this error is not important.

The second method for calculating μ from the polarization consists in measuring the polarization of the solute as a function of temperature. It should be clear that only the molecular orientation factor of the polarization should be temperature sensitive. As the temperature increases, so does the kinetic motion of the molecules; accordingly, it becomes more and more difficult to orient a molecule in the electrical field as the temperature is increased. In keeping with this conception is the observation that the dielectric constant of a polar liquid or of a polar gas decreases with increasing temperature. By plotting P against $1/T$ we can immediately distinguish a polar from a non-polar molecule. Such a plot is shown diagrammatically in Fig. 2.

In the examples we are considering, the plot of P against $1/T$ yields straight lines whose equations are

$$P = a + \frac{b}{T} \qquad 18$$

where a and b are constants.

Comparing equation 18 with equation 16, we see that

$$b = \frac{4\pi N\mu^2}{9k} \qquad 19$$

from which the dipole moment can be calculated.

TABLE 2

THE DIPOLE MOMENT OF SOME COMMON MOLECULES IN DEBYE UNITS

Substance	Dipole moment
Carbon tetrachloride	0
Methane	0
Ethylene	0
Carbon dioxide	0
Benzene	0
Chloroform	1.05
Methanol	1.68
Ethanol	1.70
Acetic acid	1.40
Water	1.85
Nitrobenzene	3.9

Dipole Moment of Zwitter Ions

The amino acids, proteins, and phospholipids are interesting as a class of polar molecules. Such molecules exist at the proper acidity of the

solution as zwitter ions. The general form of such zwitter ions can be illustrated by the simplest of the amino acids, glycine (CH_2NH_2COOH), shown in Fig. 3.

Such a marked separation of charge would be expected to produce a high dipole moment. The dipole moment of glycine can be estimated from structural considerations. The positive charge is supposed to be located at or close to the center of the nitrogen atom; the negative charge resides midway between the two oxygens of the carboxyl group. The distance between these charges is estimated to be 3.17 Å. Multiplying this distance by the electronic charge gives a dipole moment of about 15 Debye units. This, as will be noted, is much greater than any of the dipole moments shown in Table 2. All the α-amino acids have the same distribution of charges as glycine; accordingly, the dipole moment of all such amino acids should be close to 15 Debye units.

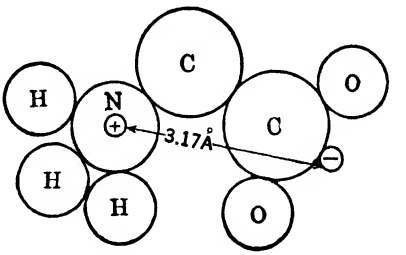


FIG. 3. Structural formula of the amino acid glycine showing distance of separation of the positive and negative charges.

The direct determination of the dipole moments of amino acids and proteins by means of dielectric-constant measurements is complicated by the fact that amino acids and proteins are, as a rule, insoluble in non-polar solvents. This necessitates the use of polar solvents such as water. The Clausius-Mosotti relation (equation 12), however, is not valid for polar solvents and cannot be used to calculate polarization from dielectric constants of such solvents.

It has been found that the dielectric constant of aqueous solutions of amino acids and of proteins is greater than for pure water. In fact, the dielectric constant increases linearly with the concentration of the zwitter ion. This increase, called the molar dielectric increment, is denoted by δ. The dielectric increments of all α-amino acids not only are positive and large but also are all very nearly the same (about 23.0).[8] The increment is also independent of the type of polar solvent used (water-alcohol mixtures, urea-water mixtures). There is also a linear increase of δ with the number of carbon atoms intervening between the NH_3^+ and COO^- groups, the increment being almost exactly 13 for each additional carbon atom.[9]

[8] J. Wyman, *J. Am. Chem. Soc.*, **56**, 536 (1934); see also *Chem. Revs.*, **19**, 236 (1936).

[9] J. Wyman and T. L. McMeekin, *J. Am. Chem. Soc.*, **55**, 908 (1933).

Wyman concludes that the polarization of a solution of a zwitter ion in a polar solvent is directly proportional to the dielectric-constant increment. This is to be contrasted with the Clausius-Mosotii relation (equation 12), which is valid for the general case and which shows polarization to be a more complicated function of the dielectric constant.

We have seen from equation 16 that polarization is a linear function of the square of the dipole moment. Combining this with Wyman's conclusions regarding the direct proportionality between the dielectric increment and polarization, we obtain for a zwitter ion in a polar solvent

$$\delta = \alpha\mu^2 - K \qquad 20$$

where α is a proportionality constant and K is the decrement in the dielectric constant due to the non-polar part of the zwitter-ion molecule. We can evaluate the magnitude of the constant α by using the value of μ calculated from structural considerations as we have already done above for glycine. We find after doing this that

$$\mu = 2.9\sqrt{\delta + K} \qquad 21$$

Usually K is small as compared with the effect due to the NH_3^+ and COO^- groups and can be neglected. Equation 21 provides a means of estimating the dipole moment of any zwitter ion dissolved in a polar solvent. Dipole moments of some zwitter ions are given in Table 3

TABLE 3

APPROXIMATE DIPOLE MOMENTS OF SOME ZWITTER IONS

Substance	Debye units
Urea (not a zwitter ion)	5
Glycine	15
α-Amino valeric acid	15
Diglycine	26
Triglycine	32
Egg albumin	250
Insulin	300
β-Lactoglobulin	750
Zein	380

The dipole moment of the proteins appears to be very large, but, when account is taken of the very great size of these molecules, proteins really show a high degree of electrical symmetry.

Conner, Clarke, and Smythe [10] find that there is a marked degree of rigidity in peptide molecules. They further conclude that, although the potential energy barriers between molecular configurations are relatively high, they are not prohibitive for crossing and, accordingly, there is still a large measure of freedom to attain a random distribution of the molecules among the possible orientations about the valence bonds.

Dielectric Constant and Frequency

Evidently, if the frequency of alternations of the charge on the condenser plates is progressively increased, a frequency will be reached at which neither the solute nor the solvent molecules will be able to rotate

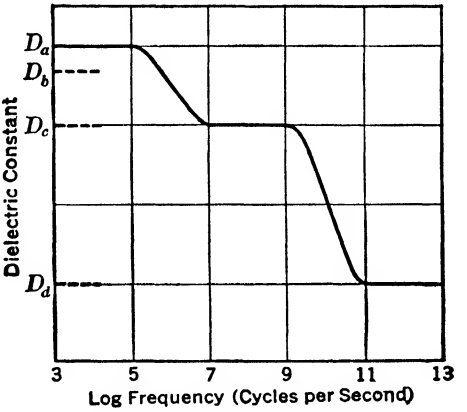


FIG. 4. Anomalous dispersion curve of the dielectric constant of an aqueous protein solution (diagrammatic).

with sufficient rapidity to respond to the alternations. If the molecules of the solute are enough larger than those of the solvent and the dielectric constant of the solution is plotted against the logarithm of the frequency, a curve such as is diagrammatically shown in Fig. 4 must be obtained. Such a curve is known as an anomalous dispersion curve.

In Fig. 4, D_a is the dielectric constant of the solution; D_b is the dielectric constant of the pure water; D_c is the dielectric constant of the solution at a frequency which is too great for the protein molecules to respond to the alternations. Note that D_c is less than D_b; the solute molecules are inert at this frequency and occupy a certain volume so that there is less water per unit volume solution than for pure water. D_d is the dielectric constant of the solution at such a high frequency that neither

[10] W. P. Conner, R. P. Clarke, and C. P. Smythe, *J. Am. Chem. Soc.*, **64**, 1379 (1942).

the solute nor solvent molecules can respond to the alternations in the sign of the charge. At this frequency only the atomic and electronic factors of polarization are operating.

It is clear that there should exist a relation between molecular size and the critical frequency at which dispersion of the dielectric constant occurs. This is indeed true, and the mathematical relations are outlined below.

The time of relaxation is a measure of the time required for the oriented dipoles to revert to random distribution if they were all oriented at zero time. The time of relaxation is related to the frequency of the field and to the dielectric constant as follows:

$$\tau = \frac{D_c + 2}{D_a + 2} \cdot \frac{1}{2\pi\nu_c} \qquad 22$$

where τ is the relaxation time and ν_c is the critical frequency which is defined as that frequency at which

$$D = \frac{D_a + D_c}{2} \qquad 23$$

If a spherical molecule is oriented by any means, electrical or otherwise, its time of relaxation is given by

$$\tau = \frac{4\pi\eta r^3}{kT} \qquad 24$$

where η is the coefficient of viscosity of the solvent, k is the Boltzmann constant, T is the absolute temperature, and r is the radius of the molecule. Combining equations 20 and 22 and rearranging, we have

$$\text{Molecular volume} = \frac{kT(D_c + 2)}{6\pi\eta\nu_c(D_a + 2)} \qquad 25$$

from which we can calculate the molecular size.

The Debye theory predicts that the decrease in the dielectric constant as the frequency is increased should be complete in 2 logarithmic units of frequency. Figure 5 is a plot of the variation of the dielectric constant of a 70 per cent aqueous ethanol solution of the protein zein against frequency.[11] As can be seen, the dispersion extends over a greater range than 2 logarithmic units of frequency. Elliott and Williams have interpreted this to mean that the zein molecule is in the shape of a prolate ellipsoid of revolution; accordingly, the experimental dispersion curve

[11] M. A. Elliott and J. W. Williams, *J. Am. Chem. Soc.*, **61**, 718 (1939).

can be resolved into two curves. The first curve, ending with D/D_e equal to 1.018, is that due to rotation of the ellipsoid about its minor axis; the one ending with D/D_e equal to 1.003 corresponds to rotation about the major axis. Elliott and Williams were able to use their data to

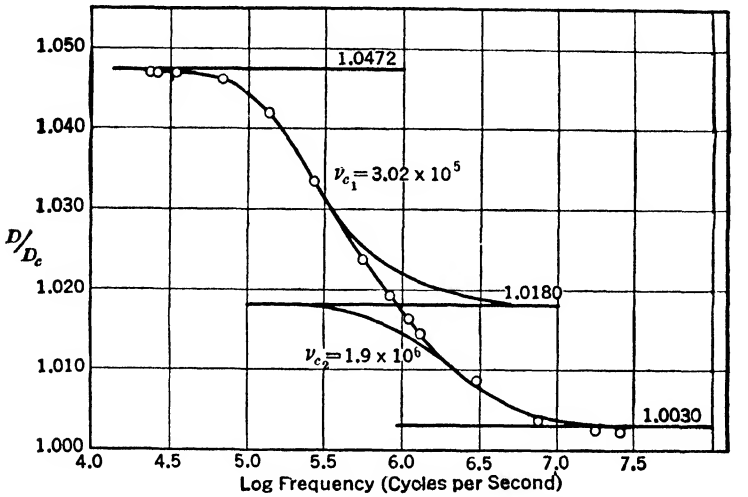


FIG. 5. Dielectric dispersion curve of zein dissolved in 70 per cent aqueous ethanol. (Elliott and Williams.)

calculate the asymmetry of the zein molecule. They find the ratio of the major to the minor axis of the molecule to be 7 to 1. Actually, this value agrees rather poorly with the ratio calculated from diffusion data, which comes out to be somewhat in excess of 20 to 1.

Chapter V

IONS IN SOLUTION

The fact that aqueous solutions of salts, of acids, and of bases are good conductors of electricity led Arrhenius [1] to the suggestion that when dissolved in water these substances exist partially as dissociated, charged particles, called ions.

Arrhenius found that in a number of instances he could formulate the dissociation of the electrolyte in terms of an equilibrium between the dissociated and the undissociated parts of the dissolved molecule. For example, for acetic acid, we have

$$HAc \rightleftarrows H^+ + Ac^-$$

which can be expressed in the form of an equilibrium constant

$$K = \frac{H^+ \times Ac^-}{HAc} \qquad 1$$

In this early work the magnitude of K was determined by measuring the electrical conductivity of a series of dilutions of the electrolyte. Today there are less ambiguous methods for the evaluation of K; these methods will be discussed in some detail in Chapter VI.

We, of our time, take the concept of ionization more or less for granted, but it can easily be imagined how revolutionary it must have seemed when it was proposed. The people of Arrhenius' day must have asked why these ionic charges did not neutralize each other; we should still ask that question.

There are, in general, two types of chemical bonds which hold atoms together (a third type, the metallic bond, will not be considered).[2] The first of these is the covalent bond, which results from the sharing of electrons by two atoms. Examples of this type are the carbon-carbon bond and the carbon-hydrogen bond. Covalent bonds do not permit dissociation of a molecule into ions. The second type is the ionic bond (sometimes called a salt linkage). The ionic bond arises from an actual transfer of electrons from one atom to another. An illustration is found

[1] S. Arrhenius, *Z. physik. Chem.*, I, 631 (1887).

[2] L. Pauling, *The Nature of the Chemical Bond*, Cornell University Press, 1940.

in ordinary sodium chloride. The metal atom, sodium, gives up an electron to the chlorine atom. This exchange leaves the sodium positively charged and imparts a negative charge to the chlorine, producing a sodium and a chloride ion, respectively. An electrostatic attraction between the sodium and chloride results. This is the ionic bond, and the attractive forces between the ions are known as coulombic forces.

The transfer of an electron from the sodium atom to a chlorine atom leaves the sodium with a completed electron shell, which is a rare-gas structure. The transfer completes the chlorine electron shell and gives it also a rare-gas structure. Such structures are very stable. In a crystal of sodium chloride, sodium and chlorine do not exist as atoms, but as ions; a similar statement is in general true of such crystals (KCl, NaBr, etc.).

The existence of ions in solution is conditioned by two factors: (1) the attraction of ions for each other, and (2) ionic hydration. The force between charged particles is given by

$$\text{Force} = \frac{Q_1 Q_2}{D d^2} \qquad 2$$

where Q_1 and Q_2 are the charges on the particles, d is the distance between the charges, and D is the dielectric constant. Water with a dielectric constant of 78.54 at 25° C. decreases the force between the ions to 1/78th of that existing in ionic crystals. Evidently this factor will enhance the solubility of ionic crystals in water. The second factor, ionic hydration, will presently be dealt with in detail. It is clear, however, that hydration will decrease the potential energy of an ion and will, accordingly, enhance the separation of ions. If the substance in question does not exist as ionic crystals but is a weak electrolyte, such as acetic acid, the two factors mentioned above are important in determining the extent of ionization of the compound. For example, decreasing the dielectric constant of water by the addition of dioxane decreases to a very marked extent the ionization of formic acid.[3]

Sizes of Ions

Since the electrostatic field around an ion extends an infinite distance, there must be an arbitrary element in the assignment of radii to ions. Table 1 gives the crystal radii of some ions. The crystal radius is defined as that distance such that the sum of the radii of any two ions is equal to the actual equilibrium interionic distance in a crystal containing the ions.

[3] H. S. Harned and R. S. Done, *J. Am. Chem. Soc.*, **63**, 2579 (1941).

TABLE 1

IONIC RADII

From *The Nature of the Chemical Bond*, by L. Pauling

Ion	Radius in Å	Ion	Radius in Å
Li^+	0.60	F^-	1.36
Na^+	0.95	Cl^-	1.81
K^+	1.33	Br^-	1.95
Rb^+	1.48	I^-	2.16
Cs^+	1.69	$SO_4^=$	1.51
Mg^{++}	0.65	NO_3^-	1.21
Ca^{++}	0.99	SCN^-	3.03
Sr^{++}	1.13		
Ba^{++}	1.35		

Interionic Attraction

It became apparent rather early that Arrhenius' idea of partial ionization did not always satisfy quantitative formulation. This was particularly true of many salts and of such strong acids as hydrochloric and sulfuric. In these substances it was found that the calculated equilibrium constant of ionization was a pronounced function of the electrolyte concentration.

Some help for the treatment of such electrolytes (known in general as strong electrolytes) was forthcoming from the empirical discovery by Lewis and Randall [4] that the effect of ions in dilute solutions was proportional, not to their molar concentration, but to their ionic strengths, and they enumerated the important rule: "In dilute solutions, the activity coefficient of a given strong electrolyte is the same in all solutions of the same ionic strength." They defined the ionic strength, μ, as equal to the sum of the ionic concentrations, each ionic concentration being multiplied by the square of the ionic valence. The whole sum is divided by 2 (concentrations are expressed in gram ions per 1,000 grams of solvent). That is

$$\mu = \frac{C_1Z_1^2 + C_2Z_2^2 + C_3Z_3^2}{2} \qquad 3$$

Consider 0.01 molar solutions of NaCl, $MgSO_4$, and H_2SO_4. These solutions have the following ionic strengths:

$$\mu \text{ for NaCl} = \frac{0.01 \times 1^2 + 0.01 \times 1^2}{2} = 0.01$$

[4] G. N. Lewis and M. Randall, *Thermodynamics*, McGraw-Hill Book Co., Inc., New York, 1923.

$$\mu \text{ for } MgSO_4 = \frac{0.01 \times 2^2 + 0.01 \times 2^2}{2} = 0.04$$

$$\mu \text{ for } H_2SO_4 = \frac{0.01 \times 1^2 + 0.01 \times 1^2 + 0.01 \times 2^2}{2} = 0.03$$

The rule given above, that solutions of all strong electrolytes at the same ionic strengths exhibit the same ionic effects, was found to be valid up to an ionic strength of about 0.1.

The principle of ionic strength is an important one for the biochemist. Frequently, for example, a study is made of the variation of some factor as a function of the *p*H of a solution, buffers being employed to yield the proper *p*H. In order, however, for the effect observed to be due only to the variation in *p*H and not obscured by secondary salt effects, it is necessary to make up the series of buffers to the same ionic strength. This does not always suffice, as many physiological responses have specific ion effects. Nevertheless, a constant ionic strength does represent the least that the biochemist can do to make his experiment as unambiguous as possible.

The failure of the Arrhenius formulation for salts, strong acids, and bases led several people to the idea that such electrolytes are at all times completely dissociated and that the anomalous effects were due to electrostatic interaction of the ions. Debye and Hückel [5] were the first to formulate this concept in a satisfactory mathematical manner. Their contribution to chemistry must be considered one of the most important in this century. We shall attempt to outline their approach:

Debye and Hückel made two assumptions: (1) that ionization of strong electrolytes was complete; and (2) that the deviation from ideal behavior was due to the electrostatic effects of the positive and negative ions on one another and to no other cause. Their derivation may conveniently be divided into three steps. First, they proceeded to derive an expression for the distribution of positive ions around a negative one.

Evidently, owing to the presence of a negative ion there will be an excess of positive ions in its immediate neighborhood. The difference in concentration between the negative and positive ions will determine the charge density in the neighborhood of a negative ion. If it is difficult for the reader to visualize such an excess of positive ions, consider the extreme example of a sodium chloride crystal. Here, a chloride ion is entirely surrounded by six sodium ions. The situation in solution naturally does not approach the regularity of a crystal.

[5] P. Debye and E. Hückel, *Physik. Z.*, **24**, 185, 334 (1923).

The concentration of positive and negative ions in the neighborhood of a negative ion was calculated by means of the Boltzmann principle, which can be expressed mathematically as

$$n = n_0 e^{W/kT} \qquad 4$$

where n_0 is the number of ions per unit volume for a homogeneous distribution (no electrostatic effects); n is the number of ions per unit volume in the presence of the electrostatic field; W is the work required to take the ions from the homogeneous distribution into a place where the electrostatic field is acting; k is the Boltzmann distribution constant; and T is the absolute temperature. The sign of the exponent is positive for a positive ion and minus for a negative one. Combining the Boltzmann expression with the knowledge that the difference in concentration of the negative and positive ions is equal to the charge density, we find (after considerable mathematical manipulation which involves the expansion of the exponential Boltzmann equation and neglecting second-order terms) that the charge density is

$$\rho = -\frac{N^2\epsilon^2\Phi}{RT}[C_1Z_1^2 + C_2Z_2^2] \qquad 5$$

where N is Avogadro's number, ϵ is the electronic charge, R is the gas constant, C_1 and C_2 are the concentration in gram ions per liter, and Z_1 and Z_2 are the respective valences, Φ is the potential existing at the location of the ion.

The second step in the derivation was to use the differential form of Poisson's equation (see Chapter IV), which relates the charge density and the potential. The charge density as given by equation 5 was substituted in the Poisson equation and the equation was integrated over the entire ionic atmosphere around the negative ion to obtain the potential. This potential when multiplied by the charge on the ion represents the electrical work required to remove the ion from its environment in the solution. When this work is summed up for all the negative as well as for all the positive ions in solution, we have the total potential energy decrease suffered by the solution due to the electrostatic charges present. Carrying out this summation we find that the total work is

$$\text{Work} = \frac{NZ_1Z_2\epsilon^2\kappa}{2D(1 + \kappa a)} \qquad 6$$

where a is the average radius of the ions and κ is the reciprocal of the distance from the center of the negative ion to the center of gravity of

the charge produced by the ionic atmosphere surrounding the ion. κ is given by

$$\kappa = \sqrt{\frac{4\pi N^2 \epsilon^2 \Sigma C_1 Z_1{}^2 + C_2 Z_2{}^2 + \cdots}{DRT}} \qquad 7$$

The third step in the derivation consists in relating the electrical work to the activity coefficient of the electrolyte. The work W is exactly equal to the free energy decrease in the system produced by the effect of the ionic charges on each other.

The free energy change in bringing a mole of a given species of ions from concentration C_0 to concentration C_1, if the ions behave in an ideal fashion, is

$$\Delta F = -RT \ln \frac{C_1}{C_0} \qquad 8$$

If the ions form an ideal solution at concentration C_0 but not at concentration C_1, the free energy change is

$$\Delta F = -RT \ln \frac{\gamma C_1}{C_0} \qquad 9$$

where γ is the activity coefficient of the ions at concentration C_1. In our particular case C_0 equals C_1; accordingly, equating ΔF and W, we have

$$-RT \ln \gamma = \frac{N Z_1 Z_2 \epsilon^2 \kappa}{2D(1 + \kappa a)} \qquad 10$$

When the values of the constants at 25° C. are substituted there results

$$\log \gamma = \frac{-0.509 Z_1 Z_2 \sqrt{\mu}}{1 + 3.3 \times 10^7 a \sqrt{\mu}} \qquad 11$$

At μ less than 0.1, equation 10 reduces to

$$\log \gamma = -0.50 Z_1 Z_2 \sqrt{\mu} \qquad 12$$

The Debye-Hückel equation has been tested and found to yield accurate values for the activity coefficients of dilute solutions of electrolytes. It has not been found to give good results in more concentrated solutions. There are several causes for the deviations of the Debye-Hückel equation at higher concentrations. A number of approximations were made in the derivation and the dielectric constant was assumed to be independent of the electrolyte concentration.

Table 2 shows the activity coefficients of some electrolytes as determined by electromotive force measurements.

TABLE 2

ACTIVITY COEFFICIENTS FROM ELECTROMOTIVE FORCE MEASUREMENTS

m	LiCl	NaCl	KCl	KOH	HCl
0.001		0.966	0.965	0.989	
0.005		0.928	0.926	0.954	
0.01	0.901	0.903	0.899	0.920	0.91
0.05	0.819	0.821	0.815	0.822	0.836
0.1	0.779	0.778	0.764	0.789	0.801
0.5	0.725	0.678	0.644	0.750	0.762
1.0	0.757	0.658	0.597	0.760	0.823
2.0	0.919	0.670	0.569		
3.0	1.174	0.714	0.571	1.062	

It will be noted from Table 2 that there is always a minimum in the activity coefficient-concentration relation. The Debye-Hückel equation cannot account for such a minimum; an additional term is needed,

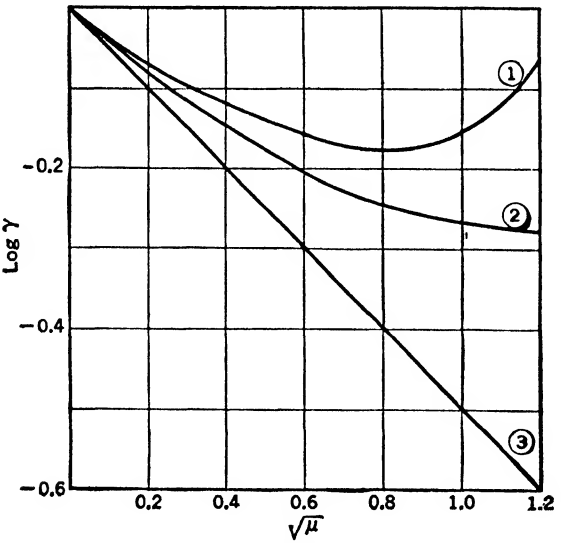


FIG. 1. Plot of the Debye-Hückel equations. (1) Equation 13. (2) Equation 10. (3) Equation 12.

a term which increases with concentration. To explain the increase in the activity coefficients a salting-out effect is postulated. Owing to the dipolar character of water, the water molecules tend to displace ions in the vicinity of any particular ion. This, of course, is nothing but an

ion hydration effect and results in an apparent increase of the ionic concentration. Including the salting-out effect, the Debye-Hückel equation becomes, for 25° C.,

$$-\log \gamma = \frac{0.509 Z_1 Z_2 \sqrt{\mu}}{1 + 3.3 \times 10^7 a \sqrt{\mu}} - B\mu \qquad 13$$

where B is an empirically determined salting-out constant. Figure 1 shows a plot of the simple and extended Debye-Hückel equations.

Although the Debye-Hückel equations were originally derived for strong electrolytes, they are equally applicable to weak and intermediate electrolytes, and the ionic strength is obtained from the degree of dissociation of the weak electrolyte. In short, the Debye-Hückel equations apply to the ionized portion of the electrolyte. This topic will be discussed in greater detail in Chapter VII.

Ion Hydration

Before we enter into the question of ion hydration, it is appropriate to indicate briefly the structure of water itself. Water is an extremely important compound for biological systems. Living tissue usually contains about 80 per cent water. The percentage may be even higher; for example, the jellyfish contains about 96 per cent water. Gortner [6] has rightly stressed the importance of water in biological systems.

The water molecule is V-shaped, the O—H distances being 0.96 Å and the H—O—H angle 105°.[7]

Liquid water is a highly complicated substance. It was believed at one time that water was made up of definite polymers which formed a neat series of monohydrols, dihydrols, trihydrols, etc. X-ray studies have demonstrated that liquid water has, essentially, a broken-down ice structure. No definite polymers are present. There is, however, a marked association between the water molecules which gives rise to a quasi-ice structure, the so-called 4-coordinated water structure. The nearest approach of water molecules in ice is 2.76 Å, which would make the radius of the water molecule 1.38 Å.

As its structure is very open, ice has a low density. The individual water molecules in ice apparently retain a considerable degree of freedom and are capable of rotation. This is shown by the relatively high dielectric constant of ice. The water molecules are bonded together by hydrogen bonds; there are four bonds per molecule at −183° C. As the temperature is raised, the average number of bonds per molecule

[6] R. A. Gortner, *Outlines of Biochemistry*, John Wiley & Sons, Inc., New York, 1938.

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

decreases. On fusion and on further increase in temperature, more bonds are broken, and at 40° C. the average number of hydrogen bonds per molecule is about two. The energy of a single hydrogen bond in water is about 4.5 large calories.

Increase in temperature decreases the extent of association of water. The addition of ions to water has the same effect upon the association of water as does increasing the temperature.[8] The enormous electrostatic field in the neighborhood of ions produces polarization in the water dipole. For example, at a distance of about 10 Å from a univalent ion, the electrostatic field has an intensity of about 14 million volts per centimeter.[9]

Attempts have been made to assign a definite number of molecules of hydration to various ions, but it is improbable that such figures have much meaning. Since the electrostatic field of an ion falls off with the distance from the ion, the water molecules are acted upon by varying forces and ionic hydration cannot, in general, be represented by a definite number of water molecules. For ions of the same valency the intensity of the electrostatic field is greater, the smaller the ions. We would anticipate, therefore, that the smaller ions would have a higher energy of hydration than the larger ones. This is found to be true.

The energy of hydration is defined as that energy which is released when ions initially in a vacuum are placed in pure water. The heat of hydration is so very nearly equal to the free energy of hydration that, considering the present accuracy of the energy values, it is pointless to distinguish between the two.

The energy of hydration of an electrolyte is equal to the difference between the energy required to vaporize a mole of the crystalline electrolyte into ions and the heat of solution. The heat of solution is easily measured by standard calorimetric techniques, but the energy of vaporization of the crystalline salt is more difficult to get at. The energy of vaporization is known as the lattice energy, and it is calculated by means of the complicated Born-Haber [10] cycle. The main idea of the Born-Haber cycle is to make use of the chemical heats of formation to deduce the lattice energy of the resulting crystalline compound. The cycle, which involves nine separate steps, will not be discussed further.

In the calculation of individual ionic hydration energies, we are faced with the difficulty of always dealing with at least one cation and one anion, and what we obtain is the sum of the effects of the cations and

[8] G. W. Stewart, *J. Chem. Phys.*, **7**, 869 (1939).

[9] R. W. Gurney, *Ions in Solution*, Cambridge University Press, 1936.

[10] M. Born, *Verhandl. deut. physik. Ges.*, **21**, 13 (1919).
F. Haber, *ibid.*, **21**, 750 (1919).

anions. A close approximation of the individual hydration energies, however, can be made. For ions of opposite sign but of the same valence, the heat of hydration can depend only on the radii of the ions; accordingly, two ions having the same radius should have the same heats of hydration. Potassium and fluoride ions very nearly satisfy this condition. The heat of hydration of KF is 191,000 calories; of this amount we assign 94,000 calories to the potassium ions and 97,000 calories to the fluoride ions. By determining the heats of hydration of various anions in combination with potassium ions and various cations in combination with fluoride ions, we can estimate the heats of a great variety of ions. Table 3 shows some ionic hydration energies obtained in this way.

Hofmeister Series

The biochemist has often concerned himself with what is known as the Hofmeister or lyotropic series. It is found that, when the effects of

TABLE 3

IONIC HYDRATION ENERGIES IN KILOGRAM CALORIES PER GRAM MOLE IONS

Ion	Hydration energy	Ion	Hydration energy
H^+	276	Sr^{++}	376
Li^+	136	Ba^{++}	346
Na^+	114	Al^{+++}	1,149
K^+	94	La^{+++}	768
Rb^+	87	Th^{++++}	1,540
Cs^+	80	Ag^+	162
NH_4^+	87	Tl^+	107
OH_3^+	130	Mn^{++}	479
OH^-	105	Fe^{++}	500
F^-	97	Co^{++}	504
Cl^-	65	Ni^{++}	516
Br^-	57	Cu^{++}	536
I^-	47	Zn^{++}	528
Be^{++}	608	Cd^{++}	462
Mg^{++}	490	Hg^{++}	480
Ca^{++}	410	Fe^{+++}	1,185

a series of cations or anions upon the salting-in or salting-out of proteins,[11] upon the viscosity of lyophilic colloids, upon electrokinetic potentials,[12] and upon a variety of other physical,[13] chemical, and biological

[11] R. A. Gortner, W. F. Hoffman, and W. B. Sinclair, *Kolloid-Z.*, **44**, 97 (1928).
[12] D. R. Briggs, *J. Phys. Chem.*, **32**, 1646 (1928).
[13] A. Frumkin, *Kolloid-Z.*, **35**, 340 (1924).

properties [14] are compared, pronounced variations are shown. Such a series of ions is known as a Hofmeister or lyotropic series. For example, the series usually given for anions is citrate > tartrate > SO_4 > acetate > Cl > NO_3 > Br > I > CNS. The series for the cations, though usually somewhat less marked than that for the anions, is generally given as Th > Al > H > Ba > Sr > Ca > K > Na > Li. Under some conditions the series is completely reversed or the ions displaced

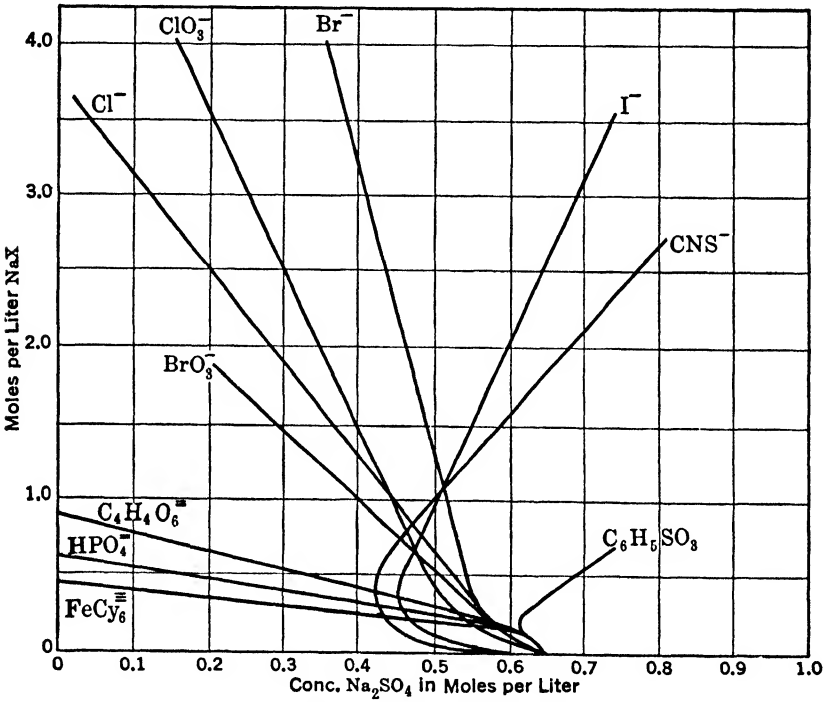


FIG. 2. Concentration of different sodium salts plotted against concentrations of Na_2SO_4, mixtures of which are necessary to bring about flocculation of agar-agar sols.

in their position. The series is not limited to aqueous systems, as Jaeger [15] has found that the values of the surface tension of molten salts at 1,000° C. fall into the series F > SO_4 > Cl > Br > NO_3 > I and Li > Na > K > Rb > Cs.

Undoubtedly, the series has its origin in the intensity of the electrostatic field around the ions, the small ions of the same valence having a more intense field than the larger ions. As pointed out previously, the more intense field of the smaller ions leads to a greater hydration, the

[14] D. Glick and C. G. King, *J. Biol. Chem.*, **94**, 497 (1931).
[15] F. M. Jaeger, *Z. anorg. allgem. Chem.*, **101**, 1 (1917).

difference in hydration being the immediate cause of the series in aqueous systems. That this is true is shown by the success of Buchner and co-workers [16] in being able to express the lyotropic series in quantitative terms and showing the direct relation to ionic hydration.

Buchner and co-workers studied the salting-out of gelatin and of agar-agar sols by Na_2SO_4 and by mixtures of Na_2SO_4 with other sodium salts. It was found that the concentration of the added sodium salt required to salt out varied linearly with the amount of Na_2SO_4 required to produce the same effect. In other words, the effect of the salts was strictly additive. This is shown in Fig. 2.

A quantitative measure of the salting-out action of the ions is expressed by the angle between the lines in the figure and the abscissas.

TABLE 4

LYOTROPIC NUMBERS OF IONS AS GIVEN BY BUCHNER ET AL.

Anion	Lyotropic number	Cation	Lyotropic number
SO_4^{--}	2.0	Li^+	115
F^-	4.8	Na^+	100
IO_3^-	6.25	K^+	75
$H_2PO_4^-$	8.2	Rb^+	69.5
BrO_3^-	9.5	Cs^+	60
Cl^-	10.0		
NO_2^-	10.1		
ClO_3^-	10.65		
Br^-	11.30		
NO_3^-	11.60		
ClO_4^-	11.8		
I^-	12.5		
CNS^-	13.25		

This permits the series to be expressed quantitatively by giving to every ion a number, N, which is defined by the relation

$$N = a \cot \alpha + b \qquad 14$$

where a and b are constants and α is the angle formed by the salt line of the ion and the abscissas shown in the figure. The angles are different for different colloids but the number N is characteristic for every ion. To fix the scale two arbitrary numbers are used. Choosing SO_4^{--} equal to 2.00 and Cl^- equal to 10, it is found that

$$N = 4.0 \cot \alpha + 10.0 \qquad 15$$

[16] A. Voet, *Trans. Faraday Soc.*, **32**, 1301 (1936).
A. Voet and E. H. Buchner, *Chem. Revs.*, **20**, 169 (1937).

This permits the evaluation of N for any anion. A similar series of numbers has been derived for the cations. The numbers given by Buchner are shown in Table 4.

It will be noted that, the smaller the value of N, the greater is the potency in salting-out; the larger the N, the greater the potency in salting-in.

That the Hofmeister series is quantitatively related to ionic hydration is shown by the linear relation between energy of hydration and the value of the Hofmeister N (Fig. 3).

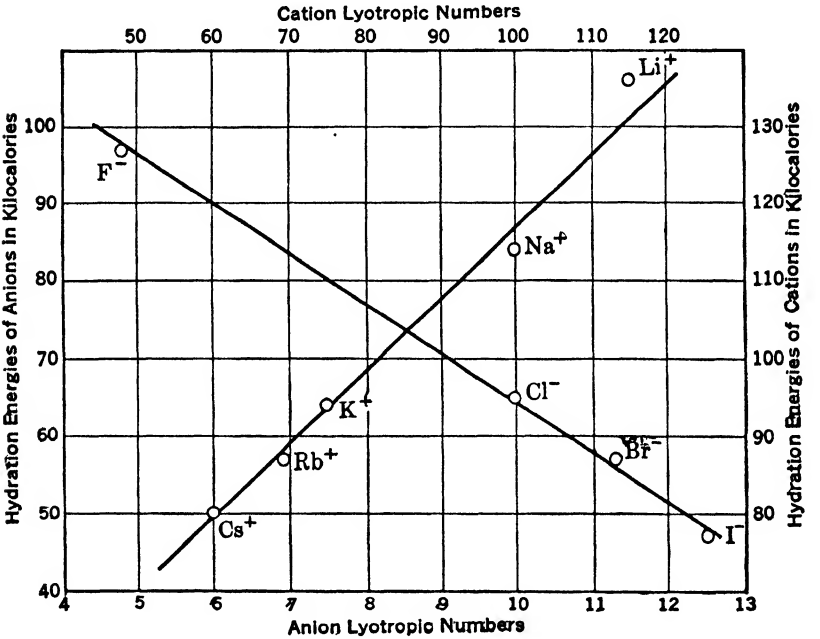


FIG. 3. Heats of hydration of ions, plotted against the lyotropic numbers.

It should be pointed out that cations and anions tend to orient water molecules in opposite directions

Thus, a cation carries its water with the hydrogens sticking outward. Such water can form hydrogen bonds with polar groups in a solute molecule and, accordingly, cations will tend to hydrate a colloid while the opposite effect is true of anions. A highly hydrated anion such as

SO_4^{--} is a precipitating ion; the highly hydrated cation such as Li^+ is a salting-in ion.

It should be remembered that the Hofmeister series is of importance only in fairly concentrated salt solutions.

Ions in Biology

Living material is, in general, very sensitive to its ionic content. The ionic concentrations of body fluids must be maintained within fairly narrow limits to sustain life. The body fluids of different classes of animals vary considerably in their absolute salt concentrations, the marine animals having a higher total concentration than land animals. The ratio of the various ions to one another is, however, approximately the same as that found in sea water. The most abundant cations are Na^+, K^+, Mg^{++}, and Ca^{++}; the anions are Cl^-, HCO_3^-, $H_2PO_4^-$, SO_4^{--}, and protein. H^+ and OH^- ions occur in extremely small concentrations. Their roles are so important, however, that a separate chapter has been reserved for their consideration. Table 5 shows the approximate ionic composition of human blood serum.

TABLE 5

IONIC COMPOSITION OF HUMAN BLOOD SERUM

Ion	Grams per liter	Moles per liter	Equivalents per liter
Na^+	3.25	0.1458	0.1458
K^+	0.19	0.0048	0.0048
Ca^{++}	0.10	0.0025	0.0050
Mg^{++}	0.027	0.0011	0.0022
HCO_3^-	1.64	0.0269	0.0269
Cl^-	3.70	0.104	0.1040
HPO_4^{--}	0.10	0.00105	0.0021
SO_4^{--}	0.19	0.00197	0.0039

Neglecting the contribution due to protein, which is small, the total ionic strength of blood is calculated to be about 0.15. This value is probably too high as some of the "ions" are undissociated. This ionic strength is just about the upper limit of the validity of the Debye-Hückel theory, i.e., without the salting-out correction. The Na^+ and Cl^- far outweigh the concentration of all the other ions put together.

The individual roles of the ions in biology are rather ambiguous and complicated. The salts have a thermodynamic role in that they help maintain the proper osmotic pressure across cell membranes. Particu-

larly is this true in regard to red cells, which show a very limited permeability to cations,[17] which, accordingly, are osmotically effective. In addition to this osmotic function, they also have specific effects.

Sodium in high concentrations is toxic, but the toxicity largely disappears if the increased concentration is accompanied by a corresponding increase in the concentration of the other ions. Sodium, apparently, increases the permeability of cell membranes. Potassium behaves in a large measure like sodium. In medical practice, however, in the treatment of edema resulting from either kidney or heart dysfunction, it is customary to remove NaCl from the diet and substitute KCl. In some cases, at least, this causes a copious loss of water through the urine. Why Na^+ should favor tissue hydration and K^+ decrease it is completely obscure at the present time.

Ion antagonism is a very complicated subject. The situation is in an unsatisfactory state, as indicated by the many publications on this topic. As ion antagonism is really not a unified phenomenon, very likely it has several causes. A striking antagonism is shown between Ca^{++} and Mg^{++}. Meltzer and Auer [18] showed that, if magnesium salt is injected into the blood stream of vertebrates, a profound anesthesia results in which the animals become motionless and insensitive to pain. If in such an anesthetized animal a trace of calcium salt is injected, there is a prompt awakening and return to sensitivity.

The anesthetic effect of magnesium ion has been the stumbling block of a number of general theories of anesthesia. Heilbrunn [19] has a general theory of anesthesia into which Mg^{++} ions fit. Heilbrunn points out that in stimulation there is a liquefaction of the cortex of cells with a release of calcium ions, followed by a calcium clotting or gelation of the cell interior. It has been found that in small concentrations organic anesthetics cause the release of calcium and thus stimulate the coagulating reaction, but in higher concentrations they prevent it and, accordingly, cause anesthesia. Magnesium acts by replacing calcium in the cortex and, when the tissue is stimulated, is released into the cell interior but is far less potent than calcium in producing coagulation and therefore has an anesthetic effect.

McLean and Hastings [20] have shown that approximately half of the calcium in normal human serum is bound to the protein and consequently is not diffusible through a collodion membrane. The other half,

[17] R. B. Dean, *Proc. Soc. Exptl. Biol. Med.*, **50**, 162 (1942).

[18] Meltzer and Auer, *Am. J. Physiol.*, **14**, 366 (1905); **16**, 233 (1906); **21**, 400 (1908).

[19] L. V. Heilbrunn, *An Outline of General Physiology*, W. B. Saunders Co., Philadelphia, Pa.

[20] F. C. McLean and A. B. Hastings, *J. Biol. Chem.*, **108**, 285 (1935).

non-protein bound, is almost totally ionized. In fact, they were able to show that the ionization of calcium, in protein-containing fluids, is determined by a chemical equilibrium between calcium and protein; as a first approximation, this equilibrium can be described by the simple equation

$$\frac{\text{Ca} \times \text{total protein}}{\text{Ca proteinate}} = K = 5.65 \qquad 16$$

where Ca^{++} is expressed in millimoles per kilogram of water and total protein in grams of protein per 100 cc. of serum. The calcium proteinate is obtained by subtracting the ionized calcium in millimoles per kilogram of water from total calcium, also expressed in millimoles per kilogram of water.

Calcium in the ionic as well as in the non-ionic forms plays important biological roles. It is involved, for example, in blood coagulation, being necessary for the activation of the thrombin system. It is involved, also, in the coagulation of milk, being in this action the precipitating ion. Calcium is a major structural element of mammals.

The formation of bones and of teeth presents interesting problems in physical chemistry. The product $(Ca^{++})^3 \times (PO_4^{---})^2$ in blood and other body fluids is normally in the neighborhood of 3.2×10^{-24}. If this product exceeds 3.2×10^{-24}, spontaneous precipitation occurs; once precipitation has been initiated, it will continue until the product is diminished to 10^{-27}. Of great significance in bone formation is the enzyme phosphatase, which occurs particularly in bone and teeth, and is capable of hydrolyzing organic phosphoric acid esters.

Low calcium results in tetany. For a clinical discussion of calcium ions, see McLean and Hastings.[21]

The specific physiological effects of ions are numerous. It has been found, for example, that Mg^{++} and Ca^{++} activate acetylcholine esterase.[22] The esterase is an enzyme which brings about the hydrolysis of the powerful physiological agent, acetylcholine. On the other hand K^+ inhibits the enzyme, and there is an antagonism between Ca^{++} and K^+ and also between Mg^{++} and K^+. The presence of Ca^{++} or Mg^{++} increases the activity of the enzyme even in the presence of K^+.

[21] F. C. McLean and A. B. Hastings, *Am. J. Med. Sci.*, **189**, 60 (1935).

[22] B. Mendel, D. Mundell, and F. Strelitz, *Nature*, **144**, 479 (1939).

Chapter VI

ELECTROMOTIVE FORCE CELLS

Electromotive force cells were discovered by Galvani about 1780. He found that, if a metallic conductor was placed in contact with a freshly dissected frog muscle and a second conductor in contact with the nerve which innervated the muscle, the muscle twitched as if alive when the two metallic conductors were connected. This is one of the first of many instances in which biological investigation has furthered the progress of the physical sciences.

The interest of the biochemist in electromotive force cells centers largely in two of their aspects. The first is their analytical usefulness; i.e., the activity of various electrolytes may be determined by means of the appropriate cell. The electrometric determination of hydrogen ions is an example of such a use to which cells have been put by the biochemist. The second is that they offer, on occasion, a close analogy to bioelectric potentials.

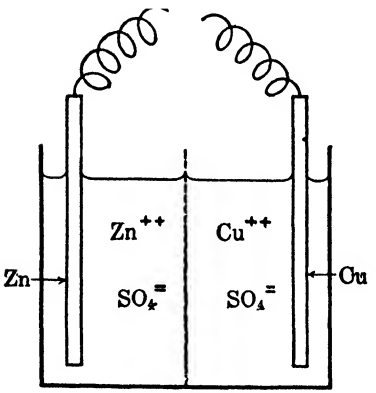


FIG. 1. Daniell cell.

One of the oldest and simplest cells is the Daniell cell, shown in Fig. 1. It consists of a strip of zinc (zinc electrode) immersed in a solution of zinc sulfate and of a strip of copper (copper electrode) immersed in a solution of copper sulfate, the two solutions being separated by a porous diaphragm.

If the two metallic electrodes are connected by an outside wire, zinc ions are formed from the metallic zinc and copper ions deposit on the copper electrode as metallic copper. This chemical reaction leaves the zinc negatively charged and produces a positive charge on the copper. Naturally, if the two metal strips are not connected, few zinc ions can be formed from the metallic zinc and very little metallic copper will be deposited. Such a process would quickly build up an electrical potential

on the plates and stop the process from proceeding. When the electrodes are connected by an outside wire, the chemical reaction involved is

$$Zn + Cu^{++} \rightarrow Zn^{++} + Cu$$

In order to measure the potential of such a cell an arrangement must be used which will draw little or no current from the cell; otherwise, the voltage which we measure will be too small and will not represent the voltage while the cell is performing reversibly. The measurement of the maximum voltage of a cell is usually accomplished with a potentiometer. A potentiometer is an instrument which can be adjusted so that the voltage of the cell which is being measured is exactly balanced with a voltage of opposite sign, so that no current flows into or out of the cell. A diagrammatic sketch of a potentiometer circuit is shown in Fig. 2.

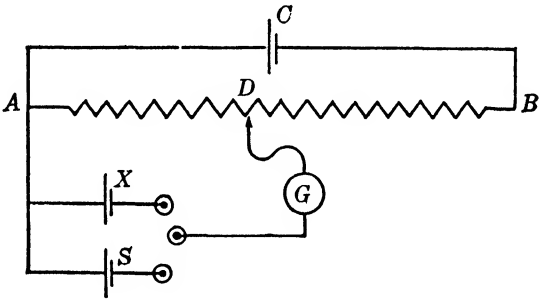


FIG. 2. Potentiometer circuit.

AB is a uniform resistance wire; C is a source of constant voltage (usually a storage battery); D is a sliding contact; G is a sensitive galvanometer; X is a cell whose electromotive force is desired; S is a standard cell whose voltage is accurately known (usually a Weston cell). The slide wire D is adjusted with cell X in circuit until no current flows through the galvanometer G. The cell S is then put into the circuit, and D is again adjusted until no current flows. The electromotive force of cell X is then calculated by proportion, using the positions of D and the value of the electromotive force of cell S.

When the voltage of the galvanic cell shown in Fig. 1 is determined in the manner described, we have an accurate measure of the free energy involved in the reaction between copper ions and the zinc. The electrical work done on a mole basis is equal to the total charge transferred from the copper to the zinc multiplied by the voltage at which the transfer is made. The electrical work is exactly equal to the free energy of the process, so that

$$\Delta F = -n\mathbf{F}E$$

where n is the number of equivalents involved (for the reaction pictured above n is 2), and **F** is the number of coulombs of electricity transferred. It has been found by experiment that to convert 1 equivalent of an element to an equivalent of ions requires 96,500 coulombs of electricity. Accordingly, **F** is equal to 96,500. E is the voltage which arises from the reaction between the zinc and copper ions. Other potentials are involved in the Daniell cell besides that due to the simple electronic transfer pictured above, but for the moment let us neglect these complicating voltages. When the above constants are substituted in equation 1, the free energy is obtained in joules. In order to convert to calories we must remember that 1 calorie is equal to 4.185 joules. ΔF in calories is, therefore, given by

$$\Delta F = -23{,}060nE \qquad 2$$

Copper in the presence of copper ions or zinc in the presence of zinc ions or, for that matter, any element in the presence of its ions is called a half cell. Obviously it is always necessary to have two half cells connected together to realize a potential which is capable of being measured. The question arises how much of the total voltage of the Daniell cell is due to the zinc electrode and how much to the copper. The actual magnitudes of these half-cell voltages cannot be measured, but the next best thing has been done. The hydrogen half cell has been selected as the standard and is defined as having zero voltage at unit activity of hydrogen ions under 760 mm. of mercury pressure of hydrogen gas and at any temperature. We have reserved a separate chapter for the consideration of hydrogen ions and the hydrogen electrode, and this topic will be discussed in more detail at that point. The voltage of all other half cells (elements in contact with a solution of their ions) is given in reference to the standard hydrogen electrode, which, as indicated above, is defined as having zero voltage. Each element has its own normal half cell and is defined as a cell consisting of the element in its standard state in a solution of its ions at unit activity and at 25° C.

The sign of the electrode presents some minor difficulties. There are a series of conventions or rules which determine the sign to be used. The author has to confess, however, that he forgets these conventions as soon as he has learned them. One can usually guess from his knowledge of chemistry what the sign of the electrode should be; the more active metallic element tends to give off cations and, therefore, becomes negative. The sign of the electrode is always that which is actually measured by the potentiometer, and potentiometers are always stamped with the proper sign.

Some Useful Half Cells

Although the hydrogen half cell is the standard reference electrode, other half cells are more convenient to use. Accordingly, the potentials of these more convenient cells have been carefully compared with that of a standard hydrogen half cell and generally are used in place of the hydrogen cell. The calomel half cell and the silver-silver chloride half cells have found wide favor in this connection. The calomel half cell consists of metallic mercury in contact with a KCl solution saturated with mercurous chloride. Actually, there are three calomel half cells, known as the saturated calomel electrode, the normal calomel electrode, and the tenth normal calomel electrode. The saturated calomel half cell employs a saturated solution of KCl; the normal and the tenth normal cells contain normal and tenth normal KCl, respectively. The potentials of the saturated and of the tenth normal calomel electrodes are:

$$\text{Buffer} \,||\, \text{Sat. KCl} \,|\, 0.1\,N\,\text{KCl} \,|\, \text{HgCl} \,|\, \text{Hg} = +0.3355$$

$$\text{Buffer} \,||\, \text{Sat. KCl} \,|\, \text{HgCl} \quad |\, \text{Hg} \quad = +0.2443$$

These voltages are given for 25° C. The calomel is usually connected to its companionate half cell by means of a saturated KCl salt bridge, the saturated KCl solution being held in an agar gel.

A silver-silver chloride electrode may be prepared in various ways.[1] One method is as follows: A platinum wire is covered with silver by electrolysis (4 volts, 1 milliampere) of a solution obtained by mixing equal volumes of 13 per cent KCN solution and 18 per cent $AgNO_3$ solution. After washing, the silver-plated platinum wire is electrolytically covered with a thin layer of AgCl, a normal solution of HCl and a current of 3.5 milliamperes being used for 20 minutes.

In addition to serving as a non-polarizable electrode in the measurement of bioelectric potentials and other potentials, the silver-silver chloride electrode can be used as a chloride electrode and the activity of a chloride-containing electrolyte can be determined with it. As we know, silver chloride is relatively insoluble; accordingly, the amount of silver ions in solution depends upon the ion product constant of silver chloride, i.e.,

$$\text{Ag}^+ \times \text{Cl}^- = K_{\text{AgCl}} \qquad 3$$

Since the voltage of a silver-silver chloride electrode is given by

$$E = E_0 + \frac{RT}{\text{F}} \ln \text{Ag}^+ \qquad 4$$

[1] W. R. Carmody, *J. Am. Chem. Soc.*, **51**, 2901 (1929); **54**, 188 (1932).
M. Randall and L. E. Young, *J. Am. Chem. Soc.*, **50**, 989 (1928).

we can substitute relation 3 in this equation and obtain

$$E = E_0 + \frac{RT}{F} \ln \frac{K_{AgCl}}{Cl^-} \qquad 5$$

or

$$E = E_1 + \frac{RT}{F} \ln \frac{1}{Cl^-} \qquad 6$$

where E_1 is a constant involving the value of the logarithm of the ion product constant of silver chloride K_{AgCl} as well as the standard electrode potential E_0 of the silver-silver chloride electrode.

Half cells of sodium and of potassium have been employed. The chemical reactivity of these elements necessitates the use of sodium or of potassium amalgam since the naked elements quickly react with water and spoil the electrode. Actually, the amalgam electrodes have a very limited biological application.

Calcium ions play very important biological roles, and it would greatly facilitate research on calcium and calcium ions if a proper calcium electrode could be devised. Calcium, like sodium and potassium, is so active that a calcium amalgam must be used in place of metallic calcium. Joseph [2] has devised a calcium electrode in which the calcium amalgam is separated from the solution under investigation by a cellophane membrane. Such an electrode seems to work as long as the solutions investigated contain no diffusible material which reacts with the amalgam. This condition is a serious objection, however, severely limiting the usefulness of such an electrode.

Joseph [3] has more recently proposed the use of lead amalgam in contact with lead oxalate and calcium oxalate. This electrode involves the solubility-product constant of both lead oxalate and calcium oxalate. The equation for this cell is

$$E = E_0 - \frac{3RT}{2F} \ln A_{CaCl_2} \qquad 7$$

where the mean activity of $CaCl_2$ is

$$A_{CaCl_2} = \sqrt[3]{A_{Ca^{++}} \times A^2_{Cl^-}} \qquad 8$$

This electrode has two requirements: (1) there must not be anions other than $C_2O_4^{--}$ which form insoluble lead salts, and (2) there must not be other cations than calcium which form insoluble oxalates.

[2] N. R. Joseph, *J. Biol. Chem.*, **126**, 389 (1938).
[3] N. R. Joseph, *J. Biol. Chem.*, **130**, 203 (1939).

Tendeloo [4] has attempted to use a thin membrane of fluorite to separate a solution of known calcium-ion concentration from a solution whose calcium-ion concentration is unknown. This arrangement is exactly analogous to that of the glass electrode which is used to determine hydrogen ions. The fluorite electrode appears to be unsatisfactory, however.[5]

McLean and Hastings [6] have developed a biological method for the estimation of calcium ions. The method consists essentially in direct comparison and matching of known solutions with biological fluids of unknown calcium-ion concentration, the criterion for equal calcium-ion concentration being the production of equal amplitude of contraction of the ventricle of the isolated heart of the frog. The contractions are recorded on a rotating smoked drum.

Actually, as pointed out in Chapter IV, the calcium which is diffusible through a collodion membrane of the proper porosity is probably almost entirely in the ionic form; accordingly, the calcium-ion concentration can be fairly closely estimated in a biological fluid by ultrafiltration through a properly prepared collodion membrane followed by a calcium determination of the ultrafiltrate.

It must be emphasized that in measuring potentials in any solution, biological or otherwise, non-polarizable electrodes are required. For example, suppose that one wished to measure the potential across a frog's skin which is bathed on both sides by KCl solutions. On first thought it might seem sufficient to plunge two platinum wires into the two KCl solutions separated by the frog's skin and measure the potential directly. This, however, is not possible. A potential difference will be observed between the platinum wires, but it will be very erratic and bear little relation to the actual potential difference between the two solutions. The proper technique is to place two silver-silver chloride electrodes in the KCl solutions, or connect the two solutions by salt bridges to two identical calomel electrodes, and measure the difference in potential between the non-polarizable electrodes. The reason why the platinum wires will not and the non-polarizable electrodes will yield the true potential difference should be clear. The platinum wires provide no mechanism for the flow of electricity from the metal to the solution or from the solution to the metal. There is no possibility of platinum ions being formed from the wire or for the deposition of ions from the solution on the wire as metallic potassium or as chlorine gas. Such an electrode is quickly polarized.

[4] H. J. C. Tendeloo, *J. Biol. Chem.*, **113**, 333 (1936).
[5] D. M. Greenberg and C. E. Larson, *J. Biol. Chem.*, **115**, 769 (1936)
R. S. Anderson, *J. Biol. Chem.*, **115**, 323 (1936).
[6] F. C. McLean and A. B. Hastings, *J. Biol. Chem.*, **107**, 337 (1934).

Concentration Cells

If, for the copper electrode in the Daniell cell, Fig. 1, we substitute another zinc electrode, and for the $CuSO_4$ solution a solution of $ZnSO_4$ at a concentration different from that in compartment 1, Fig. 1, we have what is known as a concentration cell. A potential difference can be observed between the two zinc electrodes. It is easy to see, in a qualitative manner, how such a potential could and would arise. Let us suppose that the $ZnSO_4$ solution in compartment 1 is more concentrated than that in compartment 2. Under these circumstances the tendency for zinc ions to be formed from the metallic zinc would be less in compartment 1 than in compartment 2, and therefore the zinc in compartment 1 would be negative to the zinc in compartment 2. As we have seen in Chapter I, the free energy experienced in diluting a solution from activity 1 to activity 2 is

$$\Delta F = RT \ln \frac{A_2}{A_1} \qquad 9$$

and since

$$\Delta F = -n\mathbf{F}E \qquad 10$$

we have

$$E = -\frac{RT}{n\mathbf{F}} \ln \frac{A_2}{A_1} \qquad 11$$

and at 25° C.

$$E = -\frac{0.059}{n} \log \frac{A_2}{A_1} \qquad 12$$

Knowing the activity of zinc sulfate in either compartment, we can, after measuring the potential difference between the zinc electrodes, calculate the activity of the zinc sulfate in the other compartment. It should be noted, however, that what we measure in this, as well as in other electromotive force cells, is not the individual ionic activities but a mean activity of the positive and negative ions. That is,

$$\text{Mean activity} = \sqrt{A^+ \times A^-} \qquad 13$$

It is impossible at the present time to measure individual ion activities.

The arrangement pictured above of two zinc-zinc sulfate half cells connected together is known as a concentration cell; in fact, all cells which include two half cells of the same element but at different ionic concentrations are known as concentration cells.

Liquid Junction Potential

The potential differences which we observe in concentration cells such as pictured above are not exactly given by equation 11. We have in

addition a potential difference at the junction of the two liquids which is technically known as a liquid junction potential. The liquid junction potential is essentially due to differences in ionic mobility. For example, in the concentration cell involving zinc sulfate, the zinc sulfate from the more-concentrated solution diffuses into the more-dilute solution. The mobility of the sulfate ion is about one and one-half times greater than that of the zinc ion; the result is that the concentrated zinc sulfate solution tends to be positive in respect to the more-dilute solution.

Henderson [7] derived an equation which expresses the liquid junction potential as a function of the concentration and mobility of the ions.

If the two solutions contain a single univalent salt at two concentrations C_1 and C_2, Henderson's equation reduces to

$$E_L = \frac{RT}{\mathbf{F}} \frac{u^+ - u^-}{u^+ + u^-} \ln \frac{C_1}{C_2} \qquad 14$$

where u^+ and u^- are the mobilities of cation and anion, respectively.

Naturally, if u^+ equals u^- the liquid junction potential becomes zero. This condition is nearly realized with solutions of KCl, and it is the basis for the use of KCl in salt bridges connecting half cells. This is not an entirely unambiguous procedure, but it does represent the best, under certain circumstances, that can be done.

The liquid junction potential can be greatly increased if the two electrolyte solutions are separated by a very dense membrane with small pore sizes, such as well-dried collodion membranes. Michaelis [8] found that, when 0.01 *M* KCl was separated from 0.1 *M* KCl solution by such a well-dried collodion membrane, a potential difference was established across the membrane which amounted, where the membrane pore size was very small, to 45 to 55 millivolts. The dilute solution was always positive in respect to the concentrated solution. Evidently, the rate of diffusion of the anion had been greatly diminished in comparison with that of the cation. We shall discuss the cause of this differential effect on the mobility of cations and anions in Chapter XVII; it is sufficient for our present purposes to note that it exists.

Applying the formula for the liquid junction potential to the extreme situation where the collodion membrane is completely impermeable to the chloride ion, we have

$$E_L = -\frac{u^+}{u^+} \frac{RT}{\mathbf{F}} \ln \frac{C_2}{C_1} \qquad 15$$

[7] P. Henderson, *Z. physik. Chem.*, **59**, 118 (1907).

[8] L. Michaelis, *Colloid Symposium Monograph*, V, 135 (1928); *Kolloid Z.*, **62**, 21 (1933).

and at 25° C.

$$E_L = -0.059 \log \frac{C_2}{C_1} \qquad 16$$

When C_2 equals $10C_1$, E will be equal to 59 millivolts. It can be seen that the more dense collodion membranes of Michaelis approached closely the theoretical impermeability to anions.

Bioelectric Potentials

The subject of bioelectric potentials is an important one for the physiologist and the biochemist. These potentials undoubtedly arise from a variety of causes, among which may be mentioned: (1) membrane potentials due to a Donnan equilibrium; (2) diffusion potentials, i.e., liquid junction potentials; (3) electrokinetic potentials due to motion of charged particles or fluids; and (4) phase potentials resulting from contact between a solid and liquid phase, or between two liquid phases, or between two solid phases. The potentials resulting from Donnan equilibria and from motion of charged particles or of fluids will be dealt with in later portions of this book. At the moment, let us consider the potentials arising from diffusion and from phases in contact.

The ionic composition of a living cell differs, in general, considerably from that of the surrounding fluids. It is true that living cells are in osmotic equilibrium with fluids in contact with them, but the distribution of individual ions is usually far from equal. This difference in ionic composition is maintained either through constant metabolism on the part of the cell or through selective impermeability of the cell membrane. The mammalian red cell, for example, shows a very limited permeability to cations. If a living cell is injured so that the membrane is destroyed or the metabolism of the cell interrupted, the electrolytes diffuse into or out of the cell, depending upon the direction of the concentration gradients. This diffusion of ions gives rise to the injury potentials. It has been found that injured tissue is almost always negative to the uninjured tissue, although positive injury potentials can be obtained if the solution which bathes the tissue has the required composition.

Steinbach [9] has investigated injury potentials of the adductor muscle of the common scallop. In his experiments, he exposed a cross section of the adductor muscle and measured the difference in potential between the injured and uninjured parts. The experiments were carried out by placing the tip of a silver-silver chloride electrode in the approximate center of the cut surface of the muscle and the other silver-silver chloride electrode on the shell at the base of the muscle cylinder. Injury poten-

[9] H. B. Steinbach, *J. Cellular Comp. Physiol.*, **3**, 203 (1933).

tials of 30 to 40 millivolts were found. These potentials decreased with time until in about an hour they had disappeared.

The effects of various electrolytes in contact with the injured portion were studied. Calcium chloride solutions greatly decreased the time of recovery, whereas potassium chloride solutions prolonged it. Sodium and magnesium ions were intermediate in effect between potassium and

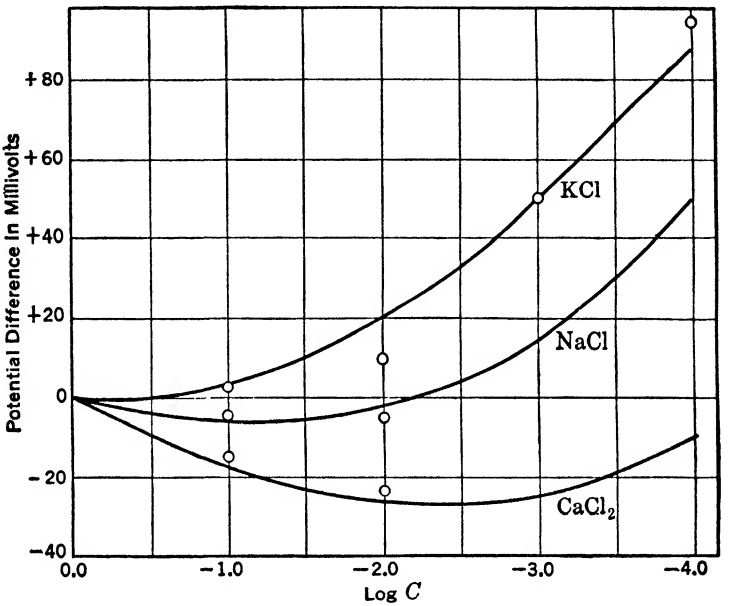


FIG. 3. Potentials of the scallop adductor muscle as a function of the electrolyte concentration of the bathing solution. The points are experimental, the potential being taken as zero at isotonic bathing solution. The curves are calculated from Henderson equation, assuming the presence of an anion in the tissue whose mobility is 10. (Steinbach.)

calcium. It is believed that the reason why calcium speeded recovery is that calcium ions bring about precipitation or coagulation of a membrane which protects the injured tissue. Later Steinbach studied diffusion potentials across uninjured scallop adductor muscle. In this instance the muscle was bathed in the electrolyte solution under investigation, and the complete electromotive force cell which he employed was

Non-polarizable electrode	Saturated KCl	Electrolyte solution	Tissue	Non-polarizable electrode

By assuming an anion in the tissue of low mobility he was able to use the Henderson liquid junction equation to predict the observed potentials. Figure 3 shows a plot of some of his data.

The observed potentials of the scallop muscle were considerably higher than could arise from the free diffusion of ordinary, inorganic electrolytes. As indicated above, Steinbach had to assume the presence of a slow-moving anion in the tissue. He believes this anion to be protein, and in a later paper [10] he studied the diffusion potentials resulting from diffusion of ions in gelatin gels. In this system he was able to show a behavior which closely paralleled that of the scallop muscle. He came to the conclusion that diffusion potentials can account quantitatively for injury potentials provided that the slow-moving anion of the tissue is considered.

Action currents are closely related to, if not identical in origin with, injury potentials. Action currents are associated with any stimulation of protoplasm. Heart waves as observed in an electrocardiogram are an example of such action currents. The most elegant and dramatic example of an action current is that produced by an electric eel. It has been suggested that action currents arise as a result of a polarization of the cell membrane. This polarization results from a differential permeability of the membrane to ions. Upon stimulation the differential permeability is momentarily lost and a flow of ions across the membrane results; a flow of current is thereby produced.

Phase Potentials

When two phases are brought into contact a difference of potential between the phases results. Two metals in contact, for example, will give rise to a potential difference; this is the basis of the thermocouple which is used for temperature measurements, the potential difference being a function of the temperature. The two metal phases show a different "solubility" for electrons. We shall see in the chapter on electrokinetics that phase potentials are of a very general occurrence and arise not only with metals in contact but also with non-metals and can involve solids, liquids, and gases. The difference in potential may be due to a flow of ions from one phase to another. Since in living tissue there are many phase separations (interfaces), such potentials are of common physiological occurrence. There are appropriate techniques for measuring such potentials, as we shall see in the chapter on electrokinetics. The question arises whether or not they can be measured by any of the techniques which we have outlined in the present chapter.

There have been a great many measurements of the electromotive cell of the type

Ag·AgCl | Salt solution | "Oil" | Salt solution | AgCl·Ag

[10] H. B. Steinbach, *J. Cellular Comp. Physiol.* **7**, 291 (1935).

in which the salt solutions contain the same electrolyte at different concentrations or they may contain different salts. By "oil" is meant some substance which is immiscible with water. Investigations along this line led to the discovery of the glass electrode.[11] Beutner has published a monograph on such measurements.[12] He felt that the potential difference observed in such an oil electrode results from a "solubility" of the ions in the oil. He emphasized the distribution coefficient of the ions between the oil and water phase. As it is not possible to obtain independent evidence for such a distribution, Beutner's theory cannot be put to test. Beutner attempted to explain injury potentials by analogy with the oil potentials, but the connection is far from clear.

Ehrensvärd and Sillen [13] believe that such oil potentials arise from an adsorption of ions at the oil surface. Their oil surfaces contained such groups as CN, CHO, OH, NO_2, Cl, NH_2. If these groups are negative they would attract cations, and if positive, anions. The adsorbed ion is supposed to be the potential-determining ion. Ehrensvärd and Sillen propose an equation for these oil potentials which they claim is in better agreement than any based upon liquid junction potentials.

Craxford, Gatty, and Rothschild [14] point out that adsorption potentials can affect the electromotive force of a cell only (1) when the interface contains a boundary across which no charged particles can pass, and (2) when diffusion of ions takes place across an interface which is not in a state of equilibrium.

Condition 1 characterizes a condenser, which could hardly be expected to yield steady and reproducible potentials. It is possible that condition 2 does influence the potential at the interface.

Dean [15] conducted some experiments on oil surfaces with and without a layer of egg albumin. He came to the conclusion that, unless the interfacial monolayer has a resistance to anions or to cations comparable with that of the bulk phase, the surface film cannot affect the potential when a steady state is attained.

It would seem that the best explanation for the oil electrode potentials is to be found in considerations which we have discussed under liquid junction potentials.

[11] M. Cremer, *Biochem. Z.*, **47**, 562 (1906).

[12] R. Beutner, *Die Entstehung elektrischer Ströme in lebenden Geweben*, Verlag von Ferdinand Enke, Stuttgart, 1920.

[13] G. Ehrensvärd and L. G. Sillen, *Nature*, **141**, 788 (1938).

[14] S. R. Craxford, O. Gatty, and Rothschild, *Nature*, **141**, 1098 (1938).

[15] R. B. Dean, *Nature*, **144**, 32 (1939).

Cells without Liquid Junctions

Although, as we have seen, liquid junction potentials have an inherent interest, they introduce uncertainties in electromotive force measurements which are disturbing. The difficulty of the liquid junction can be circumvented by making a cell without a liquid junction. For example, Nims and Smith [16] used the cell shown in Fig. 4.

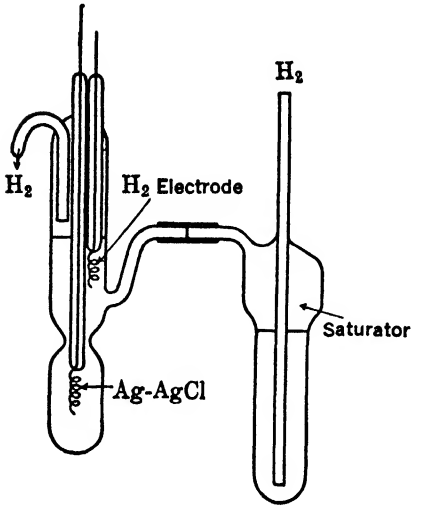


FIG. 4. Cell without liquid junctions. (Nims and Smith.)

Because of the very low solubilities of hydrogen and of silver chloride, the solution between the electrodes is practically of uniform composition, even though the hydrogen and the silver chloride must be kept in separate parts of the containing vessel. Such a cell is known as a cell without a liquid junction.

If we have two such cells as are pictured in Fig. 4 containing HCl at different concentrations, the difference in their voltages will evidently be the voltage of a true concentration cell, and we have

$$E_2 - E_1 = E = \frac{2RT}{\mathrm{F}} \ln \frac{A_2}{A_1} \qquad 17$$

where A_2 and A_1 are activities of HCl in two cells.

This is the type of measurement which is meant when reference is made to a concentration cell without a liquid junction. The distinctive feature of a cell without liquid junction is that two types of electrodes are involved, each of which must be reversible to an ion constituent in the solution contained in the cells.

[16] L. F. Nims and P. K. Smith, *J. Biol. Chem.*, **113**, 146 (1936).

Chapter VII

ACIDS AND BASES

Brönsted [1] has formulated a generalized theory of acids and bases. Instead of thinking of an acid as a substance which dissociates into hydrogen ions and a base as dissociating into hydroxyl ions, Brönsted conceives of acids as substances which give off protons, the proton being the unhydrated hydrogen ion. A base, on the other hand, is defined as a substance which can unite with protons, i.e., a proton acceptor. The Brönsted theory may be expressed by the following equation:

$$A \rightleftarrows H^+ + B$$

where A is an acid and B is a base.

The purpose of the theory is to extend the concept of acids and bases to non-aqueous media. For instance, sodium acetate dissolved in glacial acetic acid bears the same relation to acetic acid as NaOH, dissolved in water, bears to water. In both solutions a titration by a strong acid yields the solvent. For example,

$$\text{Na Acetate} + HCl \rightarrow \text{Acetic acid (solvent)} + NaCl$$

and

$$NaOH + HCl \rightarrow \text{Water (solvent)} + NaCl$$

Solvents may be divided into four classes:

1. The so-called aprotic solvents, such as benzene, toluene, paraffin, and chloroform. These solvents cannot yield or accept protons.

2. Acidic solvents such as anhydrous acetic, sulfuric, and hydrochloric acids. These solvents can yield but cannot accept protons. For example,

$$HF \text{ (acidic solvent)} + HNO_3 \text{ (base)} \rightarrow H_2NO_3^+ + F^-$$

3. Basic solvents such as anhydrous liquid ammonia and aniline. These solvents can accept protons but ordinarily cannot produce them. For example, in liquid ammonia we have

$$CH_3CONH_2 \text{ (acid)} + KNH_2 \text{ (base)} \rightarrow CH_3CONHK + NH_3$$

[1] J. N. Brönsted, *Chem. Revs.*, **5**, 231 (1928).

4. Amphiprotic solvents such as water and the alcohols. These solvents can both accept and give off protons. For example,

$$H_2O \rightarrow H^+ + OH^- \text{ (as an acid)}$$

or

$$H_2O + H^+ \rightarrow H_3O^+ \text{ (as a base)}$$

Similarly for an alcohol such as methyl alcohol

$$CH_3OH \rightarrow H^+ + CH_3O^- \text{ (as an acid)}$$

or

$$CH_3OH + H^+ \rightarrow CH_3OH_2^+ \text{ (as a base)}$$

Although our interest naturally centers on aqueous solutions, it is well to remember that the action of acids and bases in water is a particular manifestation of a more general phenomenon.

Ionization of Water

We know that water dissociates into hydrogen and hydroxyl ions, i.e.,

$$H_2O \rightleftarrows H^+ + OH^-$$

which can be expressed as

$$K_A = \frac{A_{H^+} \times A_{OH^-}}{A_{H_2O}} \qquad 1$$

where A_H, A_{OH}, and A_{H_2O} denote the activities of the hydrogen ions, hydroxyl ions, and water molecules, respectively. Since in any ordinary

TABLE 1

K_{WA} AS A FUNCTION OF TEMPERATURE (HARNED AND HAMER)

Temperature, ° C.	$K_{WA} \times 10^{14}$
0	0.115
10	0.293
20	0.681
25	1.008
30	1.471
40	2.916
50	5.476
60	9.614

aqueous solution the activity of water is so much greater than that of either the hydrogen or hydroxyl ions, we may assume that its activity is constant, or

$$A_{H^+} \times A_{OH^-} = K_{WA} \qquad 2$$

where K_{WA} is a constant which involves the dissociation constant of water as well as the activity of water.

The ionization constant of water has been accurately studied by Harned and Hamer,[2] using the appropriate electromotive force cell. Their results, given in Table 1, show how the ion activity product constant of water varies with the temperature.

The pH Scale

In living tissue, we deal with extremely dilute acids and bases. For example, the hydrogen-ion concentration of the human body is about 0.000,000,05 N. Such a number is awkward to handle; more important still is the difficulty of comparing this value with stronger or weaker acids. Sometimes such a figure is expressed in powers of 10, that is, we write 5×10^{-8} instead of 0.000,000,05. This method is still not entirely suited to our needs. The term pH was suggested by Sörensen [3] to help us out of this difficulty. Sörensen defined pH as

$$p\text{H} = -\log C_{\text{H}^+} = \log \frac{1}{C_{\text{H}^+}} \qquad 3$$

and likewise

$$p\text{OH} = -\log C_{\text{OH}^-} = \log \frac{1}{C_{\text{OH}^-}} \qquad 4$$

where C_{H} denotes the hydrogen-ion concentration and C_{OH} the concentration of hydroxyl ions.

Sörensen assumed that the concentration of hydrogen ions could be calculated from conductance measurements using Arrhenius' theory of dissociation. In this way he concluded that the hydrogen-ion concentration of a 0.1 N HCl solution was 0.09165, which corresponds to a pH of 1.038; similarly, the hydrogen-ion concentration of a 0.01 N HCl solution in 0.09 N KCl was calculated to be 0.009165 or a pH of 2.038. We now know that it is impossible to calculate accurately the hydrogen-ion concentration from conductance measurements with the methods used by Sörensen. The question is not a simple one. The correct formulation of pH would not be so important if we used the pH scale only for comparative purposes, but, since the hydrogen electrode is defined as the standard against which all other electrodes are compared, it becomes imperative for the purposes of thermodynamic calculations to formulate and calibrate the pH scale as accurately as possible.

Since the hydrogen electrode measures the hydrogen-ion activity

[2] H. S. Harned and W. J. Hamer, *J. Am. Chem. Soc.*, **55**, 2194 (1933).

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

rather than the hydrogen-ion concentration, a more useful definition of pH than that proposed by Sörensen is

$$p\text{H} = \log \frac{1}{A_{\text{H}^+}} \qquad 5$$

and likewise

$$p\text{OH} = \log \frac{1}{A_{\text{OH}^-}} \qquad 6$$

Expressions 5 and 6 are still not entirely unambiguous. As pointed out by Guggenheim,[4] single ion activities cannot be measured. What we really measure is the mean ionic activity, i.e., $\sqrt{A^+ \times A^-}$. In what is to follow, we shall speak of the activity of the hydrogen ion, although we must realize that what we are really talking about is the mean ionic activity.

The expressions for pH and pOH have been defined in terms of the activities of the hydrogen and hydroxyl ions, respectively. It now remains to calibrate the pH scale. That is to say, we must have some solution whose exact hydrogen activity is known. As we can deal with this problem more effectively after we have discussed the dissociation of weak acids, the calibration of the pH scale will be described a little later.

The meaning and nature of the pH scale is rather obvious, but perhaps it is well to point out two things about the scale. (1) It is not arithmetical; i.e., when a solution is brought from pH 7 to pH 6 the arithmetical increase in hydrogen-ion activity is only one-tenth of the increase in going from pH 6 to pH 5. (2) The scale is in inverse sense to the hydrogen-ion activity; i.e., the lower the pH, the higher is the hydrogen-ion activity.

As we have seen from Table 1, at 25° C. the ion activity product of water is very nearly 1×10^{-14}. Taking the logarithm of both sides of the expression for the ion activity product constant (equation 2) and multiplying by -1, we have very nearly

$$p\text{H} + p\text{OH} = pK_w = 14 \qquad 7$$

At pH 7 the pOH is also 7, and the reaction of water is acidic or basic to the same extent. Although absolutely pure water should have a pH of 7, the pH of the most carefully prepared water will seldom be exactly 7, owing to slight contamination.

The hydrogen ion is hydrated in solution; naked protons do not exist in water as such. The hydrated hydrogen ion is written H_3O^+ and is called the hydronium ion. This fact, however, does not change any of

[4] E. A. Guggenheim, *J. Phys. Chem.*, **33**, 842 (1929).

our considerations, and we shall continue to speak of the hydrogen ion although it is understood that we mean the hydronium ion. This is a convention to which we shall adhere.

Ionization of Weak Acids

Some acids and bases are completely dissociated, such as, for example, the first hydrogen of sulfuric, nitric, and hydrochloric acids, and sodium, potassium, and the alkali-earth hydroxides. We have already discussed in the chapter entitled "Ions in Solution" electrolytes which are completely ionized. Here we are interested in the so-called weak acids and bases, i.e., acids and bases which do not ionize completely.

Consider a weak acid such as acetic acid dissolved in water

$$HA \rightleftarrows H^+ + A^-$$

It will be evident that, though the above dissociation can be expressed in terms of an equilibrium constant, two such constants will be obtained, depending upon whether we use ionic concentrations or ionic activities. Thus, using concentrations, we have

$$K_c = \frac{C_{H^+} \times C_{A^-}}{C_{HA}} \qquad 8$$

or using activities

$$K_A = \frac{A_{H^+} \times A_{A^-}}{A_{HA}} \qquad 9$$

The relation between these constants is given by

$$K_A = \frac{\gamma_{H^+}\gamma_{A^-}}{\gamma_{HA}} \cdot K_c \qquad 10$$

where γ_{H^+}, γ_{A^-}, and γ_{HA} are the activity coefficients of the hydrogen ions, anions, and undissociated acid, respectively. The value of K_A is independent of electrolyte concentration; that of K_c will vary with electrolyte concentration. K_A is called the thermodynamic or true acid dissociation constant, and K_c is called the stoichiometric constant. It is evident that K_A is by far the more useful and fundamental constant. Not infrequently, a confusion arises as to which constant is meant, and tables of constants do not always distinguish one from the other.

The two methods most frequently used to evaluate acid and base ionization constants are (1) conductance measurement and (2) the electromotive force method. The older measurements were made almost exclusively through conductance determinations. As originally employed, this method does not yield true dissociation constants. With

modern refinements in calculations, however, this method is capable of giving quite as accurate results as the electromotive force method.

The conductance method depends upon the measurement of the conductance of a series of dilutions of a pure acid. The equivalent conductance is determined, and, since the conductance due to each ion is known, the ion product can be calculated. The technique of the calculation of the true dissociation constant from such data is described by MacInnes [5] and can be summarized as follows: As we have seen, for a weak acid

$$K_A = \frac{\gamma_{H^+} \cdot \gamma_{A^-}}{\gamma_{HA}} \cdot K_c \qquad 11$$

The activity coefficient of the anion is assumed equal to that of the hydrogen ion. The activity coefficient of the uncharged, unionized acetic acid does not depart significantly from unity. We can, therefore, write equation 11 as

$$K_A = \gamma_{H^+}{}^2 K_c \qquad 12$$

and, taking the logarithm of both sides,

$$\log K_A = \log K_c + 2 \log \gamma_{H^+} \qquad 13$$

From the Debye-Hückel theory, at 25° C. and at low ionic strengths,

$$-\log \gamma = 0.506\sqrt{C} \qquad 14$$

Substituting 14 in 13 leads to

$$\log K_A = \log K_c - 1.012\sqrt{C} \qquad 15$$

The value of K_c is determined by conductance measurements for a series of dilutions, and $\log K_c$, thus obtained, is plotted against $\sqrt{C}$.

Evidently, when C is zero, $\log K_c$ must equal $\log K_A$. The intercept on the y-axis is, therefore, equal to $\log K_A$. Such a plot for acetic acid is shown in Fig. 1.

The electromotive force method for acid dissociation constants was first formulated by Harned and Ehlers.[6] They employed a cell without a liquid junction which consisted of the following

$$H_2 \,|\, HA, NaAC, NaCl, AgCl \,|\, Ag$$

[5] D. A. MacInnes, *Cold Spring Harbor Symposia Quant. Biol.*, **I**, 190 (1931).
[6] H. S. Harned and R. W. Ehlers, *J. Am. Chem. Soc.*, **54**, 1350 (1932).

The equation for this cell is

$$E = E_0 - \frac{RT}{F} \ln \gamma_{H^+}\gamma_{Cl^-} m_{H^+} m_{Cl^-} \qquad 16$$

where m_{H^+} and m_{Cl^-} are the molalities of the hydrogen and the chloride ions, respectively. E_0 is the normal electrode potential of the cell

$$H_2 \mid HCl, AgCl \mid Ag$$

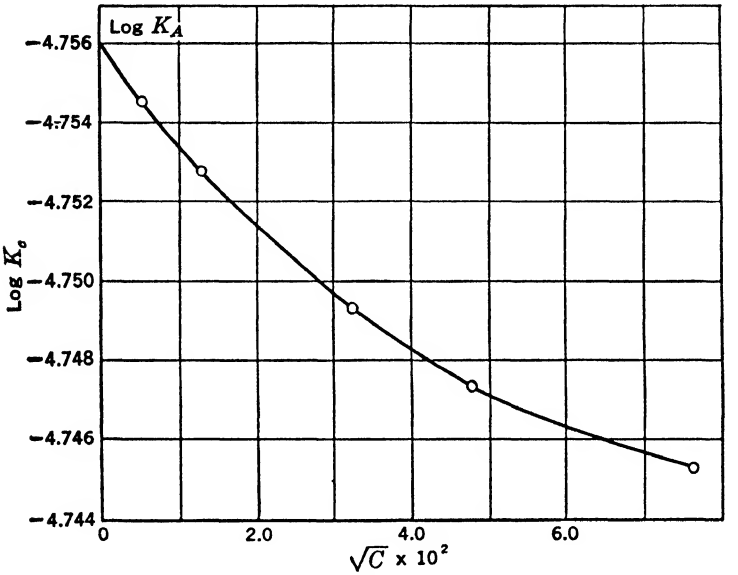


FIG. 1. Plot of log K_c obtained from conductance measurements against the square root of the acetic acid concentration.

Substituting in equation 16 the expression for the thermodynamic dissociation constant of a weak acid, and rearranging, there results

$$E - E_0 + \frac{RT}{F} \ln \frac{m_{HA} m_{Cl^-}}{m_{A^-}} = \frac{RT}{F} \ln \frac{\gamma_{H^+}\gamma_{Cl^-}}{\gamma_{H^+}\gamma_{A^-}} - \frac{RT}{F} \ln K_A \qquad 17$$

The first member of the right side of equation 17 becomes zero at infinite dilution; the activity coefficients approach unity as the concentration decreases. Accordingly, if the left side of equation 17 is plotted against the ionic strength, a very nearly straight line is obtained which can be extrapolated to zero ionic strength. The intercept at zero ionic strength is $-\frac{RT}{F} \ln K_A$, from which K_A can be computed.

The only unknown in the above is E_0, and this can be evaluated by a method suggested by Hitchcock.[7] The equation for the cell of which E_0 is the normal potential is

$$E_0 = E + \frac{4.6RT}{F} \log m_{HCl} + \frac{4.6RT}{F} \log \gamma_{HCl} \qquad 18$$

In moderately dilute solutions (see equation 13, page 81)

$$\log \gamma = -0.5\sqrt{m} + Bm \qquad 19$$

Combining 18 and 19 and rearranging, we have

$$E_0 - \frac{4.6RTB}{F} m_{HCl} = E - \frac{2.3RT}{F} \sqrt{m_{HCl}} + \frac{4.6RT}{F} \log m_{HCl} \qquad 20$$

The right side of equation 20 is plotted against m and the curve extrapolated to zero value of m. The intercept is equal to E_0.

There are thus two independent methods, conductance measurements and electromotive force methods, for the evaluation of the true dissociation constant of a weak acid. Both methods involve the principle that at infinite dilution the activity coefficient of all substances is unity and at this concentration the stoichiometric concentrations are equal to the activities, i.e. K_A equals K_c. The two methods agree quite well. Table 2 shows some thermodynamic dissociation constants determined by these methods.[8]

TABLE 2

THERMODYNAMIC DISSOCIATION CONSTANTS OF SOME WEAK ACIDS

Acid	25° C.	40° C.
Formic	1.772×10^{-4}	1.716×10^{-4}
Acetic	1.754×10^{-5}	1.703×10^{-5}
Chloroacetic	1.378×10^{-3}	1.229×10^{-3}
Propionic	1.336×10^{-5}	1.284×10^{-5}
n-Butyric	1.515×10^{-5}	1.395×10^{-5}
Phosphoric		
K_1	7.537×10^{-3}	7.152×10^{-3}
K_2	6.226×10^{-8}	6.349×10^{-8}
Bisulfate ion	12.0×10^{-3}	9.73×10^{-3}
Boric	5.80×10^{-10}	6.35×10^{-10}
Lactic	1.37×10^{-4}	1.367×10^{-4}

[7] D. I. Hitchcock, *J. Am. Chem. Soc.*, **50**, 2076 (1928).

[8] D. A. MacInnes, *The Principles of Electrochemistry*, Reinhold Publishing Corp., New York, 1939.

MacInnes and Belcher [9] have determined the first and second ionization constants of carbonic acid at 25° C. and at 38° C. using the glass electrode in place of the customary hydrogen electrode. They found K_1 to be 4.52×10^{-7} at 25° C. and 4.91×10^{-7} at 38° C., while K_2 was 5.59×10^{-11} at 25° C. and 6.25×10^{-11} at 38° C. Actually, the first dissociation constant is an apparent acid dissociation constant and, strictly speaking, is

$$K' = \frac{H^+ \times HCO_3^-}{CO_2 + H_2CO_3} \qquad 21$$

The actual dissociation constant of carbonic acid is, of course,

$$K_1 = \frac{H^+ \times HCO_3^-}{H_2CO_3} \qquad 22$$

It has been shown that, at equilibrium, the amount of dissolved CO_2 is about 1,000 times the amount of hydrated CO_2, i.e., carbonic acid; therefore (approximately)

$$K' = \frac{K_1}{1{,}000} \qquad 23$$

The true first dissociation constant of carbonic acid is, accordingly, about 4.5×10^{-4}. The hydration of CO_2 is a comparatively slow reaction, and it is only in the presence of the enzyme carbonic anhydrase that the reaction reaches equilibrium rapidly. Red cells of mammalian blood are rich in carbonic anhydrase.

Considerable work has been published on the influence of substituted groups on the ionization constants of acids. It is generally observed that the substitution of a strongly polar group greatly increases the strength of a carboxyl group. The closer the substituted polar group is to the carboxyl, the greater the influence on the dissociation of the carboxyl. For example, substituting a chlorine for a hydrogen atom in aliphatic acids gives the following pK_a values (pK_a is equal to log $1/K_a$):

Position of chlorine	α	β	γ	δ
pK_a	2.86	4.10	4.52	4.70

A special example of the effect of substitution is presented by the dicarboxylic acids. It can be shown from statistical considerations that the first dissociation constant must be at least four times greater than the second. As a matter of fact, all recent measurements have shown

[9] D. A. MacInnes and D. Belcher, *J. Am. Chem. Soc.*, **55**, 2630 (1933); **57**, 1683 (1935).

the first constant to exceed this limit considerably. Table 3 gives the ionization constants of some dicarboxylic acids.[10]

TABLE 3

IONIZATION CONSTANTS OF SOME DICARBOXYLIC ACIDS

Acid	Formula	$K_1 \times 10^5$	$K_2 \times 10^6$	K_1/K_2
Malonic	$HOOC(CH_2)COOH$	149.0	2.0	734
Succinic	$HOOC(CH_2)_2COOH$	6.4	3.3	19.2
Glutaric	$HOOC(CH_2)_3COOH$	4.5	3.8	11.9
Adipic	$HOOC(CH_2)_4COOH$	3.8	3.9	9.8
Pimelic	$HOOC(CH_2)_5COOH$	3.2	3.7	8.7
Suberic	$HOOC(CH_2)_6COOH$	3.0	3.9	7.7
Azelaic	$HOOC(CH_2)_7COOH$	2.8	3.8	7.3

The influence of substituted groups on the ionization of organic acids has been discussed by Kirkwood and Westheimer.[11]

Calibration of the pH Scale

In a previous section we defined pH as being equal to the logarithm of the reciprocal of the hydrogen-ion activity. It now remains to calibrate the pH scale. That is to say, we must have a solution of hydrogen ions whose activity is exactly known.

Most pH measurements are by necessity made with cells involving liquid junction potentials. This is due to the nature of the solutions with which the biochemist has to deal. Usually the cell employed to measure the pH is of the type

$$H_2 \cdot HA \mid \text{Sat KCl} \cdot HgCl \cdot Hg$$

and the pH is calculated by the equation

$$p\text{H} = \frac{E - E_0}{\frac{RT}{F}} \qquad 24$$

All the quantities in equation 24 are known except E_0. The problem is then to evaluate E_0. MacInnes, Belcher, and Shedlowsky [12] proceeded

[10] R. Gane and C. K. Ingold, *J. Chem. Soc.*, **1931**, 2153.

[11] J. G. Kirkwood and F. H. Westheimer, *J. Chem. Phys.*, **6**, 506 (1938).

[12] D. A. MacInnes, D. Belcher, and T. Shedlowsky, *J. Am. Chem. Soc.*, **60**, 1094 (1938).

in the following manner. A cell of the type shown above was used. A weak acid in the presence of its salt was employed, the pH of whose solution can be approximately calculated from the relation

$$p\text{H} = pK_\text{A} + \log \frac{\text{Salt}}{\text{Acid}} \qquad 25$$

The ratio of salt to acid is fixed at unity, and pH equals pK_A. The value of pK_A has been accurately determined as described in the previous section. The determination of the true pH of the solution consists, therefore, in the selection of a value for E_0 which will yield a pH equal to the thermodynamic pK_A. The pH is thus substantially defined in terms of the potential of the half cell employed.

The above workers give the following standard buffer mixtures which can be used to calibrate an electrode: A 0.01 N acetic acid solution plus an equal volume of 0.01 N sodium acetate solution yields a pH of 4.700 at 25° C. and one of 4.710 at 38° C. A 0.10 N acetic acid solution plus an equal volume of 0.10 N sodium acetate solution yields a pH of 4.640 at 25° C. and 4.635 at 38° C. These new values are about 0.03 of a pH unit higher than the Sörensen values.

Determination of Hydrogen Ions

It is not necessary to elaborate extensively on this topic; methods of measurement are adequately treated in several excellent textbooks.[13]

Two general methods are available for the determination of hydrogen ions: (1) the indicator method, and (2) the electrometric method. The indicator method depends upon the fact that various organic substances have a particular color in an acid solution. The shade or tint of the color, at least over a range of about 2 pH units, is a function of the pH of the solution. Accordingly, a series of buffers of known pH is made up and a definite amount of the selected indicator added to each. The same amount of indicator is added to the unknown solution, and the color of the unknown is compared with the colors of the buffer series. The closest matching color is found and the pH estimated directly. This method is useful for routine work where conditions can be rigorously controlled. Indicators are subject to a number of errors, however, and, in general, the electrometric method is to be greatly preferred.

[13] W. M. Clark, *Determination of Hydrogen Ions*, Williams and Wilkins Co., Baltimore, 1928.

L. Michaelis, *Hydrogen Ion Concentration*, Williams and Wilkins Co., Baltimore, 1926.

I. M. Kolthoff, *The Colorimetric and Potentiometric Determination of pH*, John Wiley & Sons, Inc., New York, 1931.

The electrometric method depends upon the measurement of the electromotive force of a cell whose potential is a function of *p*H. The hydrogen electrode is the classical electrode for this purpose. It consists of a piece of inert metal such as gold or platinum covered with platinum black by electrolyzing a solution of platinic chloride. The platinum black is saturated with hydrogen gas and maintained in this condition. The hydrogen is adsorbed on the surface of the platinum black, and for our purposes the electrode becomes a strip of hydrogen. The hydrogen electrode is put into the solution under investigation and connected with a half cell. This arrangement may or may not involve a liquid junction. If a calomel half cell is used, as is the common practice, it is connected with the solution under investigation with a potassium chloride salt bridge. This, of course, involves a liquid junction. The voltage of the hydrogen electrode–calomel electrode is measured with an accurate potentiometer. The *p*H is directly proportional to the electromotive force, and the *p*H is calculated by the formula

$$p\mathrm{H} = \frac{E - E_0}{\frac{RT}{\mathrm{F}}} \qquad 26$$

where E is the voltage measured with the potentiometer and E_0 is the contribution made by the calomel half cell. We have seen in a previous section how E_0 is evaluated by means of the ionization constant of a weak acid. Other electrodes than the hydrogen electrode can be used to measure hydrogen-ion activity.

The quinhydrone electrode is such a half cell. Quinhydrone consists of an equimolecular mixture of quinone (oxidant) and hydroquinone (reductant). It is dissolved in the unknown solution, an inert metal electrode such as platinum is inserted into the solution, and the solution is connected through a salt bridge to a calomel half cell. An oxidation-reduction potential is realized which is proportional to the *p*H. We shall see in the next chapter why such an oxidation-reduction electrode is responsive to hydrogen ions. The quinhydrone electrode cannot be used above *p*H 8.0. It also fails in the presence of oxidants or reductants.

The glass electrode is by far the most convenient electrode for the determination of *p*H. It was discovered by Cremer [14] but was developed by Haber and Klemensiewicz.[15] Dole [16] has published a very complete and useful book dealing with the glass electrode. The glass electrode

[14] M. Cremer, *Biochem. Z.*, **47**, 562 (1906).

[15] F. Haber and Z. Klemensiewicz, *Z. physik. Chem.*, **67**, 385 (1909).

[16] M. Dole, *Glass Electrode*, John Wiley & Sons, Inc., New York, 1941.

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. 2.

The voltage is a straight-line function of the *p*H. The slope of the line varies with the temperature. At 25° C. the voltage changes 59

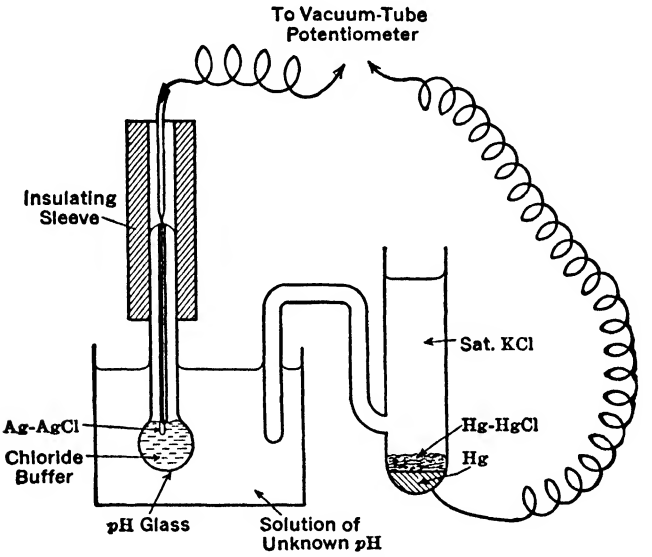


FIG. 2. Glass electrode for the measurement of *p*H.

millivolts for a change of 1 unit. It is necessary to calibrate the glass electrode with a standard buffer.

The glass membrane, though of the lowest resistance obtainable, still offers a tremendous resistance. This necessitates the use of special apparatus for the measurement of the electromotive force produced. A quadrant electrometer may be used for this purpose; the usual practice, however, is to employ a vacuum-tube potentiometer. The glass found most suitable for the membrane had the following composition: SiO_2, 72 per cent; CaO, 6 per cent; Na_2O, 22 per cent. The glass of this composition has the lowest melting point of all glasses containing these ingredients.[17]

The advantages of the glass electrode are:[16] (1) The glass electrode is independent of oxidation-reduction potentials. (2) It is not necessary to pass a gas through the solution or to add any material to it. (3) It

[17] D. A. MacInnes and M. Dole, *Ind. Eng. Chem., Anal. Ed.*, I, 57 (1929).

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 poorly buffered solutions. (6) Equilibrium is rapidly reached.

The glass electrode is subject to errors both in the extreme acid and in the extreme alkaline region. Figure 3 shows the deviations between the glass electrode and the hydrogen electrode as a function of *p*H.[18]

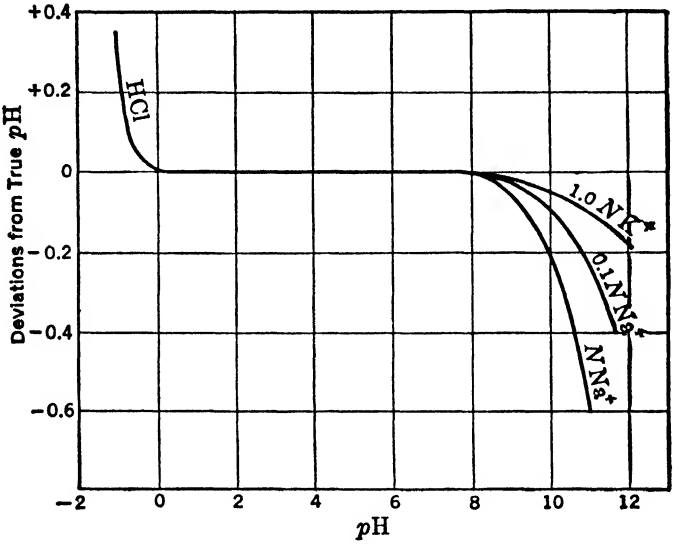


FIG. 3. Acid and base errors of the glass electrode. (Dole.)

The errors of the glass electrode are positive on the alkaline side; i.e., the electrode gives more positive potentials than it should. Anions are, apparently, without effect. Qualitatively, it is easy to see why such errors might be found on the alkaline side. In this region the concentration of cations other than hydrogen ions is about 10^9 greater than that of the hydrogen ions. This enormous excess of cations over the hydrogen ions would be expected to influence the potential. The recently developed Beckman glass shows a far smaller cation effect than the conventional glass.[19]

On the acid side, below *p*H 0, we again find deviations from the straight-line relation between *p*H and the electromotive force. In this region the errors are negative; i.e., the calculated *p*H is greater than the actual *p*H.

In order to understand the deviation in the extreme acid region, we

[18] M. Dole, *Theoretical and Experimental Electrochemistry*, McGraw Hill Book Co., Inc., New York, 1935.

[19] M. Dole, R. M. Roberts, and C. E. Holley, *J. Am. Chem. Soc.*, **63**, 725 (1941).

must have a look at the theory of the glass electrode. The potential developed across the glass membrane is believed to be due to the fact that, under ordinary circumstances, only hydrogen ions can enter the surface of the glass. In short, the potential can be treated as a diffusion potential, the magnitude of which, as we have seen in the previous chapter, is given by

$$E = \frac{u_- - u_+}{u_- + u_+} \frac{RT}{\mathrm{F}} \ln \frac{C_2}{C_1} \qquad 27$$

u_- is the mobility of the anion; if the glass membrane is impermeable to anions, u_- is zero and

$$E = -\frac{RT}{\mathrm{F}} \ln \frac{C_2}{C_1} \qquad 28$$

Accordingly, the electromotive force is dependent upon the difference in hydrogen ions as indicated. As explained previously, we do not have naked protons (H^+) in solution, but the hydrogen ions are hydrated to form H_3O^+. This is the form in which protons enter the glass membrane. The glass electrode thus acts as a water electrode since it is concerned with the transport of the hydrated proton. In the extreme acid region, the activity of the water is decreased by the high acid concentration. This decreased water activity leads to a decrease in the electromotive force of the electrode.

Haugaard [20] has studied the conductance across glass-electrode membranes. He believes that the glass electrode performs as a hydrogen electrode, owing to the ability of glass to exchange sodium ions for hydrogen ions. His picture is somewhat as follows: The glass takes up water, and the sodium salts of silicic acid dissociate. The hydrogen ions displace the sodium ions from the surface of the glass, forming a skeleton of silicic acid. There remains, however, in the center of the glass membrane a layer of intact sodium salt. With the passage of current through the membrane during the compensation of the electromotive force of the system, there is a slight movement of the salt layer towards one side or the other, depending on the direction of the current. The hydrogen ions, however, are the only ions which can enter the surface of the glass, and, accordingly, the glass electrode may be looked upon as a membrane which is permeable only to hydrogen ions.

Approximate Calculation of pH

The occasion frequently arises in experimental work when it is desirable to calculate the approximate value of the *p*H of a solution. For pure

[20] G. Haugaard, *J. Phys. Chem.*, **45**, 148 (1941).

solutions of strong acids such as hydrochloric, or solutions of these acids in the presence of salts of strong acids and strong bases such as KCl or NaCl, a fairly close value can be obtained simply by taking the logarithm of the reciprocal of the acid concentration. Thus, the *p*H of a 0.001 *N* HCl solution is very nearly 3.

The calculation of the *p*H of a pure solution of a weak acid can be made, provided that we know the ionization constant of the acid. In such a case the hydrogen-ion concentration is equal to the anion concentration, and we can write

$$\frac{(H^+)^2}{\text{Acid}} = K_a \qquad 29$$

or

$$H^+ = \sqrt{K_a \times \text{Acid}} \qquad 30$$

Taking the logarithms of both sides of equation 30 and rearranging, we obtain

$$p\text{H} = \tfrac{1}{2}pK_a - \tfrac{1}{2}\log \text{Acid} \qquad 31$$

where the acid concentration is really that of the undissociated acid, but for most purposes can be taken equal to the total acid concentration.

During the titration of a weak acid such as acetic acid by a strong base such as sodium hydroxide, the equation

$$\frac{H^+ \times A^-}{HA} = K_a \qquad 32$$

holds throughout the titration. Since, however, the acid is ionized to a very limited extent, practically all the anions come from the ionization of the salt formed by titration, so that we may, without making any considerable error, write

$$\frac{H^+ \times \text{Salt}}{\text{Acid}} = K_a \qquad 33$$

Taking logarithms of both sides of the equation and rearranging, we obtain

$$p\text{H} = pK_a + \log \frac{\text{Salt}}{\text{Acid}} \qquad 34$$

where "acid" is taken as the total amount of acid present, i.e., both dissociated and undissociated. Equation 34 is very useful and important. The *p*H of a buffer can be calculated with it. Note that, if the acid is just half neutralized, the *p*H equals the pK_a. Also note that the *p*H is independent of dilution and depends only on the ratio of salt to acid;

although this is not entirely true, it is true enough for a great many purposes. A completely analogous equation can be derived for the titration of a weak base by a strong acid.

It is a common experience that salts of strong acids and strong bases have neutral reactions. For example, NaCl, K_2SO_4, etc., dissolved in water are neutral. On the other hand, a solution of sodium acetate is alkaline, and one of ammonium chloride is acid.

Sodium acetate is practically completely ionized in solution

$$NaAc \rightarrow Na^+ + Ac^-$$

The acetate ion so formed reacts with water

$$Ac^- + H_2O \rightleftarrows HAc + OH^-$$

The acetate ion acts in this instance as a base in the Brönsted sense, since it is able to accept a proton from water.

For the acetic acid formed from the hydrolysis of sodium acetate, we can write the usual equilibrium equation

$$\frac{H^+ \times Ac^-}{HAc} = K_a \qquad 35$$

and rearranging

$$H^+ = \frac{K_a HAc}{Ac^-} \qquad 36$$

The amount of acetic acid formed equals the amount of hydroxyl ions produced, and the amount of acetate ion very nearly equals the amount of sodium acetate (salt) in solution. When this information is incorporated in equation 36, there results

$$H^+ = \frac{K_a \times OH^-}{Salt} \qquad 37$$

Since

$$OH^- = \frac{K_w}{H^+} \qquad 38$$

we have from 37 and 38

$$(H^+)^2 = \frac{K_a K_w}{Salt} \qquad 39$$

Taking logarithms of both sides of equation 39 and multiplying through by a minus one, we have

$$pH = 7 + \tfrac{1}{2}pK_a + \tfrac{1}{2}\log Salt \qquad 40$$

Similarly, for a salt of a strong acid and a weak base, for example, ammonium chloride

$$p\text{H} = 7 - \tfrac{1}{2}pK_b - \tfrac{1}{2}\log \text{Salt} \qquad 41$$

where K_b is the ionization constant of ammonium hydroxide.

Employing the same line of reasoning, we find that the pH of a solution of a salt of a weak acid and a weak base, for example, ammonium acetate, is

$$p\text{H} = 7 + \tfrac{1}{2}pK_a - \tfrac{1}{2}pK_b \qquad 42$$

The pH of a solution of such a salt is independent of concentration.

Titration Curves

If acids are titrated with bases or bases with acids and the pH of the solution is determined as the titration is carried out, a titration curve

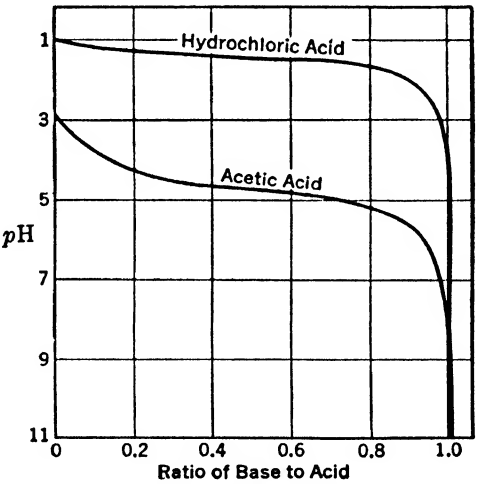


FIG. 4. Titration curves of 0.1 N acids and bases.

can be drawn. One axis of such a graph is pH and the other the amount of acid or base added. Such titration curves have shapes characteristic of the acids and bases employed. Titration curves are shown in Fig. 4.

The subject of titration curves brings us directly to the next topic, buffers.

Buffers

Buffers are defined as substances which resist changes in the pH of a system. All weak acids or bases in the presence of their salts form buffer systems. Examine the titration curve of acetic acid by sodium

hydroxide shown in Fig. 4. It will be noted that the *p*H changes least for a given addition of base at the half neutralization point. The slope of the titration curve is at a minimum here, and the buffer capacity is at a maximum. The *p*H also equals the pK_a.

Buffer capacity is defined as the infinitesimal change in *p*H produced by the addition of an infinitesimal part of an equivalent of acid or base. It is the reciprocal of the slope of the titration curve when the amount of acid or base is expressed in equivalents. Buffer capacity is usually denoted by the symbol π. In short, π equals $\frac{dB}{dp\text{H}}$ or $\frac{dA}{dp\text{H}}$. Figure 5 shows how the buffer capacities of several buffer systems vary with *p*H

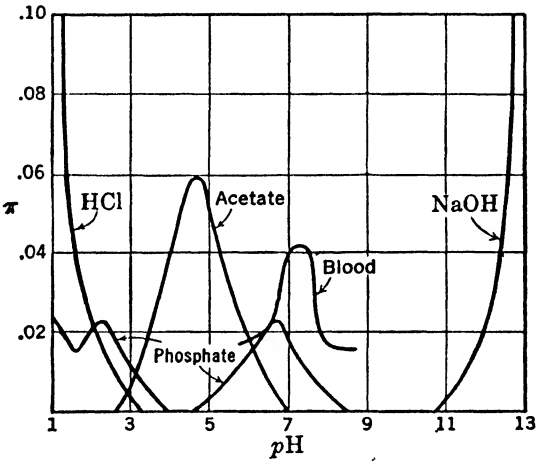


FIG. 5. Buffer capacities of several 0.1 *M* buffer systems as a function of *p*H. Also included is the buffer capacity of whole dog blood.

of the solution. Included in Fig. 5 is the buffer capacity of whole dog blood. Note that the maximum buffer capacity of blood covers the physiological range. This is as it should be.

All physiological systems are equipped with adequate buffers. In such systems proteins usually play a very important role. For example, in mammalian whole blood about 80 per cent of the buffer capacity is due to the proteins present.

Buffers are, in general, used for two purposes: (1) as a reference standard for *p*H determinations, and (2) to maintain the proper acid-base reaction of a medium such as a bacterial culture or an enzymatic reaction mixture.

The selection of the proper buffer systems for a given experimental use is a common problem. In the extreme acid region, hydrochloric

acid of the proper concentration is usually chosen. In the intermediate region of *p*H 3 to 4 phthalic acid-potassium acid phthalate can be used, and from *p*H 4 to 6 acetic acid-sodium acetate serves well. The monosodium dihydrogen phosphate (acid)-disodium monohydrogen phosphate (salt) system covers the range from *p*H 6 to 8. The more concentrated the buffer, the greater is its buffer capacity and the wider is the *p*H range it will cover.

The proper ratio of salt to acid to give the desired *p*H is calculated with the aid of equation 34. Sometimes the conditions of the experiment demand that the ionic strength of the buffer remain constant as the *p*H is varied over a certain range. In this event, the ionic strength is calculated by equation 3 of Chapter V and the concentration of the buffers is adjusted until the ionic strength is constant. In making such a calculation it is assumed that all neutral salts are completely ionized. Disodium monohydrogen phosphate ionizes as follows:

$$Na_2HPO_4 \rightarrow 2Na^+ + HPO_4^{--}$$

and the ionic strength of a 0.1 *M* solution of this salt is

$$\mu = \frac{0.1 \times 1^2 + 0.1 \times 1^2 + 0.1 \times 2^2}{2} = 0.3$$

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 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 [21] proposed a "universal" buffer which has an adequate buffer capacity over a very wide range of *p*H values. The directions for making up this buffer system are as follows: 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 cc. To each 5 cc. of this solution are added 2 cc. of an 8.5 per cent sodium chloride solution plus *a* cubic centimeters of 0.1 *N* hydrochloric acid and (18 − *a*) cubic centimeters of water. Table 4 gives *a* and the corresponding *p*H. All the *p*H values shown in Table 4 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.

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

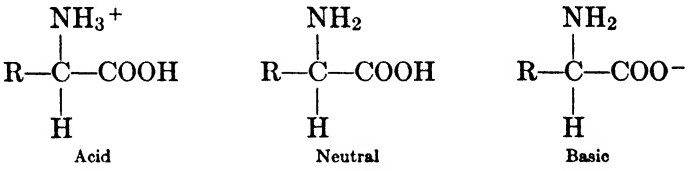
TABLE 4

VERONAL BUFFER SYSTEM (MICHAELIS)

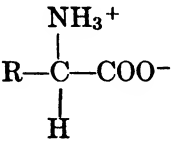
a	*p*H	*a*	*p*H
0.0	9.64	6.5	6.12
0.25	9.16	7.0	5.32
0.5	8.90	8.0	4.93
0.75	8.68	9.0	4.66
1.0	8.55	10.0	4.33
2.0	8.18	11.0	4.13
3.0	7.90	12.0	3.88
4.0	7.66	13.0	3.62
5.0	7.42	14.0	3.20
5.5	7.25	15.0	2.62
6.0	6.99	16.0	

Zwitter Ions

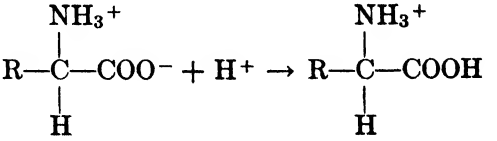
In the early days, amino acids were supposed to exist in three forms in solution, i.e., one neutral and two ionized forms



It was suggested by Adams [22] in 1916 that amino acids exist at their neutral point as what later became known as zwitter ions (also called dipolar ions). In 1923 Bjerrum [23] formulated this idea more completely. The zwitter ionic form of an amino acid is written

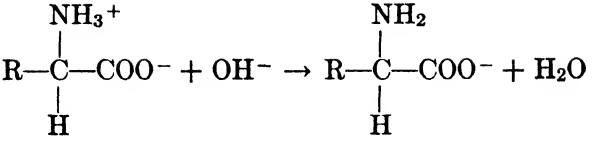


A zwitter ion when titrated with an acid gives



[22] E. Q. Adams, *J. Am. Chem. Soc.*, **38**, 1503 (1916).
[23] N. Bjerrum, *Z. physik. Chem.*, **104**, 147 (1923).

and with a base



Accordingly, the dissociation constants of zwitter ions can be formulated as

$$K_{az} = \frac{{}^{-}A^{+} \times H^{+}}{A^{+}} \qquad 43$$

and

$$K_{bz} = \frac{{}^{-}A^{+} \times OH^{-}}{A^{-}} \qquad 44$$

It is important to grasp the physical meaning of these dissociation constants. The reaction of the zwitter ion with acid really amounts to a back titration of the carboxyl group. In an acid titration of a zwitter ion, the strength of the carboxyl group is being measured and, correspondingly on the basic side, the strength of the "basic" group (NH_3^+) is being measured. If we attempt to determine the dissociation constants of a zwitter ion by the conventional method of measuring the pH at half neutralization, we do not obtain the true dissociation constants but rather the apparent constants. Equations 43 and 44 are the expressions for the true or zwitter ionic constants. These constants are related to the apparent or experimental constants as follows:

$$K_{az} = \frac{K_w}{K_b \text{ (apparent)}} \quad \text{and} \quad K_{bz} = \frac{K_w}{K_a \text{ (apparent)}} \qquad 45$$

The several lines of evidence for the existence of zwitter ions may be summarized as follows:

1. If the apparent dissociation constants as obtained from the titration curves of the amino acids are determined, they are found to be much smaller than would be expected on the basis of the chemical constitution of the amino acids. For example, the pK_a of acetic acid is 4.73, and the apparent pK_a of glycine, an amino derivative of acetic acid, is 9.60. On the basis of the zwitter ionic structure of glycine, the true pK_a is 2.31, which is in keeping with the chemical structure, whereas a pK_a of 9.60 is entirely out of line.

2. Harris [24] found that the presence of formaldehyde produced a shift of the titration curve of an amino acid in the basic region but not in the

[24] L. J. Harris, *Biochem. J.*, **24**, 1080 (1930).

acid region. It is known that formaldehyde reacts with the amino group. It can be concluded, therefore, that it is the amino group which is involved in the basic titration of an amino acid and, hence, the amino acid must exist as a zwitter ion. Figure 6 shows the results obtained by Harris on glycine.

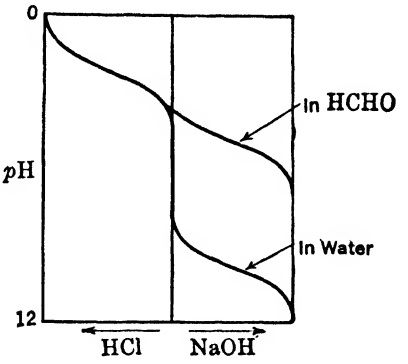


FIG. 6. Titration of glycine with and without formaldehyde. (Harris.)

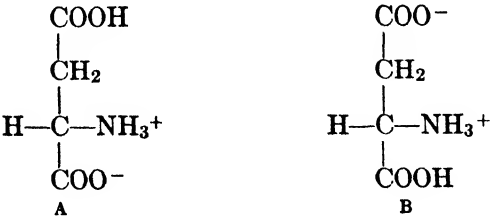
3. As we have seen in Chapter IV, the dielectric constant of aqueous solutions of amino acids is given by the relation

$$D = D_0 + \delta C \qquad 46$$

where D is the dielectric constant of the solution and D_0 that of the solvent. C is concentration of the amino acid in moles per liter. δ, the dielectric constant increment, is large and positive for all amino acid solutions. The zwitter ionic structure explains the rapid increase of the dielectric constant of amino acid solutions with increasing concentration. The large positive values for δ are the reflection of the large dipole moments of the amino acids. The large dipole moment results from the discrete separation of charges in the zwitter ionic structure.

Further arguments for the existence of zwitter ions, because of their indirect or involved nature, will not be dealt with, although among them may be mentioned the heats of neutralization of amino acids and proteins.[25] Studies on the infra-red spectrum of amino acid solutions also indicate the existence of zwitter ions.[26]

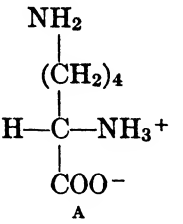
Evidently, a diamino monocarboxylic acid or a dicarboxylic monoamino acid has two possible zwitter ionic structures. For aspartic acid we might have



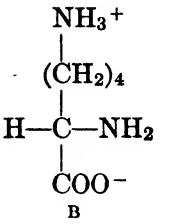
[25] H. H. Weber, *Biochem. Z.*, **189**, 381, 407 (1927); **218**, 1 (1930).

[26] J. T. Edsall, *Chemistry Amino Acids and Proteins*, edited by C. L. A. Schmidt, Chas. C. Thomas, Springfield, Ill.

Calculations show the A form to predominate. For lysine, the two possible zwitter ionic structures are



A



B

The B form with the charges widely separated is the predominating form.

The question arises whether or not any amino acid can exist as an uncharged molecule. Let us define a constant K_z such that [27]

$$K_z = \frac{^{-}A^{+}}{A} \qquad 47$$

The basic dissociation constant of the uncharged form of the amino acid is

$$K_b = \frac{H^{+} \times A}{A^{+}} \qquad 48$$

Combining equations 47 and 48 with equation 43, we find

$$K_z = \frac{K_{az}}{K_b} \qquad 49$$

The magnitude of the basic dissociation constant of the neutral form can be closely approximated by studying the ethyl ester of the amino acid. The value of K_z for all amino acids studied is in the neighborhood of 100,000, which indicates that the concentration of the zwitter ion is 10^5 to 10^6 times greater than that of the uncharged molecule.

Amino acids are not the only compounds which form zwitter ions. As shown by the work of Fischgold and Chain,[28] the phospholipids lecithin, cephalin, and sphingomyelin also exist as zwitter ions.

Isoelectric and Isoionic Points

The isoelectric point is defined [29] as that *p*H at which the net positive and negative charge on a particle or on a surface is zero. The isoionic

[27] J. T. Edsall and M. H. Blanchard, *J. Am. Chem. Soc.*, **55**, 2337 (1933).
[28] H. Fischgold and E. Chain, *Proc. Roy. Soc. (London)*, **117B**, 239 (1935).
[29] W. B. Hardy, *J. Physiol.*, **33**, 251 (1905).

point, on the other hand, is defined [30] as that *p*H at which the number of protons combined on the basic groups is equal to the number dissociated from the acidic groups. It is evident that the isoelectric and isoionic points are identical only if the zwitter ion combines with no ions other than hydrogen ions. The isoelectric point is properly measured with electrokinetic techniques and will be discussed in Chapter X. Isoionic points may be determined by comparing titration curves with varying amounts of zwitter ions. Provided that the solutions contain few diffusible ions, the slopes of the curves will vary with concentration but will intersect at the isoionic point. The isoionic point of a protein may also be determined by dialyzing about a 1 per cent protein solution until all foreign ions are removed; the *p*H of the protein will equal its isoionic point if the isoionic point is not too far removed from neutrality. A workable range is from *p*H 4.5 to 9.5. The *p*H of the isoionic point of a simple zwitter ion may be calculated by the equation of Michaelis.[31]

$$p\text{H isoionic} = \tfrac{1}{2}pK_{az} - \tfrac{1}{2}pK_{bz} + \tfrac{1}{2}pK_w \qquad 50$$

The distinction between isoionic and isoelectric points becomes a real one when one deals with a protein. The net charge on a protein is not entirely due to ionizing groups; certain ions may be adsorbed from the solution. This adsorption may shift the isoelectric point and leave the isoionic point unchanged or change it in a different sense from the change of the isoelectric point.

Dissociation constants of some amino acids are shown in Table 5.

TABLE 5

DISSOCIATION CONSTANTS OF SOME AMINO ACIDS AT 25° C.

Compound	Formula	pK_{az}	pK_{bz}
Glycine	CH_2NH_2COOH	2.31	4.28
α-Alanine	$CH_3CH_2NH_2COOH$	2.39	4.28
Leucine	$(CH_3)_2CHCH_2CHNH_2COOH$	2.34	4.36
Aspartic acid	$HOOCCH_2CHNH_2COOH$	2.08	4.20
Aspartic acid (second ionization)		3.87	

[30] S. P. L. Sörensen, K. Linderstrom-Lang, and E. Lund, *Compt. rend. trav. lab. Carlsberg*, **16**, No. 5 (1926).

[31] L. Michaelis, *Hydrogen Ion Concentration*, Williams and Wilkins Co., Baltimore, 1928.

Titration Curves of Proteins

The acid-base reactions of proteins are of importance for physiology. As previously noted, about 80 per cent of the buffer capacity of whole blood is due to the proteins present. Quite apart from this aspect of the problem, however, titration of proteins with acids and bases has revealed much which is fundamental to protein chemistry.

As is well known, proteins are made up of amino acids linked together through peptide bonds. A certain fraction of these amino acids have two acidic or two basic groups in them. The peptide bond ties up one

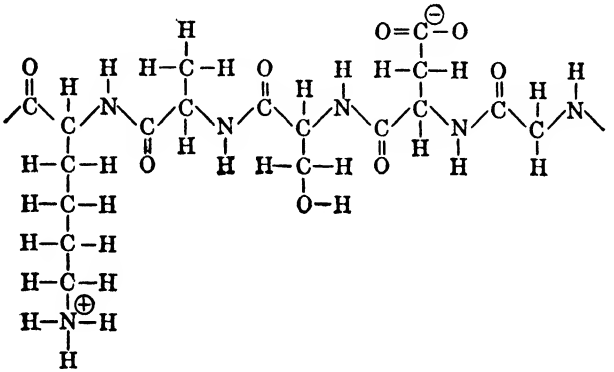


FIG. 7. Segment of hypothetical peptide chain.

amino and one carboxyl group for each amino acid. If, however, dibasic or diacidic amino acids are present, it is evident that a number of free basic and acidic groups will be present in the protein molecule. It is to these groups that the proteins owe their amphoteric properties. A segment of a hypothetical peptide chain containing exposed ionogenic groups is shown in Fig. 7.

The exposed ionogenic groups on a protein form zwitter ions; the evidence for this formation is of the same kind as we have presented for the existence of amino acids as zwitter ions.

Much careful work has been done on the acid-base binding capacity of proteins. The subject has been reviewed by Cannan [32] in a very clear and useful paper. In order to calculate the acid-base binding capacity of a protein, it is necessary to make certain assumptions. It is assumed first that the activity coefficients of the hydrogen or hydroxyl ions in the presence of the protein are the same as they would be in the absence of the protein, and second that no diffusion potentials exist at the liquid junction. It is awkward to construct a cell without a liquid junction for

[32] R. K. Cannan, *Chem. Revs.*, **30**, 395 (1942).

the study of proteins, because the protein is likely to react with the elements of the half cell.

The exact amount of acid or of base which is added to the solution of protein is known. The *p*H of the resulting solution is determined, and the hydrogen- and hydroxyl-ion concentrations are calculated. The difference between the amounts of hydrogen or hydroxyl ions added and those found by the *p*H determination gives the amount of these ions bound by the protein. This binding is usually expressed in equivalents

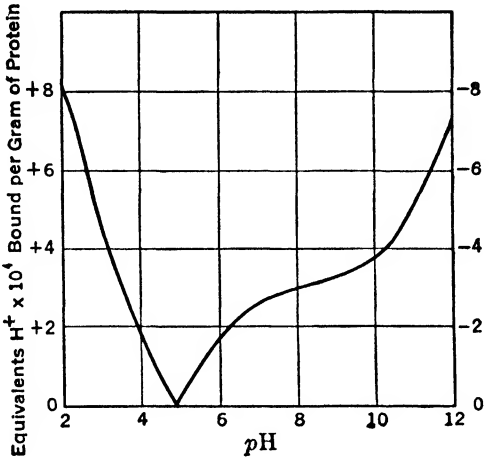


FIG. 8. Dissociation curve of egg albumin in the presence of HCl and of NaOH. (Kekwick and Cannan.)

of hydrogen ions bound per gram of protein and is denoted by the letter *h*. On the acid side of the isoionic point, *h* is positive; on the basic side, it is negative. Figure 8 shows the binding of hydrogen ions by egg albumin.[33] It is possible to locate on the dissociation curve of a protein the contributions of the exposed acidic and basic groups of the individual amino acid residues. The active acidic and basic groups of a protein are the carboxyl groups from aspartic and glutamic acids, the imidazole group of histidine, the amino group of lysine, the phenolic group of tyrosine, the sulfhydryl group of cysteine, and the guanidine group of arginine. The important groups between *p*H 2 to 11 are the carboxyl groups, the imidazole groups, and the amino groups.

Steinhardt [34] has emphasized the role which the anions play in the acid titration of proteins. The binding of anions greatly increases the amount of hydrogen ions held by the protein.

[33] R. A. Kekwick and R. K. Cannan, *Biochem. J.*, **30**, 227 (1936).
[34] J. Steinhardt, *Ann. N. Y. Acad. Sci.*, **41**, 287 (1941).

Acid-Base Catalysis

It is well known that hydrogen and hydroxyl ions catalyze the hydrolysis of esters. In fact, the rate of hydrolysis of sucrose by acids was used as a measure of their strengths before the advent of the modern methods for determining hydrogen-ion concentrations.

It soon became apparent, however, that hydrogen and hydroxyl ions were not alone in possessing catalytic powers towards esters. The substances which have such catalytic powers may be divided into two groups: (1) hydrogen ions, i.e., H_3O^+ ions, undissociated molecules of weak acids; and (2) hydroxyl ions, undissociated molecules of bases, and anions of weak acids. It can be seen, therefore, that the capacity to act as a catalyst in hydrolytic reactions is not limited to hydrogen and to hydroxyl ions but is possessed by all substances capable of accepting or donating protons. This conclusion is in keeping with Brönsted's definition of acids and bases.

It is believed that the hydrolysis of an ester involves a combination of the ester with water to form a complex. The proton acts as a catalyst by combining with this complex to produce an unstable molecule which then breaks down to yield the alcohol and acid. Thus, in order to have acid-base catalyst, two types of substances must be present: an acid capable of supplying a proton, and a base capable of accepting a proton. Water being an amphiprotic solvent, it can either accept or donate protons.

In Table 6 are shown the velocity constants for the hydrolysis of ethyl acetate in the presence of various substances. All these substances are capable of either accepting or donating protons.

Acid-base catalysis is discussed in a book by Bell.[35]

TABLE 6

VELOCITY CONSTANTS OF THE HYDROLYSIS OF ETHYL ACETATE IN THE PRESENCE OF VARIOUS ACID-BASE CATALYSTS

Catalyst	Velocity constant
Hydroxyl ions	6.5
Hydrogen ions	6.5×10^{-3}
Acetic acid (unionized)	1.45×10^{-6}
Acetate ions	2×10^{-6}
Water	negligible

[35] R. P. Bell, *Acid-Base Catalysis*, Oxford University Press, New York, 1941.

Chapter VIII

OXIDATION-REDUCTION

We shall limit ourselves in this short discussion to a consideration of electronic oxidation-reduction, attempting only to outline the general theory of such systems together with the experimental techniques involved in the use of indicators and electrometric methods. This will be followed by a brief summary of biological oxidations.

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

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

On the other hand, many reactions which are termed oxidations do not involve oxygen, such as

$$FeCl_2 + Cl \rightarrow FeCl_3$$

In the above reaction, the ferrous ion has been "oxidized" to the ferric state although no oxygen has been used. Oxidation here involves the loss of an electron

$$Fe^{++} \rightarrow Fe^{+++} + (-)$$

In the same reaction chlorine was reduced; it gained an electron. We can generalize this experience. If any component of a system is reduced, there must be a simultaneous and equivalent oxidation of some other component. On the basis of the above example, we may formulate a general definition of an oxidation or of a reduction; an oxidation involves the loss and a reduction the gain of electrons. According to this definition, a metal in the presence of its ions is an oxidation-reduction system; i.e.

$$Ag \underset{\text{Reduction}}{\overset{\text{Oxidation}}{\rightleftarrows}} Ag^+ + (-)$$

An example with which all biochemists are familiar and which we have already discussed is the hydrogen electrode. Here the hydrogen gas is the reduced and the hydrogen ions the oxidized form.

Electrode Potentials

If an inert metal such as platinum is immersed in a reversible oxidation-reduction system such as a solution of ferrous and ferric chloride, and if this electrode is connected to a hydrogen half cell at unit activity of hydrogen ions, a potential difference will be observed. It can be shown theoretically and experimentally that the observed voltage is a function of the ratio of the oxidized to the reduced form; the more of the oxidized component, the more positive will be the platinum electrode in respect to the hydrogen electrode.

We wish to indicate how the relation between the measured potential difference and the ratio of the oxidant to the reductant can be derived. Consider again the reaction

$$Fe^{++} \rightleftarrows Fe^{+++} + (-)$$

We can formulate this reaction in the usual manner of chemical equilibria

$$\frac{Fe^{+++} \times (-)}{Fe^{++}} = K \qquad 1$$

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 the escaping tendencies of the electrons in the two phases. It can be shown that the work done in transferring an equivalent of electrons from the solution into the wire is

$$\text{Work} = RT \ln \frac{e_m}{e_s} \qquad 2$$

where e_s is the concentration of electrons in the solution and is the electronic concentration in the metal, T is the absolute temperature, and R is the gas constant, equal to 1.99 calories per degree.

The work is equal to the quantity of electricity transferred multiplied by the potential at which the transfer is made. As we remember from previous considerations

$$\text{Work} = n\mathbf{F}E \qquad 3$$

where n is the number of equivalents transferred, $\mathbf{F}$ is a conversion factor, and E is the potential at which the transfer is made. Combining equations 2 and 3 and rearranging, we have

$$E = \frac{RT}{\mathbf{F}} \ln e_m + \frac{RT}{\mathbf{F}} \ln \frac{1}{e_s} \qquad 4$$

where E is the oxidation-reduction potential.

Equation 1 is substituted in equation 4 by equating e_s and $(-)$. The concentration of electrons in the metal (e_m) is a constant. The first term on the right side of equation 4 is, therefore, a constant. Combining this constant with the constant in equation 1 to form a new constant K_1, we have

$$E = K_1 + \frac{RT}{\mathbf{F}} \ln \frac{Fe^{+++}}{Fe^{++}} \qquad 5$$

When the concentration of ferric ions equals that of the ferrous, E equals K_1. E is usually written E_h, and K_1 is denoted by E_0. Then

$$E_h = E_0 + \frac{RT}{\mathbf{F}} \ln \frac{Fe^{+++}}{Fe^{++}} \qquad 6$$

This equation can be generalized in the form

$$E_h = E_0 + \frac{RT}{n\mathbf{F}} \ln \frac{Ox}{Red} \qquad 7$$

At 30° C. and logarithm to the base 10, and for one electron transfer

$$E_h = E_0 + 0.06 \log \frac{Ox}{Red} \qquad 8$$

For a reduction involving two electrons at 30° C., the equation is

$$E_h = E_0 + 0.03 \log \frac{Ox}{Red} \qquad 9$$

The oxidation-reduction potential is always referred to the normal hydrogen electrode, which, as we have seen, is defined as having zero potential. In practice, it is generally more convenient to use the calomel half cell and correct for the potential contributed by the calomel half cell. E_0 in equation 6 has considerable significance. As pointed out above, if the concentration of the oxidant equals that of the reductant, the term involving the ratio of oxidant to the reductant becomes zero and the observed voltage is equal to E_0. E_0 is known as the standard oxidation-reduction potential and is used to compare oxidation-reduction systems with one another.

The oxidation-reduction potential is a measure of the tendency a substance has to give up or to take up electrons. Stated in a different way, it is a quantitative measure of the free energy of the oxidation of a substance. Evidently, in order to have a potential established at an inert electrode, there must be an electronic transfer, although with some

reactions involving no such transfer there are certain tricks which can be resorted to and which yield an apparent potential. It should be added that the oxidation-reduction potential is an intensity factor in the same sense as temperature is an intensity factor and does not in any way indicate the reducing or oxidizing capacity of a system.

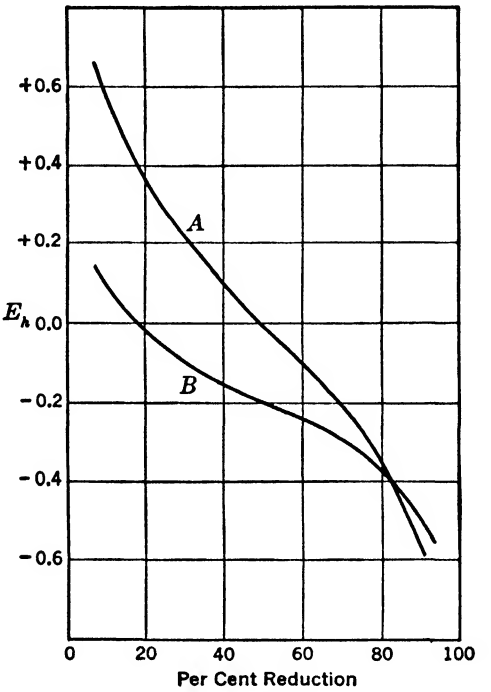


FIG. 1. Oxidation-reduction potential as a function of the percentage of oxidant present. System *A* involves the transfer of one electron; system *B* involves a two-electron transfer.

Figure 1 shows the plot of the oxidation-reduction potential as a function of percentage of oxidant present.

We see from Fig. 1 that, although system *A* has a higher normal oxidation-reduction potential than system *B* if *B* contains 90 per cent oxidant, it will tend to oxidize system *A* if that system has 90 per cent reductant and 10 per cent oxidant.

We have said nothing regarding the relative speeds of oxidation or reduction, and oxidation-reduction potentials are not capable of giving us any information on this point. Such potentials give us, in truth, a negative answer. We can say definitely that one system with a lower potential can never oxidize a system of higher potential in the same way as we can say that heat can never, without some external agency, flow

from a cold to a hot body. We cannot, however, say that a system at a higher potential will definitely 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. Sometimes an empirical relation is observed between the speed of oxidation and the oxidation potential for a series of closely related compounds. Michaelis and Smythe [1] found that in a series of organic iron compounds the autoxidizability of the ferro compounds at a given *p*H closely paralleled the normal oxidation-reduction potential of the ferro-ferri systems of the compounds. The more negative the potential, the greater was the autoxidizability.

Effect of Hydrogen Ions

The hydrogen-ion concentration of oxidation-reduction systems must usually be considered, because in most systems the reduced form can exist as an anion which can accept hydrogen ions and so become inoperative as far as contributing to the potential. Naturally, the potential of such a system is greatly influenced by the hydrogen-ion concentration.

A reduction involving the formation of an anion may be expressed as

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

and the electrode equation is

$$E_h = E_0 + \frac{RT}{\mathbf{F}} \ln \frac{\mathrm{Ox}}{\mathrm{Red}^-} \qquad 10$$

but the reduced form ionizes as

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

and

$$\frac{\mathrm{H}^+ \times \mathrm{Red}}{\mathrm{H\ Red}} = K \qquad 11$$

The total reduced form, Red, is equal to the ionized plus the unionized form

$$\mathrm{Red} = \mathrm{H\ Red} + \mathrm{Red}^- \qquad 12$$

Combining equations 11 and 12, we have

$$\frac{\mathrm{H}^+ \times \mathrm{Red}^-}{\mathrm{Red} - \mathrm{Red}^-} = K \qquad 13$$

or

$$\mathrm{Red}^- = \mathrm{Red} \frac{K}{\mathrm{H} + K} \qquad 14$$

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

Substituting equation 14 in equation 10, there results

$$E_h = E_0 + \frac{RT}{F} \ln \frac{Ox}{Red} \left(\frac{H^+ + K}{K} \right) \qquad 15$$

or

$$E_h = E_0 + \frac{RT}{F} \ln \frac{Ox}{Red} - \frac{RT}{F} \ln \frac{K}{H^+ + K} \qquad 16$$

If the ratio of the oxidant to the reductant is unity, equation 16 becomes

$$E_h = E_0 - \frac{RT}{F} \ln \frac{K}{H^+ + K} \qquad 17$$

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

$$E_h = E_0 - \frac{RT}{F} \ln \frac{K}{H^+} \qquad 18$$

or

$$E_h = E_1 - \frac{RT}{F} \ln \frac{1}{H^+} \qquad 19$$

and at 30° C.

$$E_h = E_1 - 0.06\, pH \qquad 20$$

and 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 and was given for the sake of simplicity. The general rule is that the reduction of an organic compound involves the addition of two electrons, and it was thought for a number of years that these electrons had to be transferred simultaneously. Michaelis,[2] however, has pointed out that in several organic systems there is definite evidence for stepwise electronic transfer.

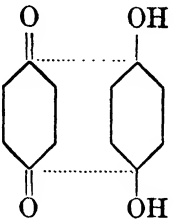
Stepwise Oxidation

Michaelis cites three criteria for deciding whether the electronic transfer occurs in one or two steps: (1) the color of the intermediate compound, (2) the study of the magnetic moment, and (3) potentiometric

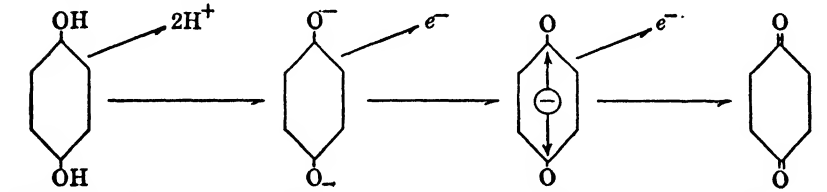
[2] L. Michaelis, *Cold Spring Harbor Symposia Quant. Biol.*, **L**, 224 (1933); **VII**, 33 (1939); *Chem. Revs.*, **16**, 450 (1935).
L. Michaelis and C. V. Smythe, *Ann. Rev. Biochem.*, **7**, 1 (1938).
L. Michaelis and M. P. Schubert, *Chem. Revs.*, **22**, 437 (1938).

oxidation titration curves. Before discussing these three criteria briefly we must fix clearly in mind what we mean by stepwise oxidation. Consider quinhydrone, for example.

As is well known, quinhydrone is an equimolecular mixture of hydroquinone and quinone. It is frequently used in a half cell to determine the hydrogen-ion concentration. It was first proposed that quinhydrone exists as a meriquinone, i.e.,



On the basis of a stepwise oxidation, we would postulate a semiquinone



Quinhydrone, however, exists only in the solid state. In solution, it is completely dissociated into hydroquinone and quinone. A semiquinone is really a free radical and usually has a distinctive color. Its appearance in appreciable amounts can be detected by the color of the solution.

It is known that all molecules containing an odd number of electrons are paramagnetic, and those with an even number are usually diamagnetic. By determining the magnetic moment of a compound in solution, the presence of a free radical is detectable. The method has been used but seldom as it is limited by the solubility and stability of the free radical; the concentration of the semiquinone in solution must be high in order to measure the magnetic properties.[3]

The most useful method for studying the formation of semiquinones is by potentiometric oxidation titrations in which the completely reduced form is titrated with a convenient oxidizing agent and the oxidation-reduction potential measured. As we have seen for a univalent oxidation (30° C.)

$$E = E_0 + 0.06 \log \frac{\text{Ox}}{\text{Red}} \qquad 21$$

[3] L. Michaelis, *J. Am. Chem. Soc.*, **63**, 2446 (1941).

and for a bivalent oxidation

$$E = E_0 + 0.03 \log \frac{\text{Ox}}{\text{Red}} \qquad 22$$

If the oxidation occurs in two distinct univalent steps, the oxidation titration curve should reveal this. Frequently, of course, there is considerable overlapping of the two oxidation steps, which makes the

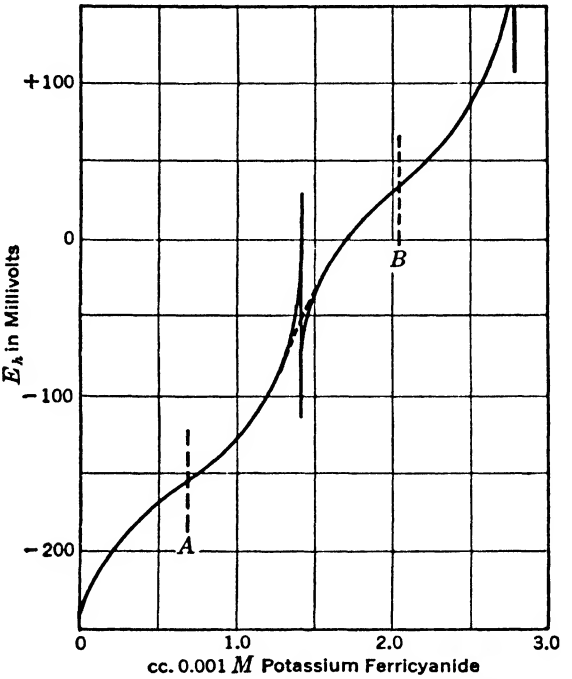


FIG. 2. Oxidation of pyocyanine with ferricyanide. *A* is the midpoint of the first electron transfer; *B* is that of the second. (Michaelis.)

analysis of the curve difficult. Figure 2 shows the titration of pyocyanine with ferricyanide. This is a clear example of two-step oxidation.

If the amount of the free radical formed is large, the two steps of the titration are very conspicuous. On the other hand, if the amount of free radical is small, the two steps blend into one curve which is that of a bivalent oxidation. By a mathematical analysis of such a curve the amount of the free radical can be estimated.

Shaffer [4] has suggested that the ability of some dyes to catalyze certain oxidation-reduction reactions is related to this property of yield-

[4] P. A. Shaffer, *J. Phys. Chem.*, **40**, 1021 (1936); *Cold Spring Harbor Symposia Quant. Biol.*, VII, 50 (1939).

ing 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, and I_3^- can accept two but not one. If a small amount of dye which is capable of stepwise oxidation-reduction is added, the reaction between Ti^{+++} and I_3^- is greatly accelerated. The dye can take electrons one at a time from Ti^{+++} and give them two at a time to I_3^-. Another example is the system thallous and thallic sulfate and ceric and cerous sulfate. Both the thallous-thallic and cerous-ceric systems give reversible, well-defined oxidation-reduction electrode potentials. The potentials obtained with these systems indicate that the thallous sulfate should almost completely reduce the ceric salt, and yet, when the two salts are mixed together in solution, no appreciable reaction takes place even when the mixture is boiled. However, if a small amount of manganese sulfate is added, the reaction proceeds rapidly.

Shaffer explains this in the following manner: the change of Ce^{++++} to Ce^{+++} involves one electron transfer; the change from Tl^+ to Tl^{+++} involves the loss of two electrons. The reaction between Ce^{++++} and Tl^+ would, therefore, require a three-body collision. As such collisions would be expected to occur very infrequently, the reaction between the thallous and ceric ions is a slow one. In the presence of Mn^{++} ions the ceric ions are reduced and the Mn^{++} ions are oxidized to Mn^{++++} in a stepwise manner through Mn^{+++}. The Mn^{++++} ions promptly oxidize the Tl^+ ions to Tl^{+++} ions. The stepwise oxidation as formulated by Michaelis is of fundamental significance for the reaction kinetics of biological oxidations.

In the discussion of the influence of the hydrogen-ion concentration on the oxidation-reduction potential, we considered the formation of an anion. The oxidized form may, however, 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.[5]

The dependence of several oxidation-reduction systems on *p*H is shown in Fig. 3. The lines in the figure correspond to the following systems: (1) hypothetical oxygen electrode, (2) *o*-phenanthroline, (3) ferrous-ferric, (4) quinhydrone, (5) ferro-ferricyanide, (6) methylene blue, (7) calomel electrode, (8) indigo monosulfonate, and (9) hydrogen electrode.

The ionization constants of the oxidation-reduction systems are equal to the hydrogen-ion concentration at the inflection points of the curves

[5] W. M. Clark, *Public Health Report*, Reprint **826**, March 30, 1923.

(see Fig. 3); that is, pH equals pK. The sign of the change of the slope of the curve at the inflection point indicates whether the dissociation belongs to the oxidant or to the reductant. When the change in this slope is negative, this change is due to the ionization of the reductant; when positive, to the oxidant. It is also possible to evaluate the dissociation constants by simple acid-base titrations.[6]

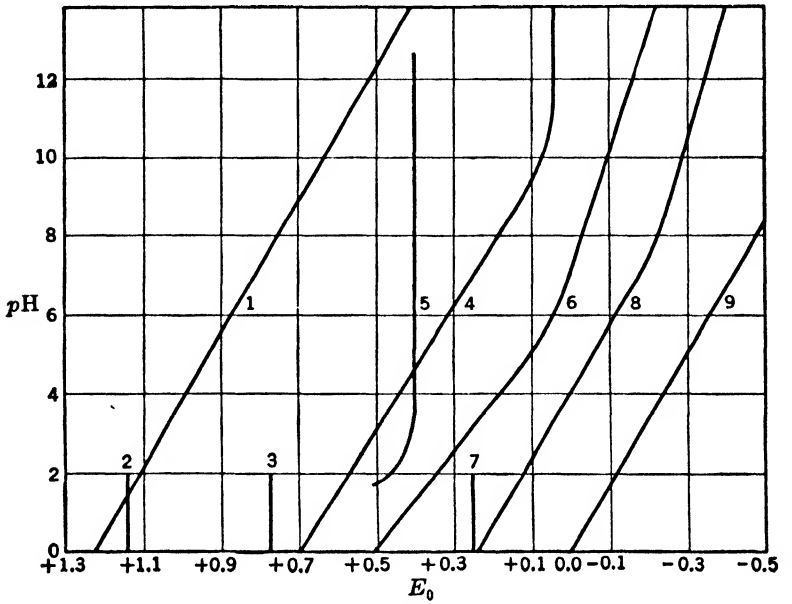


FIG. 3. Oxidation-reduction potentials as a function of pH.

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

$$A \rightleftarrows B + (+)$$

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

$$\text{Reductant} \rightleftarrows \text{Oxidant} + (-)$$

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

To summarize this theoretical discussion, we may state that the oxidation-reduction potential depends upon three conditions: (1) the innate

[6] W. L. Hall, P. W. Preisler, and B. Cohen, Paper XIV, *Public Health Report*, Supplement 71 (1928).

tendency of a system to accept or yield electrons, (2) the ratio of the oxidant to the reductant, and (3) the hydrogen-ion concentration.

Measurement

The oxidation-reduction potential can be determined by two methods, the indicator and the electrometric. For various reasons, indicators are less reliable in the measurement of the oxidation-reduction potential than for the measurement of *p*H. These two methods will now be considered.

It is no accident that a substance which undergoes oxidation and reduction easily is usually colored. Such a substance possesses loosely held electrons which can be transferred to another substance without difficulty; 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 change in color of the indicator, it is possible to assign to the system its approximate oxidation-reduction potential. The indicator method has proved of value in the study of living cells, where it has not as yet been possible to devise an electrode which would not injure the cell. Indicators are also useful in quantitative analysis in determining the end point of an oxidation-reduction reaction. It is also possible for the dye to serve as an oxidizing agent as well as an indicator. For example, Farmer and Abt [7] use 2,6-dichlorobenzenoneincophenol to estimate the ascorbic acid (vitamin C) content of small amounts of blood. The oxidized form of this indicator is blue, and the reduced form is pink. When the ascorbic acid has been completely oxidized by sufficient addition of indicator, the blue color of the oxidized form persists.

In general, five considerations govern 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 might be due to a change in *p*H and would thereby 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 so that very low concentrations can 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

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

it may have a large reserve 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 of the actual reducing intensity of the cell. Considerable lapse of time may be necessary to determine the true reducing level of the cell.

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

In determining the reducing intensity of cells with indicators, two methods are used. Either the dye is injected into the cell by a microinjection, or the dye is allowed to diffuse into the cell. The microinjection method cannot be used if the cells are small or sensitive to injury. 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 a suspension of the same cells.

Table 1 lists the more common indicators which, under the proper circumstances, may be used to advantage.

Electrometric Method

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 and more exact in its application.

The apparatus for the purpose is quite simple. The quinhydrone electrode is such a common oxidation-reduction system that it seems needless to describe it. The same arrangement is employed to measure the potential of any oxidation-reduction system, unless 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 out all oxygen in the system. The nitrogen must be purified by passing it over heated copper filings or through pyrogallol solutions. The same technique is followed in the electrometric study of cell suspensions.

Unfortunately, a number of oxidation-reduction systems do not yield stable, well-defined potentials; indeed, some reactions such as the oxidation of an aliphatic aldehyde to the acid do not give a potential of any kind. If one insists upon measuring the potential of such a system, it will be found to fluctuate greatly and will, in general, depend upon traces of oxygen or other impurities. Other reactions are sluggish and exhibit a stable potential only after several hours. Often the reasons for these

TABLE 1

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_2H-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

difficulties are obscure. Two explanations for the sluggishness suggest themselves: (1) the material is irreversibly oxidized or reduced; (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 in the molecule. Ionic reactions, as a rule, proceed with extreme rapidity.

Conant and others have studied the irreversible oxidation-reduction of organic compounds and have developed methods for determining what they term the apparent oxidation-reduction potential. This is done by choosing some easily reversible system with the oxidant and reductant in equivalent amounts and using this in conjunction with an inert electrode. The substance under investigation is added, and the potential, which is due to the reversible system, observed. It is possible by this technique to bracket the irreversible potential between the potentials of two reversible systems.

Other methods have been followed. 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 oxidized or reduced until their molecules have been activated by an enzyme or perhaps by a dye. The reduction of fumaric to succinic acid is reversible only in the presence of a succinic acid dehydrogenase and an oxidation-reduction indicator. The action of the indicator dye is probably that of a mediator; i.e., it takes the hydrogen which has been activated by the dehydrogenase and passes it on to the oxidant.

The important sulfhydryl (−SH) systems have been carefully studied by Ryklan and Schmidt [7a] who were able to show that under properly controlled conditions reversible and reproduceable potentials can be obtained for the reaction, $RSSR \rightleftharpoons RSH$.

Biological Systems

A number of systems of biological interest have been studied and, to an extent, characterized. These include the sulfhydryl, hemoglobin, cytochrome, ascorbic acid, hermidine and echinochrome, pyrocyanine, dilauric acid, various sugar systems, succinate-fumarate, adrenaline, and

[7a] L. R. Ryklan and C. L. A. Schmidt, *Univ. Calif. Pub. Physiol.*, **8**, 257 (1944).

the oxytocic hormone of the pituitary gland. Altogether, about fifteen biological pigments have been found to constitute oxidation-reduction systems, and respiratory functions have been attributed to some of these pigments. This represents a definite and worthwhile achievement. However, the number of biological systems which yield reversible potentials and are thus capable of an exact theoretical treatment are relatively few, and not long ago it appeared as though the study of oxidation-reduction potentials would soon be completed with the characterization of these systems. Later work has shown 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 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 a state of equilibrium. No living cell is in equilibrium in respect to all its oxidation-reduction systems; if it were, it would not be able to metabolize. The best that can be achieved in such experiments is a steady state, and under these conditions it is not valid to apply the electrode equations which we have given in this chapter. The field must be examined experimentally, and the reducing intensity must be related to known physiological conditions. This method of approach has already proceeded a considerable distance and is proving of special value in bacteriology.[8]

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.[9]

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

[8] L. F. Hewitt, *Oxidation-Reduction Potentials in Bacteriology and Biochemistry*, published by the London County Council, 1936.

[9] E. S. G. Barron, *Physiol. Revs.*, **19**, 184 (1939).

D. E. Green, *Mechanisms of Biological Oxidations*, Cambridge University Press, New York, 1941.

A Symposium on Respiratory Enzymes, University of Wisconsin Press, Madison, 1942.

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 involves the decarboxylation of the substrate which has been oxidized by the removal of hydrogen and hydrated with the addition of water.

Chapter IX

ELECTRICAL CONDUCTANCE

It is evident that, if electrolytes exist partly or completely as ions in solution, and if a difference of potential exists across two electrodes placed in such a solution, a current will flow between the electrodes. This flow of current is due to the migration of ions through the solution. The cations, being positively charged, travel towards the negative electrode (cathode), and the anions, having a negative charge, move towards the positive electrode (anode). The amount of current carried by an ion will depend upon its valence and upon its speed of migration. One gram equivalent of an ion will transport 96,494 coulombs of electricity.

Modes of Expression

The conductance of an electrolyte may be expressed in several ways. The specific resistance of any substance, whether in solution or not, may be defined as the resistance in ohms of a column of substance 1 centimeter long and 1 square centimeter in cross section. Specific conductance (or conductivity) is the reciprocal of the specific resistance and is expressed in reciprocal ohms or mhos. Conductance is defined as the ratio of the current flowing through a conductor to the difference in potential between its ends. Conductivity, as we have seen, can be defined as the quantity of electricity transferred across unit area per unit potential gradient per unit time.

Other ways of expressing conductance are by equivalent and by molar conductance. The equivalent conductance is the conductance in reciprocal ohms of a solution containing 1 gram equivalent of the electrolyte enclosed between two electrodes 1 centimeter apart. Similarly, molar conductance is the conductance in mhos of a solution containing 1 gram molecular weight under the same conditions. The relation between the specific conductance, equivalent conductance, and molar conductance is as follows: If κ denotes the specific conductance of a solution, Λ the equivalent conductance, μ the molar conductance, V_e the volume in cubic centimeters of solution which contains 1 equivalent of

solute, and V_m the volume which contains 1 gram molecular weight of the solute, we have

$$\mu = \kappa V_m \qquad 1$$

and

$$\Lambda = \kappa V_e \qquad 2$$

If C denotes the concentration of a solution of an electrolyte in gram equivalents per liter, we have

$$V_e = \frac{1{,}000}{C} \qquad 3$$

Substituting equation 3 in equation 2, we obtain

$$\Lambda = \frac{1{,}000\kappa}{C} \qquad 4$$

A similar relation can be developed for the molar conductance.

The specific conductance (or conductivity) of an electrolyte decreases and approaches that of the solvent as the concentration is decreased, whereas the equivalent conductance and the molar conductance increase with decreasing concentration and approach a limit which is characteristic of the solute ions. The variation of the conductance with electrolyte concentration is due to two influences: (1) increased ionization with decreased concentration, and (2) decreased interaction of the ions upon dilution. The first effect was one of the strongest supports for Arrhenius' theory of ionization. The dependence of the equivalent conductance on electrolyte concentration can be formulated quantitatively and used to evaluate the ionization constants of weak electrolytes. The second effect has been considered by Debye and Hückel [1] and in more detail by Onsager.[2] The theoretical treatment, however, is too involved for us to consider. For a fairly simple exposition of this effect, see a paper by MacInnes.[3]

The ionization "constant" of a weak electrolyte can be obtained from the equation

$$K = \frac{\Lambda^2 C}{\Lambda_0(\Lambda_0 - \Lambda)} \qquad 5$$

where C is the concentration in gram equivalents per liter, Λ is the equivalent conductance at the given concentration, Λ_0 is the equivalent

[1] P. Debye and E. Hückel, *Physik. Z.*, **24**, 185, 305 (1923).
[2] L. Onsager, *Physik. Z.* **27**, 388 (1926).
[3] D. A. MacInnes, *Science*, **86**, 23 (1937).

conductance at infinite dilution (the limiting conductance as the electrolyte is made more and more dilute), and K is the ionization constant, the value of which, however, varies somewhat with concentration. As the dilution is increased, the constant approaches the true constant more closely. (See Chapter VII for a discussion of the ionization constants of weak acids.)

The equivalent conductance of a dilute solution of a strong electrolyte (completely ionized) varies inversely as the square root of the concentration.

The specific conductance of an aqueous solution increases approximately 2 per cent for each degree rise in temperature. This is a rather complicated effect but involves principally the decrease in the viscosity of the water with increasing temperature.

Experimental Determination of Conductance

The determination of the conductance of a solution is one of the simplest and most accurate of physical measurements. A conductance cell is employed which consists of a glass vessel of the proper shape and dimensions with two platinum disks held in fixed positions in the cell. The platinum electrodes are covered with platinum black. The cell is calibrated by measuring the resistance offered to a solution of known conductance (usually 0.1 or 0.01 N KCl). The unknown is then placed in the cell, its resistance is measured, and the specific conductance of the solution is calculated. For example, suppose that the resistance of the cell filled with 0.01 N KCl is 307.9 ohms at 25° C. The specific conductance of 0.01 M KCl at 25° C. is 0.001407. Accordingly, the cell constant is 307.9×0.0014 or 0.428. In order to find the specific conductance of the unknown, we divide the cell constant by the resistance offered by the unknown solution or

$$\text{Specific conductance } (\kappa) = \frac{\text{Cell constant}}{\text{Resistance}} \qquad 6$$

The proper shapes and sizes of the conductance cell are given by Washburn [4] and by Jones and Josephs; [5] the best values for the calibrating solutions by Jones and Bradshaw.[6]

The electrical apparatus for measuring conductance makes use of a Wheatstone bridge, shown in Fig. 1.

The resistance of the cell is measured by balancing its resistance against that of the known R_1 by moving the slide along S until no sound

[4] E. W. Washburn, *J. Am. Chem. Soc.*, **38**, 2431 (1916).
[5] G. Jones and R. C. Josephs, *J. Am. Chem. Soc.*, **50**, 1049 (1928).
[6] G. Jones and B. C. Bradshaw, *J. Am. Chem. Soc.*, **55**, 1780 (1933).

is heard in the receiver T. The resistance of the cell is then calculated by the following proportion

$$\frac{R_2}{R_1} = \frac{ab}{bc} \qquad 7$$

where R_2 is the resistance of the conductance cell, R_1 the known resistance, and ab and bc refer to the position of the slide wire at balance.

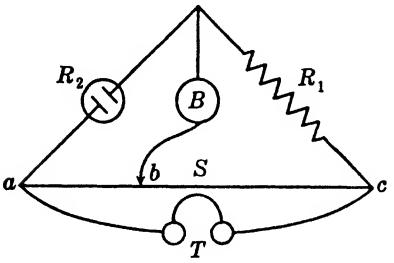


FIG. 1. Wheatstone bridge. R_1 is a known resistance; S, a graduated resistance slide wire; R_2, resistance of the conductance cell; B, a buzzer giving an alternating current (usually 1,000 cycles per second); T, a telephone receiver.

Various modifications of this simple arrangement have been employed. The electrodes of the conductance cell form a condenser with a certain capacity, and, in order to compensate for this capacity, a condenser of approximately the same capacity should be included in parallel with the known resistance. The proper design for a Wheatstone bridge is described by Jones and Josephs.[5] A very satisfactory method for solutions whose conductance is not too high is to replace receiver T by an alternating-current galvanometer, and the buzzer B by a source of 60-cycle alternating current of low voltage. The slide is fixed at the mid-point of the slide wire, and balance is obtained by varying R_1 until no deflection in the alternating-current galvanometer is observed. The resistance of the cell is then exactly equal to the known resistance.

Uses of Conductance Measurements

There are a variety of uses for conductance measurements in the biochemical laboratory. The end point of acid-base titrations can be accurately determined. The conductance shows a sharp break at neutralization. Both hydrogen and hydroxyl ions have a high mobility; accordingly, at neutralization the conductance changes sharply.

Conductance measurements furnish a very convenient way of testing for the absence or presence of electrolytes in protein preparations.

Conductance measurements have been extensively used to study permeability changes in living cells. Obviously, an increase in the conductance of a cell means that ions are passing through the cell with greater ease. Osterhout was a pioneer in the study of living-cell perme-

ability by means of conductance. Figure 2 shows some of his results on the sea alga *Laminaria agardhii*.[7]

The electrolyte concentrations used in Osterhout's experiments were such as to yield the same conductivity as sea water. In the early stages of the experiments the process is reversible; but, if there is too great a change in conductivity, either increasing or decreasing, the process is irreversible and death results. The resistance of normal *Laminaria* is

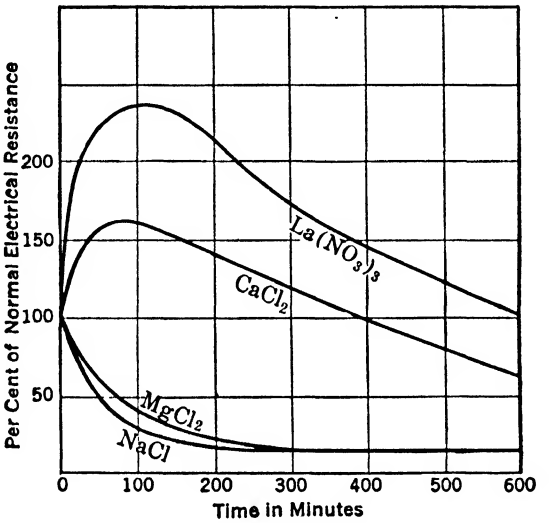


FIG. 2. Conductance of *Laminaria agardhii* in the presence of various ions. (Osterhout.)

about 10 times that of sea water, whereas the dead tissue has a resistance about equal to that of sea water. In general, Ca^{++} decreases the permeability and Mg^{++}, Na^{+}, and K^{+} increase it.

One of the difficulties of conductance measurements on cell suspensions is that current passes not only through the cells but also around them. This source of confusion is resolved by applying Maxwell's equation which describes the resistance of suspensions of conducting spheres in a conducting suspension. This equation is

$$\frac{\frac{r_1}{r} - 1}{\frac{r_1}{r} + 2} = \phi \left(\frac{\frac{r_1}{r_2} - 1}{\frac{r_1}{r_2} + 2} \right) \qquad 8$$

[7] W. J. V. Osterhout *Injury, Recovery and Death in Relation to Conductivity and Permeability*, J. B. Lippincott Co., Philadelphia, 1922.

where r is the specific resistance of the suspension, r_1 the specific resistance of the suspending liquid, r_2 the specific resistance of the suspended particles, and ϕ the ratio of the volume of the suspended particles to the total volume of the suspension.

The conductance of suspensions of ellipsoids has been discussed by Fricke [8] and more recently by Velick and Gorin.[9] These last workers derived an equation for non-conducting ellipsoids which takes the form

$$\phi = \frac{\frac{r}{r_1} - 1}{\frac{r}{r_1} - 1 + f} \qquad 9$$

where f is a form factor. For spheres it is 1.5, and for all other structures it is greater than 1.5.

To use this equation, the ratio of r to r_1 of a relatively concentrated suspension is measured. The suspension is then diluted by the addition of known amounts of the medium, and the ratio r/r_1 is determined for each mixture. The form factor f and ϕ are then obtained in the following way: Let ϕ' and r'/r_1' be the volume fraction and conductance ratios, respectively, for the most dilute member of the dilution series, and ϕ and r/r_1 the same for any other member of the series. Let V_1 be the total volume to which 1 cc. of the original suspension was diluted for the most dilute member of the series, and V the total volume containing 1 cc. of the original suspension for any other member of the series. With this arrangement, it follows from equation 9 that

$$\frac{V}{V_1} = \phi' \left(1 + \frac{f}{(r/r_1) - 1}\right) \qquad 10$$

When $\frac{V}{V_1}$ is plotted against $\frac{1}{(r/r_1) - 1}$, a straight line should be obtained with the intercept ϕ' and a slope of $\phi' f$ from which ϕ' and f are found. Velick and Gorin verified their equation with duck erythrocytes. The author attempted to apply their equation to protein solutions, but without success.

Much recent work has been done on the conductance of biological systems (individual cells, cell suspensions, and tissue cells), using varying high-frequency currents (500 to 10 million cycles). In general, as the frequency is increased, the resistance of biological systems decreases. In

[8] H. Fricke, *Cold Spring Harbor Symposia Quant. Biol.*, I, 117 (1933).
[9] S. Velick and M. Gorin, *J. Gen. Physiol.*, 23, 753 (1940).

order to understand the apparent drop in resistance of biological systems with increasing frequency, it is necessary to examine for a moment in a qualitative fashion the behavior of alternating currents through various circuits. If a condenser is connected in series with a direct-current source, there is a momentary flow of current due to the charging of the condenser. After the condenser is charged, however, the flow of current in a perfect condenser stops completely. On the other hand, if a condenser is placed in series in an alternating circuit, there is a flow in and out of the condenser and, as the frequency is increased, the apparent flow of current increases to reach a limiting value. Evidently, there will be a time lag between the maximum current and the maximum potential, since the condenser is charged before the potential reaches a maximum. For a perfect condenser (no leakage), the maximum current is a quarter cycle ahead of the maximum potential. The term reactance is applied to the ratio of the maximum potential to the maximum current.

For a pure resistance, the maximum current and maximum potential naturally coincide. In most actual circuits there is a mixture of capacity and resistance. Accordingly, the phase angle between the potential and the current may be anywhere from 0° to 90°. The ratio of the maximum values of the potential to the current in such a mixed circuit is called the impedance. In biological systems we deal, in general, with circuits involving both resistance and capacity. Such a situation can be represented by an equivalent circuit (Fig. 3).

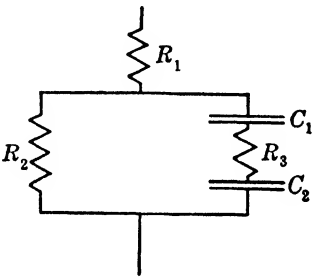


FIG. 3. Circuit equivalent to a living cell. R_1 is the resistance through the medium to the cell, R_2 resistance around the cell, R_3 resistance inside the cell; C_1 and C_2 are the capacities of the cell walls.

From what has been said, it is evident that, in such mixed circuits, as the frequency is increased the reactance tends to disappear and at sufficiently high frequencies pure resistance remains. It is possible, therefore, to analyze the biological circuit by varying the frequency. The general form of the variation of the impedance of biological systems is shown in Fig. 4.

By increasing the frequency sufficiently and wiping out the reactance due to the capacity of the cell wall, the conductances of the insides of the cells can be found. In general, they are surprisingly high. The interior of red blood cells, for example, has a conductance which is equivalent to that of 0.1 to 0.4 per cent KCl solutions. The reactance of living systems is due to the cell wall, which behaves as a condenser. In this connection

there are two possibilities: (1) the membrane acts as a pure dielectric, or (2) the membrane shows preferential permeability for cations or for anions (the membrane becomes polarized).

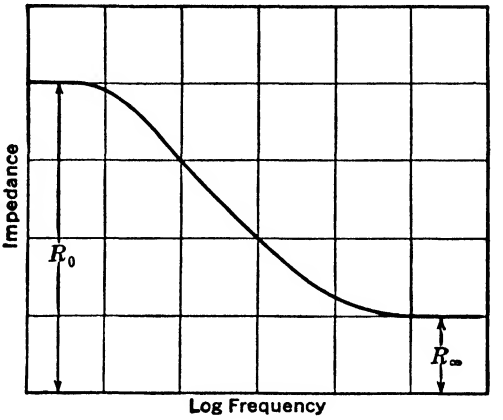


FIG. 4. Diagrammatic representation of the variation of the impedance of a biological system with frequency.

Figure 5 is a schematic representation of the lines of flow through suspensions of unfertilized and fertilized sea-urchin eggs at the indicated frequencies.[10]

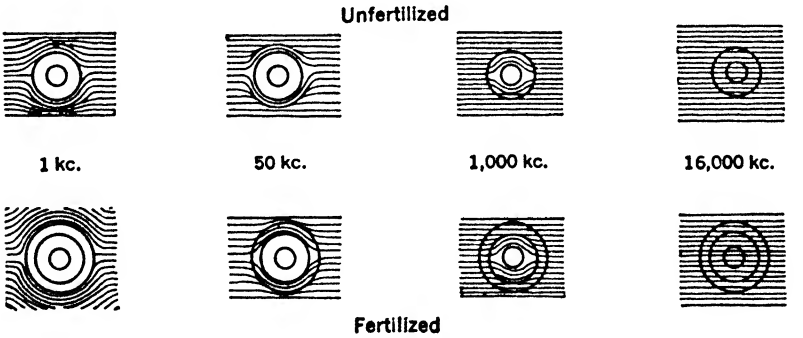


FIG. 5. Lines of current flow at the indicated frequencies through fertilized and unfertilized sea-urchin eggs. Note the fertilization membrane. (Cole.)

[10] K. S. Cole, *Trans. Faraday Soc.*, **33**, 971 (1937).

Chapter X

ELECTROKINETICS

In general, when two phases come in contact there is a separation of electrostatic charges, one phase becoming negative in respect to the other. This separation of charges gives rise to a potential difference at the interface between the phases. The purpose of this chapter is to explain the origin, measurement, and meaning of this interfacial potential.

The two standard books on electrokinetics are those by Abramson [1] and by Abramson, Moyer,* and Gorin.[2]

Origin of Interfacial Charges

It is generally considered that the three ways in which a separation of charge at the interface can arise are as follows: [3]

1. *Ionization.* Numerous substances derive their charge in part or entirely by ionization. If the particle or surface has dissociable groups, the ionization of these groups will leave the surface with a charge of sign opposite to that of the ionized group. For example, proteins owe their charge largely to the ionization of the amino and carboxyl groups. Pauli [4] in particular emphasized the role of dissociable groups in the production of charges on colloidal particles. Pauli believes that colloids consist of three parts: (1) The ionogenic complex, the ionization of which gives rise to the charge on the colloid. (2) Gegen ions, or counter ions, as they are sometimes called. The gegen ions arise from the ionogenic complex and balance the charges on the colloidal particle so that the net charge of the entire solution is zero. (3) The neutral part, which is the uncharged core of the particle. For example, in a Bredig gold sol formed by arcing

[1] H. A. Abramson, *Electrokinetic Phenomena*, The Chemical Catalog Co., Inc., New York, 1934.

* Laurence S. Moyer, a brilliant young scientist who lost his life in the service of his country on June 9, 1942.

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

[3] L. Michaelis, *Effect of Ions in Colloidal Systems*, Williams and Wilkins Co., Baltimore, 1925.

[4] Wo. Pauli, *Trans. Faraday Soc.*, **31**, 11 (1935).

two gold electrodes in a dilute electrolyte solution, the ionogenic complex is $AuCl_2H$. This complex is situated on the surface of the metallic gold particle and ionizes as

$$AuCl_2H \rightarrow AuCl_2^- + H^+$$

The hydrogen ions are the gegen ions, and the metallic gold is the neutral part. In a protein in the presence of, say, hydrochloric acid, the gegen ions are chloride ions. It is probably not true that this is a completely generalized picture of the origin of the electrostatic charges on colloidal particles; as indicated above, there are two other ways by which a surface can acquire a charge.

2. *Adsorption.* The surface can become charged by a preferential adsorption, either of cations or of anions from solution. This is of importance in connection with non-ionogenic surfaces like that of oil in contact with water. Adsorption probably also accounts, at least in part, for the charge on such surfaces as those of cellulose and of cholesterol in contact with water. Actually, in practically all such surfaces anions are preferentially adsorbed to cations, owing to the fact that anions are, in general, less hydrated than cations, so that less work is required to take the anions out of solution and adsorb them on the surface. Compare, for example, the energies of hydration of H_3O^+ which is 130,000 calories per mole and Na^+ which is 115,000 calories per mole with those of OH^- which is 105,000 calories per mole and of Cl^- which is 64,500 calories per mole.

3. *Contact.* In non-aqueous solutions no ions may be present. It is not possible for colloidal particles suspended in such a medium to acquire a charge by ionization or by adsorption; nevertheless such colloidal particles are charged, as is shown by their migration in an electrical field. The theory of the origin of such charges has not yet been formulated. The interfacial potential in such systems is undoubtedly related rather intimately to the structure of the molecules of the liquid phase, as the work of Martin and Gortner,[5] of Jensen and Gortner,[6] and of Lauffer and Gortner [7] show. It can be argued that in many so-called non-aqueous solutions there are traces of water which are responsible for the potential. Quite apart from such dubious experiments, one can demonstrate a separation of charge by rubbing a silk cloth on an amber rod; here there is certainly no possibility that water could be responsible for the charge. Non-aqueous liquids in contact with a solid are quite analogous to a silk cloth in contact with an amber rod.

[5] Wm. Mc. Martin and R. A. Gortner, *J. Phys. Chem.*, **34**, 1509 (1929).
[6] O. G. Jensen and R. A. Gortner, *J. Phys. Chem.*, **36**, 3138 (1932).
[7] M. A. Lauffer and R. A. Gortner, *J. Phys. Chem.*, **42**, 641 (1938).

Methods of Measurement

All methods for the measurement of the interfacial potential depend upon the movement of one phase in respect to another—hence the name electrokinetic potential. This potential can be measured in four ways:

1. *Electroosmosis* involves the movement of a liquid relative to a solid under the influence of an external electrical field applied tangentially to the interface.
2. *Streaming* potential involves the measurement of the potential resulting from the movement of a liquid relative to a solid in response to a mechanical force applied tangentially to the interface.
3. *Electrophoresis* or cataphoresis involves the movement of a solid phase in respect to a liquid under the influence of an external electrical field.
4. *The Dorn effect* involves the movement of a solid phase in respect to a liquid under the influence of a mechanical force. The mechanical force is usually produced by gravity, and this technique is sometimes known as the sedimentation potential method.

Electroosmosis

As indicated above, this method involves the measurement of the velocity of flow of a liquid relative to a solid phase. The motion of the liquid is produced by the application of an external electrical field tangential to the interface. It is not our purpose to give a rigorous derivation of the equations of electroosmosis, but simply to outline the general mathematical technique. This should prove profitable, because it gives an insight into the nature of the interfacial potential. As essentially the same technique is employed to derive equations for the three other methods for measuring the electrokinetic potential, this derivation will serve as a model for the others. The theory of electroosmosis was first quantitatively expressed by Helmholtz [8] and later was reviewed by Smoluchowski.[9]

It is evident that, if there is a separation of charge at the interface between a liquid and a solid, and if an external electrical field is applied tangentially to this interface, the charges residing in the liquid at the interface will move towards the electrode which carries the charge of opposite sign. Suppose that the liquid is situated in a glass capillary and that the electrodes supplying the external electrical field are connected to the two ends of the capillary. The tendency of the charge in the liquid to move and carry along the water will be resisted by the

[8] H. von Helmholtz, *Wied. Ann.*, **7**, 337 (1879).
[9] M. von Smoluchowski. *Anz. Akad. Wiss. Krakau*, **1903**, 182.

viscosity of the liquid. The resistance to flow is given by Newton's law of flow

$$\text{Resistance to flow} = \eta \frac{du}{dl} \qquad 1$$

where η is the coefficient of viscosity of the liquid, u is the linear velocity of flow of the liquid, and l is the distance normal to the wall of the capillary and extending into the liquid. The differential du/dl is thus the flow gradient normal to the wall of the capillary. If the velocity of flow varies directly with the distance, we can replace du/dl by the ratio u/l. The electrical force acting on 1 square centimeter of glass-liquid surface is

$$\text{Electrical force} = E\sigma \qquad 2$$

where E is the potential gradient of the external electrical field parallel to the axis of the capillary and σ is the charge per unit area at the glass-liquid surface. The electrical force is the force tending to produce the motion of the liquid. At a steady rate of flow the acting forces must equal the resisting forces, and

$$E\sigma = \eta \frac{u}{l} \qquad 3$$

As we have seen from Chapter IV, the relation between the charge and the potential for a parallel-plate condenser is

$$\zeta = \frac{4\pi d\sigma}{D} \qquad 4$$

where ζ is the potential across the solid-liquid interface, d is the distance between the centers of gravity of the charges in the liquid and that on the solid, and D is the dielectric constant. In using equation 4, we are implicitly assuming that the distance d is so small compared with the radius of the capillary that the electrical double layer at the interface can be treated as two parallel plates. To proceed with our derivation, we next assume that the distance d is equal to the distance l; i.e., the electrical thickness is equal to the mechanical thickness of the double layer. This allows us to substitute equation 4 into equation 3, from which we find

$$u = \frac{\zeta DE}{4\pi\eta} \qquad 5$$

The average linear velocity of flow, u, is evidently given by

$$u = \frac{V}{\pi r^2} \qquad 6$$

where V is the volume of liquid flowing through the capillary in unit time and r is the radius of the capillary. Substituting equation 6 in equation 5, we have

$$V = \frac{\zeta D E r^2}{4\eta} \qquad 7$$

The radius of the capillary r, and the potential gradient E, can be expressed in terms of the current flowing and of the specific conductance of the liquid since per unit length

$$i = E\kappa_s \pi r^2 \qquad 8$$

where i is the electric current flowing through the capillary and κ_s is the specific conductance of the liquid in the capillary. Substituting equation 8 into equation 7, we have

$$V = \frac{i\zeta D}{4\pi\kappa_s \eta} \qquad 9$$

All electrical quantities in equation 9 are to be expressed in electrostatic units, and V is to be expressed in cubic centimeters per second. As we remember from Chapter IV, 1 e.s.u. volt is equal to 299.8 ordinary volts, 1 e.s.u. charge is equal to 3.33×10^{-10} coulomb, and 1 e.s.u. resistance unit is equal to 8.988×10^{11} ohm centimeters. These factors being substituted along with the numerical value of π, equation 9 becomes

$$\zeta = 1.1295 \times 10^6 \frac{\kappa_s V \eta}{iD} \qquad 10$$

where i is to be expressed in milliamperes, κ_s in mhos, and V in cubic centimeters per second. ζ will be calculated in ordinary millivolts.

A convenient arrangement of an electroosmotic apparatus has been used by Fairbrother and Balkin.[10] The rate of migration of an air bubble along a calibrated capillary serves as a measure of the rate of flow of the liquid through a membrane inclosed between two electrodes. The capillary and membrane constitute a closed system, and any displacement of the liquid in the membrane has to produce a motion of the air bubble in the connected capillary. Non-polarizable electrodes are necessary. Equation 10 has certain serious qualifications which can be discussed best after we have considered the streaming-potential technique.

[10] F. Fairbrother and M. Balkin, *J. Chem. Soc.*, **1931**, 389.

Streaming Potential

Since, as we have seen, the application of an electrical field tangentially to the walls of a capillary will cause a liquid to flow through a capillary, the motion of a liquid by means of a mechanical pressure should produce a difference of potential between the two ends of a capillary. This, in fact, happens, and such potentials are known as streaming potentials. By means of a mathematical technique similar to the one outlined above for the derivation of the electroosmotic equation, an equation relating the streaming potential to the ζ-potential can be derived.[11] This equation is

$$\zeta = \frac{4\pi\eta\kappa_s H}{DP} \qquad 11$$

where H is the streaming potential measured across the ends of the capillary, η is the coefficient of viscosity of the liquid, P is the pressure forcing the liquid through the capillary, and κ_s is again the specific conductance of the liquid in the capillary. It is to be noted that the streaming potential, H, is independent of the length and size of the capillary. This fact is important because it permits streaming-potential measurements in membranes where the capillaries are of every degree of size. The streaming potential is also independent of the shape of the capillary. We shall see presently that, if the size of the capillary becomes very small, serious deviations from equations 9 and 11 are to be expected.

The streaming potential, H, must be measured with a high-resistance apparatus such as a quadrant electrometer or a vacuum-tube potentiometer; otherwise the charge built up at the two ends of the capillary will discharge through the measuring apparatus. The specific conductance, κ_s, is subscripted by s because it includes the surface conductance in the capillary. In short, the conductance of the liquid must be measured in the capillary or in the membrane and not in the bulk of the liquid. All electrical quantities in equation 11 must be expressed in electrostatic units, and the pressure in dynes per square centimeter. The streaming potential and the specific conductance are measured in terms of ordinary units; accordingly, equation 11 must be multiplied by the proper conversion factors

$$\zeta = \frac{4 \times \pi \times 8.988 \times 10^{11} \times \kappa_s \times \eta \times (H/299.8)}{13.53 \times 980.3 \times P \times D} \qquad 12$$

[11] H. von Helmholtz, *Wied. Ann.*, **7**, 337 (1879).
M. von Smoluchowski, *Handbuch der Elektrizität und des Magnetismus*, Vol. II, Barth, Leipzig.

Equation 12 yields ζ in electrostatic volts. Since it is customary to express ζ in ordinary millivolts, the value of ζ calculated from equation 12 must be multiplied by 299.8, and

$$\zeta = 8.513 \times 10^8 \frac{\eta \kappa_s H}{PD} \qquad 13$$

where H and ζ are both expressed in ordinary millivolts and P in centimeters of mercury pressure (density of mercury at 25° C. is 13.53).

Briggs [12] developed a streaming-potential method for membranes of packed cellulose. Starting with the work of Briggs, the Gortner group has published a long series of papers dealing with the streaming-potential technique, which they have brought to a high degree of perfection. Their papers are published for the most part in the *Journal of Physical Chemistry*.

In some important respects the streaming-potential method is superior to the electroosmotic method. It is easier to control experimentally, there is no danger from electrolysis; gold electrodes can be used to measure the potential,[13] whereas reversible electrodes must be used with electroosmosis.

Several ambiguities associated with both the electroosmotic and streaming-potential techniques will now be discussed. If the diameter of the capillaries is too small neither method will yield correct results. Bull and Moyer [14] have discussed the influence of capillary size on the streaming potential. Their arguments apply to electroosmosis as well.

In the derivation of the streaming-potential equation, it was assumed that the thickness of the double layer in respect to the radius of the capillary was so small that the electric double layer could be treated as a parallel-plate condenser. Actually, we are dealing with two concentric cylinders. The potential between two such cylinders is

$$\zeta_1 = \frac{4\pi\sigma r}{D} \ln\left(1 + \frac{d}{r}\right) \qquad 14$$

where r is the radius of the inner cylinder, and d is the distance between the cylinders or, as here, the thickness of the electrical double layer. When $r \gg d$, equation 14 reduces to that for a parallel-plate condenser

$$\zeta_2 = \frac{4\pi\sigma d}{D} \qquad 15$$

[12] D. R. Briggs, *J. Phys. Chem.*, **32**, 641 (1928).
[13] H. B. Bull, *J. Am. Chem. Soc.*, **57**, 259 (1935).
[14] H. B. Bull and L. S. Moyer, *J. Phys. Chem.*, **40**, 9 (1936).

The ratio of equation 14 to equation 15 is

$$\frac{\zeta_1}{\zeta_2} = \frac{r}{d} \ln \left(1 + \frac{d}{r}\right) \qquad 16$$

The ratio ζ_1/ζ_2 is a measure of the influence of decreasing capillary radii on the streaming potential as well as on the electroosmotic flow. Figure 1 shows a plot of this ratio against the ratio of the capillary radii to the thickness of the double layer (curve A). As the capillaries become

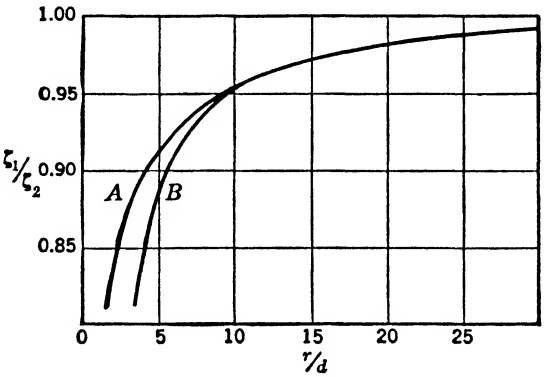


FIG. 1. Diminution in ζ produced by small values of r/d.

smaller, an additional complication enters. The double layers from the opposite walls of the capillary begin to overlap. Lens [15] has discussed this effect. When the equation of Lens is combined with equation 16, we have

$$\frac{\zeta_1}{\zeta_2} = \frac{r}{d} \ln \left[1 + \frac{d}{r} - \frac{1}{(e^{r\ d} - 1)}\right] \qquad 17$$

Plotting the ratio ζ_1/ζ_2 against the ratio r/d gives curve B in Fig. 1. Evidently, if the ratio of r/d is 5 to 1 or smaller, serious deviations from the simple theory of electroosmosis and streaming potential can be anticipated. Later in this chapter we shall discuss the method for calculating d in any specific case. It is sufficient for our present purposes to note that the thickness varies inversely with the square root of the ionic strength. Komagata,[16] who also studied the influence of capillary size on the streaming potential, came to much the same conclusion as Bull and Moyer. Several workers have reported experiments which indicate

[15] J. Lens, *Proc. Roy. Soc.* (*London*), **139A**, 596 (1933).
[16] S. Komagata, *Researches Electrotech. Lab. Tokyo*, No. **362**, 1934.

a basic reality to the above discussion.[17] It is to be concluded that equations 9 and 11 are not valid for very small capillaries. This is an important and somewhat disconcerting conclusion, since it severely limits streaming-potential and electroosmotic work on membranes.

Bikerman [18] has commented on the measurement of streaming potentials through membranes and through gels. He points out that there is evidence for a decreased streaming potential as the packed membrane swells in water. The ionic mobility and, accordingly, the electrical conductance is substantially the same through a gel as through a sol. That is, if we measure the conductance of a gelatin solution before it has gelled, we will find practically no change after the gelatin has gelled. On the other hand, the transport of water is greatly diminished by the presence of a gel. Bikerman applies this idea to the electrokinetic techniques. A liquid moving through a membrane whose capillaries have a cross section Q_1 carries with it per unit length I_1 units of electricity per second, and

$$I_1 = \frac{DP\zeta Q_1}{4\pi\eta} \qquad 18$$

This electrical current builds up a streaming potential H, which causes a return current to flow

$$I_2 = \kappa_s E Q_2 \qquad 19$$

where Q_2 is the cross section available for the movement of ions. In a steady state I_1 equals I_2; hence

$$H = \frac{D\zeta P}{4\pi\eta\kappa_s}\frac{Q_1}{Q_2} \qquad 20$$

It has been the custom to assume that Q_1 equals Q_2, from which equation 11 results, but according to Bikerman, in a swollen membrane such as that of packed cellulose or of insoluble protein in contact with water, Q_1/Q_2 is always less than unity and the calculated ζ-potential is smaller than the real ζ by the ratio Q_1/Q_2. The author agrees with Bikerman that the streaming potential is reduced by the swelling of the membrane, but he prefers to view the matter in the following way: The swelling of cellulose or of membranes of this type means that new capillaries are being formed as water penetrates between the cellulose fibers. These capillaries are of molecular dimensions, but ions can and do penetrate them and electricity is conducted through them. However, the electrical double layer on the surface of the cellulose is thicker than the radii

[17] H. L. White, F. Urban, and E. T. Krick, *J. Phys. Chem.*, **36**, 120 (1932). H. B. Bull and R. A. Gortner, *J. Phys. Chem.*, **36**, 111 (1932).

[18] J. J. Bikerman, *J. Phys. Chem.*, **46**, 724 (1942).

of most of these small capillaries. This results in serious overlapping of the double layers from the opposite walls of the membrane capillaries, with a marked reduction of the ζ-potentials in these capillaries in accord with the theory of Bull and Moyer discussed above. The net effect is to produce a shunt through which the streaming potential of the larger capillaries of the membrane is discharged, and the measured streaming potential is less than it should be. No way of circumventing this difficulty is known.

As we have seen, the motion of a liquid through a capillary builds up a streaming potential across the ends of the capillary. This streaming potential gives rise to an electroosmotic back pressure [19] which can be shown to be

$$P_{\text{Back pressure}} = \frac{8\eta\kappa_s H^2}{Pr^2} \qquad 21$$

The effect of this back pressure is to increase the apparent viscosity of the liquid in the capillary.[20] Lauffer and Gortner [21] were able to confirm the existence of such a back pressure. Except for small capillaries and high streaming potentials, it represents a very small fraction of the imposed pressure.

Surface Conductance

It has been indicated above that the specific conductance of a liquid in a capillary is increased above that in bulk; accordingly, in streaming potentials and in electroosmosis it is necessary to determine the specific conductance of the liquid in the capillaries. This is accomplished by measuring the resistance of the membrane filled with the liquid being investigated. The liquid is replaced by strong KCl solution, and the cell constant of the membrane is determined. The specific conductance of the liquid in the capillaries is then calculated in the usual manner (see Chapter IX). The increase of the conductance above that found in bulk is called the specific surface conductance. Under certain conditions the surface conductance may amount to a very large fraction of the total conductance. For example, the specific conductance of water in a certain cellulose membrane was found to be 31.3×10^{-6} mho whereas the specific conductance of this same water in bulk was 4.8×10^{-6} mho.

In an aqueous solution, surface conductance is due to at least two causes: [22] (1) the movement of the charge at the interface, which is a

[19] H. B. Bull, *Kolloid-Z.*, **60**, 130 (1932).
[20] H. A. Abramson, *J. Gen. Physiol.*, **15**, 279 (1932).
[21] M. A. Lauffer and R. A. Gortner, *J. Phys. Chem.*, **42**, 641 (1938).
[22] J. J. Bikerman, *Kolloid-Z.*, **72**, 100 (1935); *Trans. Faraday Soc.*, **36**, 154 (1940).

true electroosmosis; and (2) the accumulation of neutral salts in the neighborhood of the interface in excess of that in the bulk of the liquid. The electroosmotic part of the surface conductance was formulated by Smoluchowski.[23] It is probable that the electroosmotic factor is important at low salt concentration, but, as the electrolyte concentration is increased, the part of the surface conductance due to the accumulation of neutral salts in the neighborhood of the interface becomes increasingly larger and finally considerably outweighs the electroosmotic part.

Streaming Potential in Biology

In living systems there is a constant ebb and flow of fluids. Every motion of fluids such as the blood flow through the arteries and veins results in a streaming potential. Because the specific conductance of blood is so high, these potentials are small, probably not above 1 or 2 millivolts. In some situations, however, even this small potential may be of great importance. Miller and Dent [24] have suggested that the T wave in an electrocardiogram arises from a streaming potential which results from the forcing of blood out of the coronary arteries as a result of the powerful contraction of the heart muscles. They present a very good case for their theory. The action of $ThCl_4$ on the T wave is curious. When a dilute solution of $ThCl_4$ is injected into the femoral vein of a dog, the height of the T wave is enormously increased while the R and S waves remain unchanged. The very potent action of thorium salts on the ζ-potential is well known, and this observation indicates that the T wave is indeed related to the electrokinetic properties of the coronary arteries.

Kupfer [25] has suggested that the process of hearing may involve a streaming potential.

The Dorn Effect

Dorn [26] in 1880 discovered that particles falling through a liquid impart opposite charges to the top and bottom of the liquid. Smoluchowski formulated the theory of this effect, and Stock [27] performed some experiments of a semiquantitative nature. He was able to show that some rather large potential differences could be established by falling powders. For example, quartz powder falling through toluene produced a difference of potential of 80 volts between the top and the bottom of the

[23] M. von Smoluchowski, *Anz. Akad. Wiss. Krakau* (A), **1912**, 635.
[24] J. R. Miller and R. F. Dent, *J. Laboratory and Clinical Medicine* (in press).
[25] E. Kupfer, *J. Laryngology & Otology*, **53**, 16 (1938).
[26] E. Dorn, *Wied. Ann.*, **10**, 46 (1880).
[27] J. Stock, *Anz. Akad. Wiss. Krakau*, **1905**, 131; **1914**, 95.

toluene column. More recently Quist and Washburn [28] investigated this technique in some detail. At best the method is an awkward one, although there is no doubt that it can be made to yield results consistent with results from the other electrokinetic methods provided that the experimenter is skillful and patient enough.

Electrophoresis

If charged particles are subjected to an external electrical field, they will move towards the electrode whose sign is opposite from their own. The process was formerly called cataphoresis because, when the phenomenon was first observed, the particles were migrating to the cathode. Electrophoresis is a more general and satisfactory term. Electrophoresis is far better suited to the investigation of most surfaces than the three other electrokinetic techniques.

Smoluchowski [29] derived the following equation which expresses the relation between the velocity of migration of the particles and the ζ-potential, the potential gradient, the dielectric constant, and the coefficient of viscosity:

$$\zeta = \frac{4\pi\eta u}{DE} \qquad 22$$

where E is the potential gradient of the external electrical field, u is the velocity of the particles in centimeters per second, and η is the coefficient of viscosity in poises. To express u in microns per second, E in ordinary volts per centimeter, and ζ in ordinary millivolts, we must multiply equation 22 by the proper conversion factors. Then

$$\zeta = 1.1295 \times 10^5 \frac{\eta u}{D} \qquad 23$$

where u is the velocity of migration in microns per second per volt per centimeter. In reference to water at 25° C., equation 23 becomes

$$\zeta = 12.85u \qquad 24$$

Frequently, electrophoretic data are not calculated in terms of the ζ-potential, but are left in terms of the mobility of the particle, i.e., in terms of μ per second per volt per centimeter.

It is to be noted that equation 22 is identical with equation 5. In the derivation of equation 22 the liquid is assumed to be stationary, whereas in the derivation of equation 5 the solid was held stationary.

[28] J. D. Quist and E. R. Washburn, *J. Am. Chem. Soc.*, **62**, 3169 (1940).

[29] M. von Smoluchowski, *Anz. Akad. Wiss. Krakau*, **1903**, 182.

The relative motion is the same in the two cases, and the derivation of equation 5 is really a derivation of the equation of electrophoresis as well. The equation of Smoluchowski was derived on the basis of the following four assumptions: (1) the usual hydrodynamical equations for the motion of a viscous fluid may be assumed to hold both in the bulk of the liquid and within the double layer; (2) the presence of the colloidal particle produces a distortion of the electrical field in such a way that the electrical current passes tangentially along the surface of the particle; (3) the electrical double layer is so thin that the electrical field can be considered to be parallel to the double layer at all points; (4) the electrical field does not deform the double layer.

If the colloidal particle is very small and the electrical double layer is fairly thick, the third assumption of Smoluchowski's is certainly not valid. Henry [30] has taken account of this situation. He finds that, if the ratio of the radius of a spherical non-conducting particle to the thickness of the double layer is not less than 300, Smoluchowski's equation should hold within 1 per cent. As the ratio becomes smaller, the coefficient 4 in Smoluchowski's equation increases to a limiting value of 6, which is reached when the ratio is reduced to 0.5. If the particle is small enough and the thickness of the double layer large enough, the particle approaches the condition of an isolated charged sphere, in which event its electrophoretic mobility may be treated in the following simple manner. The force producing migration is evidently QE where Q is the charge on the particle and E is the potential gradient of the external field. The resistance to motion is the Stokes factor $6\pi r\eta u$, where r is the radius of the sphere and u is its velocity through the medium. At a steady state these two factors must be equal, and

$$QE = 6\pi r\eta u \qquad 25$$

The potential of a charged sphere is

$$\zeta = \frac{Q}{Dr} \qquad 26$$

Substituting equation 26 into equation 25, we find

$$\zeta = \frac{6\pi\eta u}{DE} \qquad 27$$

We see from these simple considerations that, in the limit, the coefficient 4 in Smoluchowski's equation should become 6. This is in accord with

[30] D. C. Henry, *Proc. Roy. Soc.* (*London*), **A133**, 106 (1931).

the conclusions of Henry. Henry also concludes that, for non-conducting particles whose size is large in comparison with the thickness of the double layer, the electrophoretic velocity is independent of the size and shape of the particles and that Smoluchowski's equation is obeyed. If, however, the thickness of the double layer is comparable with the radius of the particle, the electrophoretic velocity becomes a function of both the size and the shape of the particles. Protein molecules, in general, fall in this size range; accordingly, their mobilities should be dependent upon their size and shapes.

Henry also treated the electrophoresis of conducting particles. The current will not be tangential to the surface of a conducting particle but some of it will pass through the particle. If a cylindrical particle is large compared with the thickness of the double layer, Henry finds that

$$u = \frac{2\kappa}{\kappa + \kappa_1} \frac{D\zeta E}{8\pi\eta} \qquad 28$$

where κ is the specific conductance of the medium and κ_1 is that of the particle. When κ_1 is small in comparison to κ, equation 28 reduces to equation 22. Owing to polarization effects colloidal particles are generally non-conducting.

Measurement of Electrophoretic Mobility

Electrophoretic mobilities can be measured by two methods. One is to observe the motion of the particles directly with a microscope; this is called the micromethod. The other, which is used with particles too small to be seen with a microscope, employs the moving boundary of the colloidal solution. The colloidal solution is placed in a U tube, and the motion of the boundaries which the colloidal solution forms with an ultrafiltrate of the colloidal solution is observed.

In the micromethod a suspension of microscopically visible particles is placed in a thin glass cell, a difference of potential is established across the ends of the cell, and the motion of the particles is observed with a microscope. There is an electroosmotic flow of liquid along the surface of the glass cell, but since the cell is a closed system this liquid must return through the center of the cell. The result is that near the top and the bottom of the cell there are stationary levels where the liquid is motionless. At these levels and only at these levels the observed mobility of the particles is their true speed in respect to the liquid. At all other levels the observed speed is the resultant of the electrophoretic and the electroosmotic speeds. If the ratio of the width of the rectangular electrophoretic cell to its thickness is great enough, the stationary levels lie at

0.211 and 0.789 fractions of the total cell thickness as demanded by the theory of Smoluchowski.[31]

As the ratio of the cell width to thickness diminishes, Komagata's [32] correction must be applied, namely,

$$\text{Stationary levels} = 0.5 \pm 0.2887 \sqrt{1 + \frac{1.255}{A}} \qquad 29$$

where A is the ratio of the width of the cell to its thickness.

If the mobilities at various levels through the cell are observed and these mobilities are plotted against the distance from the ceiling of the

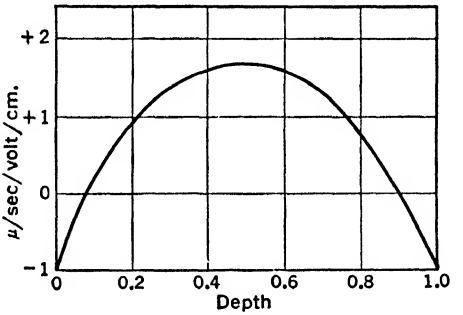


FIG. 2. Plot of the mobilities in a rectangular electrophoretic cell against the distance from the ceiling of the cell. The thickness of the cell is taken as unity.

cell at which the mobility measurements were made, a hyperbolic curve is obtained. Such a curve is shown in Fig. 2.

Moyer [33] has described in detail the apparatus and methods for the microdetermination of electrophoretic mobilities. The construction of the rectangular cell is a job for an expert glass blower. The Abramson [34] cell appears to be the best of the many cells proposed. Briggs [35] has given directions for making a microcell which he reports is satisfactory. The reader is cautioned against the use of the cylindrical microelectrophoretic cell,[36] as this cell has a number of serious objections and, in spite of its apparent advantages, is not acceptable.

The micromethod has much to recommend it. It is comparatively inexpensive, and measurements are made quickly and easily. The moving-boundary method is the fashion of the day, but in the hands of such

[31] M. von Smoluchowski, *Handbuch der Elektrizität und der Magnetismus*, Vol. 2, p. 366, Leipzig, 1921.

[32] S. Komagata, *Researches Electrotech. Lab. Tokyo*, No. 348, 1933.

[33] L. S. Moyer, *J. Bact.*, **31**, 531 (1936).

[34] H. A. Abramson and E. B. Grossman, *J. Gen. Physiol.*, **14**, 563 (1931).

[35] D. R. Briggs, *Ind. Eng. Chem., Anal. Ed.*, **12**, 703 (1940).

[36] S. Mattson, *J. Phys. Chem.*, **32**, 1532 (1928); **37**, 223 (1933).

buffered protein and then dialyzing the protein solution against the buffer until equilibrium is reached. Sufficient electrolyte must be present to render the Donnan effect unimportant. The arms of the U tube are connected to non-polarizable electrodes, and the motion of the protein-buffer boundaries is observed under the influence of a known potential gradient. If the protein is colored, mobilities can be observed directly. If the protein is uncolored, the boundaries can be observed with ultra-violet light since proteins fluoresce in ultra-violet light, or the boundaries can be photographed since proteins are opaque to ultra-violet light. This necessitates the use of a quartz U tube as glass is also opaque to ultra-violet light. A number of cells of the U tube type have been described. The cell used by Davis and Cohn [38] appears satisfactory.

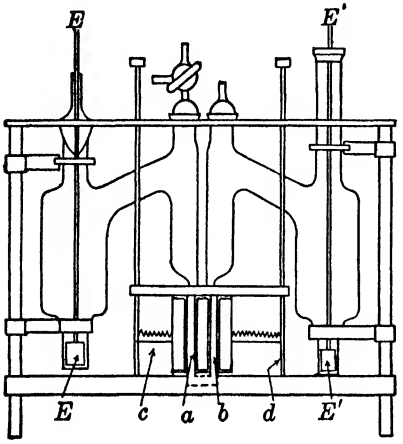


FIG. 4. The electrophoresis cell of Tiselius for quantitative study of the moving boundary and isolation of protein components. E and E' are silver-silver chloride electrodes. a and b are sections of the electrophoresis U tube. c and d are mechanical devices for moving sections a and b.

Tiselius [39] has greatly improved the moving-boundary method, making it into a very elegant tool for the electrophoretic study of proteins. By his method also it is possible to separate a protein mixture into its pure components. The improvement has been so great that it is substantially a new method. The U tube of Tiselius is divided into several sections which are fitted to each other by ground-glass joints that can be slid into and out of place by a mechanical device. The movement of the protein fractions is watched, and when a certain fraction is located in a particular section of the U tube this section is moved out and the contents isolated. As in the simple U tube, the protein solution is in the bottom half of the U. Over the protein solution is layered the buffer, whose electrolyte concentration is the same as that of the protein solution. Sharp boundaries between the protein solution and the buffer must be obtained. Figure 4 shows such an electrophoresis cell.

The most important improvement which Tiselius instituted was the method of observing protein boundaries. He used the so-called schlieren

[38] B. D. Davis and E. J. Cohn, *J. Am. Chem. Soc.*, **61**, 2092 (1939).
[39] A. Tiselius, *Trans. Faraday Soc.*, **33**, 524 (1937).

or shadow method. When light passes through a boundary separating two liquids of different indices of refraction, the light is refracted or bent in the direction of the liquid of greater refractive index. It is a common observation, for example, that a straight elongated object sticking into water appears bent to an observer. The buffer solution and the protein solution have different indices of refraction, the difference being proportional to the protein concentration; the greater the protein concentration, the greater will be the bending of the light rays which pass through the protein-buffer boundary. If an image of the electrophoresis

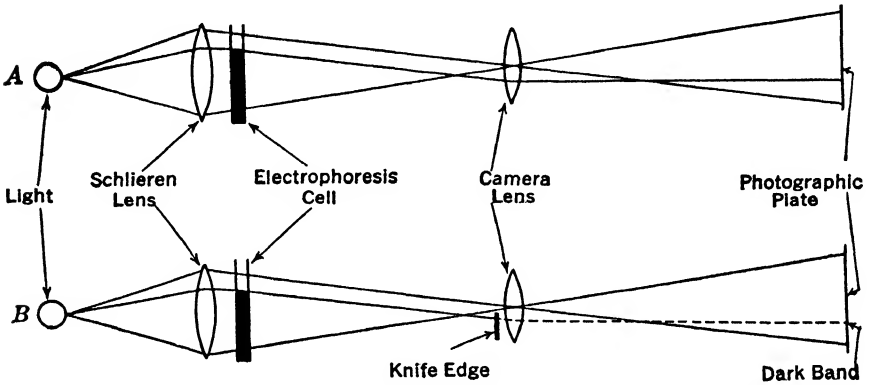


FIG. 5. Schlieren method for visualizing the position of a boundary in an electrophoresis cell. *A*, without knife edge. *B*, with knife edge in place.

cell is focused on a ground-glass plate and a knife edge is placed in the beam of light emerging from the electrophoretic cell in such a way as to intercept the deflected beam, the boundary of the protein-buffer will appear opaque and, accordingly, the position of the boundary can be accurately located on the ground-glass plate as a dark band. This optical system is diagrammed in Fig. 5.

If the protein solution contains more than one electrophoretically distinct component, boundaries for each component will appear as the components move away from each other as the result of their different electrophoretic mobilities, and these boundaries can be visualized on a ground-glass plate. Although the vertical width of the dark schlieren bands is proportional to the protein concentration, no accurate idea of the relative concentrations of the components can be conveniently achieved by this method.

Following the lead of Thovert, Philpot [40] has developed an extraordinarily ingenious optical arrangement which allows not only the boun-

[40] J. Thovert, *Ann. Physik*, (9), **2**, 369 (1914).
J. St. L. Philpot, *Nature*, **141**, 283 (1938).

workers as Abramson and Moyer the micromethod has proved of great value. The book by Abramson, Moyer, and Gorin [2] should be consulted for a description of the large number of surfaces studied by this method. A development which has greatly extended the usefulness of the micromethod was the discovery that inert particles of such substances as glass, carbon, silica, and oil droplets when suspended in a protein solution become coated with the protein. The mobilities of the coated particles

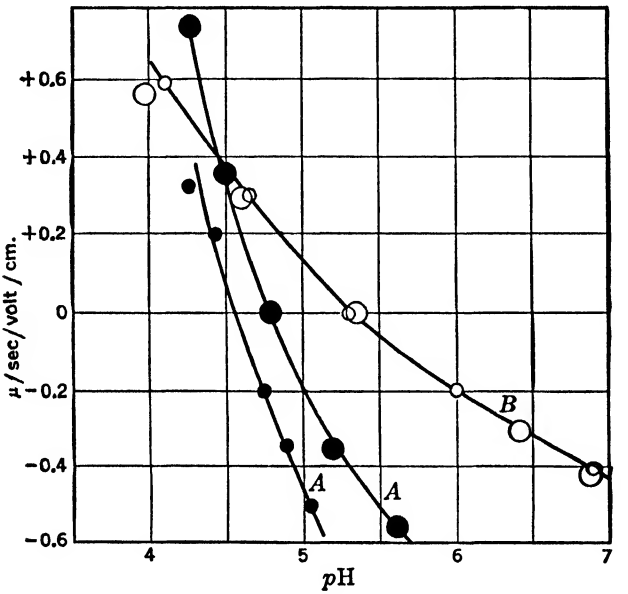


FIG. 3. Comparison of the mobilities of dissolved (small circles) and adsorbed protein (large circles). *A*. Egg albumin. *B*. Horse pseudoglobulin. (Moyer.)

are very close to, if not identical with, the mobilities of the dissolved protein as determined with the moving-boundary technique. Figure 3 shows a comparison of the moving-boundary mobilities with the mobilities of coated particles as determined with the micromethod.[37]

The identity of the adsorbed and dissolved mobilities though extremely useful has certain theoretical implications that are disturbing. According to the treatment of Henry, dissolved protein molecules fall into the region where, in general, their mobilities would be dependent upon both their size and shape. It would be anticipated, on the other hand, that proteins adsorbed on surfaces of microscopically visible particles would obey Smoluchowski's equation, that their mobilities would be independent of size and shape, and, accordingly, that the mobilities of the dis-

[37] L. S. Moyer, *J. Phys. Chem.*, **42**, 71 (1938); *J. Biol. Chem.*, **122**, 641 (1938).

solved protein, as determined by the moving-boundary method, should not agree with the mobilities of inert particles covered with protein. The answer to this seems to be that the protein molecules on the surface of the inert particle retain their own radius of curvature and that this radius of curvature determines their mobilities. This conclusion raises a question about the nature of the surface of any particle, whether covered with protein or not. Certainly surfaces except those of oil droplets are extremely irregular in terms of the dimensions which we must consider. If the thickness of the double layer is very great, the hills and valleys on the surface may be considered substantially plane, whereas, if the thickness of the double layer is made less by the addition of electrolytes, the diffuse double layer probably follows the submicroscopic contour of the surface. At this point the mobility becomes dependent on the radius of curvature of the submicroscopic hills on the surface. At this point, also, the mobility begins to depart from that predicted by the simple theory of Smoluchowski. The situation is by no means clear, however, and a serious ambiguity exists here.

The micromethod is a splendid tool for the determination of the isoelectric point of proteins. The isoelectric point is a physical constant of great utility in protein chemistry. At this point the net charge on the particle is zero, and a protein has its minimum solubility at this *p*H. Microscopically visible glass particles are placed in a fairly concentrated solution of the protein (1 or 2 per cent) and become completely coated with protein in the course of a few minutes. The suspension of glass particles is then diluted with a convenient buffer, the dilute suspension is placed in a microelectrophoresis cell, and the mobility is determined. A series of buffers of different *p*H values are used, and the mobilities of the protein-coated particles are plotted against *p*H. The best line is drawn through these points and the line interpolated to zero mobility. The *p*H at which the mobility is zero is the isoelectric point of the protein. For example, the isoelectric point of pseudoglobulin used in the experiments shown in Fig. 3 is at *p*H 5.3.

Moving-Boundary Method

The moving-boundary method employs a glass tube in the form of a letter U which contains the colloidal solution under investigation. The colloidal solution, for example protein, is placed in the lower part of the U tube and is layered over with buffer of the same electrolyte content as the protein solution. The proper electrolyte content of the buffer is obtained by ultrafiltering the buffered protein solution and using the ultrafiltrate to layer the protein solution, or more conveniently by making up the buffer to as nearly the same electrolyte content as the

buffered protein and then dialyzing the protein solution against the buffer until equilibrium is reached. Sufficient electrolyte must be present to render the Donnan effect unimportant. The arms of the U tube are connected to non-polarizable electrodes, and the motion of the protein-buffer boundaries is observed under the influence of a known potential gradient. If the protein is colored, mobilities can be observed directly. If the protein is uncolored, the boundaries can be observed with ultra-violet light since proteins fluoresce in ultra-violet light, or the boundaries can be photographed since proteins are opaque to ultra-violet light. This necessitates the use of a quartz U tube as glass is also opaque to ultra-violet light. A number of cells of the U tube type have been described. The cell used by Davis and Cohn [38] appears satisfactory.

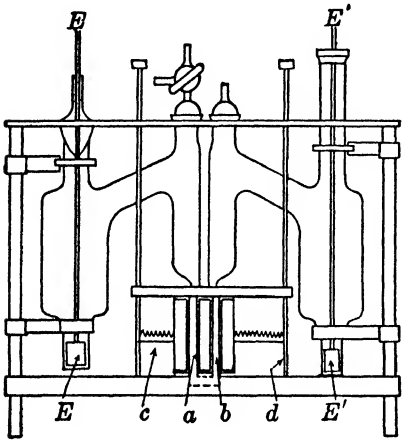


FIG. 4. The electrophoresis cell of Tiselius for quantitative study of the moving boundary and isolation of protein components. *E* and *E'* are silver-silver chloride electrodes. *a* and *b* are sections of the electrophoresis U tube. *c* and *d* are mechanical devices for moving sections *a* and *b*.

Tiselius [39] has greatly improved the moving-boundary method, making it into a very elegant tool for the electrophoretic study of proteins. By his method also it is possible to separate a protein mixture into its pure components. The improvement has been so great that it is substantially a new method. The U tube of Tiselius is divided into several sections which are fitted to each other by ground-glass joints that can be slid into and out of place by a mechanical device. The movement of the protein fractions is watched, and when a certain fraction is located in a particular section of the U tube this section is moved out and the contents isolated. As in the simple U tube, the protein solution is in the bottom half of the U. Over the protein solution is layered the buffer, whose electrolyte concentration is the same as that of the protein solution. Sharp boundaries between the protein solution and the buffer must be obtained. Figure 4 shows such an electrophoresis cell.

The most important improvement which Tiselius instituted was the method of observing protein boundaries. He used the so-called schlieren

[38] B. D. Davis and E. J. Cohn, *J. Am. Chem. Soc.*, **61**, 2092 (1939).
[39] A. Tiselius, *Trans. Faraday Soc.*, **33**, 524 (1937).

or shadow method. When light passes through a boundary separating two liquids of different indices of refraction, the light is refracted or bent in the direction of the liquid of greater refractive index. It is a common observation, for example, that a straight elongated object sticking into water appears bent to an observer. The buffer solution and the protein solution have different indices of refraction, the difference being proportional to the protein concentration; the greater the protein concentration, the greater will be the bending of the light rays which pass through the protein-buffer boundary. If an image of the electrophoresis

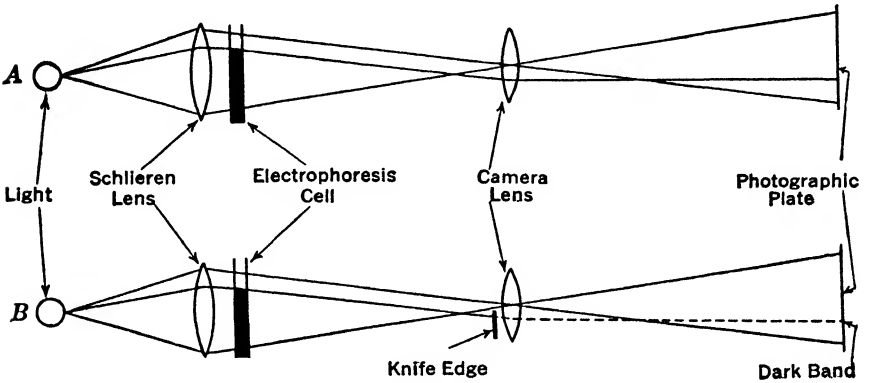


FIG. 5. Schlieren method for visualizing the position of a boundary in an electrophoresis cell. *A*, without knife edge. *B*, with knife edge in place.

cell is focused on a ground-glass plate and a knife edge is placed in the beam of light emerging from the electrophoretic cell in such a way as to intercept the deflected beam, the boundary of the protein-buffer will appear opaque and, accordingly, the position of the boundary can be accurately located on the ground-glass plate as a dark band. This optical system is diagrammed in Fig. 5.

If the protein solution contains more than one electrophoretically distinct component, boundaries for each component will appear as the components move away from each other as the result of their different electrophoretic mobilities, and these boundaries can be visualized on a ground-glass plate. Although the vertical width of the dark schlieren bands is proportional to the protein concentration, no accurate idea of the relative concentrations of the components can be conveniently achieved by this method.

Following the lead of Thovert, Philpot [40] has developed an extraordinarily ingenious optical arrangement which allows not only the boun-

[40] J. Thovert, *Ann. Physik*, (9), **2**, 369 (1914).
J. St. L. Philpot, *Nature*, **141**, 283 (1938).

daries to be accurately located, but likewise a good estimate to be made of the relative concentrations of the protein components. If the proper information is available, the absolute protein concentrations also can be determined. The Philpot optical system has been studied and improved by Svensson.[41] The optics of this system is somewhat complicated;

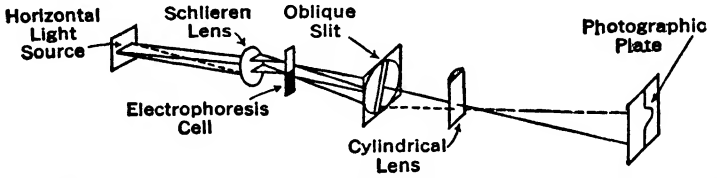


FIG. 6. Diagrammatic sketch of the cylindrical lens system for electrophoresis.

fortunately, the optics does not have to be understood to be used. The light source, a narrow horizontal slit, is focused on an oblique slit by means of a schlieren lens. The electrophoresis cell is placed as close to the schlieren lens as is practicable. The undeviated beam falls at the top of the oblique slit; the deviated beam falls lower down, and its deviation is proportional to the protein concentration. The light which suffers a downward displacement emerges from the oblique slit with a horizontal displacement relative to the undeviated beam, since all other components of the beam have been blocked by the inclined slit. Furthermore, the horizontal displacement is proportional to the original downward displacement. A camera lens collects the light from the oblique slit and passes it into a cylindrical lens. The cylindrical lens has the property of focusing light horizontally but not vertically, and, if there is no concentration gradient in the electrophoresis cell, a sharp vertical straight line will be observed on the ground-glass plate. In the presence of a concentration gradient, the line on the plate will be shifted horizontally to produce a hump at this point. Since the deviated beam has undergone a horizontal displacement due to the blocking by the oblique slit and the deviation is proportional to the protein concentration, the size of the hump will be proportional to the protein concentration. The optical arrangement for the cylindrical lens is shown in Fig. 6.

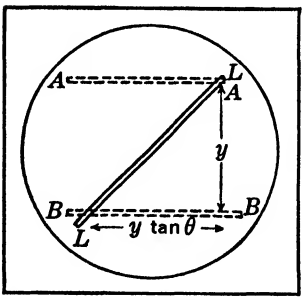


FIG. 7. Front view of the oblique slit. *LL* is the slit opening. *AA* is the position of the image of the undeviated slit. *BB* is the position of the deviated horizontal slit image due to concentration gradient.

[41] H. Svensson, *Kolloid-Z.*, **87**, 181 (1939).

It is necessary to show how the area under the curve which appears on the ground-glass plate or photographic plate is related to protein concentration. Figure 7 is a front view of the oblique slit.

Consider a point in the electrophoresis cell with the coordinate X_1 and the index of refraction gradient $n(X_1)$. The deviation of the horizontal slit image passing through this point is

$$Y_1 = aln(X_1) \tag{30}$$

where a is the thickness of the cell and l is the distance of the cell from the oblique slit. The horizontal component of the displacement of the light produced by the oblique slit is equal to $Y_1 \tan \theta$. The coordinates of a deviated point of light on the ground-glass plate at unit magnification are then

$$X = X_1 \qquad \text{and} \qquad Y = Y_1 \tan \theta \tag{31}$$

The equation of the line formed by the points will be

$$Y = al \tan \theta \frac{dn}{dX} \tag{32}$$

which when rearranged becomes

$$dn = \frac{Y\,dX}{al \tan \theta} \tag{33}$$

Integrating equation 33 between the limits X_2 and X_1, we have

$$n = \frac{1}{al \tan \theta} \int_{X_1}^{X_2} Y\,dX \tag{34}$$

Since there is a linear relation between the index of refraction and protein concentration

$$n = KC \tag{35}$$

where C is the protein concentration in per cent and K is a constant. Then

$$C = \frac{1}{Kal \tan \theta} \int_{X_1}^{X_2} Y\,dX \tag{36}$$

The integral $\int_{X_1}^{X_2} Y\,dX$ is plainly the area under the curve; if the oblique slit is inclined at 45° to the horizontal, tan θ is unity and, accordingly,

$$C = \frac{\text{Area}}{K_1} \tag{37}$$

where K_1 is a constant involving K, a, and l. The area under the curve is obtained with a planimeter. If K_1 is known, the absolute concentra-

tions of the various components can be calculated; if K_1 is unknown, the relative concentrations can be obtained by comparing the areas under the curves of the various protein components present.

Many technical details are involved in the construction and operation of the moving-boundary electrophoretic equipment. Fortunately, several excellent review articles on the subject [42] are available. One important feature is that the water bath in which the electrophoresis cell is immersed must be maintained at or near the temperature of the maximum density of the buffer in order to diminish the convection currents

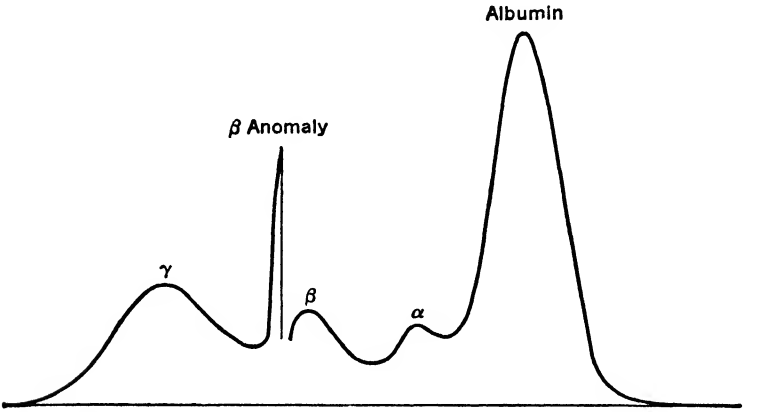


FIG. 8. Electrophoretic pattern of normal human serum at *p*H 7.8 in a veronal buffer. Descending boundary. (Cooper.)

in the cell due to heat changes. For aqueous solutions this temperature is usually in the neighborhood of 4° C.

The Tiselius electrophoretic apparatus with its subsequent refinements has been a tremendous stimulus to electrophoretic work on proteins and particularly on serum protein work. A large number of excellent papers have appeared on the application of electrophoresis to scientific and clinical studies. Since it is not possible to review this work here, the reader should consult the book by Abramson, Moyer, and Gorin.[2]

As noted above, the studies on the blood serum proteins by the electrophoretic technique have been especially valuable. In Fig. 8 is shown the electrophoretic pattern of normal human serum which had been diluted four times with a veronal buffer at *p*H 7.8. This picture was made with the cylindrical lens optical system in the author's laboratory.[43]

The picture shown in Fig. 8 is of the descending boundary; i.e., the

[42] "Electrophoresis," *Ann. N. Y. Acad. Sci.*, **39**, 105 (1939).
L. G. Longsworth, *Chem. Revs.*, **30**, 323 (1942).

[43] J. A. Cooper, unpublished results.

boundary was moving towards the protein instead of towards the buffer. The large, fast albumin component and the slower-moving α, β, and γ globulins are clearly visible. Note the peak of the β globulin. This is known as the β anomaly and is observed only on the descending and not on the ascending limb of the U tube. The occasion for the β anomaly is unknown.

In the ideal case, the patterns of the ascending and descending boundaries should be mirror images of each other. In no actual case is this symmetry observed. The presence of the β-boundary anomaly in the descending limb of the electrophoretic cell and its absence in the ascending limb has already been noted. There are other discrepancies. The concentrations of the several components as calculated from the areas under the corresponding peaks, are not identical for the two boundaries. The ϵ-boundary which is due to a buffer electrolyte gradient is found in the descending limb of the electrophoretic cell (this salt boundary has not been resolved from the γ-globulin as shown in Fig. 8). A δ-boundary which results from gradients of both buffer and protein occurs in the ascending limb.

Longsworth [44] reports that in 0.1 N sodium diethylbarbiturate buffer at pH 8.6 and with human blood plasma there is a complete separation of the ϵ- and δ-peaks from the peak due to the γ-globulin. This separation permits an accurate determination of the area under the γ-globulin peak. Longsworth also noted that in this buffer it was possible to separate a new component which occurred between the albumin and the α-globulin peaks and which he designated as α_1.

Perlmann and Kaufman [44a] have investigated the electrophoresis of human blood plasma as a function of the total protein and of the ionic strength. They found that the concentrations of the various components were functions of both the dilution of the plasma proteins and of the ionic strength. More satisfactory results appeared to be obtained if the concentration of the plasma proteins was maintained as low and the ionic strength made as high as was consistent with good electrophoretic technique.

Validity of Electrokinetic Methods

We can show that the equations for electroosmosis and for streaming potentials obey certain simple requirements. Velisek and Vasicek [45] have found that, in accord with equation 9, when the volume of liquid flowing is plotted against the current, a straight line is obtained. They

[44] L. G. Longsworth, *Chem. Revs.*, **30**, 323 (1942).

[44a] G. E. Perlmann and D. Kaufman, *J. Am. Chem. Soc.*, **67**, 638 (1945).

[45] J. Velisek and A. Vasicek, *Z. physik. Chem.*, **A71**, 281 (1934).

used membranes made of porous porcelain. Bull [46] was able to show that the plot of the streaming potential against the hydrostatic pressure also yields a straight line as demanded by equation 11. He used Pyrex-glass capillaries.

It has been shown that the methods of electroosmosis, of streaming potential, and of electrophoresis are consistent in that they yield the same calculated ζ-potential for identical surfaces in identical electrolyte solutions. The microelectrophoretic method provides a way for determining not only the electrophoretic mobility but also the electroosmotic mobility. Abramson [47] has pointed out that in a rectangular microelectrophoretic cell

$$U_e = 2(U_{1/2} - U_c) \qquad 38$$

where $U_{1/2}$ is the observed mobility at the midpoint of the cell, U_c is the electrophoretic mobility, U_e is the electroosmotic velocity of flow of the water past the cell wall. Moyer and Abramson [48] coated quartz particles and the inside surfaces of the electrophoretic cell with protein. The particles and cell walls, accordingly, had identical surfaces. Under these conditions the electroosmotic mobility of the water past the cell surface was found to be equal to the electrophoretic mobility of the coated quartz particles. This shows that electrophoresis and electroosmosis are consistent with each other. Incidentally, if the surfaces of the particles and of the inside surfaces of the electrophoretic cell are identical, U_c equals U_e and from equation 38 U_c must equal $\frac{2}{3}U_{1/2}$. This allows the electrophoretic mobility to be calculated from the particle mobility at one-half the depth of the electrophoretic cell. Experimentally, the half thickness is the most convenient place to make measurements. Bull [49] performed the same type of experiments on protein-covered surfaces as did Moyer and Abramson, and he included in his studies streaming potentials through glass capillaries covered with protein. He was able to show that streaming potential, electrophoresis, and electroosmosis are all consistent with one another.

The demonstration that electrophoresis, electroosmosis, and streaming potential are consistent with one another does not, of course, mean that the true ζ-potential is calculable from equations 9, 11, and 22. It may simply mean that the same correction factor must be applied to the three methods. For smooth capillaries of any shape whose cross-sectional dimensions are very much larger than the thickness of the electrical double

[46] H. B. Bull, *Kolloid-Z.*, **66**, 20 (1934).
[47] H. A. Abramson, *Colloid Symposium Monograph*, **8**, 289 (1930).
[48] L. S. Moyer and H. A. Abramson, *J. Gen. Physiol.*, **19**, 727 (1936).
[49] H. B. Bull, *J. Phys. Chem.*, **39**, 577 (1935).

layer, and for smooth particles of any shape whose three dimensions are all very much larger than the thickness of the double layer, it is reasonably certain that equations 9, 11, and 22 are valid. On the other hand, if the size and shape of the submicroscopic hills and valleys on surfaces influence the potential as the experiments of Abramson and Moyer on protein-covered surfaces seem to indicate, then the situation for any real surface becomes indeterminate and the true ζ-potential is not, in general, calculable from electrophoresis, electroosmosis, or streaming-potential measurements without more information than we now have. In addition to this qualification, it is fairly well established that, if the radius of the capillaries approaches the thickness of the double layer, equations 9 and 11 for streaming potential and electroosmosis are invalid. Also, if the radius of the particle approaches the magnitude of the thickness of the double layer, equation 22 for the ζ-potential from electrophoretic measurements is invalid and the relation between mobility of the particles depends both upon the shape and size of the particles. The relation between mobility and ζ-potential has been formulated for small spheres, for small circular disks, and for long thin cylinders,[2] but for no other shapes. We do not know the sizes or shapes of protein molecules in solution, and it follows that it is quite impossible to calculate the ζ-potential of a protein molecule; only approximations of it can be made.

Since it has been shown experimentally that the methods of electrophoresis, electroosmosis, and streaming potential are equivalent to one another, it would appear that the best way of expressing electrokinetic data is by means of a function common to all three methods. Mobility in μ per second per volt per centimeter is widely used to express electrophoretic data. It is proposed to use this term also for electroosmotic and streaming-potential data. This is quite a reasonable method of expression for electroosmosis and would have a definite physical meaning. It would have no direct physical meaning for streaming potential but would bring this technique into line with the other methods without the use of the ζ-potential.

The numerical equivalence of these expressions using ordinary electrical units and expressing P in centimeters of mercury is

$$7.54 \times 10^3 \frac{\kappa_s H}{P} = 10.00 \frac{\kappa_s U}{i} = u \qquad 39$$

where H is expressed in ordinary volts, i in amperes, U in cubic centimeters per second, κ_s in mhos, and u in μ per second per volt per centimeter. $\kappa_s H/P$ is for streaming potential; $\kappa_s U/i$ is for electroosmosis; and u is for electrophoresis.

Structure of the Electrical Double Layer

Helmholtz, who did the first important theoretical work on electrokinetics, considered that an electrical double layer of fixed thickness existed at the surface. This layer was assumed to be of the order of one molecule in thickness and could be represented by two parallel flat plates. He further considered the dielectric constant to have the value of unity. Gouy [50] later pointed out that the double layer at the interface could not be rigid, compact, and of fixed thickness, but must be diffuse. He pictured the situation as follows: At very low temperatures ions are rigidly adsorbed on the surface, thus forming a true Helmholtz layer, but as the temperature is raised, say to room temperature, a large fraction of the adsorbed ions "vaporize" from the surface. However, these ions are not able to escape entirely from the influence of the charge on the wall. This ionic atmosphere in the immediate neighborhood of the surface is fairly dense, and at greater distances from the surface the density diminishes until the net charge density is zero. Not all the ions in the diffuse double layer are of opposite sign from those on the wall, but the ions of opposite sign do predominate. In view of the fact that the diffuse layer extends some distance into the water phase, it is necessary to consider the dielectric constant; this constant is assigned the value it has in pure water. Such was Gouy's picture. On the basis of this concept of the double layer, he was able to derive equations relating the net charge density to the distance between the center of gravity of the net charge in the diffuse layer and the center of gravity of the charges fixed on the solid surface and to the potential of the double layer. He also formulated an equation for the thickness of the double layer. Thus in 1910 an adequate mathematical theory of the double layer was available, but it was not until several years after the advent of the Debye-Hückel theory of strong electrolytes that the importance of these ideas for electrokinetic phenomena was realized.[51]

The same type of mathematical approach was used by Gouy as was later employed by Debye and Hückel (see Chapter V). By means of the Boltzmann principle, the charge density of the diffuse layer was calculated. This was substituted in Poisson's equation and the equation integrated to yield a relation between charge density and the ζ-potential. The electrostatic charge per square centimeter of surface is given by

$$\sigma = \sqrt{\frac{DRT}{2{,}000\pi}} \sqrt{C_c(e^{-Z_c\epsilon\zeta/kT} - 1) + C_a(e^{Z_a\epsilon\zeta/kT} - 1)} \qquad 40$$

[50] L. Gouy, *J. phys.*, **9**, 457 (1910).

[51] E. F. Burton, *Colloid Symposium Monograph*, **4**, 132 (1926).

H. A. Abramson, *Electrokinetic Phenomena*, The Chemical Catalog Co., Inc., New York, 1934.

where the subscript c applies to cations and the subscript a to anions. In water at 25° C.,

$$\sigma = 1.765 \times 10^4 \sqrt{C_c(e^{-0.039Z_c\zeta} - 1) + C_a(e^{0.039Z_a\zeta} - 1)} \qquad 41$$

where ζ is expressed in ordinary millivolts and σ is obtained as the number of electrostatic charges per square centimeter. For uni-univalent salts in water at 25° C.,

$$\sigma = 3.53 \times 10^4 \sqrt{C} \sinh 0.039\zeta \qquad 42$$

where sinh is the hyperbolic sine. Tables of hyperbolic functions are available in most physical and chemical handbooks.

The thickness of the double layer is considered to be the distance between the center of gravity of the fixed layer on the solid surface and that of the diffuse layer. It is equal to the reciprocal of κ of the Debye-Hückel theory, and

$$d = \frac{1}{\kappa} = \sqrt{\frac{1{,}000DRT}{8\pi\epsilon^2 N^2 \mu}} \qquad 43$$

where d is the thickness of the double layer and μ is the ionic strength. In water at 25° C.

$$d = \frac{3.05 \times 10^{-8}}{\sqrt{\mu}} \text{ centimeter} \qquad 44$$

Note that the thickness of the double layer is independent of the potential and charge of the layer and at any given temperature depends only on the ionic strength. The value of the constants and variables to be used in equations 40 and 43 are given in Table 1.

TABLE 1

CONSTANTS AND VARIABLES TO BE USED IN EQUATIONS 40 AND 43

Function	Name	Value
ϵ	Elementary charge	4.80×10^{-10} e.s.u.
k	Boltzmann constant	1.38×10^{-16} erg per degree
N	Avogadro's constant	6.023×10^{23}
$R(kN)$	Molar gas constant	8.31×10^7 ergs per degree
D	Dielectric constant	78.54 water at 25° C.
π		3.1416
T	Absolute temperature	
Z_c	Valence of cation	
Z_a	Valence of anion	
C_c	Moles cations per liter	
C_a	Moles anions per liter	

As a first approximation, if the double layer is considered as a parallel-plate condenser, then

$$\zeta = \frac{4\pi d\sigma}{D} \qquad 45$$

It is evident from equation 45 that a variation of ζ may be a reflection of a change of d or of σ or of both. It was thought at one time that electrolytes precipitate colloids by discharging the particles. Actually, as a general picture this is erroneous.[52] The addition of electrolytes decreases d but increases σ. This is to be anticipated; the more ions present, the greater the adsorption to be expected. The coagulating effect of electrolytes on colloids is to be explained as due to a decrease in the ζ-potential. Other complicating factors will be considered duly in Chapter XI.

Table 2 shows the variation of the thickness of the double layer with various concentrations of different types of electrolytes.

TABLE 2

THICKNESS OF THE ELECTRICAL DOUBLE LAYER IN THE PRESENCE OF ELECTROLYTES

Concentration moles per liter	Thickness of double layer in Å		
	Uni-univalent	Uni-divalent	Di-divalent
10^{-5}	1,000	560	480
10^{-3}	100	...	...
10^{-1}	10	5.6	4.8

Proteins and the Diffuse Double Layer

Abramson [53] pointed out that there must be a connection between the electrophoretic mobility of a protein and the titration curve of a protein, and he formulated the following rule: "In solutions of the same ionic strength, the electric mobility of the same protein at different hydrogen-ion activities should be directly proportional to the number of hydrogen (hydroxyl) ions bound." He reasoned that the protein molecule in solution can be pictured as representing two concentric spheres, the inner sphere being the surface of the protein molecule, and the outer sphere the locus of the center of gravity of the net charge in diffuse ionic layer.

[52] H. B. Bull and R. A. Gortner, *J. Phys. Chem.*, **35**, 309 (1931).
[53] H. A. Abramson, *J. Gen. Physiol.*, **15**, 575 (1932).

The integration of Poisson's equation for such a situation leads to

$$\zeta = \frac{Q}{Dr\left(\frac{r}{d} + 1\right)} \qquad 46$$

and since for small spheres where $d > r$

$$\zeta = \frac{6\pi u\eta}{D} \qquad 47$$

we have upon substituting equation 47 into equation 46

$$u = \frac{Q}{6\pi\eta r\left(\frac{r}{d} + 1\right)} \qquad 48$$

If the value of d is maintained constant by means of a series of buffers of different *p*H but of the same ionic strength, the mobility of a spherical protein molecule will be directly proportional to the charge. Ordinarily, as we have seen, the mobility is a function of both the charge and the thickness of the double layer. The net charge on a protein molecule, as a first approximation, is equal to the sum of the hydrogen ions bound or to the sum of the hydroxyl ions bound. This information can be obtained from a titration curve of the protein with acids and with bases. Abramson compared the charge calculated by means of equation 48 with that calculated from titration data. The mobility of the protein was plotted as a function of the *p*H. The titration curve was plotted on the same graph by shifting the coordinates of the titration curve until it passed through zero at the isoelectric point as determined by electrophoresis. The mobility data were then adjusted by selecting one single mobility and determining the factor by which it must be multiplied to bring it on the titration curve. All other mobilities were multiplied by this same factor. By making these adjustments, remarkable agreement between the mobility and titration curves was obtained. This proportionality has been confirmed by Longsworth,[54] who was able to demonstrate proportionality between the mobility and the titration curve of egg albumin from *p*H 3 to 12. Cannan, Palmer, and Kibrick [55] have also succeeded in showing a satisfactory relation between these two factors for β-lactoglobulin from cow's milk over a wide range of *p*H values.

The calculation of the actual charge on a protein molecule from mobility measurements is not simple. The shape and size of the molecule

[54] L. G. Longsworth, *Ann. N. Y. Acad. Sci.*, **41**, 267 (1941).

[55] R. K. Cannan, A. H. Palmer, and A. Kibrick, *J. Biol. Chem.*, **142**, 803 (1942).

must be known; furthermore, the relation of the mobility to the shape, size, and charge for this particular shape and size must be known. Abramson, Moyer, and Gorin have carried out these calculations for egg albumin on the assumption that this molecule is rodlike with a ratio of length to breadth of 4 to 1. The relation between the charge on the molecule and the titration curve is also uncertain. Steinhardt [56] has found that anions as well as hydrogen ions are bound in the acid region. This means that the charge calculated by considering only the hydrogen ions bound will be in excess of that actually present. Gorin and Moyer report that they avoided this difficulty by extrapolating the experimental titration curve to zero ionic strength and to zero protein concentration.

No ambiguity is associated with the calculation of the thickness of the double layer by means of equation 43. This equation is valid irrespective of the shape and sizes of particles. It is also independent of the ζ-potential.

[56] J. Steinhardt, *Ann. N. Y. Acad Sci.*, **41**, 287 (1941).

Chapter XI

SURFACE ACTIVITY

Inasmuch as in any living system countless surfaces and interfaces are present, the subject of surface activity is of interest to the biologist and the biochemist. Unfortunately, the study of surfaces in living or, for that matter, dead tissues is very difficult, and as yet the subject has been explored to a limited extent. The full development of this field must await the more subtle approaches of the future. It is the purpose of this chapter to enumerate and describe the properties of surfaces as they are known to the physical chemist with the hope that such an exposition will stimulate the biologist to complete his side of the picture. Readers interested in the romantic aspects of this subject should read *Flatland* by Abbott;[1] our discussion will be of a more prosaic nature.

Surface Tension

It is a familiar observation that a sewing needle, if placed carefully on a clean water surface, will float; the water behaves as if it had a skin. The surface of water is really under a considerable tension; the needle does not have sufficient mass to overcome this tension, and as a result it floats.

The physical origin of surface tension may be pictured somewhat as follows: In the body of the liquid the molecules are acted upon in all directions by the attractive forces of other liquid molecules. The molecules in the surface, however, are attracted only downward and sideways. The molecules in the body of the liquid thus have a lower potential energy than those on the surface, and, accordingly, in order to bring a molecule from the interior of the liquid to the surface, work must be done. Any extension of the surface requires the expenditure of energy.

Surface tension can be expressed in one of two ways, which are numerically and dimensionally equivalent, namely, in dynes per centimeter or in ergs per square centimeter, i.e., in force per unit length or in work per unit area. The dimensions of surface tension are M/t^2.

The surface tension of pure liquids is independent of extension: i.e., no matter how greatly the surface is extended, the work required to

[1] Rev. E. A. Abbott, *Flatland*, Little, Brown & Co., Boston, 1926.

produce an additional square centimeter of surface is the same as it was *initially*. Rubber is a bad analogy for the surface tension of pure liquids because, as is well known, the force required to extend rubber is proportional to the extension. Under some conditions, however, liquid surfaces do show elasticity, i.e., the surface tension is proportional to the extension. The surfaces of dilute solutions of surface-active substances as well as surfaces of spread monomolecular films show elasticity. Such surfaces will be considered presently.

The thickness of the surface film of a pure liquid is undoubtedly a function of the nature of the liquid. By such a thickness we mean, of course, the transition distance between the uniform conditions of the gaseous phase to the uniform conditions of the liquid phase. McBain et al.[2] report from optical measurements that the least possible thickness of the water surface is 2 to 3 Å.

Measurement of Surface Tension

Numerous methods are available for the measurement of surface tension. In general, they may be divided into two classes, static and dynamic. Some methods do not belong to either class but are intermediate between the truly static and truly dynamic methods. The static surface tension is the equilibrium tension; the dynamic tension is the tension of a liquid before the surface film has had time to form. The time required for the attainment of surface equilibrium of pure liquids is exceedingly small, probably of the order of a millionth of a second. The time required for a solution in which the solute molecules are of moderate size to reach equilibrium is also exceedingly short. It is only when one deals with solutions of colloidal materials that the time required for the formation of the surface is appreciable. We shall confine our attention to methods which give or are supposed to give a static surface tension. The most important and useful of these static methods are the capillary-rise method, the du Noüy ring method, the Wilhelmy balance method, and the drop-weight method.

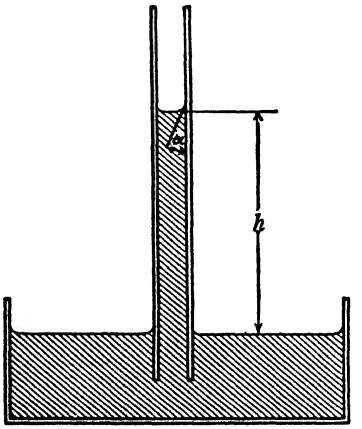


FIG. 1. Rise of a liquid in a capillary.

[2] J. W. McBain, R. C. Bacon, and H. D. Bruce, *J. Chem. Phys.*, **7**, 818 (1939).

The capillary-rise method is a truly static one. It consists in measuring the height of rise of the liquid in a capillary. Consider a capillary of radius r, into which ascends a liquid to height h, above the surface of the liquid. This liquid is to make an angle of contact α with the walls of the capillary (see Fig. 1). It is evident that the force pulling the liquid up is $2\pi r\sigma \cos \alpha$, while the force pulling downward is $\pi hr^2 g\rho$, where σ is the surface tension, g is the acceleration of gravity, and ρ is the density of the liquid. At equilibrium the upward and downward directed forces must equal each other, and

$$\sigma = \frac{hrg\rho}{2 \cos \alpha} \qquad 1$$

In practice only those liquids are employed which completely wet the surface of the capillary. Under this condition, the angle of contact is zero and cos α is unity. The difficulty with the method is that the bore of the capillary must be exactly uniform—a requirement hard to meet. It is not so important that the inside of the capillary be an exact cylinder; i.e., the cross section of the capillary may be slightly elliptical without seriously affecting the accuracy of the measurement. The capillary-rise method is entirely unsuited for the measurement of the surface tension of protein solutions; the protein surface films tend to stick in the capillary and prevent the solution from reaching its equilibrium height. This method, accordingly, is severely limited in its application to the determination of the surface tension of biological fluids. Critical papers dealing with this method have been published by Richards and Carver,[3] by Harkins and Brown,[4] and by Jones and Frizzell.[5]

Jones and Ray [6] have devised a differential method based on the capillary-rise technique which is capable of extreme accuracy. The apparatus consists of a wide tube and a narrow one in the shape of a U. The liquid is brought to a fixed mark in the narrow tube by the addition of liquid to the larger tube. The main feature is that the difference in levels of the narrow and wide menisci is found by weighing the total amount of liquid in the apparatus. The level of the liquid in the smaller tube being fixed, it is not necessary to insist upon extreme uniformity of the diameter of the smaller tube. With this apparatus Jones and Ray were able to obtain some interesting and unexpected results on the surface tension of aqueous solutions of inorganic electrolytes. Previous to their work it was thought that inorganic electrolytes, throughout

[3] T. W. Richards and E. K. Carver, *J. Am. Chem. Soc.*, **37**, 1656 (1915).
[4] W. D. Harkins and F. E. Brown, *J. Am. Chem. Soc.*, **41**, 503 (1919).
[5] G. Jones and L. D. Frizzell, *J. Chem. Phys.*, **8**, 986 (1940).
[6] G. Jones and W. A. Ray, *J. Am. Chem. Soc.*, **59**, 187 (1937).

their entire concentration range, raised the surface tension of water.[7] Jones and Ray found characteristic minima in the concentration-surface tension curves at very low electrolyte concentrations. Dole [8] was able to formulate a theory of this effect. Langmuir [9] has also proposed a theory of the Jones-Ray effect in which he suggested that the minima in the surface tensions were only apparent and concluded that they resulted from the electrostatic charges on the surface of the glass capillary. The recent results of Long and Nutting [10] would appear to support Langmuir's interpretation. However, this question must be regarded as unsettled.

The du Noüy ring method has the advantage of simplicity, and the complete apparatus is commercially available. The method consists in determining the force required to detach a platinum ring from the surface of a liquid. The simple theory requires that the surface tension shall be the force required to pull the ring out of the surface divided by twice the circumference of the ring. The distance is twice the circumference because two films are involved, an inner and an outer one. Actually, the surface tension as measured by the tensiometer is sensitive to the diameter both of the ring and of the wire. Large errors may be introduced by the use of rings of improper diameter. In spite of the acknowledged errors to which this method is subject, it is surprising how frequently quite accurate results are obtained with the du Noüy tensiometer. Good references to this method are papers by Harkins and Jordan,[11] by Freud and Freud,[12] and by Hauser.[13]

Dole and Swartout [14] have devised a differential method involving the use of two large platinum rings. One ring dips into pure water and the other into the solution whose surface tension is to be measured. The rings are supported from the ends of the beam of an analytical balance. The liquids are simultaneously lowered, and the ring which is detached is noted. Weights are added or subtracted from the balance until a small change of weight will cause one or the other rings to break from the surface. By comparing this weight with the weight found when both pans contain pure water, an accurate value of the relative surface tension of the solution is obtained. Dole and Swartout investigated the

[7] L. Onsager and N. N. T. Samaras, *J. Chem. Phys.*, **2**, 528 (1934).

[8] M. Dole, *J. Am. Chem. Soc.*, **60**, 904 (1938).

[9] I. Langmuir, *Science*, **88**, 430 (1938).

[10] F. A. Long and G. C. Nutting, *J. Am. Chem. Soc.*, **64**, 2476 (1942).

[11] W. D. Harkins and H. F. Jordan, *J. Am. Chem. Soc.*, **52**, 1751 (1930).

[12] B. B. Freud and H. Z. Freud, *J. Am. Chem. Soc.*, **52**, 1772 (1930).

[13] A. E. Hauser, H. E. Edgerton, B. M. Holt, and J. I. Cox, Jr., *J. Phys. Chem.*, **40**, 973 (1936).

[14] M. Dole and J. A. Swartout, *J. Am. Chem. Soc.*, **62**, 3039 (1940).

surface tension of aqueous solutions of electrolytes and obtained the characteristic minima in the surface tension-concentration relation at low electrolyte concentrations as reported by Jones and Ray.[6]

The Wilhelmy balance method was devised by Wilhelmy [15] a number of years ago and revived by Harkins and Anderson.[16] The apparatus consists simply of a thin strip of glass dipping into the liquid under investigation and suspended from an arm of an analytical balance (a chainomatic balance is most convenient). The weight of the dry slide in air, and its weight while dipping into the liquid, are determined. After the buoyancy correction of the liquid displaced by the slide is added, and the weight of the slide in air subtracted from the weight when dipping in the liquid, the surface tension can be directly calculated. The net pull of the surface on the slide in grams is multiplied by the acceleration of gravity and divided by twice the length of the slide. The result is the surface tension of the liquid. The Wilhelmy method involves no rupture of the surface, and the surface is disturbed to a very small extent. This method is highly recommended. The author has found it to be of special value in the measurement of the surface tension of protein solutions and of other surface-active colloidal materials. A requirement of this method is that the slide must be completely wet by the liquid (zero contact angle).

The drop-weight method depends upon the principle that the surface tension along with the density of a liquid determines the size of a drop forming at the end of a capillary. The larger the surface tension, the bigger the drop is before it breaks from the tip of the capillary and falls. These simple considerations yield the following equation:

$$\sigma = \frac{wg}{2\pi r} \qquad 2$$

where w is the weight of the drop in grams, g is the acceleration of gravity, and r is the radius of the capillary tip. The trouble is that not all the drop on the tip of the capillary falls. In addition to this complication Hauser [17] has found by means of ultra fast motion pictures that, when such a drop falls, it is accompanied by a number of smaller drops. These difficulties lead to serious errors if equation 2 is used to calculate the surface tension. Harkins and Brown [18] found empirically that a more

[15] L. Wilhelmy, *Ann. Physik*, **119**, 177 (1863).
[16] W. D. Harkins and T. F. Anderson, *J. Am. Chem. Soc.*, **59**, 2189 (1937).
[17] H. E. Edgerton, E. A. Hauser, and W. B. Tucker, *J. Phys. Chem.*, **41**, 1017 (1937).
[18] W. D. Harkins and F. E. Brown, *J. Am. Chem. Soc.*, **41**, 499 (1919).
W. D. Harkins, *Nature*, **117**, 690 (1926).

accurate expression for surface tension as determined by the drop-weight method is

$$\sigma = \frac{wg}{2\pi f(r/V^{1/3})} \qquad 3$$

where V is the volume of the drop and f indicates that the surface tension is a function of $r/V^{1/3}$. Harkins and Brown studied the corrections to be applied, and their paper should be consulted. The drop-weight method is probably the least satisfactory of the common methods for determining surface tension.

Other satisfactory methods for the determination of surface tension are the sessile drop method [19] and Sugden's [20] maximum bubble pressure method.

Table 1 shows the values of the surface tension of some common liquids.

TABLE 1

SURFACE TENSION OF SOME COMMON LIQUIDS IN DYNES PER CENTIMETER AGAINST AIR

Substance	Temperature, ° C.	Surface tension
Acetic acid	20	27.6
	50	24.7
Acetone	20	23.7
	60	18.6
Benzene	20	28.88
	50	25.0
n-Butyric acid	20	26.8
	50	24.0
Ethanol	20	22.3
	50	19.8
Methanol	20	22.6
	50	20.1

Table 2 gives the surface tension of water in dynes per centimeter against air at various temperatures.

[19] A. W. Porter, *Phil. Mag.*, **25**, 752 (1938).
H. V. Tartar, V. Sivertz, and R. E. Reitmeier, *J. Am. Chem. Soc.*, **62**, 2375 (1940).
[20] S. Sugden, *J. Chem. Soc.*, **121**, 858 (1922).
T. H. Hazlehurst, *J. Chem. Education*, **19**, 61 (1942).

TABLE 2

SURFACE TENSION OF WATER IN DYNES PER CENTIMETER AGAINST AIR

Temperature, ° C.	Surface tension
0	75.6
10	74.22
15	73.49
18	73.05
20	72.75
25	71.97
30	71.18
40	69.56
50	67.91
60	66.18

Surface Tension and Vapor Pressure

The surface of a liquid may be convex (drops of liquid), concave (liquid inside a capillary tube), or plane. These situations are diagrammed in Fig. 2.

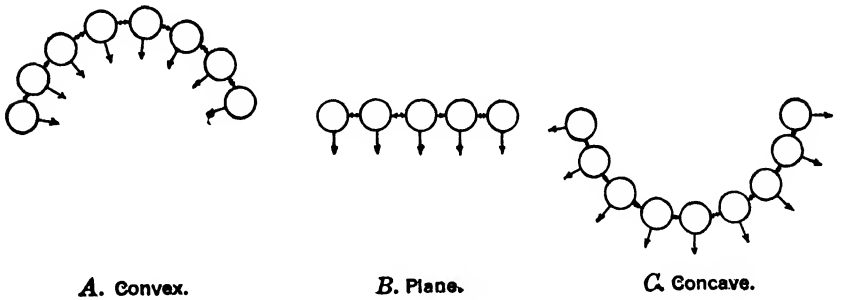


FIG. 2. Convex, concave, and plane liquid surfaces.

In *A* the molecules of the surface are attracted by other liquid molecules to a less extent than they are in a plane surface; in *C* the surface molecules are attracted to a greater extent by other liquid molecules than they are in a plane surface. We would expect, therefore, that the vapor pressure of the liquid in a convex surface would be greater than the normal vapor pressure of the liquid (plane surface). On the other hand, the vapor pressure of a liquid with a concave surface should be less than the normal vapor pressure. These expectations are fulfilled,

and the relation is quantitatively expressed by Thomson's [21] equation

$$P = P_0 + \frac{2\sigma P_0 M}{RT\rho r} \qquad 4$$

where P_0 is the normal vapor pressure of the liquid, σ is the surface tension, M is the molecular weight of the vapor, R is the gas constant which must be expressed in ergs per degree per mole (8.31×10^7), T is the absolute temperature, ρ is the density of the liquid, and r is the radius of the drop of liquid or the radius of the capillary. For a convex surface, r is positive; for a concave surface, it is negative. With water at 25° C. equation 4 reduces to

$$P = 23.756 + \frac{25 \times 10^{-7}}{r} \qquad 5$$

For a capillary with a radius of 10^{-6} cm. the vapor pressure of water would be reduced about 10 per cent.

Energy of a Surface

The total energy of a surface is made up of two terms, the free surface energy and the heat energy. The free surface energy is evidently equal to the surface tension in ergs per square centimeter multiplied by the total area of the surface. The heat term is defined as the heat which must be supplied to an expanding surface to maintain a constant temperature in the surface. The entire surface energy per unit area is the sum of these terms, or

$$u = \sigma - T\frac{d\sigma}{dT} \qquad 6$$

where $d\sigma/dT$ is the rate of change of the surface tension with a change of temperature. This differential ($d\sigma/dT$) is always negative; that is to say, surface tension always decreases with increasing temperature. This negative sign cancels the negative sign in equation 6 to give a sum.

According to Eötvös, and to Ramsay and Shields, the surface tension is the following function of temperature

$$\sigma(MV)^{2/3} = K(T_c - T - 6) \qquad 7$$

where $(MV)^{2/3}$ is proportional to the molecular surface, M is the molecular weight, and V is the specific volume of the liquid. T_c is the critical temperature of the liquid, and T is the temperature at which the surface tension is measured. These temperatures are expressed in degrees centigrade. K is a constant having the value 2.12 for normal unassociated

[21] W. Thomson, *Phil. Mag.*, **42**, 141, 448 (1871).

liquids. If the liquid is associated, the constant is less than 2.12; i.e., the molecular weight is larger than that used in the calculation based on the assumption of no association. Table 3 shows some Eötvös constants.

TABLE 3

EÖTVÖS CONSTANTS FOR SOME COMMON LIQUIDS

Substance	Eötvös constant
Acetone	1.9
Benzene	2.22
Carbon tetrachloride	2.21
n-Octane	2.30
Phenol	1.85
Water	1.04
Butyric acid	1.65
Caprylic acid	2.12
Lauric acid	2.56
Palmitic acid	2.92
Stearic acid	3.04

Hunten and Maass [22] have suggested that, if a constant below 2.12 indicates association of the liquid, a constant above 2.12 indicates that the molecule occupies less area at the surface than a normal molecule. This can be true only if the molecule is elongated and oriented at the surface. It seems probable that the suggestion of Hunten and Maass is at least qualitatively correct.

Addition of a Solute

A substance dissolved in a liquid will do one of two things to the surface tension of that liquid: it will either raise or lower it. If the addition of the solute raises the surface tension, the solute is said to be capillary inactive; if it lowers the surface tension, the solute is capillary active. Most inorganic electrolytes in appreciable concentrations are capillary inactive. Soaps, proteins, and a great many other organic compounds are capillary active.

Szyszkowski [23] proposed the following empirical equation to describe the surface-tension lowering produced by the addition of capillary-active materials

$$\frac{\sigma_0 - \sigma}{\sigma} = B \log \left(\frac{C}{A} + 1 \right) \qquad 8$$

[22] K. W. Hunten and O. Maass, *J. Am. Chem. Soc.*, **51**, 153 (1929).
[23] B. von Szyszkowski, *Z. physik. Chem.*, **64**, 385 (1908).

where σ_0 is the surface tension of the pure liquid and σ is the surface tension of the solution, C is the concentration of the solute in moles per liter of solution, and A and B are constants. Szyszkowski's equation serves as a basis for comparison of the capillary activity of solutes. B is nearly the same for all members of an homologous series of organic

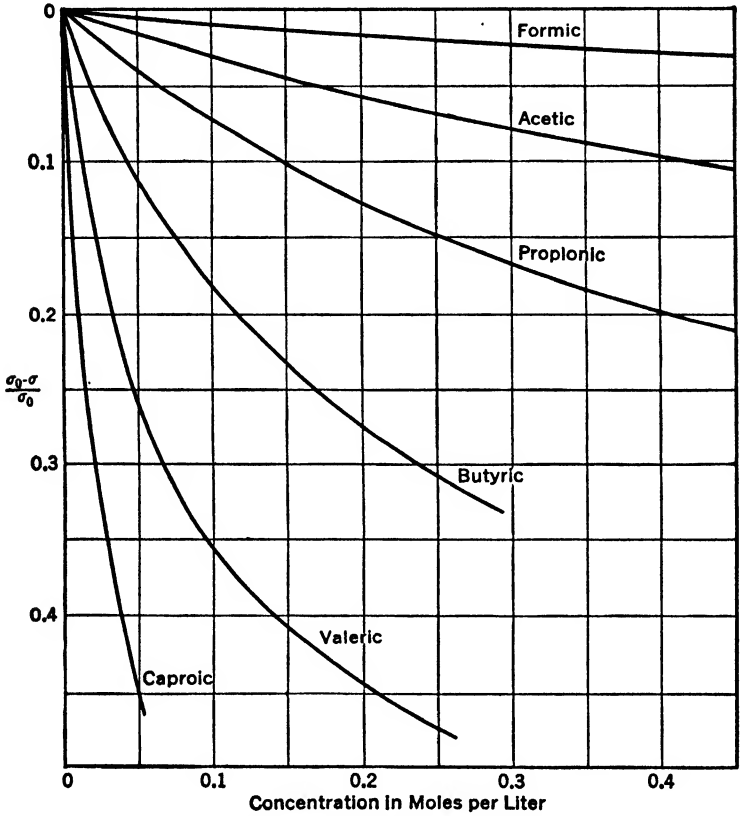


FIG. 3. Surface-tension lowering of aqueous solutions of fatty acids (15° C.).

compounds, and we can choose C equal to A for each member of the homologous series. Under this condition the surface-tension lowering $\left(\frac{\sigma_0 - \sigma}{\sigma_0}\right)$ will be the same for each member of the series. The values of C required to produce a given lowering of the surface tension become, therefore, an inverse measure of the capillary activity of the solute. By the same token the reciprocal of A is a direct measure of the capillary activity. In Fig. 3 the surface tension of aqueous solutions of the homologous series of straight-chain fatty acids is plotted against concentration.

Table 4 shows some numerical values for the Szyszkowski constants A and B obtained from the data plotted in Fig. 3.

TABLE 4

THE SZYSZKOWSKI CONSTANTS FOR SOME FATTY ACIDS (15° C.)

	Acid	B	A	$1/A$	A/A_1
Formic	HCOOH	0.29	1.38	0.72	....
Acetic	CH_3COOH	0.29	0.35	2.86	3.97
Propionic	CH_3CH_2COOH	0.37	0.165	6.07	2.12
Butyric	$CH_3CH_2CH_2COOH$	0.40	0.051	19.6	3.25
Valeric	$CH_3CH_2CH_2CH_2COOH$	0.41	0 0150	66.7	3.40
Caproic	$CH_3CH_2CH_2CH_2CH_2COOH$	0.41	0.0043	232.5	3.48

In the last column of Table 4 the ratio A/A_1 gives the capillary activity of a fatty acid divided by the capillary activity of the next lower member of the series. Note that the capillary activity of this homologous series of fatty acids increases geometrically with the addition of CH_2 groups in the carbon chain. This is, in effect, a quantitative statement of the so-called Traube's rule.

Frequently, the biological effects of an homologous series increase in geometrical proportions with increasing length of the carbon chain. For example, the toxicity of the straight-chain aliphatic alcohols forms such a geometrical series. (Methanol is an exception to this rule; properly speaking, it does not belong to this homologous series.) The finding of such a series is taken as presumptive evidence that the action of alcohols is a surface one. It is not conclusive, however, since the distribution coefficient of an homologous series between oil and water also follows a geometrical series; the alcohols may be acting on fatty tissue.

Belton and Twidle [24] have determined the surface tension of the amino acids glycine, alanine, and valine as a function of pH. Glycine and alanine increased the surface tension of water slightly; valine left it practically unchanged. The variation of the surface tension of aqueous solutions of these amino acids varied somewhat with pH, but the variation was not pronounced. Judging by the action of these amino acids, we may conclude that amino acids are, in general, capillary inactive.

Adsorption at a Liquid Surface

If the concentration of a dissolved substance is higher at a surface than in the bulk of the solution, positive adsorption is said to have

[24] J. W. Belton and A. H. H. Twidle, *Trans. Faraday Soc.*, **36**, 1198 (1940).

occurred. If, on the other hand, the concentration of the solute is less at the surface than in the bulk of the solution, negative adsorption has taken place. A little reflection will reveal that there must be a relation between the surface-tension raising or lowering and the amount of solute adsorbed. We have seen that the surface tension is a measure of the work required to bring molecules from the interior of the liquid to the surface. If positive adsorption has taken place, it means that less work is required to take a solute molecule to the surface than a solvent molecule and, accordingly, the surface tension of the solution must be less than that of the pure solvent. This reasoning is exactly expressed by Gibbs' adsorption equation,

$$\frac{d\sigma}{dC} = -\frac{a}{C}\frac{dP}{dC} \qquad 9$$

where a is the concentration of the solute at the surface in excess of that in solution, C is the bulk concentration of the solute, $d\sigma/dC$ is the rate of change of the surface tension with increasing concentration, and dP/dC is the rate of change of the osmotic pressure of the solute with increasing concentration. If the ideal gas laws are assumed to apply to the osmotic pressure of the solute, that is, if

$$P = CRT \qquad 10$$

we obtain the differential

$$\frac{dP}{dC} = RT \qquad 11$$

Substituting equation 11 in Gibbs' adsorption equation, we have

$$a = -\frac{C}{RT}\frac{d\sigma}{dC} \qquad 12$$

If the surface tension decreases with increasing concentration, the differential $d\sigma/dC$ is negative and, accordingly, a is positive. On the other hand, if σ increases with increasing concentration, a is negative. As we have noted, inorganic electrolytes, in general, increase the surface tension of water. They are, therefore, negatively adsorbed. It has been said that if the Ancient Mariner had known his physical chemistry he need not have suffered from thirst; he could have skimmed the surface of the ocean and drunk it!

Gibbs' adsorption equation has been tested in an ingenious manner by McBain and Humphrey.[25] The apparatus employed consisted of a microtome blade attached to the front of a small carriage. The carriage

[25] J. W. McBain and C. W. Humphrey, *J. Phys. Chem.*, **36**, 301 (1932).

was provided with wheels and traveled at 35 feet per second over the surface of the solution, which was contained in a long trough. The microtome blade sliced off the top layer of the surface to a depth of 0.05 to 0.1 mm. The top layer of the solution was thus collected and analyzed, and its concentration was compared with that in the bulk of the solution. Their results showed satisfactory agreement with Gibbs' adsorption equation.

Langmuir [26] combined the differential form of Szyszkowski's equation with that of Gibbs. This enabled him to calculate the amount of material adsorbed from the constants of Szyszkowski's equation. The differential form of Szyszkowski's equation, after conversion to natural logarithms, is

$$d\sigma = -0.434B\sigma_0 \frac{dC}{C + A} \qquad 13$$

Substitution of this expression in Gibbs' adsorption equation results in

$$a = \frac{0.434B\sigma_0}{RT} \frac{C}{C + A} \qquad 14$$

For dilute solutions equation 14 reduces to

$$a = \frac{0.434B\sigma_0}{RT} \frac{C}{A} \qquad 15$$

and for concentrated solutions to

$$a = \frac{0.434B\sigma_0}{RT} \qquad 16$$

Langmuir applied these equations to aqueous solutions of fatty acids. For fatty acids higher than propionic acid B equals 0.41. Accordingly, for concentrated solutions of fatty acids

$$a = \frac{0.434 \times 0.41 \times 72}{8.32 \times 10^7 \times 293}$$

$$= 5.25 \times 10^{-10} \text{ mole per square centimeter}$$

or

$$a = 5.25 \times 10^{-10} \times 6.027 \times 10^{23}$$

$$= 31.6 \times 10^{13} \text{ molecules per square centimeter}$$

The area occupied by one fatty acid molecule is, therefore, 31.7×10^{-16} square centimeter. Since all the fatty acids above propionic acid give

[26] I. Langmuir, *J. Am. Chem. Soc.*, **39**, 1848 (1917).

the same value for the constant B, at saturation of the surface the acids must be oriented vertically in the surface. In this position all fatty acids would be expected to occupy the same area since their cross sections are the same.

These calculations along with his studies on spread monomolecular films of longer-chain acids led Langmuir to his picture of the orientation of polar molecules at a surface. In Chapter IV we discussed polar molecules from the standpoint of their electrical asymmetry, the dipole moment being a measure of this asymmetry. Here we wish to introduce another concept of polar molecules. It is essentially a concept of solubility. Hydrocarbons are characteristically insoluble in water. If groups such as —COOH, —OH, —NH_2, and —SO_3H are substituted into a hydrocarbon, they tend to make the hydrocarbon water-soluble. Such groups are called hydrophilic groups (water-loving). The hydrocarbon portion is said to be hydrophobic (water-hating). If the hydrocarbon chain is small and the attraction of the hydrophilic group of water strong enough, the resulting molecule will be completely water-soluble as, for example, acetic and propionic acids. At a water-air surface such polar molecules are situated with the polar group (hydrophilic group) sticking into the water and with the hydrocarbon chain lying flat on the surface. At higher concentration of adsorbed fatty acids the hydrocarbon chain is oriented vertically to the surface. We will discuss this subject in more detail when we describe spread monomolecular films.

Surface Tension of High-Molecular-Weight Compounds

The surface tension of solutions of colloidal surface-active substances present several complicated features. For one thing, the time required for attainment of equilibrium in many substances is very long, and it is really doubtful that in some a true equilibrium tension is ever observed. Hauser and Swearingen [27] investigated the aging of the surfaces of egg albumin; their results show that the final surface-tension values are approached very slowly. Figure 4 shows the change of the surface tension of solutions of egg albumin in $N/150$ sodium acetate buffer at pH 4.9.[28]

The slow attainment of the final value for the surface tension of such solutions is definitely not due to the slow rate of diffusion of the protein molecules to the surface. It has been calculated [29] that, if every egg albumin molecule which diffused to the surface remained in the surface,

[27] E. A. Hauser and L. E. Swearingen, *J. Phys. Chem.*, **45**, 644 (1941).
[28] H. B. Bull, unpublished data.
[29] H. Neurath and H. B. Bull, *Chem. Revs.*, **23**, 391 (1938).

the surface of a 0.05 per cent egg albumin solution would be completely saturated in 0.58 second. There are several possible reasons for the slow approach to equilibrium. Possibly one of the most important is that the native egg albumin is really capillary inactive, and it is only after the native molecule has been surface-denatured that it becomes capillary active. The transfer of the native molecule into the surface where it it can surface-denature encounters an energy barrier which must be overcome before the denatured, capillary-active egg albumin can be formed. This requires time.

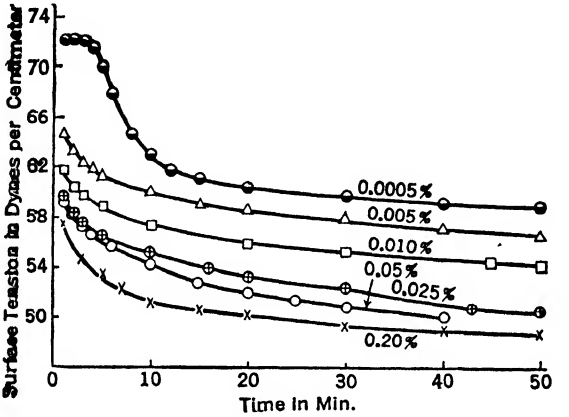


FIG. 4. Surface tension of egg albumin solutions at *p*H 4.9 in *N*/150 sodium acetate buffer as a function of time. (Wilhelmy balance method.)

The situation at the surface of a quiescent protein solution is probably highly complicated. The surface of egg albumin solutions was studied [30] with simple apparatus which consisted of a rotating porcelain drum dipping into an egg albumin solution. The amount of insoluble, surface-denatured protein was determined as a function of the speed of rotation of the drum. The results indicated a layer of flat, denatured molecules on the surface which was about 10 Å thick. Under this layer was adsorbed a layer of native molecules about 45 Å thick.

With proteins like egg albumin, it is doubtful that an equilibrium exists between the denatured molecules in the film and the protein solution; the denatured molecules are so insoluble that none of them can return to the liquid. Therefore it is meaningless to attempt to apply Gibbs' adsorption equation to such a system.

Hydrogen ions, in general, have a pronounced effect upon the surface tension of proteins, a minimum in the surface tension being observed in the isoelectric zone.

[30] H. B. Bull, *J. Biol. Chem.*, **123**, 17 (1938).

Some colloidal substances such as soaps and the long-chain sulfated alcohols exhibit a distinct minimum in the surface tension-concentration curve.[31] Figure 5 shows the relative surface tension of aqueous solutions of laurylsulfonic acid as a function of concentration.[32]

In terms of the Gibbs equation, the portion of the curve to the right of 1.5 grams per liter (Fig. 5) would indicate a decrease in adsorption with increasing concentration. Such a decrease has not been found

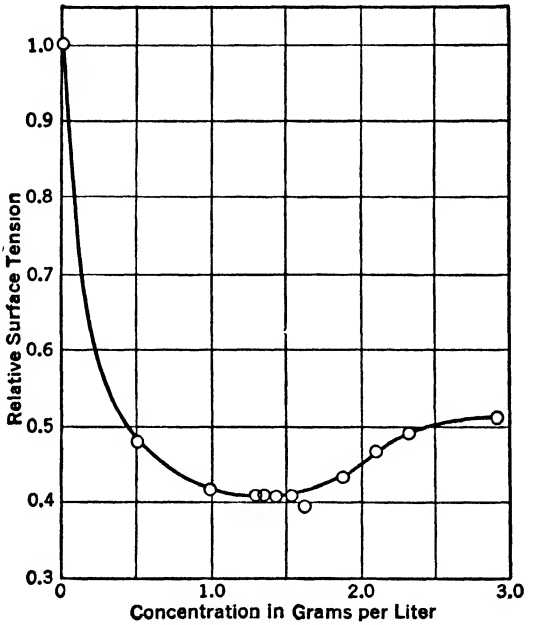


FIG. 5. Surface tension of solutions of laurylsulfonic acid as a function of the concentration at 25° C. (Wilhelmy balance method.)

experimentally. The differential dP/dC must have become negative at higher concentrations (see equation 9). This would mean, in terms of a physical picture, that the laurylsulfonic acid particles are undergoing association with micelle formation and that the rate of association increases with concentration (above 1.5 grams per liter).

Du Noüy [33] and other workers have reported rather sharp maxima and minima of the surface tension of protein solutions as the protein

[31] J. W. McBain, J. R. Vinograd, and D. A. Wilson, *J. Am. Chem. Soc.*, **62**, 244 (1940).

[32] H. B. Bull, unpublished data.

[33] P. L. du Noüy, *Surface Equilibria of Biological and Organic Colloids*, The Chemical Catalog Co., Inc., New York, 1926.

concentration is increased. They have attempted to calculate molecular dimensions from these maxima and minima on the assumption that the irregularities in the surface tension are due to different molecular orientations. The author is of the opinion that these maxima and minima are artifacts and that this attempt to evaluate molecular dimensions is completely in error.

Interfacial Tension

Up to this point we have confined ourselves to consideration of the solution-air surface. In general, the same treatment applies to a liquid-liquid interface. The interfacial tension is a measure of the energy required to bring molecules from within one liquid to the interface plus the energy needed to take molecules from the other liquid and put them in the interface. The interfacial tension is, in general, much lower than the sum of the surface tensions of the two immiscible liquids. The reason is that the potential energy of a molecule at the interface is lower than at the air-liquid surface. This, in turn, is due to the fact that more of the attractive forces of the molecules have been satisfied at an interface than at a liquid-air surface. The relation between the interfacial tension and the surface tensions of the two liquids is given by the approximate Antonoff's rule,

$$\sigma_{AB} = \sigma_A' - \sigma_B' \qquad 17$$

where σ_A' is the surface tension of liquid A saturated with liquid B, and σ_B' is the surface tension of liquid B saturated with A. For a critical discussion of Antonoff's rule, see Harkins.[34]

The same techniques can be used for the measurement of interfacial tensions as have been described for measurement of surface tensions. Appropriate modifications have to be made, and the determinations are usually somewhat more troublesome.

Surfaces of Living Cells

The interfacial tension of some living cells has been determined, and Harvey [35] states two general conclusions that can be drawn from such measurements as have been made: (1) the tension is very low, less than 1 dyne per centimeter; and (2) the cell surface has elastic properties. The centrifuge method can be used for egg cells which contain oil globules (oil globules are less dense than the cell fluid) and yolk granules (yolk granules are more dense than the cell fluid). In this method the

[34] W. D. Harkins, *Proc. Natl. Acad. Sci. U. S.*, **5**, 569 (1919).
[35] E. N. Harvey, *Trans. Faraday Soc.*, **33**, 943 (1937).

cells are placed in a centrifuge microscope, and the centrifugal force just sufficient to separate the cell into two halves is determined. This force is equated to the interfacial tension forces around the circumference of the cell, and the interfacial tension is calculated.

Another method used by Cole [36] is to measure the flattening of the cell produced by a microbeam of gold 6 μ thick and 180 μ wide. The cell was photographed with and without the microbeam, and the corresponding cell radii were measured. From the flattening of the cell by the microbeam, the interfacial tension of the cell was calculated. It was found that the cell surface of the unfertilized *Arbacia* egg was elastic with a very low tension. When the tensions were extrapolated to the uncompressed state, a value of 0.08 dyne per centimeter was found. The fertilized *Arbacia* egg with its fertilization membrane was, however, quite rigid, and no measurement of the interfacial tension was possible.

Ascherson,[37] as long ago as 1840, suggested that cell membranes were analogous to the "skin" which forms around oil droplets suspended in a solution of "albumin" (protein). It is entirely possible that the structural element of animal cell membranes is indeed fiber protein which gives the cell surface its elastic properties. Parpart and Dziemian,[38] for example, have analytical evidence that the mammalian red blood cell membrane contains a collagenlike protein.

Solid-Liquid Interface

If a drop of liquid is placed on a solid surface, the liquid may spread over the solid and completely wet it, or it may remain as a drop. The situation is diagrammed in Fig. 6.

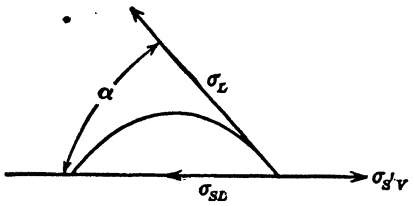


FIG. 6. The resolution of forces acting on a drop of liquid at a solid surface.

The angle α is called the contact angle. The magnitude of this angle depends upon the resultant of the forces acting along the surface of the solid. That is,

$$\sigma_{S'V} = \sigma_{SL} + \sigma_L \cos \alpha \qquad 18$$

where $\sigma_{S'V}$ is the surface tension of the solid saturated with the vapor of the liquid, σ_L is the surface tension of the liquid, and σ_{SL} is the interfacial tension of the solid-liquid interface.

There is considerable confusion in the literature in regard to the term

[36] K. S. Cole, *J. Cellular Comp. Physiol.*, **1**, 1 (1932).

[37] F. M. Ascherson, *Arch. Anat. Physiol.*, **1840**, 44.

[38] A. K. Parpart and A. J. Dziemian, *Cold Spring Harbor Symposia Quant. Biol.*, **8**, 17 (1940).

adhesion tension. Harkins and Livingston [39] have considered the whole problem of wetting of solids by liquids and have pointed out in a precise manner the source of our confusion. We shall follow their treatment. The work required to separate a unit area of a liquid from a unit area of a solid is clearly

$$\text{Work of adhesion} = \sigma_{SA} + \sigma_L - \sigma_{SL} \qquad 19$$

where σ_{SA} is the surface tension of the solid in a vacuum. The source of confusion arose in attempting to identify σ_{SA} and $\sigma_{S'V}$. It is quite obvious that these two terms are very far from being identical. In general, $\sigma_{S'V}$ is very much less than σ_{SA}. Equating these two terms, as was done in the literature, would allow us to substitute equation 18 into equation 19. We would then have per unit area

$$\text{Adhesion tension} = \sigma_L + \sigma_L \cos \alpha \qquad 20$$

This expression, however, is true only if we consider the solid surface saturated with vapor. The expression has been so misused that Harkins and Livingston propose that the term adhesion tension be abandoned.

The free energy required to transfer molecules from the vapor state to the solid surface is a quantitative measure of the affinity of the solid for the vapor. The free energy involved in this process can be measured by a study of the amount of the vapor adsorbed as a function of the vapor pressure of the gas and is given by the equation

$$\Delta F = \frac{RT}{M} \int_0^x \frac{a}{x}\, dx \qquad 21$$

where a is the weight of the vapor adsorbed per given weight of dry solid, x is the relative vapor pressure of the vapor, and M is its molecular weight. The vapor has been assumed to obey the ideal gas laws.

In order to integrate equation 21, a/x is plotted against x and the area under the curve determined. Unfortunately, it is difficult to measure the area under the curve because for small values of x the curve swings upward too steeply and appears to approach infinity when x is equal to zero. To circumvent this difficulty a modification of the method used by Boyd and Livingston [40] is suggested.

The region under the curve which results from a plot of a/x against x is divided into segments by drawing vertical lines at regular and convenient intervals to the x-axis. Successive areas under the curve to the right and to the left of $x = 0.5$ are measured with a planimeter. The

[39] W. D. Harkins and H. K. Livingston, *J. Chem. Phys.*, **10**, 342 (1942).

[40] G. E. Boyd and H. K. Livingston, *J. Am. Chem. Soc.*, **64**, 2383 (1942); see especially H. K. Livingston, Ph.D. thesis, University of Chicago, 1941.

area under the curve from $x = 0$ to $x = 0.5$ cannot, of course, be measured with a planimeter, and indeed it is this area which we are seeking and have resorted to this complicated procedure to obtain. The areas to the left of the mid-vertical line are given negative signs and those to the right are assigned positive signs. These areas are then plotted against x, and the resulting curve is extrapolated to $x = 0$ and to $x = 1$. The free energy of adsorption to any given relative vapor pressure can then be found by adding the ordinate corresponding to x to the ordinate found by extrapolation to $x = 0$ and multiplying this sum by RT/M. The values of a, the amount of vapor adsorbed, corresponding to a series of x values are then obtained from the adsorption isotherm where the amount adsorbed is plotted against the relative vapor pressure. This allows us to plot ΔF against amount of vapor adsorbed. If ΔF is calculated for the solid in the presence of its saturated vapor, we have a direct measure of the difference $(\sigma_{SA} - \sigma_{S'V})$, and, provided that we know the surface area of the solid, we can express ΔF in terms of this surface-tension difference. Practically, in order to study the amount of vapor adsorbed by a solid, a large surface is necessary; otherwise the weight of the vapor adsorbed will be so small that accuracy will be impossible. Finely powdered or porous solids are employed. If the adsorption has been studied at two different temperatures, the heat of the adsorption reaction can be calculated by the equation

$$\Delta H = \frac{T_2\Delta F_1 - T_1\Delta F_2}{T_2 - T_1} \qquad 22$$

where the temperatures are absolute. It is assumed that ΔH is constant in the temperature interval employed. Boyd and Harkins [41] have discussed the heat of wetting of crystalline powders by a variety of liquids. They found that the heat of wetting depended almost entirely on the nature of the liquid and not on that of the powder, provided that only polar liquids are compared. They were able to calculate the total energy of wetting.

A large number of techniques have been devised for the measurement of contact angles. The fact that so many techniques have been employed indicates the difficulty of such measurements. For one thing, the receding and advancing angles are usually not equal to each other. Surface contaminations influence the magnitude very profoundly. The solid may also undergo slow changes due to the presence of the liquid. For example, a fresh surface of solid paraffin wax is quite hydrophobic with a large contact angle against water. If, however, the paraffin wax is

[41] G. E. Boyd and W. D. Harkins, *J. Am. Chem. Soc.*, **64**, 1190, 1195 (1942).

immersed in water for several days, the surface becomes hydrophilic and gives a contact angle of zero or almost zero. Harkins [42] recommends the use of a tilting plate. Bartell and Bristol [43] have devised a suitable method.

Mudd Interfacial Technique

Mudd and co-workers [44] have developed a technique for the study of the wetting characteristics of cell surfaces. The aqueous suspension of bacteria is placed on a microscope slide adjacent to a drop of oil. When a cover slip is lowered over this preparation, an oil-water boundary is formed which advances across the microscope field. When the boundary reaches a bacterium, one of three things may happen: (1) if the surface of the bacterium is hydrophilic, it remains in the water phase; (2) if the surface is hydrophobic (oilophilic), it passes easily into the oil phase; and (3) if it is neither pronouncedly hydrophilic or hydrophobic, it tends to remain at the oil-water interface.

Adsorption at Solid Surfaces

A great deal of experimental and theoretical work has been done on adsorption, both at solid-gas and at solid-liquid surfaces. Adsorption is of great technical importance, and adsorbents of various kinds are extensively used in chemical and biochemical laboratories.

Adsorption can be quite simply demonstrated.[45] To a series of dilutions of propionic acid is added 1-gram samples of activated charcoal (Norite), the volumes of the propionic acid being 75 cc. The flask containing the charcoal and acid is shaken for 30 minutes at 25° C. The solutions are filtered, and the concentrations of the propionic acid in the filtrates are determined with standard base. The amount of acid adsorbed is then calculated by subtracting the concentration of the acid in the filtrate from that in the original solution. The results of such an experiment are shown in Fig. 7.

Freundlich [46] proposed an empirical adsorption isotherm to describe the relation between the amount of solute adsorbed and its concentrations. His equation is

$$\frac{a}{m} = \alpha C^{1/n} \qquad 23$$

[42] W. D. Harkins and F. M. Fowkes, *J. Am. Chem. Soc.*, **60**, 1511 (1938); **62**, 3377 (1940).

[43] F. E. Bartell and K. E. Bristol, *J. Phys. Chem.*, **44**, 86 (1940).

[44] S. Mudd and E. B. H. Mudd, *J. Expll. Med.*, **40**, 633 (1924).

[45] E. R. Linner and R. A. Gortner, *J. Phys. Chem.*, **39**, 35 (1935).

[46] H. Freundlich, *Kapillarchemi*, Academische Verlagsgesellschaft, Leipsig, 1930.

where a is the amount of solute adsorbed, m is the weight of the adsorbent in grams, C is the equilibrium concentration of the solute, and α and $1/n$ are constants. Equation 23 may be converted into logarithmic form

$$\log \frac{a}{m} = \log \alpha + \frac{1}{n} \log C \qquad 24$$

If Freundlich's equation is obeyed the plot of log a/m against log C should give a straight line. The slope of this line is equal to $1/n$, and the intercept on the log a/m axis is equal to log α.

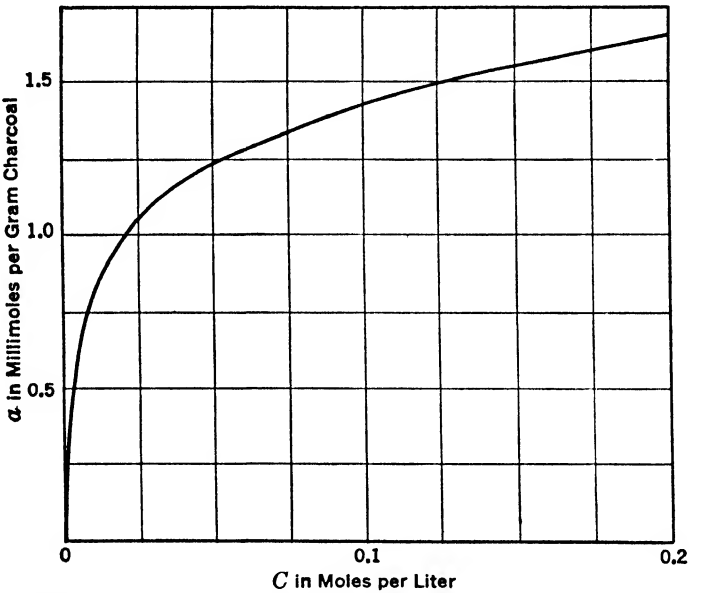


FIG. 7. Adsorption of propionic acid by charcoal at 25° C. (Linner and Gortner.)

Langmuir [47] was able to derive an adsorption equation from theoretical consideration by equating the rate of evaporation of a gas from a solid surface to the rate of condensation.

The rate of condensation on a solid surface is proportional to μ, the number of molecules striking each square centimeter of surface per second, and to $(1 - \theta)$, the fraction of the surface not covered by the adsorbed molecules, and to ϕ, the fraction of the molecular collisions which are inelastic. The rate of evaporation is equal to $\nu\theta$, where ν is the rate of evaporation when all the surface is occupied and θ is the fraction of the surface already occupied. When equilibrium is estab-

[47] I. Langmuir, *J. Am. Chem. Soc.*, **40**, 1361 (1918).

lished the two rates are equal; that is

$$\mu\phi(1 - \theta) = \nu\theta \tag{25}$$

and

$$\theta = \frac{\phi\mu}{\nu + \phi\mu} \tag{26}$$

If X is the number of molecules per square centimeter at a given gas pressure and X_0 is the number of molecules per square centimeter at saturation,

$$\theta = \frac{X}{X_0} \tag{27}$$

Substituting the value of θ in equation 26 and rearranging, we have

$$X = \frac{X_0(\phi\mu/\nu)}{1 + (\phi\mu/\nu)} \tag{28}$$

Since ϕ/ν is constant and μ is proportional to the gas pressure P, and X is proportional to the number of gram molecules adsorbed per gram of adsorbent, we have

$$a = \frac{\alpha_1\beta_1 P}{1 + \alpha_1 P} \tag{29}$$

where a is the amount of solute adsorbed and α_1 and β_1 are the proportionality constants. We can also apply the Langmuir equation to the adsorption at a liquid-solid interface. In this case we replace pressure by concentration to obtain

$$a = \frac{\alpha\beta C}{1 + \alpha C} \tag{30}$$

Equation 30 can be rearranged to give

$$\frac{C}{a} = \frac{1}{\alpha\beta} + \frac{C}{\beta} \tag{31}$$

If C/a is plotted against C, a straight line should be obtained, the slope of which is $1/\beta$ and the intercept is $1/\alpha\beta$. At concentrations sufficient to saturate the adsorbing surface, we know from equation 30 that

$$a = \frac{\alpha\beta C}{\alpha C} = \beta \tag{32}$$

from which we conclude that β has the same dimensions as a, which, as we realize, has the dimensions of concentration per unit area or L^{-2}.

Substituting this dimension for β in equation 30 we find α to be dimensionless; α is related to the energy of adsorption.

Linner and Gortner [45] made a comprehensive study of the adsorption of thirty-one organic acids on Norite (a decolorizing charcoal). The constant β of Langmuir's equation furnished an excellent means by which to compare the adsorbability of the various acids. It is evident from the above discussion of the meaning of β that $1/\beta$ is proportional to the area occupied by the acids on the surface of the charcoal. They found $1/\beta$ to vary in a steplike fashion between even- and odd-numbered carbon atoms in an homologous series. It is to be recalled that the variation between the physical properties of odd- and even-numbered carbon atoms is due to their crystal structure; [48] steplike variations are not found when the acids are in the liquid state. The observed variation of $1/\beta$ was, accordingly, accepted by Linner and Gortner as evidence that the acids exist at the surface of Norite as surface crystals; i.e., the acids were oriented at the charcoal surface in an orderly arrangement.

The adsorption of water vapor on a porous solid such as silica gel or on solid proteins usually follows a typical course. The plot of the amount of water adsorbed against the aqueous vapor pressure gives an S-shaped curve. The adsorption of water vapor on purified collagen at 25° C. is shown in Fig. 8.[49]

The first steep portion of the adsorption curve (Fig. 8) is an adsorption in the Langmuir sense and represents the filling-up of a monomolecular layer of water molecules on the collagen surface. The second rise is a type of capillary condensation. As we have already pointed out, the vapor pressure of a liquid is decreased on a concave surface (see equation 4). The theory of capillary condensation has never been satisfactorily and quantitatively tested in this respect. Brunauer, Emmett, and Teller [50] have treated this problem in some detail and have proposed an interesting and stimulating theory of such S-shaped adsorption curves. Essentially, they generalized Langmuir's adsorption theory to include multilayer adsorption. They considered the rates of evaporation and of condensation, not only of a unimolecular adsorbed layer but also of n such layers. They attribute the heat of adsorption of all but the first adsorbed layer to the heat of liquefaction of the vapor. The author is not prepared to admit this simple interpretation of the heat of adsorption of the layers beyond the first layer; the situation is more complex than that. This interpretation, however, is not an essential part of their theory. All that has to be agreed upon is that a heat term is in-

[48] W. E. Garner and E. A. Ryder, *J. Chem. Soc.*, **127**, 720 (1925).

[49] H. B. Bull, unpublished data.

[50] S. Brunauer, P. H. Emmett, and E. Teller, *J. Am. Chem. Soc.*, **60**, 309 (1938).

volved in the multilayer adsorption. From these considerations, they were able to derive the following equation to describe the adsorption of vapors on free surfaces.

$$V = \frac{V_m C P}{(P_0 - P)[1 + (C + 1)P/P_0]} \qquad 33$$

where V is the volume of adsorbed gas at the gas pressure P, V_m is the volume of gas adsorbed when the entire adsorbent surface is covered with

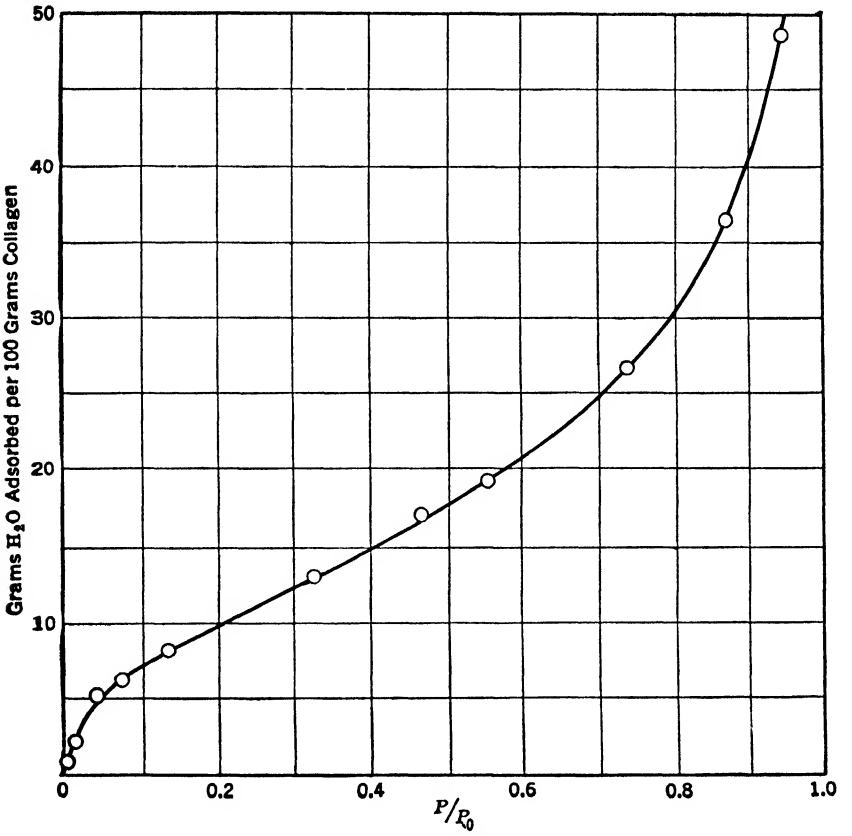


FIG. 8. Adsorption of water vapor on collagen at 25° C.

a complete unimolecular layer, P_0 is the vapor pressure at saturation pressure of the gas, C is a constant which is very nearly given by the relation

$$C = e^{(H_1 - H_2)/RT} \qquad 34$$

where H_1 is the heat of adsorption of the first unimolecular layer directly on the adsorbent and H_2 is the heat of adsorption of other layers.

Brunauer, Emmett, and Teller believed H_2 to be the heat of liquefaction of the gas. At low gas pressures equation 33 reduces to

$$V = \frac{V_m C P}{P_0[1 + (CP/P_0)]} \qquad 35$$

which, as will be recognized, is a form of the Langmuir equation. For the purpose of evaluating constants and testing equation 33, it is thrown into the form

$$\frac{P}{V(P_0 - P)} = \frac{1}{V_m C} + \frac{C - 1}{V_m C} \cdot \frac{P}{P_0} \qquad 36$$

A plot of $P/[V(P_0 - P)]$ against P/P_0 should give a straight line whose intercept is $1/V_m C$ and whose slope is $(C - 1)/V_m C$. The constants V_m and C can thus be evaluated. As we have noted, V_m is the volume of adsorbate required to form a complete unimolecular adsorbed layer. The constant V_m thus becomes a measure of the surface area of the solid. Emmett [51] discusses the use of V_m to evaluate the surface area of fine powders. In order to do this we must be able to assign the proper dimensions to the adsorbed molecule. There is some ambiguity in this assignment. Do the adsorbed molecules exist at the surface as gases, liquids, or solids? He proposes the equation

$$\text{Area per molecule} = 1.585 \times 10^{-6} \left(\frac{M}{\rho}\right)^{2/3} \qquad 37$$

where M is the molecular weight of the gas, and ρ is density of the solidified or liquefied gas. This equation is derived on the assumption that the molecules are hexagonally close-packed in the solidified and in the liquefied gas.

If the thickness of the adsorbed film cannot exceed some finite number, n, owing to the dimensions of the capillaries, then we have instead of equation 33 the following extended equation

$$V = \frac{V_m C\, P/P_0}{(1 - P/P_0)} \left[\frac{1 - (n + 1)(P/P_0)^n + n(P/P_0)^{n+1}}{1 + (C - 1)\, P/P_0 - C(P/P_0)^{n+1}}\right] \qquad 38$$

In order to evaluate the constants in equation 38, we plot as before $P/[V(P_0 - P)]$ against P/P_0. We use the linear portion of the curve to find the constants C and V_m. These two constants along with the experimental data are substituted in equation 38, and the best average value

[51] P. H. Emmett, *Advances in Colloid Science*, Interscience Publishers, Inc., New York, 1942.

for n is determined. Figure 9 shows the plot of the data on collagen given in Fig. 8 according to the method of Brunauer et al.

If we assign the value of 10.9 square Å for the area occupied by one water molecule on the surface, we can calculate the surface area of the collagen. One cubic centimeter of water covers an area of $\frac{10.9 \times 6.023 \times 10^{23}}{18 \times 10^{20}}$ or 0.36×10^4 square meters of surface. From Fig. 9 the value of V_m is 9.53 cc. per 100 grams of collagen. Accordingly, the surface area of 100 grams of collagen is $9.53 \times 0.364 \times 10^4$ or 3.48×10^4 square meters.

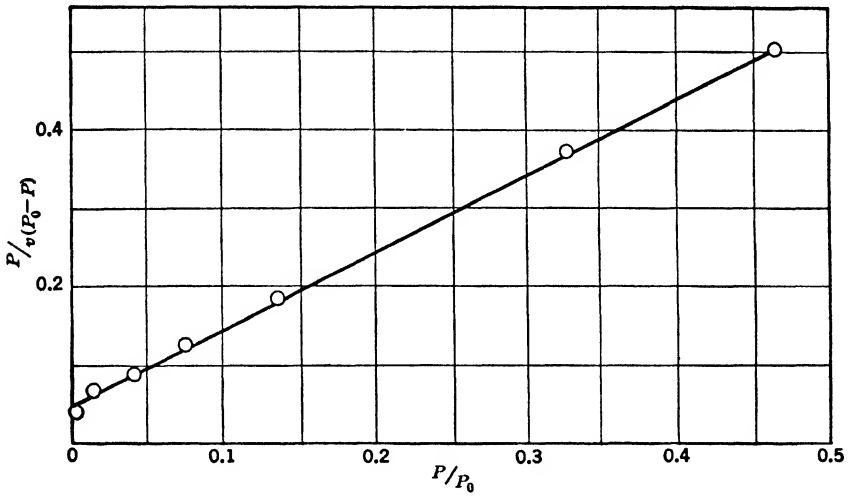


FIG. 9. The adsorption data of water on collagen at 25° C. plotted to obtain the constants V_m and C.

Brunauer, Deming, Deming, and Teller [52] have extended the above theory to more complicated types of gas adsorption.

There is no evidence that the adsorption of a solute from a solution on a solid surface ever exceeds one layer of molecules so that the Brunauer theory has no meaning for this situation.

Negative Adsorption from Solution

If the concentration of the adsorbate is increased sufficiently, a concentration will be reached at which it exceeds the concentration of the solvent. It is then a question which is actually the adsorbate. Bartell

[52] S. Brunauer, L. S. Deming, W. E. Deming, and E. Teller, *J Am. Chem. Soc.*, **60**, 1723 (1940).

and Scheffler [53] have studied adsorption from a binary mixture of methyl alcohol and benzene on carbon. Figure 10 shows some of their results.

At high concentrations of benzene the methyl alcohol is preferentially adsorbed; at high concentrations of methyl alcohol, the benzene is adsorbed. What this really means, of course, is that at high benzene concentrations the concentration of methyl alcohol at the carbon surface exceeds that in the liquid phase, whereas at high alcohol concentrations

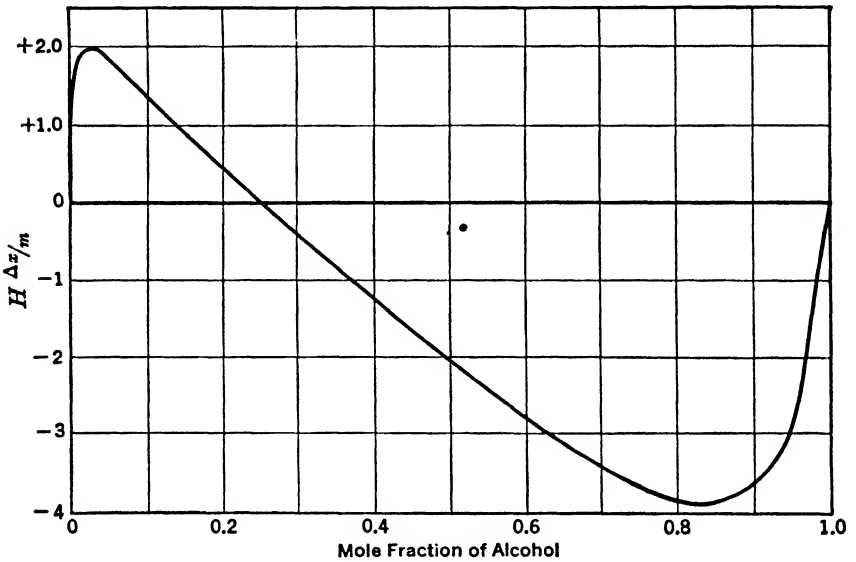


FIG. 10. Adsorption of a binary mixture of methyl alcohol and benzene on carbon. H is the total number of millimoles in solution, ΔX is the mole fraction change due to adsorption, and m is the weight of adsorbent.

the benzene concentration at the surface exceeds that in solution. At zero adsorption, their concentrations at the surface bear the same ratio to each other as their concentrations in solution. This conclusion leads us to the dictum that, whenever there is a positive adsorption of one substance from a binary mixture, there must likewise be a negative adsorption of the other component.

Chromatographic Analysis

Adsorption has numerous laboratory applications, one of the more important of which is chromatographic analysis.

A suitable adsorbent is packed in a vertical glass tube, and a solution of the substance which it is desired to separate into its various compo-

[53] F. E. Bartell and G. H. Scheffler, *J. Am. Chem. Soc.*, **53**, 2507 (1931).

nents is poured into the tube. The component with the greatest affinity for the adsorbent is adsorbed first; the component with the next greatest affinity is adsorbed next; and so on. There is at this stage considerable overlapping of the bands of the adsorbents in the column. The chromatogram is now developed by passing some of the pure solvent into the column. Each substance dissolves slightly as the solvent passes through the column but is readsorbed lower down. The net effect is to remove the less readily adsorbed material from the upper layers and carry it further down in the column. The column finally shows a separate band for each substance. This technique has been very helpful in separating carotene pigments from one another. A book [54] has been published on chromatographic analysis.

For colored materials, the simple type of chromatographic analysis described above suffices. If, however, the adsorbates are colorless, there is some difficulty in identifying the location of the bands. Tiselius [55] has developed a technique which not only serves to locate the bands but also provides an estimate of their relative concentrations. The developed chromatogram is transferred without mixing to a rectangular glass cell by slowly pouring solvent through the adsorbent column. The rectangular cell is connected to the chromatographic column, and the eluted adsorbates occupy the same relative positions in the rectangular glass cell as they did in the chromatographic column. The optical system of Philpot which employs a cylindrical lens is now used to visualize the boundaries of the adsorbates (the cylindrical lens system was described in the chapter on electrokinetics).

Adsorption of Proteins on Solid Surfaces

In the chapter on electrokinetics we pointed out that the microelectrophoretic technique for measuring the mobility of proteins involved the adsorption of proteins from solution on inert particles such as those of powdered quartz, the powdered quartz being of sufficient size to be visible under a microscope. Considering the potential interest and importance of the subject, there has been little quantitative work on the adsorption of proteins. Among papers which may be referred to are ones by Hitchcock,[56] by Palmer,[57] and by Dow,[58] who worked in Hitchcock's laboratory.

[54] H. H. Strain, *Chromatographic Adsorption Analysis*, Interscience Publishers, Inc., New York, 1942.

[55] A. Tiselius, *Advances in Colloid Science*, Interscience Publishers, Inc., New York, 1942.

[56] D. I. Hitchcock, *J. Gen. Physiol.*, **8**, 61 (1925); **10**, 179 (1926).

[57] A. H. Palmer, *J. Gen. Physiol.*, **15**, 551 (1932).

[58] P. Dow, *J. Gen. Physiol.*, **19**, 907 (1936).

Lindau and Rhodius [59] investigated the adsorption of gelatin and of egg albumin on powdered quartz. Although these workers found the adsorption of egg albumin to follow Langmuir's adsorption isotherm, they also report that the egg albumin was irreversibly adsorbed on the quartz. They discovered that at about 50 per cent coverage of the quartz surface by egg albumin, the surface suddenly became pronouncedly hydrophobic. It is the belief of the author that change of the hydrophilic properties of the surface reflects a surface coagulation of the egg albumin; i.e., the egg albumin molecules on the surface react with each other. It is at this point that the adsorption of egg albumin becomes irreversible.

Adsorption and Chemical Binding

For many years the cause of adsorption at solid-gas or at solid-liquid surfaces was clouded in mystery. People spoke of adsorption as though it were completely different from chemical binding. It is now believed that there is no fundamental difference between adsorption and chemical binding; identical types of forces are operating. The apparent lack of a stoichiometric relation in adsorption reactions is due to the varying affinities of the adsorbent molecules for the adsorbate molecules. The adsorbent molecules immediately on the surface are more or less completely exposed, as, for example, the molecules on the sharp edges of the molecular hills of the surface. Other molecules on the surface are less exposed. This results in a graded series of affinities which leads to the typical parabolic relation between concentration and the amount adsorbed. In general, it is believed that adsorptive forces extend only about one molecular distance away from the surface, so that it is to be anticipated that, unless capillary condensation takes place, the adsorbed film will be unimolecular at saturation of the surface.

Physical chemists find it convenient to distinguish two types of adsorption. The physical or van der Waals adsorption is characterized by a low heat of adsorption, whereas activated or chemisorption has a much larger heat of adsorption. The distinction is of value in discussing the adsorption of gases on solids.

Spread Monomolecular Layers

When a small amount of an insoluble substance is placed on a clean water surface, one of two things may happen. The substance may remain as a compact mass, leaving the rest of the surface clean, or it may spread over the surface to form a very thin film.

The first experiments on spreading were made many years ago. In

[59] G. Lindau and R. Rhodius, *Z. physik. Chem.*, **A172**, 321 (1935).

1890 Lord Rayleigh [60] spread olive oil on water and measured the thickness of the spread film. He found it to be about 16 Å thick.

Pockels [61] discovered that spread surface films could be pushed along with barriers and that such films could be compressed by moving the barriers together. She also measured the surface tension as a function of the area of the film and found that, at complete expansion, the surface tension was independent of the area of the film, but below a certain critical area the surface tension dropped rapidly as the area was decreased.

Langmuir [62] introduced new techniques and interpretations into this field of investigation. He used for the first time a floating mica strip to register the film "pressure." The fatty acids with which Langmuir worked were applied to the surface by first dissolving them in a volatile solvent such as benzene or petroleum ether and dropping the solution on the clean water surface. The petroleum ether evaporated and left the pure fatty acids spread on the surface. Langmuir measured the area occupied by a given weight of straight-chain fatty acids, and, knowing the number of fatty acid molecules present, he was thus able to calculate the area per fatty acid molecule.

The fact that the fatty acid molecules contained the hydrophilic —COOH group suggested that in compressed films the molecules were oriented with these groups sticking into the water and that the hydrocarbon chain pointed into the air. Langmuir further suggested that in the uncompressed film the molecules lie flat on the surface. The films of fatty acids containing less than 14 carbon atoms per molecule dissolved when compressed. Evidently the buoyancy of the hydrocarbon chain was not sufficient to overcome the attraction of the —COOH group for water.

Suggestions that molecules were oriented at air-water surfaces were made previous [63] to and simultaneous [64] with Langmuir's publication, but Langmuir's demonstration of molecular orientation at a surface was so clear cut and beautiful that credit is usually given him for the discovery.

The surface films of a large number of substances have been investigated principally by Harkins, Adam, and Gorter. We shall now outline the results from some of the studies on the surface films of fatty acids and of proteins.

In this discussion we shall talk about the film "pressure." It must

[60] Lord Rayleigh, *Proc. Roy. Soc.* (*London*), **47**, 364 (1890).
[61] A. Pockels, *Nature*, **43**, 437 (1891).
[62] I. Langmuir, *J. Am. Chem. Soc.*, **39**, 1848 (1917).
[63] W. B. Hardy, *Proc. Roy. Soc.* (*London*), **86A**, 610 (1912); **88A**, 303 (1913)
[64] W. D. Harkins et al., *J. Am. Chem. Soc.*, **39**, 354 (1917); **39**, 541 (1917).

be realized that what we really mean is the difference in surface tension between the clean surface on one side of the mica float and that of the film-covered surface on the other side of the float. A true film pressure would involve a negative surface tension.

Long-chain aliphatic compounds can exist in four types of films:[65]

1. Condensed film, in which the molecules are closely packed and steeply oriented to the surface.

2. Liquid-expanded films, which are still coherent but occupy a much larger area than condensed films. These films can form on the surface a separate phase, a gaseous film which is in equilibrium with the liquid phase.

3. Vapor-expanded films, which are similar to liquid-expanded films but have less cohesion and do not show a region of constant surface pressure.

4. Gaseous or vapor film, in which the molecules are separate and show independent motion. Adsorbed films of soluble, capillary-active substances are nearly always gaseous.

We thus see that there is a close analogy between phase transitions in two dimensions (surface films) and phase transitions in three dimensions.

A single substance, under the proper conditions of temperature and surface pressure, may be obtained as a condensed film, as either a liquid- or vapor-expanded film, and as a gaseous film. Phase transitions are shown by myristic acid spread on water at 18° C. See Fig. 11.

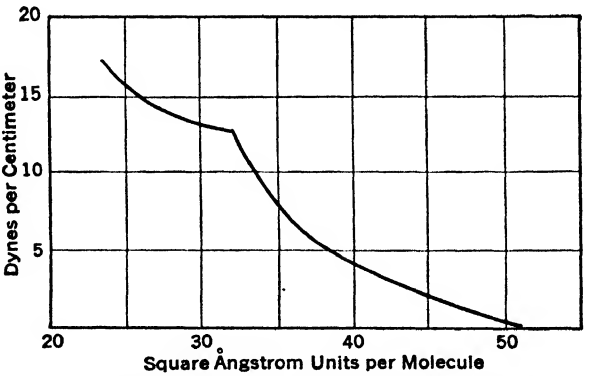


FIG. 11. Compression curve for myristic acid film on water at 18° C.

The portion of the curve at low pressures represents the compression of a liquid-expanded film while the section at higher pressures shows the compression of a condensed film.

[65] N. K. Adam, *Physics and Chemistry of Surfaces*, Oxford University Press, New York.

A large number of spread films have been studied extensively; studies on sterol films [66] and on lecithin films [67] may be mentioned.

Physics and Chemistry of Surfaces," by N. K. Adam, contains an exhaustive review of work on spread films.

Spread Protein Films

Devaux [68] was, apparently, the first to spread protein on water and to study the properties of such spread films.

Following the lead of Langmuir, Adam, and Harkins, Gorter began a series of studies on spread protein films. Gorter's work has been followed by numerous investigations from other laboratories. The fact that, before spreading, many proteins are water soluble while the spread film is water insoluble suggests that proteins undergo drastic molecular changes during spreading. Another notable fact arising from film work on proteins is that the intrinsic differences between native proteins largely disappear upon spreading. Thus, the thicknesses of spread protein films at 10 dynes per centimeter "pressure" are all about the same. Likewise the area per given weight of protein is approximately the same for all proteins. In the native state proteins differ greatly in their molecular dimensions.

Methods for Spreading Proteins

Two general methods are used for spreading proteins: (1) The protein is dissolved in a solvent, usually water, although a 60 per cent solution of isopropyl alcohol in water has been employed. Whatever the solvent, the solutions of the proteins must be quite dilute. The solution is dropped on the surface from a vertical pipette or is blown over the surface from a horizontal pipette. (2) The dry, powdered protein is placed on the surface, and the protein spreads spontaneously. The first method, using a solvent, seems to give a better performance.

The time required for complete spreading of protein films is usually appreciable, sometimes many hours being required. The time is shortest when the solution upon which the film is being spread has a *p*H that corresponds to the isoelectric point of the protein. Increasing the salt concentration of the buffer upon which the spreading is being done also decreases the spreading time. Under favorable conditions, spreading should be essentially complete in 15 minutes.

The film balance for protein work is the same as that used with fatty acids and other surface films. The area of the spread film is decreased

[66] J. F. Danielli and N. K. Adam, *Biochem. J.*, **28**, 1583 (1934).

[67] J. B. Leathes, *Lancet*, **208**, 803, 853, 957, 1019 (1925).

[68] H. Devaux, *Proc. verb. soc. sci. phys. nat. Bordeaux*, November, 1903.

by means of a movable barrier, and the film "pressure" is determined with a torsion balance connected to a mica float. A complete apparatus, accurate enough for most purposes, can be bought from the Central Scientific Company of Chicago; it is called a hydrophile balance.

The pressure-area curves for the proteins investigated show considerable similarity. Below 0.2 dyne per centimeter pressure, the pressure is largely independent of the area. At pressures between 0.2 and 2 dyne per centimeter the slope of the pressure-area curves increases rapidly. Above 5 dynes per centimeter the slope is large but constant as the area

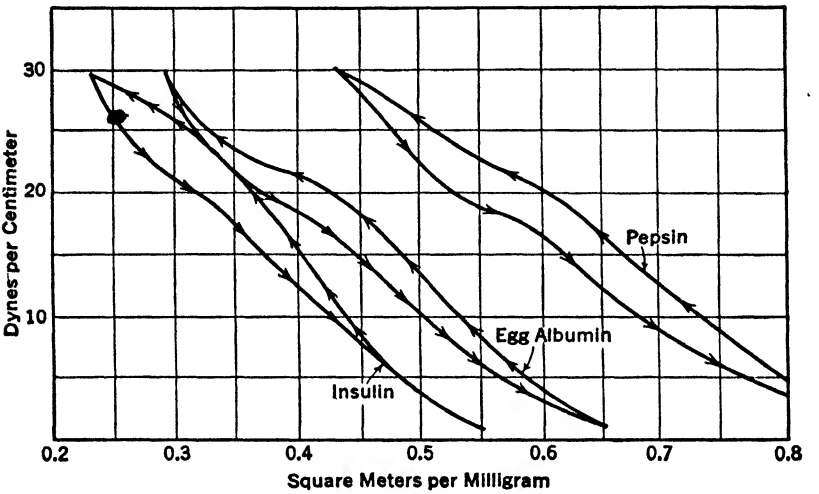


FIG. 12. Reproducible force-area curves for aged monolayers of several proteins. (Langmuir and Waugh.)

is decreased. At about 20 dynes per centimeter the slope begins to decrease with increasing pressure. The film gels at a pressure of about 5 dynes per centimeter.

Langmuir and Waugh [69] found that, as a spread film of protein is compressed, a soluble component whose molecular weight was about 1,000 was forced out of the film. These workers also report that, after the protein film had been compressed and allowed to age, the films could be expanded and compressed any number of times and the compression-expansion curves followed a well-defined and reproducible cycle. Some of their results are shown in Fig. 12.

It will be noted that the aged films of Langmuir and Waugh occupy somewhat smaller areas per unit weight than unaged films.

Protein films as well as other spread monolayers have been charac-

[69] I. Langmuir and D. F. Waugh, *J. Am. Chem. Soc.*, **62**, 2771 (1940).

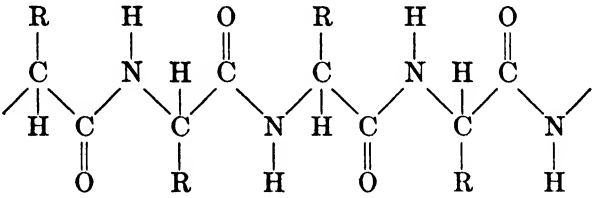
terized by extrapolating the approximately linear portion of the force-area curves to zero pressure and calling this area the area occupied by the film at close packing; i.e., this area when divided by the number of molecules in the film is supposed to yield the area per molecule of spread material. Although this is the customary procedure, it has no physical justification. As pointed out by Bateman and Chambers,[70] a more logical method is to determine the coefficient of compressibility of the spread film. The compressibility is then plotted against the film area, and the point at which the compressibility passes through a minimum is assumed to be the area of the maximally compressed film (without film rupture). The coefficient of compressibility is given by the relation

$$\text{Coefficient of compressibility} = -\frac{1}{A}\left(\frac{dA}{dF}\right) \qquad 39$$

where A is the area of the spread film and F is the surface pressure in dynes per centimeter.

Structure of Protein Monolayers

Proteins are made up of amino acids joined together through the peptide linkage. This produces a long peptide chain containing several hundred amino acid residues



In a native protein, this polypeptide chain is coiled or contracted into something approaching a spherical molecule. When spread in a monolayer on the surface, the native molecule is unfolded to yield a more or less stretched chain. In the completely expanded film (low film pressures) the R groups of the chain are probably lying flat on the surface. At higher pressures, the amino acid side chains (the R groups) orient themselves vertically to the surface. In most proteins there is a fairly close balance between the hydrophilic and hydrophobic R groups, so that, upon orientation, about half of them, that is, the hydrophilic groups, would stick into the water phase while the other half would point upwards. As the film is compressed, water is squeezed out of it. The decrease in area of a compressed protein monolayer is thus the

[70] J. B. Bateman and L. A. Chambers, *J. Chem. Phys.*, **7**, 244 (1939).

result of two factors: (1) orientation of the R groups, and (2) a dehydration of the film.

There are two ways of rendering the outline of a surface film on a surface visible. The older technique was to sprinkle the clean surface with powdered talc. The talc is insoluble and clearly visible. When the fatty acid or protein is spread, it pushes the talc along the surface so that the area covered by the film is clearly outlined by the talc on the unspread area.

Langmuir and his co-workers have developed indicator-oils to outline the spread area. It will be recalled that water covered by a thin film of oil frequently shows iridescent colors. Langmuir et al. heat ordinary motor oils to produce a degree of oxidation. The longer and more intensely the oils are heated, the thinner the films they give when spread on the surface. These oils are then used to outline the spread film in the same way that talc is used. Schaefer [71] also reports that interesting and characteristic patterns are produced if indicator oil is placed in the center of a compressed protein film. The film tears in a most spectacular manner.

Mixed Films

Neurath [72] investigated the interaction between myristic acid and egg albumin. He found that, at a certain ratio of fatty acid to the protein, the protein film was greatly expanded. Rideal and co-workers [73] have studied the associations and interactions between proteins and other surface-active materials. The interactions which they observed were followed both by pressure-area measurements and by film potentials. Film potentials are determined by having a reversible electrode dipping into the buffer under the spread film and another electrode a small distance above the film. The air between the electrode and the film is rendered conducting by means of a small amount of radioactive material. The electrostatic potential across the film is measured and is spoken of as the film potential.

Injection of a surface-active substance under a surface film was found to change both the film potential and the film pressure. If the injected molecules are anchored to the film by polar groups and do not penetrate the film, the film potential is altered but not the film pressure. Thus gallic acid, tannic acid, or certain dyes are adsorbed beneath the protein

[71] V. J. Schaefer, *J. Phys. Chem.*, **46**, 1089 (1938).
[72] H. Neurath, *J. Phys. Chem.*, **42**, 39 (1938).
[73] E. G. Cockbain and J. H. Schulman, *Trans. Faraday Soc.*, **35**, 716 (1939).
E. K. Rideal and J. H. Schulman, *Nature*, **144**, 100 (1939).
E. Stenhagen and E. K. Rideal, *Biochem. J.*, **33**, 1591 (1939).

monolayer without penetrating it. If, however, in addition to polar group interaction there is a strong attraction between the hydrophobic groups of the injected material and the hydrophobic groups of the film molecules, the injected molecules penetrate the film. Such penetration can be readily detected by a considerable increase of surface pressure which is well above the possible film pressure of either of the components individually. If the interaction between the molecules of two components of a mixed film is weak, lateral compression will force one molecular species out of the film. Thus, a stable and rigid cholesterol-gliadin film liquefies suddenly at pressures above 20 dynes, revealing all the characteristics of a cholesterol film. This displacement of the gliadin from the film is reversible when the pressure is decreased. Interestingly enough, they found that, in general, those substances which hemolyze red blood cells also penetrate protein films.

Built-up Monolayers

Blodgett [74] has described a technique by which successive monolayers of fatty acids may be built up on solid surfaces. The monomolecular film is transferred from the water surface to glass or metal slides by raising the slide slowly out of water upon which a monolayer has been spread. A fatty acid film is deposited on a calcium bicarbonate buffer. The *p*H of the calcium bicarbonate buffer is adjusted to the needs of the experiment by bubbling CO_2 through the buffer. Two types of deposition are described by Blodgett. The *Y*-deposition is obtained on the up, as well as on the down, trip of the slide; the *X*-deposition takes place only on the down trip of the slide. The governing factor in the type of deposition is the *p*H of the buffer. The *Y*-deposition is obtained at *p*H 7 and lower, and *X*-deposition at *p*H 9 and above. During deposition, a constant pressure is maintained on the film by means of a "piston oil." The piston oil spreads with an approximately constant pressure. Table 5 lists some piston oils used for this purpose.

TABLE 5

"PISTON OILS" AND THEIR FILM PRESSURE

Piston oil	*F* in dynes per centimeter
Tricresyl phosphate	9.5
Rape-seed oil	10.5
Castor oil	16.5
Neat's-foot oil	19.0
Oleic acid	29.5

[74] K. B. Blodgett, *J. Am. Chem. Soc.*, **56**, 495 (1934); **57**, 1007 (1935).

If the index of refraction of the glass slide differs sufficiently from that of the deposited film, beautiful interference colors may be produced. The color depends upon the thickness of the film and upon the difference of index of refraction between the glass and the deposited film. As the index of refraction of fatty acids is about 1.4 and that of ordinary glass is also about 1.4, no color results from films deposited on ordinary glass. Fatty acid films deposited on slides of optical glass having an index of refraction of 1.64 give beautiful and intense colors. The best colors, however, are exhibited by films deposited on polished metal slides such as those of chromium.

Y-films which are deposited on the up as well as down trip should have the methyl groups adjacent to methyl and carboxyl groups adjacent to carboxyls. *X*-films, on the other hand, should have the methyls of one layer adjacent to the carboxyls in the next layer. Actually, X-ray studies show [75] the *X*- and *Y*-films to have the same crystal structure.

Bikermann [76] has suggested that the deposited film ceases to be a monolayer and becomes aggregates or microcrystals when the film is deposited on a solid from a liquid surface. He came to this conclusion through deposition ratio studies. The deposition ratio is the ratio of the area of the spread film on the liquid to its area when deposited.

Deposited Protein Films

Langmuir and Schaefer,[77] using essentially the same technique as described above for the deposition of fatty acid films, have deposited protein monolayers on solid surfaces. The slide must be conditioned with a layer of calcium or barium stearate before deposition of the protein; otherwise the protein film deposited on the up trip out of the surface will peel off on the down trip through the surface.

Astbury, Bell, Gorter, and Van Ormondt [78] have deposited up to seventeen hundred successive layers of egg albumin on chromium-plated metal slides. These films have a tendency to tear parallel to the direction in which the slide was moved through the liquid surface. The films were birefringent (transmitted polarized light) when viewed perpendicular to the surface, the slow vibration being parallel to the direction of movement of the slide. The thickness of these films, measured by direct mechanical means, was 9.5 Å per monolayer, which is in excellent agreement with the results obtained with the surface balance. Unfortunately, no deposition ratios for the films are reported, and without this informa-

[75] C. Holley, *Phys. Rev.*, **53**, 534 (1938).
[76] J. J. Bikermann, *Proc. Roy. Soc.* (*London*), **170A**, 130 (1939).
[77] I. Langmuir and V. Schaefer, *Chem. Revs.*, **24**, 181 (1939).
[78] W. T. Astbury, F. O. Bell, E. Gorter, and J. Van Ormondt, *Nature*, **142**, 33 (1938).

tion a comparison with the thickness of the film on the liquid surface is really not justified.

Langmuir and Schaefer have deposited what they call *S*-films of proteins on solids. These protein films are produced by placing a drop of protein solution on a wet slide upon which had previously been deposited a stearic acid film conditioned with aluminum chloride or thorium nitrate. The protein solution, after being placed on such a slide, is washed off; the *S*-film remains on the slide.

Harkins, Fourt, and Fourt [79] have reported a very striking example of specific adsorption. A *S*-film of beef catalase was placed on a chromium-plated slide and the slide exposed to a serum containing beef anticatalase from rabbits. The anticatalase deposited on the slide. It was then possible to deposit another layer of catalase. Successive alternating layers of catalase and anticatalase were thus built up. If the catalase film was exposed to serum not containing anticatalase, little or no protein was taken up by the slide. Incidentally, the deposited catalase retained its ability to decompose H_2O_2 even when it was covered by a layer of anticatalase.

Films at Oil-Water Interfaces

Spread films have been studied at oil-water interfaces. Askew and Danielli [80] used a modified film balance technique. On the other hand, Alexander, Teorell, and Aborg [81] spread the material at the interface in a fixed area and measured the change in film pressure by a surface-tension technique. The pressure was varied by increasing the amount of material at the interface. This last technique appears to be more convenient than the modified film balance. No very unexpected results were reported. The films tended to be more expanded than at the air-water surface.

[79] W. D. Harkins, L. Fourt, and P. C. Fourt, *J. Biol. Chem.*, **132**, 111 (1940).

[80] F. A. Askew and J. F. Danielli, *Trans. Faraday Soc.*, **36**, 785 (1940).

[81] A. E. Alexander and T. Teorell, *Trans. Faraday Soc.*, **35**, 727 (1939).
A. E. Alexander, T. Teorell, and C. G. Aborg, *Trans. Faraday Soc.*, **35**, 1200 (1939).

CHAPTER XII

COLLOIDAL SOLUTIONS

A "colloid" is a state of matter and not a kind of matter, as is shown by the fact that, if the proper technique is used, practically any substance may be brought into a colloidal state. Strictly speaking, the term "colloid" should be used only as an adjective to define a physical state of matter.

In the older literature will be found a distinction between colloids and crystalloids, the assumption being 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. The limits of size are rather arbitrarily chosen to be from 1 mμ (10^{-7} cm. or 10 Å) to 0.1 μ (10^{-5} cm.) in diameter. All particles within this range of dimensions are known as colloids whether they are molecularly dispersed or not. Proteins, therefore, are included as colloids.

Colloids may be formally divided into eight classes:

1. Gas dispersed in a liquid—foams.
2. Gas dispersed in a solid—aero gels.
3. Liquid dispersed in a gas—mists.
4. Liquid dispersed in a liquid—emulsions.
5. Liquid dispersed in a solid—pearl, opal.
6. Solid dispersed in a gas—smoke.
7. Solid dispersed in a liquid.
   (*a*) Hydrophobic suspensoids (suspensoids)—gold sol.
   (*b*) Hydrophilic suspensoids (emulsoids)—protein in water.
8. Solid dispersed in a solid—ruby glass, black diamond.

In any colloidal system, the continuous phase is known as the dispersion medium and the discontinuous phase as the dispersed phase.

It is evident that not all the classes of colloids listed above are of interest to the biochemist. In living matter, we deal, in general, with only two of these classes, i.e., emulsions and hydrophilic suspensoids. Although hydrophobic suspensoids have little direct biological signifi-

cance, they give us an insight into general colloidal behavior, and, accordingly, we shall include then in our discussion.

Hydrophobic Suspensoids

The metallic suspensoids such as gold sols and silver sols are representatives of this group. There are various ways by which suspensoids may be produced. In general, these methods of preparation may be divided into two classes: (1) by crystallization or condensation, and (2) by solution or dispersion. These relations are diagrammed in Fig. 1.

FIG. 1. General methods for preparing hydrophobic suspensoids.

The details of these methods of preparation need not concern us here; any good textbook on colloidal chemistry will supply a wealth of information.

As we have indicated in the chapter on electrokinetics, metallic sols are not simply metallic particles suspended in water but have a more complex structure, as Pauli [1] has pointed out. Such a colloidal complex consists of three parts: (1) the neutral, metallic core; (2) the ionogenic complex; and (3) the gegen or counter ions.

Stability of Hydrophobic Suspensoids

An outstanding and common property of colloidal solutions is their lack of stability and their sensitivity to physical and chemical conditions. The hydrophobic suspensoids exhibit this instability to a rather marked degree. They are particularly sensitive to electrolytes, and at sufficient electrolyte concentration they promptly precipitate. This instability in respect to electrolytes indicates the important role of the electrostatic potential on the hydrophobic particles. The connection between the electrostatic potential (ζ-potential) of the particles and their stability has long been recognized.[2]

Hydrophobic suspensoids exhibit two types of precipitation or coagulation, rapid and slow precipitation. Rapid precipitation occurs when the ζ-potential has been reduced to such a point that it is no longer of

[1] Wo. Pauli, *Trans. Faraday Soc.*, **31**, 11 (1935).
[2] W. B. Hardy, *Proc. Roy. Soc.* (*London*), **66**, 110 (1900).

any importance in maintaining stability. Then every collision between particles results in a cohesion, and, accordingly, the problem resolves itself into a calculation of the number of collisions. This has been done by Smoluchowski,[3] who derived the following equation:

$$N = \frac{N_0}{1 + t/t_{\frac{1}{2}}} \qquad 1$$

where N is the number of particles per cubic centimeter at any time t, and N_0 is the original number of particles per cubic centimeter. $t_{\frac{1}{2}}$, the time required to reduce the number of particles by one-half, is

$$t_{\frac{1}{2}} = \frac{3\eta}{N_0 4kT} \qquad 2$$

where η is the coefficient of viscosity of the dispersions medium, k is Boltzmann's distribution constant (1.38×10^{-16} erg per degree), and T is the absolute temperature.

Thus, according to Smoluchowski, the rate of rapid precipitation of a hydrophobic suspensoid is independent of size, shape, and other specific characteristics of the particles. Smoluchowski's equation has been found to hold for suspensoids of gold, of kaolin, and of quartz. If, however, the sol is pronouncedly polydispersed, it coagulates at a faster rate than is predicted by Smoluchowski. Rod-shaped particles also show a much higher rate of precipitation than spherical particles. It has been observed, for example, that the very asymmetric particles of vanadium pentaoxide sol coagulate thirty times faster than Smoluchowski's equation[4] predicts.

Slow coagulation takes place when the ζ-potential has been reduced below a certain critical value. The potential, though partially effective in preventing coagulation, is not completely so. In this event only a fraction of the collisions between particles are fruitful in producing cohesion between the particles. The probability of a collision resulting in cohesion varies between one and zero. When the probability is one, rapid coagulation results; when it is zero, the sol is completely stable.

The critical potential of a suspensoid is a concept originated by Powis.[5] The critical potential is defined as that value of the ζ-potential above which the sol is relatively stable and below which it rapidly coagulates. The rule is usually valid when coagulation is brought about

[3] M. von Smoluchowski, *Physik. Z.*, **17**, 557, 583 (1916).
See also H. Muller, *Cold Spring Symposia Quant. Biol.*, **I**, 60 (1933).
[4] G. Wiegner and C. E. Marshall, *Z. physik. Chem.*, **140**, 1 (1929).
[5] F. Powis, *J. Chem. Soc.*, **109**, 734 (1916).

with divalent or trivalent ions, but KCl and NaCl produce precipitation before the ζ-potential has been reduced below the critical potential. For a discussion of the critical potential, see Briggs.[6]

The influence of the valence of the precipitating ion is expressed by the Schulze-Hardy rule, which states that the effective ion in the precipitation of sols by electrolytes is the ion of opposite sign to the charge on the colloidal particle. Furthermore, the precipitating power of the effective ion increases greatly with increasing valence of the ion. In Table 1 is shown the concentration of the electrolyte in millimoles per liter required to precipitate a negatively charged copper ferrocyanide sol.[7]

TABLE 1

PRECIPITATION VALUES OF CATIONS FOR A NEGATIVELY CHARGED COPPER FERROCYANIDE SOL (SEN)

Salt	Precipitating value in millimoles per liter
KCl	35.6
NaCl	92.5
$BaCl_2$	0.46
$SrCl_2$	0.54
$MgSO_4$	0.76
$Al_2(SO_4)_3$	0.058
$Ce(NO_3)_3$	0.034
$Th(NO_3)_4$	0.038

Even though the valence of the precipitating ion is undoubtedly of great importance, it is found that ions of the same valence do not behave alike and the alkaloid as well as dye ions follow no rule at all in regard to valence.

If the precipitating effect of univalent ions on hydrophobic sols is studied in detail, a definite lyotropic series is observed. Papada [8] investigated the precipitating power of the chlorides of several cations on negatively charged sols. He found that the precipitating power of these cations could be arranged in the order Cs > Rb > K > Na > Li. Voet and Balkema,[9] who studied the influence of potassium salts on a positively charged gold sol, found the series Cl > Br > I > CNS.

[6] D. R. Briggs, *J. Phys. Chem.*, **34**, 1326 (1930).
[7] B. Sen, *J. Phys. Chem.*, **29**, 517 (1925).
[8] N. Papada, *Kolloid-Z.*, **9**, 137 (1911).
[9] A. Voet and F. Balkema, *Rec. trav. chim.*, **52**, 371 (1933).

Levine [10] and Langmuir [11] have considered independently the role of the electrical forces in the stability of colloidal solutions. Langmuir concerned himself with the attractive and repulsive forces between infinite plane particles, whereas Levine discussed such forces between spherical and cylindrical particles. Their approach was essentially that of Debye and Hückel in their calculation of the forces between ions in solution. Their arguments are highly mathematical and cannot be given in detail.

When two particles of a hydrophobic suspensoid approach each other, an attractive force exists which is electrostatic in nature; the force is attractive in spite of the fact that the particles carry the same sign of charge. It arises from a tendency of the two double layers to fuse. In short, there is an attraction between the first particle and the double layer of the second particle and between the second particle and the double layer of the first particle. That such an attractive force must exist can be seen from the analogy with a crystal such as sodium chloride. Sodium and chlorine exist in a crystal as ions, but the crystal does not fall apart; the sodium ions in the crystal are bound together by the chloride ions. There is, in addition to the attractive force between the colloidal particles, a repulsive force. The two particles have the same sign of charge and, accordingly, repel each other. The magnitude of this repulsive force increases rapidly as the particles approach close together. We have, therefore, both attractive and repulsive forces operating between the colloidal particles, and both of these forces are electrostatic. The net force acting between these two particles is, of course, the resultant of these two forces. When the particles have approached very closely, the relatively powerful van der Waals attractive forces begin to operate and cohesion results. Figure 2 [12] shows the energy relations calculated by Levine for two gold particles suspended in 0.001 *M* KCl as a function of the distance between the particles.

Note the minimum in the $F_a + F_r$ curve. The position and depth of this minimum depend only on the product

$$\tau = a\lambda = \frac{a\sqrt{C}}{3.04 \times 10^{-8}}$$

where a is the radius of the particle, C is the concentration of univalent electrolyte, and λ is the thickness of the ionic double layer. In most spherical particles the depth of the energy minimum is smaller than the thermal energy kT of the particles, and, accordingly, the heat energy is

[10] S. Levine, *Proc. Roy. Soc.* (*London*), **A170**, 145 (1939); *Trans. Faraday Soc.*, **35**, 1125 (1939); **36**, 215 (1940); *Phil. Mag.*, **29**, 105 (1940); *J. Phys. Chem.* **46**, 239 (1942).

[11] I. Langmuir, *J. Chem. Phys.* **6**, 873 (1938).

[12] S. Levine, *Proc. Roy. Soc.* (*London*), **A170**, 145 (1939).

sufficient to kick the particles out of this energy trough. The trough can be important, however, particularly in asymmetric particles where the surface of interaction between the particles is greatly increased, and we shall deal with this situation in our discussion of colloidal structures. Note also the maximum in the $F_a + F_r + F_v$ curve. For the particles

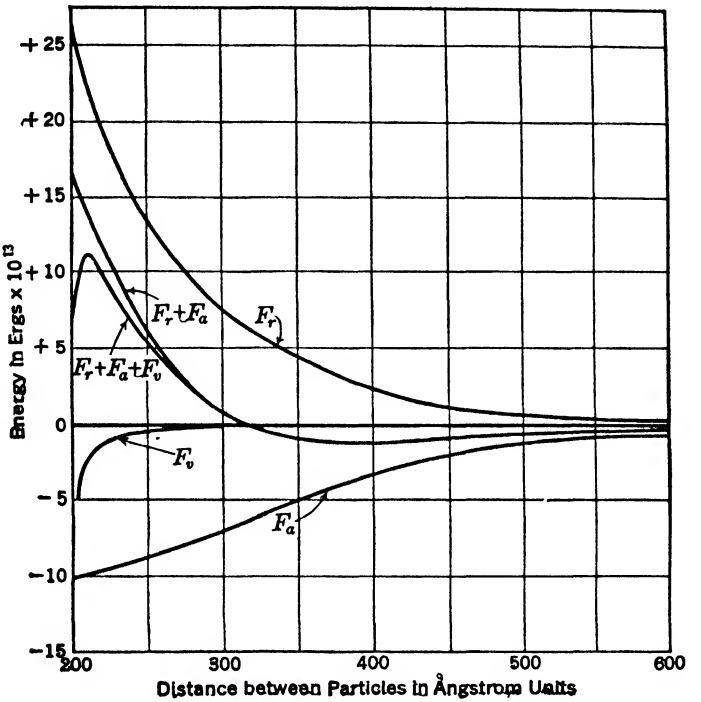


FIG. 2. Energy relations between two colloidal gold particles as they approach the radii of the spherical particles 10^{-6} cm. and their charge is 100 e.s.u. They are suspended in 0.001 *M* KCl solution. F_r is the electrostatic energy of repulsion, F_a is the electrostatic energy of attraction, and F_v is the energy resulting from the van der Waals forces. (Levine.)

to cohere, they must hurdle this energy hump. The height of this hump is determined by the electrolyte concentration and by the magnitude of the ζ-potential. The greater the electrolyte concentration, the lower is the hump; the higher the ζ-potential, the higher the hump. It is the height of this hump which determines the stability of hydrophobic colloids. Neither Levine nor Langmuir has considered the influence of hydration on the stability of suspended particles, and, indeed, no quantitative theory of the influence of hydration on stability has been forthcoming. However, we shall presently attempt to deal with this aspect of the problem at least in a qualitative fashion.

Optical Properties of Hydrophobic Suspensoids

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 various other sols. The light is scattered from individual particles, and the cone of light which is visible is known as the Tyndall cone.

With particles whose diameter is considerably larger than the wavelength of the light used, the scattering is brought about by simple reflection from the surface of the particles. With a metallic particle, the reflection is selective, so that certain wavelengths are reflected to a greater

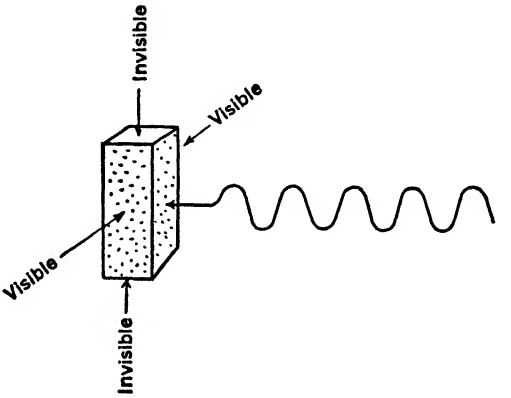


FIG. 3. Directions in which plane-polarized light is visible after passing into a colloidal solution. The plane of vibration of the incident light is in the plane of the paper.

extent than others and the suspension exhibits a color. On the other hand, if the particle is a non-conductor of electricity, the reflection is almost independent of the wavelength. If, however, the particles are small in comparison to the wavelength of the incident light, a phenomenon takes place which is more akin to refraction, and we can apply the Rayleigh equation.

If plane-polarized light is the source of light, i.e., light whose vibrations take place in only one plane, the light will be scattered only in the plane at right angles to the plane of polarization. This situation is diagrammed in Fig. 3.

If the incident light is unpolarized, the Tyndall effect is visible in all directions. The scattered light, however, is polarized in a plane which is vertical to the beam of incident light.

The intensity of the light scattered at right angles to the plane of

polarization of the incident light is given by Rayleigh's equation, which is

$$I = \frac{I_0 9\pi^2 V^2 N}{\lambda^4 X^2}\left(\frac{n_1^2 - n^2}{n_1^2 + 2n^2}\right)^2 \qquad 4.$$

where I_0 is the intensity of the incident light, V is the volume of the individual spherical particles, N is the number of particles per cubic centimeter, λ is the wavelength of the light, X is the distance between the Tyndall cone and the observer, n_1 is the index of refraction of the particles, and n is the index of refraction of the dispersion medium. The Rayleigh equation is valid for spherical particles whose diameter is less than about one-third the wavelength of the incident light. The particles must also be non-conducting (dielectrics).

We see from an examination of Rayleigh's equation that the intensity of the light scattered from a Tyndall cone increases with (1) increasing intensity of incident light, (2) decreasing wavelength of incident light, (3) number of particles, (4) decreasing particle size, and (5) difference between the indices of refraction of the particles and of the medium. If the two indices are equal, the Tyndall effect vanishes completely. The Tyndall effect superficially resembles fluorescence, but actually the two phenomena are quite different. Fluorescent light is not polarized, and the distribution of the wavelength of emitted light is not determined by the size of the particles as in Tyndall scattering.

It is noteworthy that, the shorter the wavelength of the incident light, the greater is the intensity of the Tyndall light. This means that blue light will be scattered to a greater extent than red light. This gives rise to the so-called Tyndall blue, instances of which are the blue of the sky, of tobacco smoke, of skimmed milk, and of cholesterol sols in water. There is no pigment in blue eyes; the color is a Tyndall blue.

The problem of light scattering by suspended particles has a number of ramifications. It appears that a great deal of useful information about the size and shape of the particles can be had by a detailed study of the optical properties of the colloidal systems.[13, 14]

Protein molecules in solution give only a very slight or no Tyndall cone in the visible wavelengths, but in the ultra-violet there is, as would be anticipated, some scattering. Some measurements have been made on the light scattering of protein solutions,[15] but, considering the importance of the problem, the field has been explored to a very limited ex-

[13] C. E. Barnett, *J. Phys. Chem.*, **46**, 69 (1942).

[14] C. R. Hoover, F. W. Putnam, and E. G. Wittenberg, *J. Phys. Chem.*, **46**, 81 (1942).

[15] P. Putzeys and J. Prosteaux, *Trans. Faraday Soc.*, **31**, 1314 (1935).

tent. It would appear to the author that this would be a most profitable subject for investigation.

The Ultramicroscope

A direct application of the Tyndall effect is the ultramicroscope. An intense, horizontal beam of light is focused in a cell containing the sol. The point of focus of the incident light is observed with an ordinary microscope. The background of the microscope field is dark, and the particles appear as bright points of light which are in rather violent Brownian motion. No idea of the shape or size of the particles can be obtained directly by the ultramiscroscope since the particles are not actually visible, being far too small to be resolved by a microscope. The resolving power of a microscope is limited in the end by the wavelength of light used, the minimum distance which can be resolved being one-half of the wavelength of light. This means that the smallest particle visible in the very best miscroscope is about 2,000 Å in diameter. An estimate of the particle size by means of the ultramicroscope can be made by the simple expedient of counting the number of particles in a given volume. Then, the total weight of suspended material as well as its density being known, the average particle size can be calculated. It is claimed that particles as small as 17 Å in diameter can be distinguished by means of the ultramicroscope.

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 properly designed condensers for the microscope. A somewhat different lens system is that of Spierer, which is another example of a dark-field microscope. Seifriz [16] and also Thiessen [17] have applied the Spierer lens to an examination of biological structures.

Color of Hydrophobic Suspensoids

Many sols, such as those of gold, silver, and arsenic trisulfide, give beautiful and very startling colors. The reason for these varied colors is still obscure. It is known that color is very intimately related to particle size, and a theory of the relation between particle size and color has been proposed by Mie.[18] The color depends not only on the size of the particle, but also on its shape; an asymmetric particle will give a color which is equivalent to that of a spherical particle of a larger mass.

[16] Wm. Seifriz, *J. Phys. Chem.*, **35**, 118 (1931).
[17] R. Thiessen, *Ind. Eng. Chem.*, **24**, 1032 (1932).
[18] G. Mie, *Ann. Physik*, (4) **25**, 377 (1908); *Kolloid-Z.*, **2**, 129 (1907).

Hydrophilic Suspensoids

From a biological point of view, hydrophilic suspensoids are by far the most important class of colloids. All proteins belong to this class, as do the carbohydrate gums such as the pentosans. These substances along with the electrolytes maintain tissue hydration. In contrast to the hydrophobic suspensoids and in spite of their relatively large size, they form remarkably stable solutions. They owe their stability both to their hydration and to their electrostatic charge.

Water of Hydration (Bound Water)

The water content of protoplasm in an actively growing condition is about 85 per cent of the total weight of the tissue. In a dormant state the water content is greatly reduced, often to one-third or less of the total tissue weight. This percentage of about 30 to 35 per cent water appears to be a critical region for protoplasm. The change in behavior of protein solutions and of protoplasm as the water content is reduced below 30 to 35 per cent is made evident in various ways (heats of imbibition, imbibition pressure, freezing points, expansion on freezing, etc.). This indicates a restriction in the freedom of motion of the water molecules when less than about 35 per cent by weight is associated with a protein.

The question of water of hydration or bound water was rightly emphasized by Gortner,[19] who was extremely active over a number of years with this problem. He developed techniques for its measurement, and we owe him much for his labors.

Since protein is the principal water binder of tissue, we will center our attention on this class of substance. Proteins bind water through their polar groups. It is, in fact, possible to estimate the quantity of water bound from a knowledge of the amino acid content of a protein. This has been done by Sponsler, Bath, and Ellis [20] in an extremely lucid consideration of the problem.

It will be recalled from our previous discussions that a protein molecule is made up of amino acids linked together through the peptide bond. This produces a peptide chain which may contain several hundred amino acid residues. Many of these residues have hydrophilic groups in their side chains. In an actual protein molecule, the peptide chain is folded in some manner, yet unknown, to produce in most instances a molecule with a low order of asymmetry. The folding of the peptide chain appears, however, to leave the polar side chains exposed. The peptide

[19] R. A. Gortner, *Outlines of Biochemistry*, John Wiley & Sons, Inc., New York (1938).

[20] O. L. Sponsler, J. D. Bath, and J. W. Ellis, *J. Phys. Chem.*, **44**, 996 (1940).

bonds are probably also at least partially exposed. There are, therefore, two types of hydrophilic centers in a protein molecule: (1) the polar side chains and (2) the oxygen and nitrogen atoms of the peptide bond. The number of water molecules which can be coordinated by various hydrophilic groups is shown in Table 2.

TABLE 2

NUMBER OF WATER MOLECULES COORDINATED BY HYDROPHILIC GROUPS (SPONSLER)

Group	Number of water molecules coordinated	
	Theoretical	Experimental
H_2O	4	4
—OH	3	3
—COOH	4–5	4
=O	2	2
$—NH_2$	3	3
—NH	2	..
=N—	1	..

Hydration centers (=O and =NH) of the backbone of the peptide chain are capable of coordinating two water molecules each, but space restrictions must reduce the number actually coordinated and it is estimated that only about half of the theoretical number are bound to the backbone. The hydrophilic end groups of the amino acid side chains have fewer space restrictions and would be expected to participate to almost their full capacity. Sponsler estimates that in gelatin the total average amount of water bound by one amino acid residue (including the amount bound on the peptide backbone as well as on the side-chain polar groups) is 2.6 water molecules. It is to be anticipated that this estimate will be somewhat different for other proteins and will depend primarily upon the number of polar groups in the amino acid side chains and on the tightness of pack of these chains in the protein molecule.

The best evidence indicates that water is bound to the hydrophilic groups by means of hydrogen bond formation. A hydrogen bond may be visualized as consisting of a hydrogen atom which is strongly associated with two electronegative atoms and thus acts as a bridge to hold them together. The hydrogen bond is essentially an electrostatic bond and, relatively speaking, acts through a considerable distance. When the hydrogen atom is about equidistant between the two electronegative atoms, the bond has its maximum strength. Any displacement of the hydrogen atom from this position leads to a weakening of the bond.

The stronger the bond, the shorter it is. Hydrogen bonds form rather readily between atoms of fluorine, oxygen, nitrogen, and chlorine, and the strength of the bond decreases in the order of elements named. Its formation is probably confined to these four atoms because they alone are electronegative enough to attract a hydrogen atom with sufficient strength. Though the energy of formation of a hydrogen bond varies, it is usually of the order of 5,000 calories, and the distance between the electronegative atoms is around 3 Å.

Since bound water is held by so many different groups exhibiting a variety of binding strengths, it is not surprising that the various experimental methods available for its estimation frequently show rather poor agreement. It is very difficult and often ambiguous to say that such and such quantity of water is bound.

Bound water is defined in terms of the experimental technique by which it is measured. Some fourteen different techniques are available, but not quite as many definitions, as some of the techniques are equivalent. In order to gain an understanding of the situation, we shall select for our discussion the most commonly used of these techniques.

1. Bound water can be defined as that water which remains unfrozen at some specified temperature below zero. The temperature usually chosen is −20° C. Several techniques are based upon this definition. Two of them are:

(*a*) The expansion of the system upon lowering the temperature through the freezing point is measured in a dilatometer. As water freezes, it expands, and the amount of water which freezes is calculated from the expansion. This amount of water is subtracted from the total water present to give the amount of bound water. Figure 4 shows some of the data of Jones and Gortner [21] on a 10.06 per cent gelatin gel.

(*b*) The second technique for the determination of bound water based upon the amount of water remaining unfrozen at temperatures below −20° C. involves the amount of heat absorbed by a system as the temperature is raised from −20° C. to 1° C. If allowance is made for the specific heats and for the fact that each gram of ice absorbs 80 calories of heat as it melts, the amount of water frozen at −20° C. can be calculated. This amount is then subtracted from the total water to give the bound water.[22]

Kistler [23] has criticized the above techniques on the grounds that it is very easy to undercool water to −20° C. or lower without its freezing. Such undercooled water would be calculated as bound water. For

[21] I. D. Jones and R. A. Gortner, *Colloid Symposium Monograph*, **9**, 410 (1931).
[22] W. Robinson, *Colloid Symposium Monograph*, **5**, 199 (1928).
[23] S. S. Kistler, *J. Am. Chem. Soc.*, **58**, 901 (1936).

example, Kistler emulsified water in toluene using aluminum stearate as the emulsifier and found that 19 per cent of the water was unfrozen at −36.7° C. Such water would be calculated as bound water, although it is obvious that no bound water is present. This was an example of undercooling; and, undoubtedly, freezing techniques are open to serious objection on this account. It is still highly important biologically that living tissue can be greatly undercooled without the formation of ice

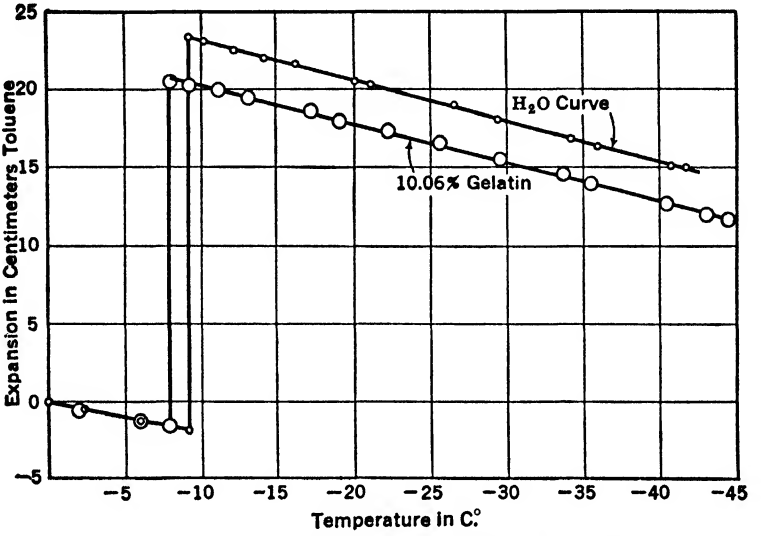


FIG. 4. Expansion of a 10.06 per cent gelatin gel upon freezing. (Jones and Gortner.)

crystals, to whatever this may be due. The planters of winter wheat in Minnesota and Saskatchewan will testify to this.

2. Bound water has also been defined as the amount of water in a system which is not available to act as a solvent. Several techniques involve this definition; only two of them will be considered.

(*a*) The freezing point of the system containing the hydrophilic substance is determined. A certain amount of a solute, such as glucose, is added, and the freezing point is redetermined. The total amount of water present and the fact that a molar solution of glucose has a freezing point 1.86° below that of water being known, the concentration of the glucose can be calculated. The concentration based upon the total amount of water present is known. The difference between the actual and calculated concentration of glucose gives the amount of water which was bound to the hydrophilic colloid and not available as a solvent.[24]

[24] R. Newton and R. A. Gortner, *Botan. Gaz.*, **74**, 442 (1922).

(*b*) The vapor-pressure lowering of water produced by the addition of a solute such as glucose is determined. The vapor-pressure lowering which should be produced by this amount of glucose if all the water were free is known. It is observed that the vapor-pressure lowering is greater than that calculated. The conclusion is drawn that a certain amount of the water is bound. This quantity of water can be calculated from the vapor-pressure lowering actually produced. Hill [25] developed a very delicate method for the determination of the vapor pressure of water. It is based upon a measurement of the temperature change experienced when water evaporates, the rate of evaporation being proportional to the vapor pressure. Hill applied his method to the estimation of the water bound by muscle and found very little water binding. Roepke [26] has reported on the recent refinements in the Hill vapor-pressure method.

Briggs [27] has described a simple direct method for measuring the vapor pressure of water associated with hydrophilic substances. The apparatus is called an isotenoscope. The vapor pressure is measured directly in centimeters of mercury. Briggs applied his method to the measurement of the vapor pressure of water associated with moist samples of casein, agar, fibrin, and cellulose. Figure 5 shows the activity of water associated with isoelectric casein as a function of the water content.

In a second paper Briggs [28] discusses some of the theoretical aspects of bound water. It is evident from Fig. 5 that in order to define bound water one has to specify the activity below which water is said to be bound. For example, at $-20°$ C. the activity of water is 0.822. Accordingly, if we lower the temperature of a colloidal system to $-20°$ C. and allow it to remain until equilibrium has been reached, all the water whose activity is unity down to 0.822 will freeze and will be calculated as free water. The remaining water with an activity of 0.822 and lower will not freeze and will be calculated as bound water. The same kind of arbitrary element exists when bound water is defined as the water which will not act as a solvent to an added solute. Briggs also emphasizes the important roles played by electrolytes associated with the colloid. This aspect of the problem, however, can best be discussed when we consider the osmotic pressure of colloids.

Defining bound water on the basis of its solvent powers is somewhat unfortunate for another reason. Usually no account is taken of the

[25] A. V. Hill, *Proc. Roy. Soc.* (*London*), B106, 477 (1930).
[26] R. R. Roepke, *J. Phys. Chem.*, 46, 359 (1942).
[27] D. R. Briggs, *J. Phys. Chem.*, 35, 2914 (1931).
[28] D. R. Briggs, *J. Phys. Chem.*, 36, 367 (1932).

amount of solute bound; in fact, it is difficult to take into account. The binding of the added solute by the colloid can give rise to considerable error. If, for example, the water and solute were bound in the same proportion as that in which they exist in solution, the techniques which involve the solvent properties of water would indicate no bound

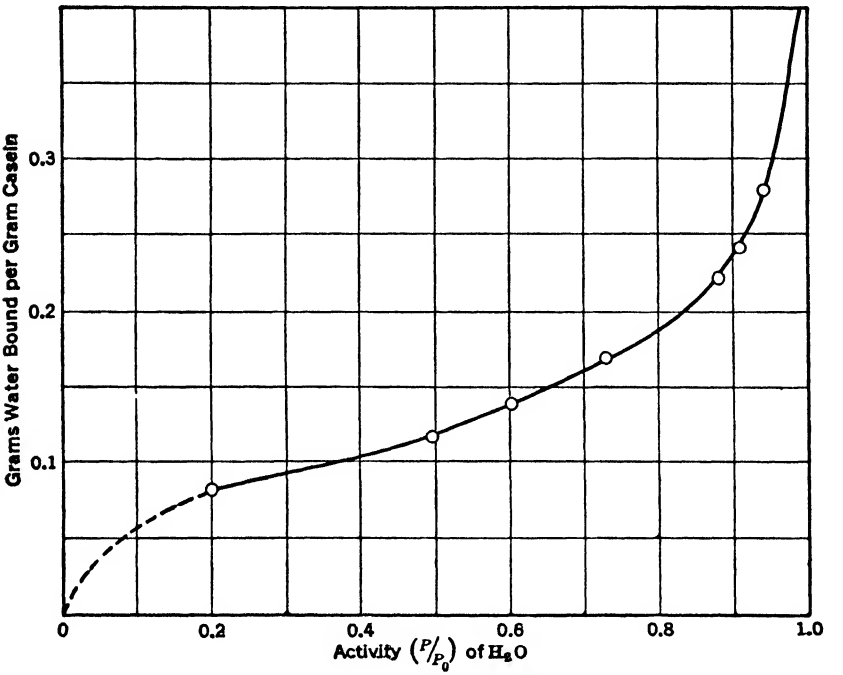


FIG. 5. Activity of water associated with isoelectric casein at 25° C. (Briggs.)

water at all, no matter how much the activity coefficient of the water had been lowered.

There has been a great deal of heated argument on the subject of bound water, and its possible physiological significance has been discussed at length. Some workers have even denied its existence.[29] Others have claimed that it is of little or no physiological significance.[30] Where the truth lies is hard to say.

Since the amount of water bound to a colloid depends primarily on the number and nature of the polar groups present in the colloid, it is to be anticipated that the colloid can be subjected to almost any treat-

[29] D. M. Greenberg and M. M. Greenberg, *J. Gen. Physiol.*, **16**, 559 (1933).

[30] A. V. Hill and P. S. Kupalov, *Proc. Roy. Soc.* (*London*), **B106**, 445 (1930).
A. Grollman, *J. Gen. Physiol.*, **14**, 661 (1931).
K. C. Blanchard, *Cold Spring Harbor Symposia Quant. Biol.*, **8**, 1 (1940).

ment which will leave the polar groups unchanged without materially altering the amount of bound water. The principal hydrophilic colloids of tissue are proteins, and the nature of the tissue proteins cannot be greatly changed and life maintained. In tissue, however, one is not dealing with an isolated protein system. Along with proteins and a great many other components there are electrolytes. Electrolytes cause proteins to swell or to shrink, depending upon the osmotic relations. Protein hydration must be considered part of the general picture of tissue hydration. Hill and other workers were concerned with the errors which might be introduced in physiological calculations by the presence of bound water. This aspect of the problem can be approximately summarized as follows: The proteins of wet tissue make up about 10 per cent of the total weight and the total water about 80 per cent. If the protein binds water to the extent of about 0.35 gram of water per gram of dry protein, then about 4.5 per cent of the total weight of water will be bound. Few physiological calculations are sufficiently exact to be seriously upset by an error of this magnitude.

As a final note of summary in regard to the bound-water problem, it may be stated that bound water may or may not contribute greatly to the understanding of physiology and pathology, but its importance for the understanding of protein reactions is extreme.

Stability of Hydrophilic Suspensoids

Hydrophilic suspensoids owe their stability, in general, to two factors: (1) hydration of the polar groups, which we have just considered; and (2) the ζ-potential. We have already elaborated on the role of the ζ-potential in the stability of hydrophobic suspensoids. The same arguments apply to the stability of hydrophilic suspensoids.

It is easy to see in a qualitative fashion why hydration should increase the stability of a colloidal suspension. The polar groups of the colloidal particles which would be most potent in bringing about cohesion between the particles are blocked by water molecules. The water held on the polar groups of one particle has no greater attraction for water on another particle than for the free water in the system. Accordingly, no energy is gained by the coalescence of two particles, and the system is stable.

The relation between the two stability factors of hydration and the ζ-potential can be diagrammed as Bungenberg de Jong has done [31] and as is shown in Fig. 6.

The protein solution is stable under all conditions except when its ζ-potential is below the critical potential and it is in a dehydrated state.

[31] H. G. Bungenberg de Jong, *Rec. trav. chim.*, **43**, 35 (1924).

Dehydration can be brought about by high salt concentration (such as Na_2SO_4 or $(NH_4)_2SO_4$) or by alcohol, tannin, etc. This is a rather simple picture to which a protein chemist would have some objections. However, it is true enough for our purposes and serves to illustrate the point which we wish to make, i.e., the role of hydration and of the ζ-potential in maintaining stability. A quantitative theory of the influence of hydration on stability has not yet been forthcoming.

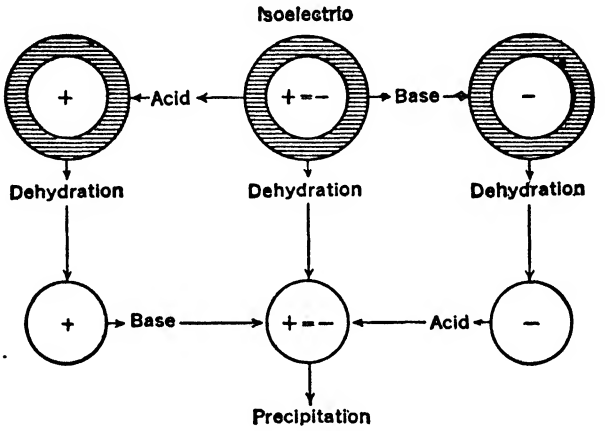


FIG. 6. Relation between the influence of hydration and the ζ-potential on the stability of a protein. Hydration is indicated by a hatched shell around the particle. The protein is originally hydrated and at its isoelectric point. (Bungenberg de Jong.)

In a gold sol, one deals with a distribution of particle sizes. From a molecular point of view, if gold particles are regarded as molecules, there are as many different kinds of molecules as there are gold particles present. To speak of solubility in respect to a gold sol has no meaning. If, however, we consider such proteins as highly purified egg albumin, β-lactoglobulin from milk, pepsin, etc., solubility can be rigorously defined in the same sense as the solubility of NaCl or $BaSO_4$ or any other pure chemical.

As was pointed out in the second chapter, the molar free energy of a solute in equilibrium with some of its solid phase is equal to that of the solid. It follows that the activity of a solute in a saturated solution at a given temperature is a constant. Accordingly, by measuring the solubility as a function of the salt content of the solution, the variation of the activity coefficient of the solute can be determined. This method of study has been widely applied by physical chemists to numerous substances whose solubility is limited. Exactly the same type of study can be done on proteins and other high-molecular-weight substances, and

the variation of the activity coefficient of such substances in solution can be measured as a function of the electrolyte content of the solution. Mellanby,[32] as a result of such studies on the serum globulin, was able to anticipate the principle of ionic strength formulated in 1921 by Lewis and Randall. Mellanby found that dilute solutions of ions with equal valences, whether positive or negative, were equally effective

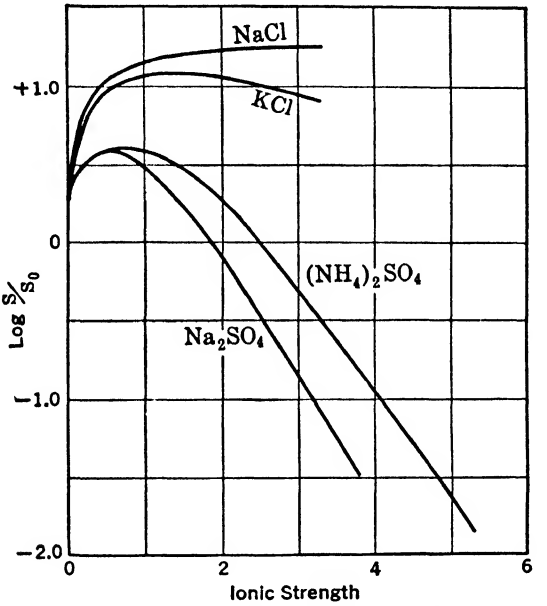


FIG. 7. Solubility of carboxyhemoglobin in aqueous salt solutions. S_0 is the solubility in pure water and S is the solubility in the salt solution. (Data of A. A. Green.)

in increasing the solubility of the globulin, and the efficiencies of ions of different valences were directly proportional to the square of their valences.

The influence of salts on the solubility of proteins is rather complicated. In general, dilute salt solutions increase the solubility of proteins. Indeed, it will be recalled that euglobulins are insoluble in pure water but become soluble in dilute salt solutions. As the salt concentration is increased beyond a certain point, however, protein solubility decreases. In this region a lyotropic series of the ions is usually rather conspicuous. In Fig. 7 is shown [33] the variation of the solubility of carboxyhemoglobin as a function of the ionic strength at 25° C.

[32] J. Mellanby, *J. Physiol.*, **33**, 338 (1905).
[33] E. J. Cohn, *Chem. Revs.*, **19**, 241 (1936).

Scatchard and Kirkwood [34] and later Kirkwood [35] extended the Debye-Hückel theory of interionic attraction to dipolar ions. They concluded that the activity coefficient of a zwitter ion should decrease in the presence of a neutral salt and that the negative logarithm of the activity coefficient of the zwitter ion should be directly proportional to the ionic strength of the solution. Their conclusions apply only to dilute salt solutions.

Since proteins are ampholytes, it is to be anticipated that the solubility of proteins would be very sensitive to *p*H changes. This is true

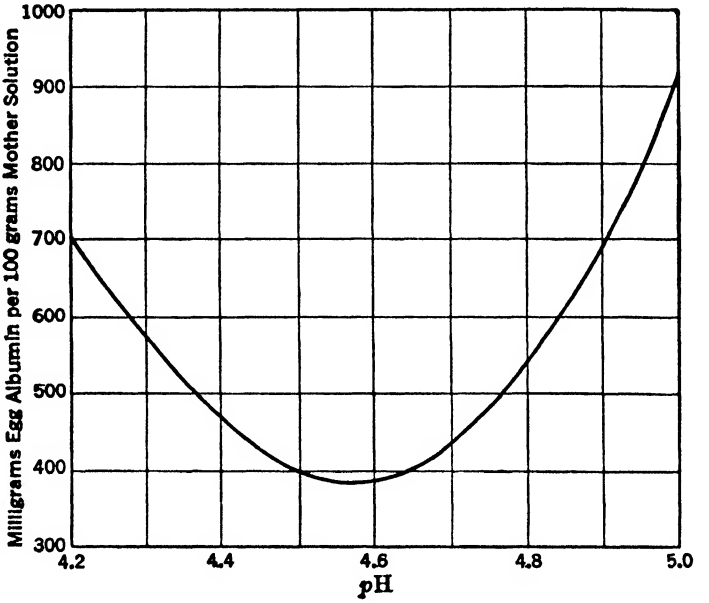


FIG. 8. Solubility of egg albumin in the presence of 25.95 grams of $(NH_4)_2SO_4$ per 100 grams H_2O as a function of *p*H.

to a pronounced degree, and it is found that the solubility is least at the isoelectric point of the protein. Sörensen,[36] in a comprehensive study of the solubility of egg albumin, investigated among other things the influence of *p*H on the solubility of this protein. Figure 8 shows the plot of some of his data.

It is easy to offer a qualitative explanation of the influence of *p*H on the solubility of a protein. The solubility of any substance is the resultant of the balance of the attraction of the solute molecules for each

[34] G. Scatchard and J. G. Kirkwood, *Physik. Z.*, **33**, 297 (1932).
[35] J. G. Kirkwood, *J. Chem. Phys.*, **2**, 351 (1934).
[36] S. P. L. Sörensen, *Compt. rend. trav. lab. Carlsberg*, **12**, 372 (1917).

other, which tends to prevent solution, and the attraction of the solvent molecules for the solute, which tends to promote solution. At the isoelectric point the attraction of the protein molecules for one another is maximal. If the *p*H is shifted away from the isoelectric point, the protein acquires a net charge. This decreases the attraction of the protein molecules for each other, and consequently the solubility of the protein tends to increase.

The solubility of proteins turns out to have a very practical significance; it can be used to test the purity of these substances.[37] The

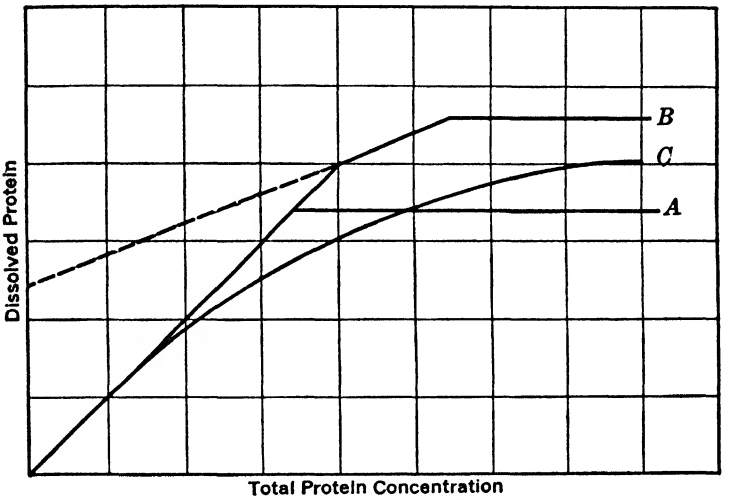


FIG. 9. Diagrammatic solubility curves. (*A*) Solubility curve of a single pure protein. (*B*) Solubility curve of a mixture of two proteins. (*C*) Solubility curve of solid solution of two or more protein components.

organic chemist uses the melting point as a classical means of testing for purity. For obvious reasons this method is not available to the protein chemist. The best substitute for him is a solubility curve.

In order to measure the solubility of a protein, varying amounts of solid protein are placed in a series of test tubes. The test tubes are completely filled with solvent and the tubes stoppered. No air is allowed in the tubes because of danger from surface denaturation of the protein. The tubes are rotated slowly and are stirred by marbles which are placed in them along with the protein. The time of stirring in a constant-temperature bath necessary to attain equilibrium must be determined by experiment. The solutions are then filtered, and the amount of protein in the filtrate is determined. The amount of protein in solution is then

[37] R. M. Herriott, *Chem. Revs.*, **30**, 413 (1942).

plotted against the total amount of protein present. Figure 9 shows diagramatically the results which may be obtained.

It is well known that, for a pure chemical compound, the solubility is independent of the amount of solid phase present. This is shown in Fig. 9*A*. No solid phase is present until the break in the line is reached at which point no more of the added solid dissolves.

Stability in Biology

The stability of hydrophilic colloids has many curious and important biological aspects.

It is well known that, if a protein foreign to an animal is injected into that animal, the animal's serum acquires the property of precipitating the foreign protein. The injected protein is called an antigen, and the protein in the serum which can precipitate the antigen is called an antibody. This is an example of two hydrophilic substances precipitating each other. The antibody molecule exactly "fits" the antigen molecule; polar groups of the antigen and antibody must be able to come in contact in order for the antigen and antibody to unite. As we have seen, ordinary and non-specific colloidal precipitation is brought about by dehydration and decrease in the ζ-potential. In the precipitin reactions we have superimposed upon these two factors the additional factor of the type of mosaic pattern of polar groups on the antigen and antibody molecules. For a discussion of the quantitative relation between antigen and antibody in the precipitin reaction, see papers by Heidelberger and by Kendall.[38] A stimulating discussion of the precipitin reaction is also given by Pauling.[39]

Non-specific agglutination of bacteria is an example of colloidal precipitation. If, for example, the bacteria have a hydrophilic surface as do acid-fast bacteria and the smooth form of a certain strain of a flagellate intestinal bacteria, the electrokinetic potential on the bacteria will be of little importance in determining the stability of suspensions of these bacteria. On the other hand, in the more hydrophobic strains, the ζ-potential will be of extreme importance.[40] The hydrophilic and hydrophobic properties of the surface of bacteria can be determined by means of the Mudd interfacial technique (see Chapter XI).

The Gold Number

The so-called protective action of hydrophilic colloids on the stability of hydrophobic suspensoids has received considerable study and is,

[38] M. Heidelberger, *Chem. Revs.*, **24**, 323 (1939); F. E. Kendall, *Ann. N. Y. Acad. Sci.*, **43**, 85 (1942).

[39] L. Pauling, *J. Am. Chem. Soc.*, **62**, 2643 (1940).

[40] S. Mudd, *Cold Spring Harbor Symposia Quant. Biol.*, **I**, 65 (1933).

among other things, the basis for a medical test of diagnostic value. It has been known for a long time that hydrophobic suspensoids, when treated with various hydrophilic suspensoids, become much more stable towards electrolytes. In 1857 Michael Faraday demonstrated his gold sol to the Royal Society. The sol was placed in the museum of the Society and remains there to the present day. Such longevity in a gold sol is astonishing. Investigation has shown that Faraday had added a protective colloid to his gold sol!

Colloidal silver and colloidal silver oxide have pronounced bactericidal properties, and the "Argyrol" of medicine is a colloidal silver protected by protein split-products.

The protective action of a hydrophilic substance is usually expressed in terms of the gold number. The gold number is defined as that weight of hydrophilic colloid in milligrams which is just insufficient to prevent a change in color from red to violet when 1 cc. of a 10 per cent sodium chloride solution is added to 10 cc. of a red gold sol to which the hydrophilic substance has been added. Gold numbers found by Gortner [41] for several hydrophilic substances are shown in Table 3.

TABLE 3

GOLD NUMBERS OF SOME HYDROPHILIC SUBSTANCES

Substance	Gold number
Gelatin	0.005–0 0125
Egg albumin *	0.08–0.10
Protalbinic acid	0.15–0.20
Lysalbinic acid	0.10–0.125
Gum arabic	0.10–0.125
Dextrin (British gum)	125–150
Soluble starch	10–15

* Highly purified egg albumin has a gold number of about 7. The presence of globulins greatly lowers the gold number.

The gold number of the cerebrospinal fluid was introduced into medicine a number of years ago as a diagnostic aid. Actually more valuable information is obtained by plotting out a curve for the spinal fluid. A series of ten test tubes is prepared, in the first of which is placed 1 cc. of the spinal fluid which has been diluted 1 to 10 with 0.4 per cent NaCl solution. In the next tube is placed 1 cc. of a 1 to 20 dilution, in the third 1 cc. of a 1 to 40 dilution, and so on. To each tube is added 5 cc.

[41] R. A. Gortner, *J. Am. Chem. Soc.*, **42**, 595 (1920).

of a red gold sol, and the mixtures are shaken. The tubes are observed after 24 hours. The several colors are plotted as ordinate against the concentration of the spinal fluid as abscissa. A curve is obtained which is characteristic for certain types of diseases. The colors are numbered

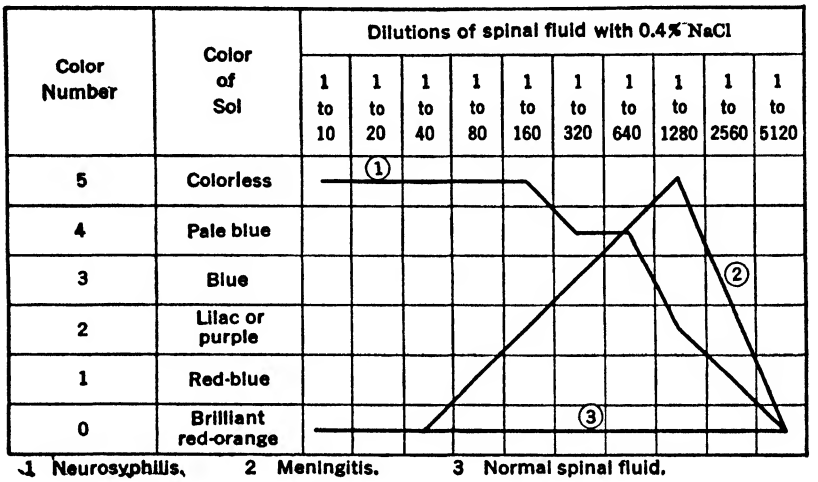


FIG. 10. Characteristic gold sol curves for cerebrospinal fluid.

as follows: cherry red, 0; bluish red, 1; reddish blue, 2; deep blue, 3; pale blue, 4; colorless, indicating complete coagulation, 5. Figure 10 shows several such curves.

If the concentration of the hydrophilic substance is quite low, it is sometimes found that hydrophilic material not only does not protect the hydrophobic colloid but may even sensitize it to electrolytes. Sometimes this sensitizing action can be explained as a flocculation of oppositely charged colloids; however, it has happened that a hydrophobic suspensoid is sensitized by a hydrophilic colloid bearing the same sign of charge. The true explanation of such action seems to lie along lines first suggested by Zsigmondy. The two colloids exhibit an appreciable tendency to unite irrespective of their charge. At low concentrations of the hydrophilic colloid, the hydrophilic colloid will be enveloped by the hydrophobic colloid, whereas, in the presence of an excess of hydrophilic colloid, the hydrophilic colloid

FIG. 11. Diagrammatic representation of the conditions for sensitization and protection of a hydrophobic sol. The circles represent hydrophilic particles; the squares, hydrophobic particles.

envelops the hydrophobic colloid. The first situation leads to sensitization; the second, to protection. This relation is diagrammed in Fig. 11.

EMULSIONS

Emulsions are a special class of colloids. Their size range is considerably greater than that indicated in the first part of this chapter for colloids in general, their particles ranging from 0.5 μ to 10 μ in diameter.

An emulsion is a suspension of one liquid in another liquid, the two liquids being immiscible. Water is usually one of these liquids, and we may distinguish two kinds of emulsions, (1) a water-dispersed-in-oil emulsion and (2) an oil-dispersed-in-water emulsion. By "oil" is meant any non-aqueous liquid which is immiscible with water.

Emulsions may be stabilized with an emulsifier or they may be made without an emulsifier. An emulsifier is a third substance which, when added to the system previous to the emulsification, stabilizes the emulsion when it is formed. There are two classes of emulsifiers, those which tend to stabilize an oil-in-water emulsion and those which tend to stabilize a water-in-oil emulsion.

With or without an emulsifier, emulsions are unstable and cream after a time. Naturally, emulsions not containing an emulsifier tend to be quite unstable although emulsions containing 0.01 per cent of oil suspended in water have been studied. Their stability depends principally on the electrokinetic potential. In fact, it was with experiments on such emulsions that Powis developed his ideas about the critical potentials of colloids.

Emulsifiers

Emulsifiers are of diverse nature. The requirement for an emulsifier is that it form a stable tough film at the interface between oil and water. If the emulsifier is more easily wetted by oil than by water, it will tend to produce a water-in-oil emulsion; if it is more easily wetted by water, it will give rise to an oil-in-water emulsion. The type of emulsion produced by some emulsifiers depends upon its past history; if it has been wet previously by oil, it will yield a water-in-oil emulsion, whereas, if it has been wet by water, an oil-in-water emulsion will result. In the formation of either type of emulsion using almost any emulsifier both oil-in-water and water-in-oil emulsions appear simultaneously; but the type of emulsion favored by the particular emulsifier is the more stable and, after a time, it greatly predominates. Table 4 gives a classification of some common emulsifiers with the type of emulsion favored.

TABLE 4

EMULSIFIERS

Oil-in-water	Water-in-oil
1. Alkali soaps.	1. Magnesium and heavy metal soaps.
2. Naphthenic acids.	2. Naphthenic acids and sulfonic acids if insoluble in water.
3. Sulfonic acids.	
4. Proteins.	3. Proteins which have been treated with phenol or tannin.
5. Agar.	
6. Saponin.	4. Gum dammar.
7. Fine divided powders of basic sulfates of Cu, Ni, Fe, also $CaCO_3$.	5. Powdered carbon, mercuric sulfide, lead sulfide.
8. Lecithin.	6. Cholesterol.

The more finely divided powders are, the better emulsifiers they make, at least up to a certain point. If they are too fine their emulsifying action decreases.

The stability of an emulsified emulsion depends principally on the state of the film formed at the oil-water interface by the emulsifier. Interfacial tension seems to have but 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 Brownian motion of the small-sized emulsion particles brings about coalescence. Emulsions are made by a mechanical agitation of the two phases in the presence of the emulsifier.[42]

Phase Reversal

It is often possible to bring about phase reversal of emulsions stabilized by soaps. Soaps of the alkali metals favor an oil-in-water emulsion. On the other hand, the soaps of the polyvalent metals favor a water-in-oil type. A reversal is noticed if we change the ratio of the "monovalent" soap (Na) to the "divalent" soap (Ca).

Table 5 illustrates the behavior of mixtures of olive oil and water in the presence of varying amounts of sodium hydroxide and calcium chloride.[43]

[42] T. R. Briggs, *J. Phys. Chem.*, **19**, 478 (1915); **24**, 120 (1920).
[43] G. H. A. Clowes, *J. Phys. Chem.*, **20**, 407 (1916).

TABLE 5

PHASE REVERSALS OF EMULSIONS

Cc. 0.1 *M*	Cc. 0.1 *M* $CaCl_2$			
NaOH	0.25	0.5	0.75	1.0
1.0	*R* *	*W–O*	*W–O*	*W–O*
2.0	*O–W*	*R*	*W–O*	*W–O*
3.0	*O–W*	*O–W*	*R*	*W–O*
4.0	*O–W*	*O–W*	*O–W*	*R*

* *R* denotes reversal; *W–O*, the water-in-oil type of emulsion; *O–W*, the oil-in-water type of emulsion.

During reversal of an emulsion, often one type of emulsion persists to some extent while the other is being formed, so that compound emulsions result. An oil droplet which itself is part of an oil-in-water emulsion may contain a water-in-oil emulsion. Thus Seifriz [44] was able to prepare five emulsions in one—as he put it, like a nest of Chinese boxes.

Emulsions in Biology

Animals and plants are often faced with the necessity of dealing with oil or fat in their economy. Oil and fat being insoluble in water, living tissue has had to evolve special methods of transportation. Fat is transported in the blood as fine particles called the chylomicron emulsion by Gage and Fish.[45] These workers conducted an extremely interesting series of studies on the absorption and transportation of fat in blood. The chylomicrons range from 0.5 to 1 μ in diameter; they increase greatly in number and slightly in size during digestion and absorption of fat.

A medical problem of serious proportions is presented by fat embolism. Emboli come from fractured bones and occasionally as a result of surgical operations. Frequently they are fatal. There is evidence to believe that fat embolism may arise in some instances from the loss of the emulsifying power of the blood, resulting in the breaking of the chylomicron emulsion. Davis and Goodchild [46] have commented on this phase of the problem. It has been reported that, if gum acacia is injected into the blood stream of a dog, large quantities of oil also may be injected into the blood stream without embolism; the gum acacia acts as an emulsifier for the oil.

[44] Wm. Seifriz, *Am. J. Physiol.*, **66**, 124 (1923).
[45] S. H. Gage and P. A. Fish, *Am. J. Anat.*, **34**, 1 (1924).
[46] H. I. Davis and C. G. Goodchild, *J. Chem. Education*, **13**, 478 (1936).

Chapter XIII

VISCOSITY AND THE FLOW OF LIQUIDS

The study of the viscosity of colloidal solutions has led to much of value. It is important to know what information viscosity can give us and what it cannot. The methods available for its measurement are remarkably simple; almost too simple, because they are apt to be abused.

Viscosity is the resistance experienced by one portion of a liquid moving over another portion of the same liquid. The unit of viscosity, the poise, is defined as that viscosity such that unit force per unit area is required to cause two parallel liquid surfaces of unit area and unit distance apart to slide past one another with unit velocity, i.e.,

$$\text{Force} = \frac{\eta A u}{d} \qquad 1$$

where A is the area of the liquid surfaces at a distance d apart, u is the velocity difference in centimeters per second between the two liquid planes, and η is the coefficient of viscosity. The dimensions of viscosity are, therefore, $Mt^{-1}L^{-1}$.

Viscosity may also be defined in terms of the energy required to maintain a constant velocity gradient. The flow of a liquid is characterized by the fact that its maintenance requires the expenditure of energy at a rate which is proportional to the volume of the liquid and to the square of the velocity gradient. The constant of proportionality varies from one liquid to another and is called the coefficient of viscosity. This coefficient is identical, both numerically and dimensionally, with the coefficient of viscosity defined above in equation 1.

Occasionally and for special purposes another kind of viscosity is used. This is known as the kinematic viscosity and is equal to η/ρ, where ρ is the density of the liquid. The dimensions of kinematic viscosity are $t^{-1}L^2$.

Theory of Flow of Pure Liquids

Eyring[1] has proposed an ingenious molecular theory of the flow of liquids. This theory regards viscous flow as related to vaporization.

[1] H. Eyring, *J. Chem. Phys.*, **4**, 283 (1936); *Advances in Colloid Science*, Interscience Publishers, Inc., New York, 1942.

In order for a liquid to flow, holes must be available for the molecules to move into. The production of such holes is akin to vaporization; accordingly, there should exist a relation between the energy of activation of the viscosity of a liquid and the heat of vaporization of the liquid. The flow of a liquid is regarded as a rate process, and from its dependence on temperature the energy of activation is calculated in the usual manner (Chapter III). Experimental results show that such a relation does indeed exist.

TABLE 1

VISCOSITIES OF SOME COMMON LIQUIDS IN POISES

Liquid	Viscosity	Temperature in °C.
Benzene	0.00763	10
	0.00654	20
	0.00567	30
n-Butanol	0.02948	20
	0.01782	40
Ethanol	0.01752	10
	0.01716	20
	0.01681	30
Glycerol	25.18	8.1
	13.87	14.3
	8.30	20.3
	4.94	26.5
Heptane	0.00416	20
	0.00341	40
Methanol	0.00596	20
	0.00456	40
Water	0.015188	5
	0.013077	10
	0.011404	15
	0.010050	20
	0.010000	20.20
	0.008937	25
	0.008007	30
	0.007225	35
	0.006560	40

With liquids containing molecules capable of hydrogen-bond formation, the energy of activation of viscosity decreases markedly with increasing temperature. The viscosity activation energy of these substances consists not only of the energy required to break what may be termed "physical bonds," but also of the energy required to break the hydrogen bonds. As the temperature is increased, the number of hydrogen bonds

in a liquid diminishes and, accordingly, the energy of activation of viscosity also decreases. It is the presence of a network of hydrogen bonds which accounts for the very high viscosities of such compounds as glycerol and water.

In Table 1 are shown the coefficients of viscosity of several common liquids.

Flow of Liquids Through Capillaries

The fundamental law governing the flow of liquids through capillaries is that of Poiseuille, which, in the integrated form, is expressed as

$$V = \frac{\pi r^4 P}{8\eta l} \qquad 2$$

where V is the volume of liquid flowing in cubic centimeters per second, r is the radius of the capillary in centimeters, P is the difference in pressure between the two ends of the capillary and is expressed in dynes per square centimeter, η is the coefficient of viscosity in poises, and l is 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. The critical pressure depends upon the geometry of the system as well as upon the density and viscosity of the liquid. The ratio of the inertial to the frictional force is $(\rho/\eta)ua$, where u is the fluid velocity and a is a characteristic linear dimension of the system. For capillaries, a is taken as the radius of the capillary. The expression $(\rho/\eta)ua$ is called the Reynolds number[2] and is denoted by R. It is dimensionless and is independent of the units used. For a capillary and with the use of equation 2 the Reynolds number is

$$R = \frac{\rho r^3 P}{8\eta^2 l} \qquad 3$$

If R for a capillary is equal to or exceeds about 1,000, turbulent flow will appear and a departure from Poiseuille's law will be noted. For other systems the Reynolds numbers are different.

Methods

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

[2] O. Reynolds, *Trans. Roy. Soc.* (*London*), A174, 935 (1883).

under the influence of gravity is determined. Such an apparatus is diagrammed in Fig. 1.

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. 1). 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

$$\eta = C\rho t - f\left(\frac{\rho V}{lt}\right) \qquad 4$$

where C is a constant which involves the length and radius of the capillary. 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

$$\eta_r = \frac{\eta}{\eta_0} = \frac{\rho t}{\rho_0 t_0} \qquad 5$$

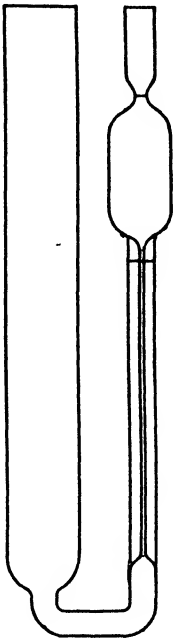


FIG. 1. Ostwald viscometer.

where η is the coefficient of viscosity of the liquid whose viscosity we wish to measure, η_0 is the coefficient of viscosity of the standard liquid, ρ_0 is the density of the standard liquid, and t_0 is 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 viscosity is known. If the two measured viscosities bear the correct ratio to each other, and if the unknown viscosity has approximately 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 convenient to be able to calculate beforehand the dimensions of the capillary which will be acceptable for a given series of measurements. Directions for doing this have been given by Grüneisen [3] and have been clearly reviewed by Bungenberg de Jong.[4]

The mean linear velocity of flow of a liquid in a capillary is evidently

$$u = \frac{V}{\pi r^2} \qquad 6$$

Substituting the value of V from equation 6 into the Poiseuille equation 1, and remembering that $hg\rho$ equals P, where h is the difference in liquid level in a viscometer and g is the acceleration of gravity, we have

$$u = \frac{hg\rho r^2}{8\eta l} \qquad 7$$

Grüneisen found that the linear velocity must not exceed certain critical values if the departure from Poiseuille's law is not to be greater than a certain per cent. Grüneisen based his calculations on water at a temperature of 10° C. The author has recalculated his results to apply to water at 25° C. Table 2 gives the critical velocities of water at 25° C. for the indicated capillary dimensions. In order to calculate the critical velocities from the data in Table 2 for any other temperature and any other liquid, the velocities in Table 2 must be multiplied by ρ'/η' where ρ' is the density relative to water at 25° C. and η' is the viscosity relative to water at 25° C.

To illustrate the use to which the results in Table 2 can be put, consider water at 25° C. and a capillary 10 cm. long and 0.50 mm. in diameter. Substituting the appropriate values in equation 7, we find

$$u = 3.4h \qquad 8$$

At h equal to 10 cm., the linear velocity will be 34.4 cm. per second. Referring to Table 2, we find that such a capillary has a critical velocity of 27.4 cm. per second for a flow error not exceeding 1 per cent. Accordingly, we conclude that in our case the flow error will be slightly in excess of 1 per cent. One has a certain margin of error in applying Table 2. As the value of h constantly diminishes as the liquid flows through the viscometer, the flow error will be somewhat smaller than that calculated. Also, the Ostwald viscometer is used to compare the flow of a standard liquid with that of another liquid, and the flow errors from the two

[3] E. Grüneisen, *Wiss. Abhandl. physik.-tech. Reichsanstalt*, **4**, 153 (1904).

[4] H. G. Bungenberg de Jong, *First Report on Viscosity and Plasticity*, Nordemann Publishing Co., Inc., New York, 1939.

liquids tend to cancel each other. In designing a viscometer, however, it is better to forget about the margin of error and design the instrument so that its calculated flow error does not exceed the other errors involved in the viscosity determination. Actually, kinetic flow errors begin to be appreciable at much lower pressures than the pressures at which turbulent flow begins (equation 3).

TABLE 2

CRITICAL VELOCITIES OF WATER AT 25° C. AND THE PERCENTAGE FLOW ERROR ASSOCIATED WITH THESE CRITICAL VELOCITIES

	Diameter, millimeters	Critical velocity for 0.1% error, centimeters per second	Critical velocity for 1.0% error, centimeters per second	Critical velocity for 5.0% error, centimeters per second
	0.40	8.9	44.5	124
	0.45	6.5	34.3	92
l = 10 cm.	0.50	5.0	27.4	75
	0.55	3.8	22.2	61
	0.60	3.0	17.3	50
	0.40	29.8	113	226
	0.45	21.6	88	179
l = 18 cm.	0.50	16.5	67	147
	0.55	12.8	53.5	122
	0.60	10.0	43.5	104

In addition to the kinetic correction mentioned above, the Ostwald viscometer is subject to three other errors: (1) drainage error, (2) working volume correction, and (3) surface-tension effects. The ideal design of a viscometer strikes a compromise with all these disturbing factors.

Jones and Stauffer [5] have shown that the drainage error is negligible except for very viscous liquids or for viscometers with very short times of outflow.

The surface-tension correction may be important for liquids whose surface tension differs greatly from that of the calibrating liquid.[6] Bull [7] found that solutions of egg albumin showed some very anomalous flow relations which he attributed to surface films of this protein. The working volume correction refers to error inherent in placing a precise volume of liquid in the viscometer. Ordinarily the error involved here is much smaller than the other types of errors.

[5] G. Jones and R. E. Stauffer, *J. Am. Chem. Soc.*, **59**, 1630 (1937).
[6] G. Jones and H. J. Fornwalt, *J. Am. Chem. Soc.*, **60**, 1683 (1938).
[7] H. B. Bull, *J. Biol. Chem.*, **133**, 39 (1940).

Anyone who plans to use the Ostwald viscometer for accurate work should consult the excellent series of papers by Jones and co-workers in the *Journal of the American Chemical Society*. The paper by Bungenberg de Jong [4] to which reference has already been made is also helpful. In conclusion of this discussion, 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. Björnstahl and Snellman [8] 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 [9] 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. 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 [10] formulated his well-known equation for the viscosity of a suspension of spherical particles. The equation is

$$\eta = \eta_0 \left[\frac{1 + 0.5\phi}{(1 - \phi)^2} \right] \qquad 9$$

[8] Y. Björnstahl and O. Snellman, *Kolloid-Z.*, **86**, 223 (1939).
[9] W. R. Hess, *Kolloid-Z.*, **27**, 154 (1920).
[10] A. Einstein, *Ann. Physik*, **19**, 289 (1906); **34**, 591 (1911).

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

Guth [11] has outlined the assumptions upon which the Einstein equation is based. They are as follows: (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 so small that the Reynolds number of the particle is much less than unity; that is

$$R\rho = \frac{\rho}{\eta_0} au \ll 1 \qquad 10$$

where a is the radius of the particle, u is the linear velocity of streaming of the medium, ρ is the density of the particle, and η_0 is the viscosity of the medium. (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

$$\frac{\eta}{\eta_0} = 1 + 2.5\phi + 4\phi^2 + 5.5\phi^3 \qquad 11$$

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

$$\frac{\eta}{\eta_0} = 1 + 2.5\phi \qquad 12$$

which is the form in which one frequently sees the equation written. Eirich, Bunzl, and Margaretha [12] 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 [13] have considered the interactions of the flow lines around neighboring particles. The attractive and repulsive forces

[11] E. Guth, *Kolloid-Z.*, **74**, 147 (1936).

[12] F. Eirich, M. Bunzl, and H. Margaretha, *Kolloid-Z.*, **74**, 276 (1936).

[13] E. Guth and O. Gold, *Phys. Revs.*, **53**, 322 (1938).

between particles were considered insignificant. Thus, with the exception of item 6 listed above, their assumptions were identical with those of Einstein. They obtained

$$\eta = \eta_0(1 + 2.5\phi + 14.1\phi^2 + \ldots) \qquad 13$$

This equation, tested [14] 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 equation 13 as

$$\frac{\eta}{\eta_0} - 1 = 2.5\phi + 14.1\phi^2 \qquad 14$$

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 14 by ϕ, we have

$$\frac{\eta_{sp}}{\phi} = 2.5 + 14.1\phi \qquad 15$$

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 should 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 should 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 [15] 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

$$\left(\frac{\eta_{sp}}{\phi}\right)_{\phi \to 0} = \frac{\ln \eta_r}{\phi} \qquad 16$$

Correspondingly, the weight intrinsic viscosity is given by

$$\left(\frac{\eta_{sp}}{C}\right)_{C \to 0} = \frac{\ln \eta_r}{C} \qquad 17$$

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

[14] H. F. Smith, *Rubber Chem. Tech.*, **15**, 301 (1942).
See also E. Guth, *Proc. 5th Intern. Congr. Applied Mechanics*, 1938.

[15] E. O. Kraemer, *The Ultracentrifuge*, by Svedberg and Pedersen, Oxford University Press, New York, 1940.

Particle Asymmetry and Viscosity

It was realized rather 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 particle 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. Jeffery [16] was the first to attempt a theoretical approach to the problem. His equations, however, were left in an indeterminate form and cannot be subjected to experimental test. Later, equations relating viscosity and particle asymmetry were derived by Eisenschitz,[17] by Kuhn,[18] by Guth,[19] by Simha,[20] by Huggins,[21] and by Burgers.[22]

Eisenschitz considered the viscosities of prolate ellipsoids of revolution with and without Brownian motion. The equation of Kuhn was derived on the assumption that the suspended particle could be represented by a rigid chain of balls separated from each other by four times the radius of the balls. Such a model is certainly unrealistic for most suspended particles and it would be surprising if the equation based upon such a model yielded valid results. Huggins considered the chemical bond distances and valence angles and derived equations for kinked as well as unkinked chains. Although his equations seem to be in reasonable agreement with experiment, they have a limited application to the general problem of the relation between particle asymmetry and viscosity. Both Guth and Simha considered the influence of the asymmetries of prolate and of oblate ellipsoids of revolution on the viscosity of suspensions of such ellipsoids. Burgers considered the influence of the asymmetry of prolate ellipsoids of revolution on the viscosity. In the next chapter, which deals with diffusion, we shall apply an approximate test to some of these equations. We can anticipate that discussion with the statement that Eisenschitz' equation appears to be entirely

[16] G. B. Jeffery, *Proc. Roy. Soc.* (*London*), **A102**, 161 (1922).
[17] R. Eisenschitz, *Z. physik. Chem.*, **A163**, 133 (1933).
[18] W. Kuhn, *Kolloid-Z.*, **76**, 258 (1936).
[19] E. Guth, *Kolloid-Z.*, **75**, 15 (1936).
[20] R. Simha, *J. Phys. Chem.*, **44**, 25 (1940).
[21] M. L. Huggins, *J. Phys. Chem.*, **42**, 911 (1938); **43**, 4 (1939).
[22] J. M. Burgers, *Second Report on Viscosity and Plasticity*, Nordemann Press, New York, 1938.

inadequate and that Simha's equation, though far from exact, gives the best results. The other equations lie between these extremes.

In a situation such as this, it is the feeling of the author that an ounce of experiment is worth a pound of theory.

Staudinger [23] approached the problem of asymmetry from an experimental point of view. Using a series of long-chain hydrocarbons, alcohols, and acids dissolved in such inert solvents as benzene or carbon tetrachloride, he found the specific viscosity of the solutions to be a function of the number of groups of "submolecules" contained in the chain. For a paraffin hydrocarbon, the submolecule is the CH_2 group; for cellulose, it is the glucose anhydride residue. He proposes the equation

$$\eta_{sp} = KnC \qquad 18$$

where K is a constant for a given series of compounds, n is the number of submolecules per molecule, and C is the concentration of the solution in submolecules per liter. For example, for a hydrocarbon, C would be equal to the concentration in grams per liter divided by 14, the molecular weight of the submolecule. It is of interest that an equation in this form has been derived by Huggins [20] for hydrocarbon chains kinked at random. The Staudinger relation has been extensively investigated by numerous other workers and has proved of great value in many technical and industrial problems. This approach, however valuable it may be to industry, still does not contribute greatly to the understanding or even description of the general problem of the relation between asymmetry of colloidal particles and viscosity.

Eirich, Margaretha, and Bunzl [24] studied the influence of the asymmetry of microscopic silk fibers on the viscosities of suspension fibers. Though their experimental conditions are far from ideal, these are the only model experiments we have on hand. Probably the most serious objection to their work is that the silk particles were so massive that the influence of the Brownian motion was not sufficient to prevent considerable orientation of the particles in the streamlines. Careful examination of their data leaves one with the impression that the results are only of a semiquantitative nature. They point out that, as is true of any continuous variable, the relation between viscosity and concentration can be represented by a power series, and, in general,

$$\eta_{sp} = A\phi + B\phi^2 + C\phi^3 + \cdots \qquad 19$$

where A, B, C, etc., are coefficients which depend upon the asymmetries of the particles. We have seen, for example, that for spherical particles

[23] H. Staudinger, *Ber.*, **68**, 2320 (1935).

[24] F. Eirich, H. Margaretha, and M. Bunzl, *Kolloid-Z.*, **75**, 20 (1936).

A is 2.5 and B is 14.1 (Guth). In order to evaluate the coefficients, η_{sp}/ϕ is plotted against ϕ for each asymmetry. The intercept on the η_{sp}/ϕ-axis is equal to A, and the slope of the line is equal to B. It appears to the author that the viscosities of suspensions whose asymmetries are 11 to 1 and 17 to 1 are the most reliable. Figure 2 shows a plot of the constants A and B against the corresponding asymmetries.

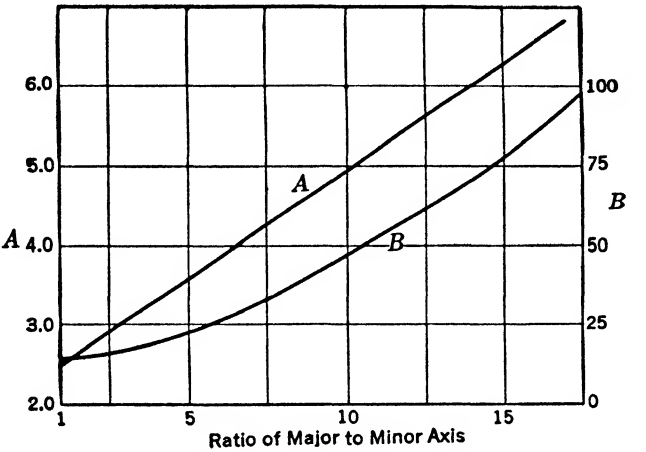


FIG. 2. Plot of the first and second coefficients of the series relating concentration and viscosity against the asymmetries of the suspended particles (data of Eirich, Margaretha, and Bunzl).

By means of Fig. 2 and equation 19, it should be possible to formulate the viscosity as a function of concentration for any asymmetry within limits of the graph. The results in Fig. 2, however, should not be taken too seriously, although they do give an approximate idea of what is to be expected.

Viscosities of Protein Solutions

The question of asymmetry of particles is particularly vital to the protein chemist. There is no way at the present time of obtaining unambiguous values for the asymmetries of protein molecules. From the above discussion, it is clear that a relation should exist between the asymmetries of protein molecules and the viscosities of their solutions. This realization has prompted considerable work on the viscosity of dilute protein solutions. The problem is not a simple one. In addition to the innate difficulty of finding a relation between particle asymmetry and viscosity there are two other difficulties. No one knows what form the asymmetries of protein molecules assume; i.e., are they cylinders, disks, prisms, prolate ellipsoids, oblate ellipsoids, etc.? Also unknown is

the influence of hydration. In order to make any sense out of viscosity measurements, we must know the volume occupied by the dispersed phase. It is an easy matter to measure the volume occupied by a given weight of dry protein, but, when dissolved in water, proteins take up an unknown amount of water and, accordingly, occupy a larger volume dissolved than in their dry state. The manner in which hydration increases the effective volume is also unknown. For example, is the swelling uniform over the whole molecule, or does it center about the major or about the minor axis of asymmetry? We have very inconclusive answers to these questions. After we have discussed diffusion, we shall return to these matters again.

Treffers [25] reported that the viscosity of a protein solution could best be expressed in terms of fluidity. Fluidity is the reciprocal of viscosity. The relation between fluidity and concentration, according to Treffers, is

$$f_r = \frac{f}{f_0} = (1 - KC) \qquad 20$$

where f_r is the relative fluidity and is equal to $1/\eta_r$. f is evidently equal to $1/\eta$, and f_0 is equal to $1/\eta_0$. C is the concentration of the protein in grams per 100 cc. of solution. Treffers expressed concentration in grams per 100 cc. solvent. From a prior reasoning one would anticipate that a better relation would be obtained by expressing concentration in terms of the volume of the solution and not in terms of the volume of the solvent. This is, indeed, found to be true. K is a constant which is dependent on the nature of the protein. Bingham and Roepke [26] have recently found this relation to hold for solutions of fibrinogen. Equation 20 is evidently equivalent to the expression

$$\eta_r = \frac{1}{(1 - KC)} \qquad 21$$

If we expand equation 21 into a power series, we have

$$\eta_r = 1 + KC + 2K^2C^2 + 3K^3C^3 + \cdots \qquad 22$$

This is a rather important and interesting conclusion. If we know the first coefficient of the series, we can immediately write down the coefficients for any other terms. Since ϕ is proportional to C and since for a spherical particle the volume intrinsic viscosity is equal to 2.5, we have the following series for a suspension of spherical particles:

$$\eta_r = 1 + 2.5\phi + 12.5\phi^2 + 46.9\phi^3 + \cdots \qquad 23$$

[25] H. P. Treffers, *J. Am. Chem. Soc.*, **62**, 1405 (1940).

[26] E. C. Bingham and R. R. Roepke, *J. Am. Chem. Soc.*, **64**, 1204 (1942).

Compare this series with that previously described by Guth (equation 13).

This approach to the problem emphasizes the significance of the first coefficient in describing the viscosity of any suspension. In it is included the influence of both hydration and asymmetry.

It is of particular interest that the coefficient K enters in equation 22 in exactly the same fashion as C. In view of this, we conclude that asymmetry has the effect of increasing the apparent concentration of the suspension. It is not significant that viscosity can be expressed in terms of a power series but it is significant that it can be expressed in terms of this particular series.

It should be recalled that the first coefficient of the series (K) is equal to the weight intrinsic viscosity, if weight concentrations have been used, or is equal to the volume intrinsic viscosity, if volume concentrations have been employed.

Table 3 shows the weight intrinsic viscosities of a number of well-defined proteins.

TABLE 3

WEIGHT INTRINSIC VISCOSITIES OF SOME WELL-DEFINED PROTEINS

Protein	Intrinsic viscosity	Reference
Pepsin (most soluble)	3.88	27
Egg albumin	3.90	7
CO-hemoglobin	4.15	28
Serum albumin (horse)		
McMeekin fraction	3.72	27
Kekwick fraction	4.27	27
Helix hemocyanin *p*H 6.67	5.28	28
β-Lactoglobulin	4.08	27
Homarus hemocyanin	4.79	28
Seroglycoid (horse)	5.05	27
Amandin	5.24	28
Octopus hemocyanin	6.77	28
Thyroglobulin	7.42	28
Helix hemocyanin	14.05	28
Fibrinogen	22.95	26

Electroviscous Effect

Falkenhagen and Dole [29] found that the viscosity of dilute electrolyte solutions could be expressed by the relation

$$\eta = \eta_0(1 + A\sqrt{C}) \qquad 24$$

[27] H. Neurath, G. R. Cooper, and J. O. Erickson, *J. Biol. Chem.*, **138**, 411 (1941).
[28] A. Polson, *Kolloid-Z.*, **88**, 51 (1939).
[29] H. Falkenhagen and M. Dole, *Physik. Z.*, **30**, 611 (1929).

where C is the equivalent concentration of the electrolyte and A is a constant involving the viscosity of the liquid, the dielectric constant, the temperature, and the valence and the mobility of the ions. This equation was derived on the assumption that the increase in viscosity arises from the work required to disrupt the ionic atmosphere around the ions and to move the positive and negative ions relative to each other.

The electroviscous effect due to charged colloidal particles was first treated by Smoluchowski,[30] who, shortly before his death, published an equation without derivation which was supposed to describe the electroviscous effect. Later Krasny-Ergen [31] reconsidered the question and derived an equation which is substantially the same as Smoluchowski's differing from it only in the magnitude of a numerical coefficient. The Krasny-Ergen equation is

$$\eta_r = 1 + 2.5\phi\left[1 + \frac{3}{2}\frac{1}{\kappa\eta_0 a^2}\left(\frac{D(\Gamma_i - \Gamma_a)}{2\pi}\right)^2\right] \qquad 25$$

where κ is the specific conductance of the dispersion medium, η_0 is its viscosity, D is the dielectric constant of the medium and a is the radius of the spherical particle. According to Krasny-Ergen, Γ_i is the potential difference between the surface of the particle and some point in the liquid where the charge density is zero, and Γ_a is the potential difference between the surface of the particle and the interior of the particle. Thus $\Gamma_i - \Gamma_a$ is the total difference in potential between the interior of the particle and the interior of the liquid. If this is true, $\Gamma_i - \Gamma_a$ cannot be replaced by ζ, which is the potential across the diffuse double layer only. This appears to be an error of definition, however, and it is the custom to assume that $\Gamma_i - \Gamma_a$ is the potential across the double layer and thus equal to the ζ-potential. The Krasny-Ergen equation involves all the assumptions of the Einstein equation in addition to the following: (1) the charge on the small sphere is uniformly distributed over the surface of the spheres; (2) the particles are non-conducting; (3) the thickness of the double layer is thin in comparison with the radius of the particle; (4) the streaming potential set up by the motion of the particles in respect to the liquid must exert no back pressure on the liquid (no electroosmotic effects).

It is very difficult to test the equation of Krasny-Ergen. It is easy enough to show that the viscosity of solutions of such substances as gum arabic and of various proteins decreases upon the addition of electrolytes when the electrolyte concentration is low. For example,

[30] M. von Smoluchowski, *Kolloid-Z.*, **18**, 194 (1916).
[31] W. Krasny-Ergen, *Kolloid-Z.*, **74**, 172 (1936).

Briggs [32] found, for sodium gum arabic, that, when he plotted $1/\eta_{sp}$ against $\zeta^2/\kappa(\eta - \eta_0)$, straight lines were obtained for the various concentrations of gum. Such straight lines are predicted by the equation of Krasny-Ergen. In all such systems, however, one is faced with the osmotic swelling of the hydrophilic colloid resulting from a Donnan equilibrium. Qualitatively, all electroviscous measurements in the literature can as readily be interpreted in terms of swelling as in terms of the electroviscous effect. One source of confusion in these tests is the failure to realize that equation 25, even if valid, applies only to very dilute suspensions since all higher terms of concentration than the first have been neglected.

Bull [33] studied the viscosity of egg albumin solutions and measured the specific conductance of the medium as well as the electrophoretic mobility of molecules of egg albumin. The approximate radius of the molecule of egg albumin in solution is known, so that all the necessary information is available to test the Krasny-Ergen equation. The measured viscosities were from 0.7 to 10 per cent of those predicted by the Krasny-Ergen equation. It is true that this still is not an unambiguous test of the equation, since the thickness of the double layer was of the order of magnitude of the size of the albumin molecule. Whether the equation is valid or not has not been settled but the author is deeply suspicious of it.

The interest of the experimenter usually lies not in a study of the electroviscous effect as such but only as a source of ambiguity; he wishes to know how to eliminate it. The most direct way of avoiding an electroviscous effect is to increase the concentration of electrolyte until a point is reached at which the measured specific viscosity is independent of salt concentration. In doing this the hydration of the colloid has been changed, but there is no other way out.

Viscosity in Biological Systems

Wherever there is a transfer of liquids in biological systems, thought must be given to the viscosity of those liquids. A classic example of such a transfer is the motion of blood in the arteries, veins, and capillaries of higher animals. Analysis shows that the velocity of flow of blood through the great aorta is so great at times that the flow approaches a condition of turbulence.

Nurnberger [34] has made a rather complete study of the viscosity of

[32] D. R. Briggs, *J. Phys. Chem.*, **45**, 866 (1941).

[33] H. B. Bull, *Trans. Faraday Soc.*, **36**, 80 (1940).

[34] J. I. Nurnberger, thesis submitted to the Graduate School of Northwestern University, Chicago, Illinois, 1942.

blood in a number of pathological systems. The red cells naturally contribute greatly to the viscosity of blood. At 37° C. the viscosity of whole human blood varies between 3 and 4 centipoises and that of the plasma is about 1 centipoise. Nurnberger used an Ostwald viscometer in his measurements but found it expedient to apply an external pressure to force the blood through the capillary. He found a striking dependence of blood viscosity on the rate of flow, the viscosity decreasing with increasing rate of flow. It is probable that this decrease was due to an orientation of the red cells in the flow gradient at the higher velocities of flow.

A streaming motion is frequently observed in protoplasm. A very elegant example of protoplasmic streaming is shown by the myxomycete *Physarum polycephalum*. Here streaming is accompanied by changes in the contour of the plasmodium as a whole. Kamiya [35] has devised a simple and effective apparatus for studying this flow. Two masses of protoplasm connected by an unbroken strand of protoplasm are placed in separate air-tight compartments. The difference in air pressure between the two compartments, which is just sufficient to prevent the flow of the protoplasm, is determined. This counter pressure is then taken as a measure of the force which is responsible for the streaming of the protoplasm. The counter pressure varies in a rhythmic manner with time, usually between the limits +10 cm. and −10 cm. of water pressure. The length of the rhythm varied, but the cycle is of the order of two minutes. Kamiya concluded that the peculiar wave forms he observed in the counter pressure with time were due to interference of several rhythmic processes occurring simultaneously. It is not known where or how the energy is derived for protoplasmic streaming. It can be surmised that the streaming fulfills the same function as the blood flow in the higher animals.

It is evident that the measurement of the viscosity of cell interiors presents difficulties. In the first place, the viscosity of a cell interior is not uniform throughout; furthermore, its value is greatly dependent on the physical and chemical changes. Any technique which is used to measure this viscosity is very likely to change the viscosity. One method is to study the rate of fall of starch grains in the interior of suitable cells. Stokes' law is then applied.[36] Stokes' law is expressed by the equation

$$V = \frac{2g(\rho' - \rho)a^2}{9\eta} \qquad 26$$

[35] N. Kamiya, *The Structure of Protoplasm*, Wm. Seifriz, Iowa State College Press, Ames, Iowa, 1942.

[36] L. V. Heilbrunn, *An Outline of General Physiology*, W. B. Saunders Co., Philadelphia, Pa., 1937.

where V is the velocity of fall of a spherical particle, g is the acceleration of gravity, ρ' is the density of the spherical particle, ρ is the density of medium, a is the radius of the particle, and η is the coefficient of viscosity of the medium. It was found that in some cells the rate of fall of the starch grains was only 8 times slower in the cell than in water. In other cells the ratio was somewhat higher, and in injured cells, the grains, at times, failed to fall.

Other cells not containing starch grains can be centrifuged and the rate of fall of smaller inclusions studied. Many cells can stand high centrifugal forces without injury. When the speed of centrifugation is varied the shearing force is varied and the viscosity is studied as a function of shear. As far as could be determined with this rather crude technique, the viscosity of cell interiors is independent of rate of shear.

Stream Birefringence

As is well known, ordinary light vibrates in all planes. If such light is passed through a calcite crystal in the proper direction, the beam of light is broken into two beams. This can be demonstrated by viewing an object through a calcite crystal. Upon rotation of the calcite crystal, a double vision is observed to appear. It will be discovered that there is one direction in which light can be passed through a calcite crystal without being broken into two beams. This direction corresponds to the optic axis of the crystal. Calcite, like a number of other crystals, is anisotropic, the index of refraction being different in two directions through the crystal. Both beams of light emerging from such a crystal are plane polarized, and the planes of vibration of the two beams are at right angles to each other. The magnitude of double refraction is measured by the difference between the index of refraction of the ordinary ray (n_o) and that for the extraordinary ray (n_e). Light whose plane of vibration is parallel to the optic axis is called the extraordinary ray; and light whose plane of vibration is perpendicular to it is known as the ordinary ray. If $n_e > n_o$, the double refraction is positive; if $n_e < n_o$ it is negative. The double refraction of muscle as well as of various other materials is negative.

Double refraction of an object can be detected by placing it between two crossed Nicol prisms. Light passing through the first Nicol (the polarizer) is plane polarized. When it enters the anisotropic object the plane-polarized beam is broken into two rays, one vibrating parallel to the optic axis of the object and the other at right angles to it. If white light has been used, the field, as viewed through the second Nicol (the analyzer), is seen to be beautifully colored and there will be no position of the second Nicol at which all the light can be extinguished. Since the

velocity of light is inversely proportional to the index of refraction and since each wavelength of light has its own index of refraction through a medium, the velocities of the various wavelengths in the two beams of light through the double refracting object are all different, so that, on emergence of the light, some of the wavelengths will differ in phase and their intensity will be diminished, whereas others will be in phase and will be reinforced. The net effect is that the object appears highly colored and the color varies as the analyzer is rotated. Monochromatic light, when used to view an anisotropic object between crossed Nicols, gives a cross with dark brushes. This is known as the cross of the isocline and will be discussed presently. Light emerging from the anisotropic object is said to be elliptically polarized because the amplitudes of the waves of the two beams of plane-polarized light, when combined, describe an ellipse.

If a suspension of long asymmetric particles, such as those of vanadium pentoxide, is placed between crossed Nicols, the field will be completely dark. If, however, streaming motion is produced in the suspension, the suspension becomes double refracting as is evidenced by a light field which persists as long as the suspension is in motion. This type [37] of birefringence is known as stream double refraction or stream birefringence.

Double refraction of flow may be due to the orientation of either optically isotropic or optically anisotropic particles. Double refraction due to orientation of optically isotropic particles varies with the index of refraction of the medium and becomes zero when the index of refraction of the medium is made the same as that of the particles. The stream double refraction of optically anisotropic particles also varies with the index of refraction of the medium, but it never becomes zero; it passes through a minimum.

There are two general experimental approaches to the study of stream double refraction. One method used by Lauffer and Stanley [38] and by Lauffer [39] in their study of the stream double refraction of solutions of tobacco mosaic virus protein consists simply in observing a capillary tube between crossed Nicols. The solution was forced back and forth in the capillary by means of a pump arrangement. The observation was carried out with a polarizing microscope. Such a microscope is equipped with a polarizer below the object and an analyzer above the eyepiece. Stream double refraction was measured by the lightness of the field.

[37] H. Freundlich, F. Stapelfeldt, and H. Zocher, *Z. physik. Chem.*, **114**, 161, 190 (1924–25).

[38] M. A. Lauffer and W. M. Stanley, *J. Biol. Chem.*, **123**, 507 (1938).

[39] M. A. Lauffer, *J. Phys. Chem.*, **42**, 935 (1938).

The intensity of light, which is proportional to the stream double refraction, was measured with a photoelectric cell. The tobacco mosaic virus protein, being rodlike with considerable asymmetry, gives a very pronounced stream double refraction which in the more dilute solutions was found to be proportional to the protein concentration. Unfortunately, not very much information can be obtained by this simple method of study. About all one can really do with it is to show the presence of stream double refraction and to investigate the influence of concentration.

The other method is better adapted to quantitative studies. In this method two concentric cylinders are employed; the inner one is stationary; the outer one contains the suspension and is rotated. The

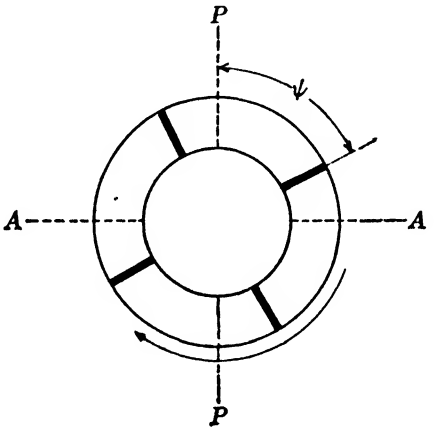


FIG. 3. The cross of isocline which is due to the orientation of particles in the streamlines between a rotating and a stationary cylinder. *PP* and *AA* are the planes of vibration of light transmitted by the polarizer and analyzer, respectively.

velocity of flow of the liquid at the boundary of the inner cylinder is zero; that at the boundary of the outer cylinder is equal to the velocity of motion of the outer cylinder. The velocity of flow thus varies from zero to some finite value, and the velocity gradient is uniform in the direction normal to the surface of the inner cylinder. The solution between the two cylinders is viewed between crossed Nicol prisms or polaroids with monochromatic light. The liquid is isotropic and dark when at rest. Upon rotation of the outer cylinder (or inner one, for that matter), stream double refraction is observed. Figure 3 is a diagrammatic representation of the state of affairs.

The dark brushes or arms of the cross mark the regions where the optic axis of the flowing liquid is parallel to the plane of polarization of

either the analyzer or of the polarizer. ψ, the angle of isocline, corresponds to an angle $90° - \psi$ to the streamline. The angle of isocline is always the larger of the two angles between the cross and the plane of polarization of either the analyzer or the polarizer. The angle between the optic axis and the streamlines—known as the extinction angle, χ—is always between 0° and 45°. Thus, the extinction angle is the complement of the angle of isocline, the sum of the extinction angle and the angle of isocline being 90°. Either of these two angles may be used to define the position of the optic axis. The other type of measurement is the determination of the double refraction $(n_e - n_0)$. This is done by means of a compensator, which is a birefringent plate with a phase difference of a quarter wavelength between the two components vibrating parallel and perpendicular to its optic axis. This compensator is set with its optic axis perpendicular or parallel to PP or AA (Fig. 3), and the analyzer is rotated until the light is extinguished. This angle of rotation is a measure of $(n_e - n_0)$.

The velocity gradient causes a continual rotation of the suspended particles. A prolate ellipsoid rotates most rapidly when the major axis is perpendicular to the streamlines and most slowly when this axis is parallel to the streamlines. If the major axis is very much greater than the minor axis, the particles spend nearly all their time with the major axis parallel to the streamline. In disklike ellipsoids, the short axis tends to orient perpendicular to the streamlines.

Robinson [40] reports an interesting study of the stream double refraction of tobacco mosaic virus protein. This study was paralleled by an investigation of the viscosity of the solutions. Robinson used a Couette viscometer, which consists, as we have seen, of two concentric cylinders. The inner cylinder is stationary and suspended by a fine wire; the outer one contains the suspension and is rotated. Robinson viewed the solution between the inner and outer cylinders with crossed Nicol prisms. He was thus able to make simultaneous optical and viscosity measurements. He found that, with increasing speeds of rotation of the outer cylinder, the stream double refraction increased. At the same time the viscosity decreased. The increased double refraction with increasing speeds is evidence of increased orientation of the particles of tobacco mosaic virus. Increasing orientation evidently leads to a decrease in viscosity. This is to be anticipated, since there is less interruption of the flow lines by the orientated particles.

It is evident that quantitative relations should exist between particle asymmetry and the extent of stream double refraction as measured both by extinction angle and by $(n_e - n_0)$. The relation, however, is about as

[40] J. R. Robinson, *Proc. Roy. Soc.* (*London*), **A170**, 519 (1939).

complicated as the relation between viscosity and asymmetry. It involves the determination of the rotary diffusion constant, which, for a sphere, is given by

$$\theta = \frac{1}{2\tau} = \frac{kT}{8\pi\eta r^3} \qquad 27$$

where τ is the relaxation time of the spherical molecule, k is Boltzmann's constant, η is the coefficient of viscosity of the medium, and r is the radius of the sphere. For an ellipsoid with three semi-axes, a, b, and c, of different lengths, three different rotary diffusion constants are required to describe the rotation of the molecule. Simultaneous stream double refraction and viscosity studies can give valuable information. Unless the particles are pronouncedly asymmetric, enormous flow gradients must be employed in order to observe stream double refraction. If stream double refraction is observed, one has definite proof that asymmetric particles are present. Likewise, if a decrease in viscosity is observed with increasing flow gradient and if gel formation can be ruled out, one also has proof that asymmetric particles are present.

Edsall has written a very useful and clear review of streaming birefringence.[41] He brings to the subject a wealth of experience.

[41] J. T. Edsall, *Advances in Colloid Science*, Interscience Publishers, Inc., New York, 1942.

Chapter XIV

DIFFUSION

It can be said fairly that the diffusion of substances through cell membranes as they pass into and out of cell interiors is a fundamental biological problem. A separate chapter has been reserved for the treatment of this problem (Chapter XVII). Quite apart, however, from the physiological aspects, diffusion measurements are capable of yielding valuable information about molecules. It is this phase of the problem which we wish to consider at this point.

Diffusion was the subject of early study. In 1855, Fick enumerated his first law of diffusion, and shortly afterwards Thomas Graham, the father of colloid chemistry, distinguished between "crystalloids" and "colloids" on the basis of diffusion. Colloidal substances were not supposed to diffuse, in contrast to crystalloidal substances. We now know that the difference in diffusibility is one of degree and not of kind. The large colloidal particles simply diffuse more slowly.

It is not our purpose to enter into an extended exposition of the theory of diffusion. This is treated in some detail in the review by Williams and Cady.[1] We should, however, have clearly in mind a physical picture of diffusion. Consider Fig. 1, which is taken from the excellent review of protein diffusion by Neurath.[2] The solution at concentration C_0 is in contact with the pure solvent. Diffusion takes place only in a horizontal direction. Under these conditions

$$\text{Rate of diffusion} = -DA\frac{dC}{dx} \qquad 1$$

where A is the cross-sectional area of the diffusion column and dC/dx is the concentration gradient. D is the diffusion constant which is characteristic of the solute molecules in question.

If the slopes of the concentration-distance curves shown in the center section of Fig. 1 are plotted against the distance, a family of curves is obtained as is shown in the bottom section of Fig. 1. These curves have the shapes of Gaussian distribution curves and are identical with one

[1] J. W. Williams and L. C. Cady, *Chem. Revs.*, **14**, 171 (1934).

[2] H. Neurath, *Chem. Revs.*, **30**, 357 (1942).

another in respect to their areas. Curves of this type follow the equation

$$\frac{dC}{dx} = \frac{C}{2\sqrt{\pi Dt}} e^{-x^2/4Dt} \qquad 2$$

where t is the elapsed time after the forming of the boundary between the solvent and solution, and C is the concentration at any distance x

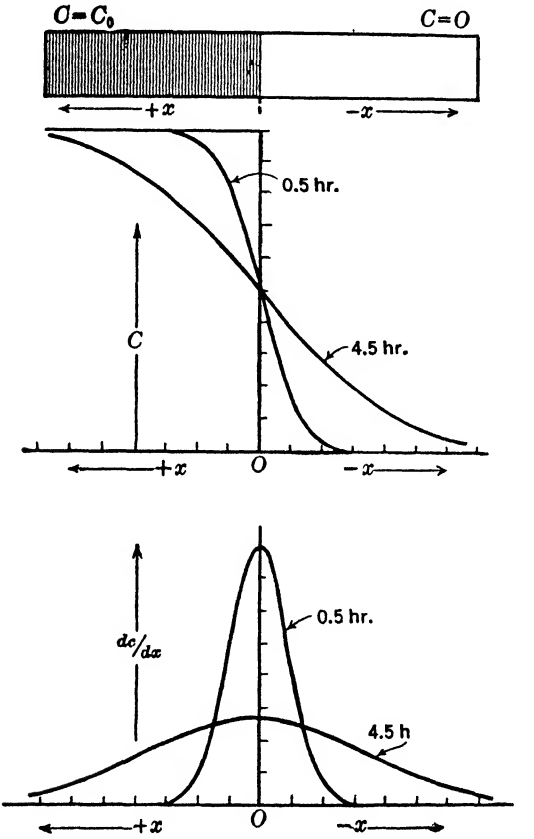


FIG. 1. Relation between concentration and distance of migration in diffusion column. Top section: graphical illustration of the diffusion column. Center section: relation between concentration and distance of migration. Bottom section: relation between concentration gradient and distance of migration. The curves as drawn refer to diffusion times of 0.5, and 4.5 hours, respectively. (Neurath.)

from the original boundary. The sign of x is negative in the direction of the solvent and positive towards the solution. e is the base of the natural logarithms. We shall describe presently how this equation can be used to evaluate the diffusion constant.

Measurement of Diffusion Constants

Of the various ways of measuring diffusion constants, we shall content ourselves with a discussion of only two. The first method is based upon the chemical analysis of two solutions of different concentration of the solute separated by a porous disk; we shall call it the porous-disk method. The second method involves a measurement of the manner in which the index of refraction of light varies at a boundary between the solution and the solvent; we shall call it the refractometric method.

The porous-disk method was developed largely by Northrop and Anson [3] and by McBain and Liu.[4, 5]

By definition, the diffusion constant is the quantity of material which diffuses per second across a surface 1 sq. cm. in area under a unit concentration gradient; hence

$$D = \frac{dQ}{A\ dt\ dC/dx} \qquad 3$$

where dQ is the quantity of diffusion material which passes across a plane of area A in time dt under a concentration gradient dC/dx. Substituting dimensional quantities in equation 3, we find the dimensions of the diffusion constant to be L^2t^{-1}. Diffusion constants are, accordingly, expressed in square centimeters per second.

Where a relatively concentrated solution on one side of the porous disk diffuses into the pure solvent on the other side and where the compartments on the two sides of the disk are well stirred, we may neglect the differentials in equation 3 and write

$$D = \frac{h}{A}\frac{Q}{tC} \qquad 4$$

where h is the effective distance through which the solute diffuses, A is the effective area of the pores of the disk, Q is the amount of the material which diffuses in time t, and C is the concentration of the solution. The ratio h/A differs for each membrane which is used but is constant for the same membrane regardless of the substance which diffuses through it. This ratio is called the membrane constant; it may be evaluated by the relation:

$$\text{Membrane constant} = \frac{h}{A} = \frac{Dt}{QV} \qquad 5$$

[3] J. Northrop and M. L. Anson, *J. Gen. Physiol.*, **12**, 543 (1928–29); **20**, 575 (1937).
[4] J. W. McBain and T. H. Liu, *J. Am. Chem. Soc.*, **53**, 59 (1931).
[5] J. W. Mehl and C. L. A. Schmidt, *University of California Publications in Physiology*, **8**, 165 (1937).

where QV is the number of cubic centimeters of the initial solution which contains an amount of substance that has diffused through the membrane in time t. Since the diffusion constant of a number of substances is known, they may be used to evaluate h/A of the particular membrane. With this value of h/A, the diffusion constant of the unknown may be calculated. Actually, it is not necessary to know the absolute amount of the original material which has passed through the membrane, but only the percentage of the amount. If the quantity of material contained in 1 cc. of the concentrated solution is taken as unity and the amount which has diffused is expressed in this unit, i.e., as the number of cubic centimeters of the concentrated solution containing the quantity which has diffused, equation 3 may be simplified still further to

$$D = \frac{hQ_{cc}}{At} \qquad 6$$

where Q_{cc} is the number of cubic centimeters of the concentrated solution which contains an amount of the substance equal to the amount which has diffused. For example, if it were found in a study on the diffusion of glucose from a 0.1 M solution into pure water that 0.001 mole of glucose had diffused in a given time, Q_{cc} would evidently be 10; i.e., 10_{cc} of the 0.1 M glucose would contain 0.001 mole of glucose. The porous-disk apparatus is diagrammed in Fig. 2.

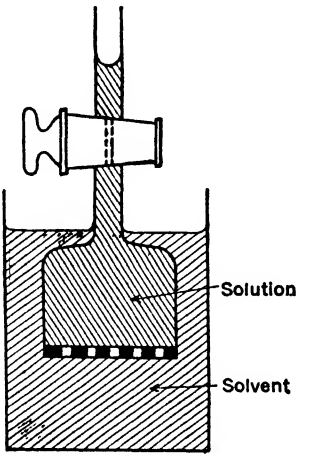


FIG. 2. Diagram of the porous-disk apparatus for the study of diffusion.

The porous-disk method is much more difficult experimentally than it appears. The author has had no personal experiences with it, but he has been told by those who have that it is next to impossible to obtain accurate results with it. It has, however, the tremendous advantage that very dilute solutions of biologically active materials can be used and no knowledge of the absolute concentrations is needed.

The refractometric method depends upon the observed fact that the index of refraction of a solution is proportional to the concentration of the solute. An apparatus is used in which the refraction experienced by a beam of light in passing through the boundary between the solvent and solution can be measured accurately. From this information the change of the concentration gradient with distance from the boundary is determined (see bottom section of Fig. 1). This gradient is measured

after appropriate time intervals, and this furnishes us with sufficient information to calculate the diffusion constant.

Diffusion measurements by the refractometric method require that the initial boundary between the solution and solvent be sharp and well defined. The Neurath [6] diffusion cell appears to meet this requirement more conveniently and exactly than any other cell. "The cell consists of two U-shaped stainless-steel blocks which can be slid horizontally past each other by a screw arrangement. Two optically flat rectangular glass windows are pressed against the vertical faces by the top and bottom clamps. The lower compartment, when separated from the top one, is filled with the solution through the left-hand tube; the upper compartment is filled with the solvent through the right-hand tube. After temperature equilibrium, the sharp boundary is formed by bringing the two compartments into vertical alignment. The dimensions of the cell proper are as follows: 1.7 cm. in the direction of the optical axis, 0.5 cm. wide and 5 cm. high."

There are two convenient methods of measuring the displacement of the light after it has passed through the solution-solvent boundary in the diffusion cell. The cylindrical-lens system accomplishes this measurement very conveniently. This system has already been discussed under electrophoresis (Chapter X). The other method, which probably is capable of a greater degree of precision than the cylindrical-lens system, although it is a very tedious method, is to photograph an accurately ruled scale through the boundary. The appropriate lens system has to be used to visualize the scale. The scale is also photographed in the absence of the boundary, and the displacement of the scale lines is then measured with a microcomparator. The scale-line displacements are then plotted against the positive and negative distances from the original boundary. Both the scale-line displacement and the cylindrical-lens system yield a curve of the type shown in the bottom section of Fig. 1.

The problem of obtaining D from the curve shown in the bottom section of Fig. 1 involves the solution of equation 2 for this constant. Actually, there are four ways of proceeding to do this. Of these ways only the maximum ordinate method will be considered. At x equal zero, equation 2 becomes

$$\left(\frac{dC}{dx}\right)_{x=0} = y_m = \frac{C}{2\sqrt{\pi Dt}} \qquad 7$$

[6] H. Neurath, *Science*, **93**, 431 (1941).

where y_m is the maximum ordinate. Since the area under the curve is proportional to the concentration of the solute, we can substitute the area for C and rearrange the equation to obtain

$$D = \frac{A^2}{4\pi t(y_m)^2} \qquad 8$$

The area A is determined by graphical integration using a planimeter. We can plot t in seconds against $(1/y_m)^2$ and a straight line should be obtained whose slope is $A^2/4\pi D$, from which D can be calculated. The D so found must be multiplied by the magnification factor of the lens system. The magnification factor is determined by photographing a standard scale and comparing the known dimensions with those found. There are a number of details of diffusion measurements which we have not considered; for these, and for a splendid treatment of diffusion in general, the reader is referred to the review by Neurath.[2]

The uniformity of the material in respect to particle size can be studied by the refractometric method. It is assumed that the curve showing the

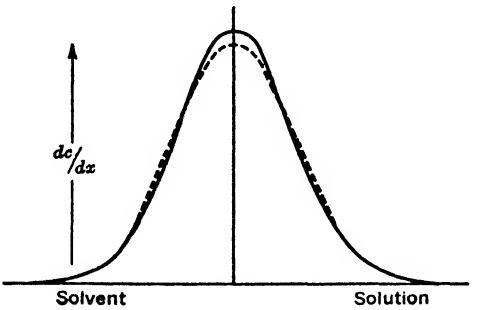


FIG. 3. Comparison of an ideal Gaussian distribution curve with the diffusion curve obtained from measurements on a 6.8 per cent serum albumin solution at *p*H 7.6 which had been denatured by heating for 30 minutes at 70° C. Polydispersity is most clearly indicated by the difference in maximum ordinate of the ideal curve (broken line) and the experimental curve (solid line). (Neurath.)

displacement of light as a function of the distance from the boundary may be treated as an ideal distribution curve. The identity of the experimental curve with the normal distribution curve indicates monodispersity, whereas deviations from the normal dispersion curve can be interpreted to mean that the material is polydispersed. See Fig. 3.

If there is considerable interaction between the solute particles leading to a certain amount of gel structure in the solution, a skewed curve is obtained when the data are plotted as described above. Such a curve is

shown in Fig. 4, which was obtained on a 1 per cent solution of tobacco mosaic virus protein.

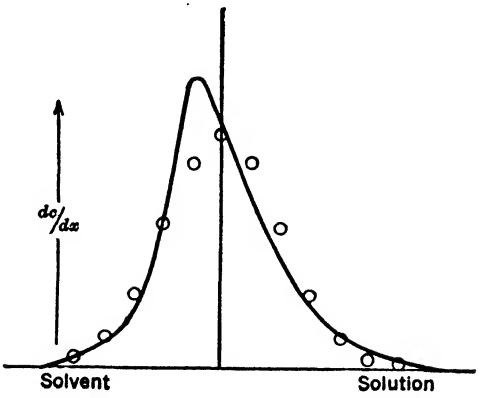


FIG. 4. Comparison of an ideal Gaussian distribution curve with the diffusion curve as obtained from the measurements on a one per cent solution of tobacco mosaic virus protein. The open circles indicate the position of the ideal curve; the solid line refers to the experimental curve plotted in normal coordinates. (Neurath.)

Temperature Influences

The diffusion constant measured at a certain temperature can be calculated for any other temperature by the equation [7]

$$D_{T_1} = D_{T_2} \frac{T_1}{T_2} \frac{\eta_{T_2}}{\eta_{T_1}} \qquad 9$$

where D_{T_1} and η_{T_1} are the diffusion constant and viscosity of the solvent at temperature T_1, and D_{T_2} and η_{T_2} are the corresponding values at temperature T_2. The temperature is expressed in degrees absolute. For diffusion in water at 20° C. and at 25° C. we have

$$D_{25} = 1.146 D_{20} \qquad 10$$

Concentration Effects

The electrostatic charge on colloidal particles can produce anomalous diffusion rates of those particles. In order to circumvent this effect, salts are added to the solvent. It is reported that, if the ionic strength is about 0.1, the electrostatic influence is negligible.

Inasmuch as it is common practice to refer diffusion constants to that in pure water at 20° C. or 25° C., it is necessary to correct for the

[7] A. Polson, *Kolloid-Z.*, **87**, 149 (1939).

influence of the salts on the solvent. The correction usually employed is

$$D_x = D_y \frac{\eta_x}{\eta_y} \qquad 11$$

where D_x and D_y are the diffusion constants measured at constant temperature in solvents of viscosity η_x and η_y.

Diffusion constants are not independent of the concentration of the solute. If the concentration dependence is great enough, a skewed diffusion curve such as is shown in Fig. 4 will be obtained. If the curve is not skewed, it is safe to neglect the influence of the solute concentration on the rate of diffusion.

Results

The diffusion constant of a protein must be known before its molecular weight can be calculated from the rate of sedimentation in the ultracentrifuge. As a consequence of this, the diffusion constants of a large number of proteins have been measured. Reports on these measurements can be found in *The Ultracentrifuge* by Svedberg and Pedersen [8] and in the review by Neurath.[2] Table 1 shows the diffusion constants of a few proteins along with their volume intrinsic viscosities.

TABLE 1

DIFFUSION AND VISCOSITY DATA OF PROTEINS AT 25° C.

(The volume intrinsic viscosities are calculated from the partial specific volumes of the proteins in solution.)

Protein	$D \times 10^7$	D_0/D	η_{sp}/ϕ
Pepsin B	10.0	1.120	5.17
Egg albumin	8.8	1.158	5.20
Pepsin A	9.2	1.210	5.94
Serum albumin (horse)			
McMeekin fraction	7.4	1.210	4.97
Kekwick fraction	7.0	1.270	5.69
Helix hemocyanin *p*H 6.67	1.6	1.241	7.05
β-Lactoglobulin	8.4	1.258	5.44
Homarus hemocyanin	3.2	1.274	6.39
Seroglycoid (horse)	6.9	1.280	6.73
Amandin	4.2	1.284	6.98
Octopus hemocyanin	1.9	1.383	9.03
Thyroglobulin (pig)	3.0	1.425	9.90
Helix hemocyanin *p*H 8.6	2.4	1.886	18.73

[8] T. Svedberg and K. Pedersen, *The Ultracentrifuge*, Oxford University Press, New York, 1940.

In Table 1, D is the measured diffusion constant in square centimeters per second and D_0 is the diffusion constant of a spherical anhydrous molecule of the same molecular weight.

Diffusion of Spherical Colloidal Particles

We are here concerned with the diffusion of particles whose volume is much greater than that of the molecules of the solvent. For a discussion of the diffusion of solute molecules whose sizes are of the same order of magnitude as those of the solvent, see Stearn et al.[9]

Sutherland [10] and Einstein [11] almost simultaneously published papers dealing with the diffusion of large spherical molecules into a medium of small molecules, and they were able to derive an equation relating the diffusion constant of the spherical molecules to their size. The genius of their derivation was the realization that the force tending to produce diffusion was the osmotic force, $RT\ dC/dx$, while the resisting force was Stokes' factor, $6\pi r u \eta N$, where r is the radius of the spherical molecule, u is the velocity of motion through the medium, and N is Avogadro's constant. In a steady state the resisting force must equal the osmotic force; hence

$$6\pi r u \eta N = RT \frac{dC}{dx} \qquad 12$$

but from equation 1

$$u = -DA \frac{dC}{dx} \qquad 13$$

Then per unit area of cross section

$$6\pi r \eta D N = RT \qquad 14$$

and

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

Since we are dealing with a spherical molecule, equation 15 may be written

$$D = \frac{RT}{6\pi \eta N \sqrt[3]{3MV/4\pi N}} \qquad 16$$

where MV is the cubic centimeters of molecular volume, i.e., the gram-molecular weight divided by the density.

[9] A. E. Stearn, E. M. Irish, and H. Eyring, *J. Phys. Chem.*, **44**, 981 (1940).
[10] W. Sutherland, *Phil. Mag.*, (6) **9**, 781 (1905).
[11] A. Einstein, *Ann. Physik*, (4) **17**, 549 (1905); *Z. Elektrochem.*, **14**, 337 (1908).

Substituting numerical values for the constants in equation 16, we have for spherical molecules in water

$$D^{20°} = \frac{28.82 \times 10^{-6}}{\sqrt[3]{MV}} \text{ square cm. per sec.} \qquad 17$$

$$D^{25°} = \frac{33.06 \times 10^{-6}}{\sqrt[3]{MV}} \text{ square cm. per sec.} \qquad 18$$

where $D^{20°}$ and $D^{25°}$ are the diffusion constants of spherical molecules at 20° C. and at 25° C., respectively.

The diffusion of a spherical molecule is thus inversely proportional to the cube root of the volume of the molecule in solution. If the molecule is hydrated, the diffusion constant will be reduced and the ratio of the anhydrous diffusion constant to the hydrated diffusion constant will be

$$\frac{D_0}{D_H} = \sqrt[3]{\frac{MV_H}{MV_A}} \qquad 19$$

where D_0 is the diffusion constant of the anhydrous spherical molecule, D_H is that of the hydrated spherical molecule, MV_H is the hydrated cubic centimeters molecular volume, and MV_A is the anhydrous cubic centimeters molecular volume.

Diffusion of Asymmetric Colloidal Particles

The influence of particle asymmetry on diffusion is somewhat more complicated than the effect of hydration last considered. The problem is a good deal simpler, however, than the relation between particle asymmetry and viscosity which was discussed in the last chapter. Qualitatively, it is easy to see why asymmetry should decrease the diffusion constant of a molecule. 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.

Gans,[12] in a highly mathematical paper, derived expressions for the resistance experienced by prolate and oblate ellipsoids of revolution moving through a viscous medium. Herzog, Illig, and Kudar [13] substituted these expressions in the general diffusion equations of Riecke and obtained two diffusion equations, one of which relates the diffusion constant of a prolate ellipsoid of revolution to its asymmetry and volume and the other relates these factors for an oblate ellipsoid of revolution.

[12] R. Gans, *Ann. Physik*, **87**, 935 (1928).

[13] R. O. Herzog, R. Illig, and H. Kudar, *Z. physik. Chem.*, **167A**, 329 (1933).

Shortly after Herzog, Illig, and Kudar published their derivation, Perrin [14] published an independent derivation of these equations. It is reassuring that Perrin's equations can be reduced to forms identical with those of Herzog, Illig, and Kudar.

The equation for a prolate ellipsoid of revolution can be expressed as

$$\frac{D_0}{D} = \frac{2(1 - b^2/a^2)^{1/6}}{\left(\frac{b}{a}\right)^{2/3} \ln \frac{1 + (1 - b^2/a^2)^{1/2}}{1 - (1 - b^2/a^2)^{1/2}}} \qquad 20$$

where D_0 is the diffusion constant of a spherical particle with the same volume as that of the asymmetric particle, D is the diffusion constant of the asymmetric particle, and a is the major axis and b the minor axis of the prolate ellipsoid of revolution.

The corresponding equation for an oblate ellipsoid of revolution is

$$\frac{D_0}{D} = \frac{(1 - .b^2/a^2)^{1/6}}{\left(\frac{b}{a}\right)^{2/3} \text{arc sin} \left(1 - \frac{b^2}{a^2}\right)^{1/2}} \qquad 21$$

where a is the major and b the minor axis of the oblate ellipsoid of revolution.

Figure 5 shows a graph of a/b against D_0/D for prolate and for oblate ellipsoids using equations 20 and 21.

Evidently, if one knew the asymmetry of a protein molecule, the hydration could be calculated from diffusion studies, or, if the hydration were known, the asymmetry could be calculated. Since neither the asymmetry nor the hydration is known, we are at an impasse.

So far no elegant way for getting by this impasse has been suggested. Various values for the volume hydration have been assumed, and the asymmetries of the protein molecules have been calculated on this basis. This is manifestly an unsatisfactory solution to the problem. The volume hydrations are at best only intelligent guesses. There is also no way of knowing where the volume increase due to hydration occurs on the protein molecule. For example, if the hydration is centered around the major axis, the asymmetry might be decreased.

Cooper and the author [15] suggest one way of escaping from this dilemma. It is not a very graceful way, but it is better than none at all. It can be outlined as follows: Plot D_0/D against the volume intrinsic viscosities for the various proteins. A line drawn through these points

[14] F. Perrin, *J. phys. radium*, 7, 1 (1936).

[15] H. B. Bull and J. A. Cooper, In Press.

should intersect the x-axis for D_0/D equal to unity at 2.5, since we know the volume intrinsic viscosity should be equal to 2.5 for a spherical molecule. We find that this cannot be done, and we conclude that the incorrect value for the specific volume of the proteins has been used. We adjust the specific volumes of the proteins until the line relating D_0/D and η_{sp}/ϕ does intersect the x-axis at 2.5 when D_0/D is unity. In order to accomplish this an average volume hydration of 28 per cent has been assumed. This value is not in disagreement with other physical

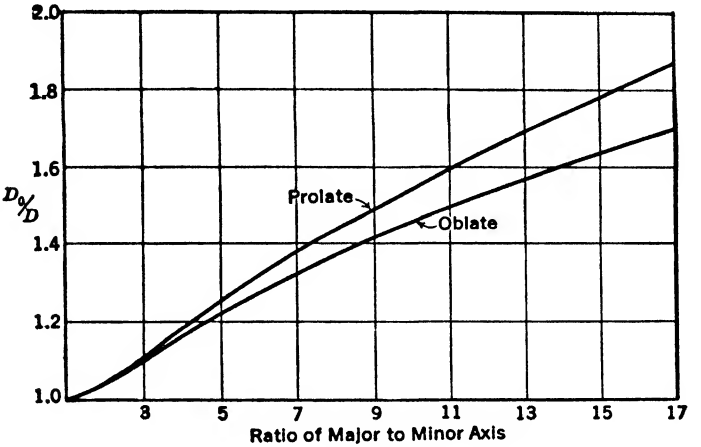


FIG. 5. Relation between asymmetry of prolate and of oblate ellipsoids of revolution and their diffusion constants as calculated from the equations of Herzog, Illig, and Kudar.

measurements. The approximate hydration being known, it is now possible to calculate the approximate asymmetries of the molecules using equations 20 and 21.

It is also possible to formulate a test of the various equations relating viscosity and the asymmetries of suspended particles (see Chapter XIII). This test was carried out in the following manner: The asymmetries corresponding to a series of assumed intrinsic viscosities were calculated by means of the various viscosity equations. The corresponding D_0/D values for these asymmetries were calculated by means of the equations of Herzog, Illig, and Kudar. The D_0/D values so found were then plotted against the assumed intrinsic viscosities. The experimental curve relating D_0/D and the intrinsic viscosities for 28 per cent volume hydration was plotted. The extent to which the theoretical curves corresponding to the various viscosity equations agree with the experimental curve is a measure of their validity. These plots are shown in Fig. 6.

It is evident from Fig. 6 that Burger's equation gives the poorest agreement with the experimental curves. The deviations are so great as more or less to discredit this equation. No viscosity equation is valid over the whole range, but Simha's equation for prolate ellipsoids of revolution is best for small viscosities. As explained in Chapter XIII, Kuhn's equation was derived on the basis of a rigid chain of balls, and modern ideas of protein structure certainly do not permit such a model

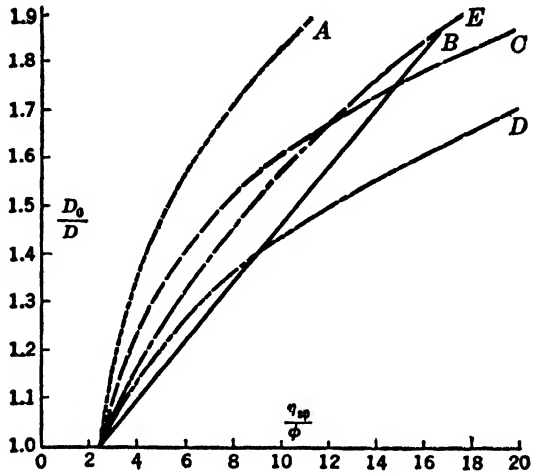


FIG. 6. *B* is the least-square line for protein molecules with 28 per cent volume hydration. *A* is calculated from Burger's equation, *C* from Kuhn's equation, *D* from Simha's equation for prolate ellipsoids, and *E* from Simha's equation for oblate ellipsoids. (Bull and Cooper.)

for the protein molecule. Any agreement between Kuhn's equation and experiment is purely coincidental.

Two ambiguities are associated with the test shown in Fig. 6. The validity of the diffusion equations of Herzog, Illig, and Kudar has been assumed, and, secondly, the protein molecules have been assumed to have shapes required by the viscosity equations. It is only if both of these assumptions are valid that the results shown in Fig. 6 become a true test of the viscosity equations. If the first assumption is valid but not the second, it can be said that, though the viscosity equations cannot be used to describe the viscosity of solutions of protein molecules, the equations might be valid for models for which they were derived. If the first assumption is untenable, the test becomes indeterminate.

Chapter XV

THE ULTRACENTRIFUGE

It is neither necessary nor desirable to enter into a detailed discussion of the ultracentrifuge. The recent book on the ultracentrifuge by Svedberg and Pedersen [1] supplies a wealth of information. In addition, there have been several reviews by authorities in the field. One of the more recent of these reviews is by Pickels.[2]

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 [3] 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 other colloidal molecules.

Methods

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 [4] have developed an air-driven ultracentrifuge which is not so elaborate as the Svedberg type and its cost of construction is less. McBain [5] 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 opti-

[1] T. Svedberg and K. O. Pedersen, *The Ultracentrifuge*, Oxford University Press, New York, 1940.

[2] E. G. Pickels, *Chem. Revs.*, **30**, 341 (1942).

[3] T. Svedberg and J. B. Nicols, *J. Am. Chem. Soc.*, **45**, 2910 (1923).

[4] J. W. Beams, F. W. Linke and P. Sommer, *Rev. Sci. Instruments*, **9**, 248, (1938).

[5] J. W. McBain *Chem. Revs.*, **24**, 289 (1939).

cal 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. These systems have already been described.

With any type of ultracentrifuge or of optical system the determination of the molecular weight can be carried out in one of two ways. The speed of the ultracentrifuge can be made very large and the rate of sedimentation measured. Such a determination can usually be completed in a few hours. The other type of determination consists in rotating the centrifuge at a more moderate speed and allowing the solution to reach an equilibrium, the solution being more concentrated towards 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, and

$$\text{Force acting} = M\omega^2X - M\omega^2X\bar{V}_2\rho \qquad 1$$

or

$$\text{Force acting} = M(1 - \rho\bar{V}_2)\omega^2X \qquad 2$$

where M is the gram-molecular weight, ω is the angular velocity in radians per second, X is the distance from the axis of rotation, $\bar{V}_2$ is the partial specific volume of the particles, and ρ is the density of the solution. In a steady state, the net centrifugal force must equal the resisting forces of the medium. These resisting forces will vary with the hydration and the asymmetry of the particles as well as with the velocity of migration of the particles, and

$$\text{Resisting forces} = f_s\frac{dX}{dt} \qquad 3$$

where f_s is the frictional force per mole per unit speed. Equating the net centrifugal force and the frictional forces, we have

$$M(1 - \rho\bar{V}_2)\omega^2X = f_s\frac{dX}{dt} \qquad 4$$

the expression $\dfrac{dX/dt}{\omega^2X}$ is the sedimentation rate per unit field of force and is denoted by S. Accordingly,

$$M = \frac{f_sS}{(1 - \rho\bar{V}_2)} \qquad 5$$

Diffusion involves exactly the same frictional forces as sedimentation in the ultracentrifuge; accordingly

$$D = \frac{RT}{f_s} \qquad 6$$

where R is expressed in ergs per degree and is equal to 8.315×10^7. Equation 5 then becomes

$$M = \frac{RTS}{D(1 - \rho\bar{V}_2)} \qquad 7$$

We see from equation 7 that, in order to calculate the molecular weight from rate-sedimentation studies, we must know S the sedimentation per unit field of force, D the diffusion constant, ρ the density of the solution, and $\bar{V}_2$ 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_0. The ratio f/f_0 was at one time known as the asymmetry ratio. It must be realized, however, that although f/f_0 is related, in a complicated way, to the molecular asymmetry, the relation is not a direct one. It can be seen that

$$\frac{f}{f_0} = \frac{D_0}{D} \qquad 8$$

and f/f_0 can be substituted for D_0/D in the equations of Herzog, Illig, and Kudar to estimate the molecular asymmetry. Molecular hydration confuses this relation between f/f_0 and asymmetry as it does in diffusion.

The other way of proceeding 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

$$M = \frac{2RT \ln C_2/C_1}{(1 - \rho\bar{V}_2)\omega^2(X_2^2 - X_1^2)} \qquad 9$$

Note that equation 9 is in the form of a Boltzmann distribution, i.e.,

$$C_2 = C_1 e^{M\psi/RT} \qquad 10$$

where ψ is a "gravitational potential" and is equal to $(1 - \rho\bar{V}_2)\omega^2(X_2^2 - X_1^2)$.

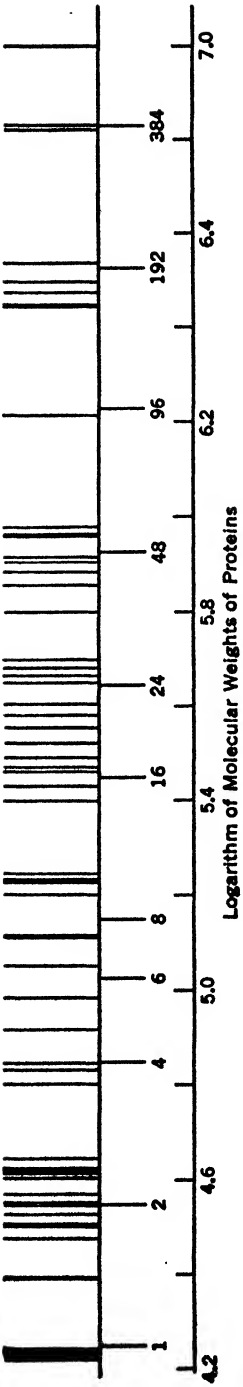


FIG. 1. A logarithmic "spectrum" of the molecular weights of proteins. The lines of numbers beginning with 1 and ending with 384 are the supposed number of 17,600 units in proteins, i.e., the molecular-weight classes. The lower line of numbers is an arithmetic scale of logarithms; the top line indicates the position of the logarithm of the protein molecular weights.

The equilibrium-sedimentation is in a sense equivalent to an osmotic-pressure determination. Its theoretical background is somewhat 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 as satisfactory as the rate-sedimentation method.

The general theory of the ultracentrifuge seems to be sound enough, and the ambiguities which exist would be expected to give rise to second-order errors. The ultracentrifuge gives anhydrous molecular weights. The author is not yet completely satisfied with the treatment accorded hydration, and it is possible that hydration does cause small errors. The difficulty, as the author sees it, is involved in the uncertain density of the water of hydration; there has been a volume contraction incidental to hydration with the result that the density of water associated with the colloid is greater than that in bulk.

Svedberg and his co-workers have repeatedly claimed that the molecular weights of proteins fall into certain classes. For example, 17,600 is believed to be the unit molecular weight, and the molecular weights of all proteins are supposed to be whole-number multiples of this unit. The first class is $1 \times 17{,}600$ or 17,600; the second class is $2 \times 17{,}600$ or 35,200; the third class is $4 \times 17{,}600$ or 70,400 (no proteins are listed which contain three units); the fourth class is $6 \times 17{,}600$ or 105,000; etc. The most massive protein according to Svedberg contains 384 of the 17,600 units. The idea

of molecular-weight classes seems to have been more or less generally accepted and has inspired more than one theory of protein structure. It is exceedingly important to decide whether or not the classification of proteins on this basis is a valid one.

Clearly, if the division of the proteins into molecular weights such as outlined above is in accord with reality, there must be some guiding principle back of protein synthesis which determines the molecular size irrespective of the type of animal or plant tissue in which the proteins originated. In short, the idea of molecular-weight classes throws the burden on the chemical nature of the amino acids. It would seem most unlikely that such a molecular-weight classification could be true. For example, what guiding principle could exist which would make zein from corn (molecular weight 40,000) and egg albumin from a hen (molecular weight 45,000), differing as they do in amino acid content, belong to the same molecular-weight class?

The author regards the classification of proteins on a molecular weight basis with extreme reservation. In Fig. 1 is shown a logarithmic "spectrum" of the molecular weights of proteins.[6]

As far as can be told by inspection of Fig. 1, there are no molecular-weight classes. There is an apparent tendency for certain molecular weights to cluster around 17,000, but no one knows how many "proteins" have smaller molecular weights than 17,000; the clustering may simply indicate that there are a large number of small molecular-weight proteins in nature. The author is not impressed by finding that the higher molecular weights seem to be approximate multiples of 17,600; all numbers, if they are large enough, are approximate multiples of 17,600.

Partial Specific Volume

At various points in this book we have used the term partial specific volume. As a knowledge of the partial specific volume is necessary to complete the calculation of the molecular weights from ultracentrifugation data, it seems appropriate to discuss in some detail the meaning and measurement of partial specific volumes at this time.

The partial specific volume of a solute may be defined as that volume increase suffered by a very large volume of solution upon the addition of 1 gram of solute. A more useful way of expressing the partial specific volume is

$$\text{Partial specific volume } (\bar{V}) = \frac{dV}{dw} \qquad 11$$

where dV is the infinitesimal increase in the volume of a solution due to the addition of an infinitesimal weight of the solute.

[6] H. B. Bull, *Advances in Enzymology*, Vol. 1, p. 1, Interscience Publishers, Inc., New York, 1941.

The partial specific volume may be determined experimentally by plotting the weights of the contents of a pycnometer for several concentrations against the weight concentration of the solute expressed as weight fraction. The weight fraction is 0.01 times the weight per cent concentration. The best smooth curve is drawn through the points of the plot and the slope of the line determined. This gives dm/dw_2, where m is the weight of the contents of the pycnometer and w_2 is the weight fraction of the solute.

The calculation of the partial specific volume from dm/dw_2 is somewhat involved. It will be necessary to derive a relation between these two factors.

The weight fraction of the solute is

$$w_2 = \frac{g_2}{g_1 + g_2} \qquad 12$$

where g_1 is the weight of the solvent and g_2 is the weight of the solute. The specific volume of the solution is evidently

$$V_{sp} = \frac{V}{g_1 + g_2} \qquad 13$$

where V is the volume of the pycnometer.

If we assume for the sake of simplicity that g_1 remains constant, we have upon differentiating equation 12

$$dw_2 = \frac{(g_1 + g_2)dg_2 - g_2 dg_2}{(g_1 + g_2)^2} \qquad 14$$

$$= \frac{g_1 dg_2}{(g_1 + g_2)^2} \qquad 15$$

Upon differentiating equation 13 and maintaining g_1 constant

$$dV_{sp} = \frac{(g_1 + g_2)dV - V dg_2}{(g_1 + g_2)^2} \qquad 16$$

$$= \frac{dV}{(g_1 + g_2)} - \frac{V dg_2}{(g_1 + g_2)^2} \qquad 17$$

Dividing equation 17 by equation 15, we have

$$\frac{dV_{sp}}{dw_2} = \frac{dV}{(g_1 + g_2)} \frac{(g_1 + g_2)^2}{g_1 dg_2} - \frac{V dg_2}{(g_1 + g_2)^2} \cdot \frac{(g_1 + g_2)^2}{g_1 dg_2} \qquad 18$$

$$= \frac{dV(g_1 + g_2)}{g_1 dg_2} - \frac{V}{g_1} \qquad 19$$

Multiplying equation 19 by $g_1/(g_1 + g_2)$ we have

$$w_1 \frac{dV_{sp}}{dw_2} = \frac{dV}{dg_2} - \frac{V}{g_1 + g_2} \qquad 20$$

Remembering that $V/(g_1 + g_2)$ is equal to V_{sp} and also that dV/dg_2 is by definition equal to the partial specific volume of the solute, we have

$$w_1 \frac{dV_{sp}}{dw_2} = \overline{V}_2 - V_{sp} \qquad 21$$

Also we have

$$V_{sp} = \frac{V}{m} \qquad 22$$

where m is the total weight of the contents of the pycnometer, i.e., m is equal to g_1 plus g_2. Differentiating equation 22

$$dV_{sp} = -\frac{V dm}{m^2} \qquad 23$$

Dividing equation 23 by dw_2, we have

$$\frac{dV_{sp}}{dw_2} = -\frac{V}{m^2}\frac{dm}{dw_2} \qquad 24$$

Since

$$w_1 = 1 - w_2 \qquad 25$$

we have from equation 21

$$(1 - w_2)\frac{dV_{sp}}{dw_2} = \overline{V}_2 - V_{sp} \qquad 26$$

Substituting equation 24 in equation 26 there results

$$(1 - w_2)\frac{V}{m^2}\frac{dm}{dw_2} = \overline{V}_2 - V_{sp} \qquad 27$$

And from equations 22 and 27 we have

$$\overline{V}_2 = V_{sp} - \frac{(1 - w_2)}{m} V_{sp}\frac{dm}{dw_2} \qquad 28$$

Since V_{sp} is equal to $1/\rho$, where ρ is the density of the solution, equation 28 becomes

$$\overline{V}_2 = \frac{1}{\rho} - \frac{(1 - w_2)}{\rho m}\frac{dm}{dw_2} \qquad 29$$

Equation 29 can be used to calculate the partial specific volume of the second component. In order to do this we need, as can readily be seen,

the density of the solution ρ, the weight fraction of the second component in the pycnometer w_2, the total weight of the contents of the pycnometer m, and the slope of line relating the total weight of the contents of the pycnometer to the weight of the solute (dm/dw_2).

It will be found for many dilute solutions, particularly those of proteins, that the partial specific volume is independent of the concentration. In this event, the apparent partial specific volume may be used in place of the true partial specific volume. The apparent specific volume of the solute is defined as

$$\bar{V}_a = \frac{V - W_1 V_1}{W_2} \qquad 30$$

where V is the total volume of the solution. W_1 is the weight of the solvent, V_1 is the specific volume of the solvent, and W_2 is the weight of the dissolved solute. The apparent specific volume of the solute can be determined directly from a density measurement using a pycnometer. The appropriate data are then substituted in equation 30. The apparent partial specific volume $\bar{V}_a$ is related to the partial specific volume $\bar{V}_2$, as is a ratio of finite increments to the corresponding differential coefficient.

The partial specific volume is a measure of the volume of solvent displaced by the solute. It does not represent and is not directly related to the actual volume occupied by the hydrated solute molecules in solution. Indeed, the actual volume occupied by the hydrated solute molecules in solution is a fuzzy concept. We have discussed this problem in the chapters on viscosity and on diffusion without arriving at an unambiguous answer.

Chapter XVI

OSMOTIC PRESSURE

Some properties of solutions depend primarily on the number of solute molecules present. Such properties are known as colligative properties and are to be contrasted with properties which depend on the nature of the solute molecules and which are called constitutive properties. Some colligative properties of solutions are the vapor-pressure lowering of the solvent, the freezing-point depression of the solvent, and the osmotic pressure.

Vapor-Pressure Lowering

As is well known, water at any given temperature and pressure exerts a definite vapor pressure. If the temperature is raised sufficiently, a temperature will be found at which the aqueous vapor pressure equals the atmospheric pressure and the water boils. If some solute is added to the water, the vapor pressure of water is found to be less than that of pure water at the same temperature. The boiling point of water is, accordingly, increased by the addition of the solute.

Vapor pressure and vapor-pressure lowering by the addition of a solute were the subject of considerable study during the last part of the last century and the first part of this century. Raoult, as a result of his experiments, proposed a relation which bears his name. This relation between the vapor pressure of a solution and the composition of the solution can be expressed as

$$\frac{P_0 - P}{P_0} = \frac{n_2}{n_1 + n_2} = N_2 \qquad 1$$

where P_0 is the vapor pressure of the pure solvent and P that of the solution, n_1 is the number of moles of solvent and n_2 the number of moles of the solute. N_2 is the mole fraction of the solute. By rearranging equation 1 it is seen that

$$\frac{P}{P_0} = \frac{n_1}{n_1 + n_2} = N_1 \qquad 2$$

where N_1 is the mole fraction of the solvent. Raoult's equation has been found to hold quite well for dilute solutions. There is a progressive

departure from it as the concentration of the solute is increased. Suppose that we make up a molal solution of some substance in water. The mole fraction of the solute in such a solution is, as we have seen in Chapter II, equal to 0.0177. The vapor pressure of pure water at 25° C. is 23.75 mm. of mercury. Substituting these data in equation 1, we find

$$P_0 - P = 23.75 \times 0.0177$$

$$= 0.420 \text{ mm. Hg} \qquad 3$$

The molecular weight of a compound may be determined by measuring the vapor-pressure lowering of the solvent produced by the addition of a solute. If w_2 grams of solute of molecular weight M_2 are dissolved in w_1 grams of solvent whose molecular weight is M_1, then the respective number of moles of each are

$$n_2 = \frac{w_2}{M_2} \quad \text{and} \quad n_1 = \frac{w_1}{M_1} \qquad 4$$

The Raoult equation becomes

$$\frac{P_0 - P}{P_0} = \frac{w_2/M_2}{w_1/M_1 + w_2/M_2} \qquad 5$$

In dilute solutions where $n_1 >> n_2$, equation 5 reduces to

$$\frac{P_0 - P}{P_0} = \frac{w_2 M_1}{M_2 w_1} \qquad 6$$

Suppose that we dissolve 1 gram of a substance whose molecular weight is 10,000 in 100 grams of pure water, what would be the vapor-pressure lowering?

At 25° C. we have

$$P_0 - P = \frac{23.75 \times 18}{10{,}000 \times 100} \qquad 7$$

$$= 0.000427 \text{ mm. Hg}$$

The vapor-pressure lowering produced by the material with a molecular weight of 10,000 is much too small to be of use as a method for the determination of its molecular weight. The mere trace of a low-molecular-weight impurity would invalidate such measurements. Vapor-pressure lowering is quite unsuited for the determination of the molecular weights of proteins.

The reason why a solute reduces the vapor pressure of the solvent and also why this lowering is proportional to the mole fraction of the solute may be seen from the following argument: The rate of exit out of the

pure solvent is equal to some constant a, while the rate into the surface of the pure solvent is bP_0, where b is a constant and P_0 is the vapor pressure of the pure solvent. The equilibrium constant must equal the ratio of these two rate constants, or

$$K_e = \frac{a}{bP_0} \qquad 8$$

In a solution, the rate out must be proportional to the ratio of the number of solvent molecules to the total number present. The rate out is, accordingly, $an_1/(n_1 + n_2)$, and the rate into the surface is bP, where P is the vapor pressure of the solvent over the solution. The equilibrium constant then becomes

$$K_e = \frac{an_1}{bP(n_1 + n_2)} \qquad 9$$

The two equilibrium constants are equal to each other; therefore

$$\frac{a}{bP_0} = \frac{an_1}{Pb(n_1 + n_2)} \qquad 10$$

and

$$\frac{P}{P_0} = \frac{n_1}{n_1 + n_2} \qquad 11$$

which is a statement of Raoult's law.

There are various ways of measuring the vapor pressure of liquids. We have already discussed this topic to some extent in Chapter XII and shall not elaborate here.

The activity of the solvent can be readily determined from vapor-pressure studies, since it is equal to the ratio of P to P_0.

Depression of the Freezing Point

When a solute is added to a liquid, the freezing point of the solution is found to be less than that of the solvent. As we have seen in Chapter XII, this is one of the methods for estimating bound water. The experimental technique of measuring freezing-point depressions has been developed and refined. An accurate and sensitive thermometer is used by which the freezing point of the solution can be read to a thousandth of a degree. To understand the relation between freezing-point lowering and the vapor-pressure lowering consider Fig. 1.

$T_0 - T$ is evidently the freezing-point lowering and $P_0 - P$ is the vapor-pressure lowering of the solution produced by the addition of the solute. For dilute solutions the vapor-pressure curves are almost parallel, and under these circumstances $(T_0 - T)/(P_0 - P)$ will be equal

to a constant. Substituting $E(T_0 - T)$ for $(P_0 - P)$ in Raoult's law, we have

$$T_0 - T = EP_0 \cdot \frac{n_2}{n_1 + n_2} \qquad 12$$

where E is the proportionality constant relating freezing-point lowering and vapor-pressure lowering. For dilute solutions equation 12 reduces to

$$T_0 - T = \frac{Ew_2}{M_2} \cdot \frac{M_1}{w_1} \qquad 13$$

For purposes of molecular-weight calculations equation 13 is usually written

$$M_2 = K_f \frac{1{,}000w_2}{\Delta T w_1} \qquad 14$$

where ΔT is equal to $T_0 - T$, w_1 is the weight of the solvent, w_2 the weight of the solute, and K_f is a constant whose value varies from solvent

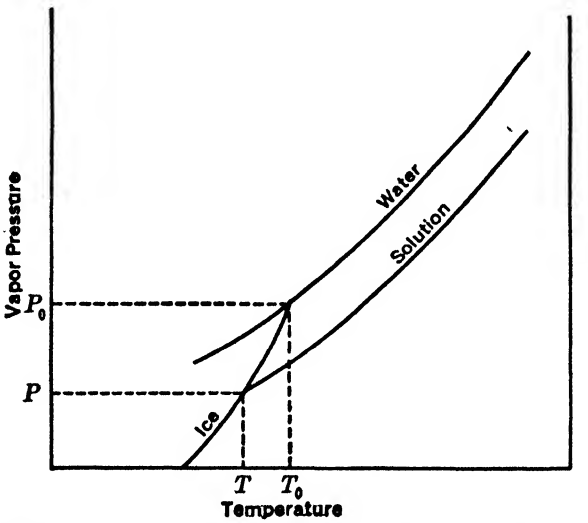


FIG. 1. Relation between the vapor-pressure lowering and freezing-point lowering of a solution.

to solvent. For water it is equal to 1.86°. A gram of pure protein with a molecular weight of 10,000 dissolved in 100 grams of pure water would give a freezing-point lowering of

$$\Delta T = \frac{1.86 \times 1{,}000 \times 1}{10{,}000 \times 100} \qquad 15$$

$$= 0.00186°$$

Quite evidently freezing-point depressions cannot be used to determine the molecular weights of high-molecular-weight compounds.

Osmotic Pressure

If a solution is separated from the pure solvent by a membrane which is permeable to the solvent but impermeable to the solute, the solvent will diffuse into the solution. This diffusion can be prevented by imposing a counter pressure on the solution. The exact pressure required to

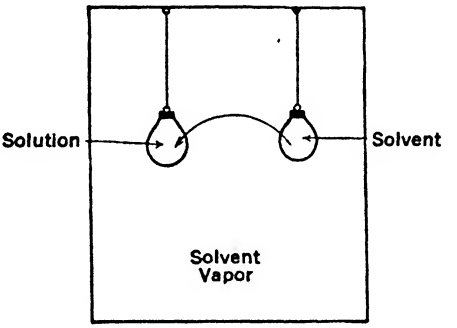


FIG. 2. Solution and solvent are inclosed in separate collodion sacks. The sacks are suspended in a sealed container, and solvent distils into the solution.

prevent this diffusion is known as the osmotic pressure. This phenomenon is another instance of the contribution of the biological sciences to physical chemistry; the first osmotic-pressure studies were made by the botanist W. F. P. Pfeffer in 1877.

Let us attempt to understand as clearly as we can the occasion for osmotic pressure. Consider Fig. 2.

Since the vapor pressure of the solvent is greater than that of the solution, the vapor will distil from the solvent into the solution. A hydrostatic pressure inside the collodion sack containing the solution will develop as the result of this distillation. A time will be reached when the hydrostatic pressure in the solution sack prevents the further distillation of vapor from the solvent to the solution. The system is then in equilibrium. The work required at this point to transfer 1 gram molecular weight of the solvent to the solution is $P_h \bar{V}_1$, where P_h is the hydrostatic pressure inside the solution sack and $\bar{V}_1$ is the partial molar volume of the solvent. Since the system is at equilibrium and the transfer is to be conducted reversibly at constant temperature and pressure, the work done is equal to the free energy change. During this transfer we have changed one mole of solvent from a vapor pressure P_0 to a vapor pres-

sure P, and the free energy of this process as we have seen from Chapter II is

$$\Delta F = RT \ln \frac{P_0}{P} \qquad 16$$

Equating the free energies of the two processes we have

$$P_h \overline{V}_1 = -RT \ln \frac{P}{P_0} \qquad 17$$

Substituting the value of P/P_0 from equation 1, there results

$$P_h \overline{V}_1 = -RT \ln \left(1 - \frac{n_2}{n_1 + n_2}\right) \qquad 18$$

If n_2 is very small compared with n_1 (dilute solution), the logarithm of $\left(1 - \frac{n_2}{n_1 + n_2}\right)$ is very nearly equal to $-\frac{n_2}{n_1 + n_2}$. Then

$$P_h \overline{V}_1 = RT \cdot \frac{n_2}{n_1 + n_2} \qquad 19$$

P_h is evidently the osmotic pressure and $\overline{V}_1$ is very nearly equal to M_1/ρ_1, where M_1 is the molecular weight of the solvent and ρ_1 is its density. Since n_2 is assumed to be small compared to n_1, it can be neglected in the denominator of equation 19. Including all this in equation 19 and rearranging, there results

$$P_{0p} = \frac{\rho_1 RTn_2}{M_1 n_1} \qquad 20$$

where P_{0p} is the osmotic pressure of the solution. Since $M_1 n_1$ equals the total weight of the solvent present in the solution, we have

$$P_{0p} = \frac{\rho_1 RTn_2}{w_1} \qquad 21$$

n_2 is the number of moles of the solute dissolved in weight w_1 of the solvent. Selecting w_1 equal to 100 grams

$$P_{0p} = \frac{\rho_1 RTC}{100 M_2} \qquad 22$$

where C is the number of grams of the solute dissolved in 100 grams of solvent and M_2 is the molecular weight of the solute (n_2 is equal to C/M_2).

If the osmotic pressure is expressed in centimeters of water, R has the value

$$R = \frac{22{,}414 \times 76.0 \times 13.597}{273.1} \qquad 23$$

$$= 8.48 \times 10^4 \text{ cc. cm. } H_2O \text{ per degree}$$

where 22,414 is the molecular volume at 76.0 cm. of mercury pressure and 13.597 is the density of mercury at 0° C. Substituting numerical values in equation 22, we have at 25° C. the working equation

$$P_{0p} = \frac{2.528 \times 10^5 \rho_1 C}{M_2} \qquad 24$$

If 1 gram of protein whose molecular weight is 10,000, is dissolved in 100 grams of water at 25° C.

$$P_{0p} = \frac{2.528 \times 10^5 \times 0.9970 \times 1.00}{10{,}000}$$

$$= 25.20 \text{ cm. } H_2O \qquad 25$$

A high-molecular-weight compound can evidently yield an osmotic pressure which is appreciable and can be conveniently measured. From calculations which we have made we see that 25.20 cm. H_2O osmotic pressure is equivalent to a freezing-point depression of 0.00186° C. and to a vapor-pressure lowering of 0.00058 cm. H_2O. Of these three methods only the osmotic pressure is suited to the determination of the molecular weight of proteins.

Methods for Osmotic-Pressure Measurements

Osmotic-pressure measurements are not conducted in the indirect manner shown in Fig. 2. Actually the solution is placed in a sack made of collodion or some other suitable substance, and the sack is immersed in the solution. The pressure developed inside the sack is measured by an appropriate manometer. The principle, however, is the same; solvent molecules distil from the solvent into the solution, causing a pressure to develop inside the sack. At equilibrium this pressure is equal to the osmotic pressure.

Osmotic-pressure measurements, as we have noted, are particularly suited to the investigation of high-molecular-weight compounds. Membranes can easily be found which will permit passage of water and of buffer molecules as well as inorganic ions but will not allow protein molecules to pass through. If the molecular weight of the protein is not too great, a reasonably dilute solution will yield an osmotic pressure which can be measured with considerable accuracy.

Several arrangements have been used to measure osmotic pressure. Among these may be mentioned the method of Burk and Greenberg,[1] of Pauli and Fent,[2] of Oakley,[3] and of Bourdillon.[4] The author [5] has developed an arrangement which permits accurate measurements, and the time required to reach equilibrium is remarkably short. The apparatus is diagrammed in Fig. 3.

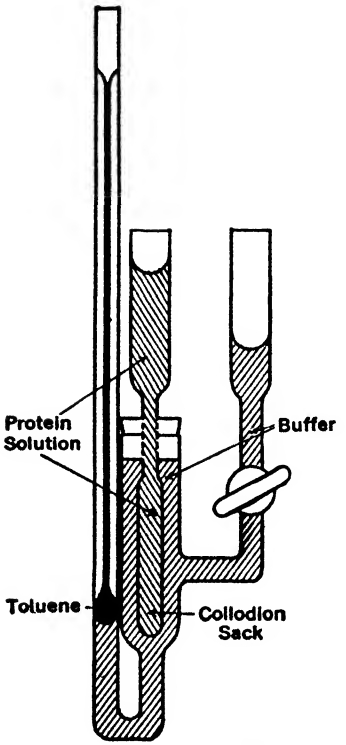


FIG. 3. Osmotic-pressure apparatus.

The apparatus is filled, as shown in Fig. 3, and set in a water bath at constant temperature. The stopcock is open, and the levels of the buffer, of the protein solution, and of the toluene in the capillary are determined with a cathetometer. The stopcock is closed, and the buffer moves into the protein solution in response to the vapor-pressure difference between the solvent and the solution. This movement forces the level of the toluene capillary down until the hydrostatic pressure developed prevents the further distillation of the solvent into the solution. The level of the toluene capillary is then read with the cathetometer, and the osmotic pressure is calculated as being equal to the difference in level between the buffer and protein solution multiplied by the density of the protein solution. To this is added the difference between the initial and equilibrium levels of the toluene multiplied by the density of the toluene. As the capillary bore is about 0.2 mm. in diameter, the amount of solvent transferred from the buffer to the protein solution is so small that it produces no significant change in protein concentration.

Osmotic Pressure of Protein Solutions

In the early work on protein osmotic pressure, confusing results were obtained, and the molecular weights calculated from these results were

[1] N. F. Burk and D. M. Greenberg, *J. Biol. Chem.*, **87**, 197 (1930).
[2] Wo. Pauli and P. Fent, *Kolloid-Z.*, **67**, 288 (1934).
[3] H. B. Oakley, *Trans. Faraday Soc.*, **31**, 136 (1935).
[4] J. Bourdillon, *J. Biol. Chem.*, **127**, 617 (1939).
[5] H. B. Bull, *J. Biol. Chem.*, **137**, 143 (1941).

very discordant. The difficulty arose from a number of sources. Among these may be mentioned: (1) inadequate experimental technique, (2) failure to appreciate the importance of the electrical charges on the protein molecules, (3) incorrect manner of expressing protein concentrations, and (4) impure protein preparations.

For the purposes of calculating the molecular weights of proteins from osmotic pressure, the conditions for measurement and calculations are exacting. Either the experimental arrangement must be so designed that the capillary rise of the liquid due to surface tension either cancels out of the calculation, or the capillary rise must be corrected for. The time of attainment of equilibrium cannot be too long or the protein will be subject to decomposition. The measurement must be made in the presence of sufficient electrolyte to wipe out all electrical effects due to the electrostatic charges on the protein molecules. The degree of purity of the protein cannot be overemphasized. Only well-defined, highly purified proteins should be used. If impure proteins are studied, the experimenter is wasting his time.

In Fig. 4 is shown the osmotic pressure of egg albumin and of serum globulin in urea solutions as a function of protein concentration.[1]

Note that over the concentrations covered the relation between osmotic pressure and the concentration of egg albumin is a linear one as demanded by the simple theory outlined above. A number of proteins, however, show a departure from a linear relation at quite low concentrations. Serum globulin shown in Fig. 4 is an example of such a protein. One can deal with this situation in a purely empirical manner and write equation 22 to read

$$P_{0p} = \frac{g\rho_1 RTC}{100M_2} \qquad 26$$

where g is called the osmotic coefficient and approaches unity as concentration approaches zero. One then plots P_{0p}/C against C, extrapolates the line to zero concentration, and calculates the molecular weight from this ratio and equation 22. It is intriguing, however, to inquire into the reason or reasons for the departure from linearity. The first point of attack is to question the validity of Raoult's law. Fowler and Rushbrooke,[6] by statistical methods, have derived equations for the activities of the constituents of a dilute binary solution in which the solute molecules are elongated, each occupying twice the volume occupied by a single solvent molecule, the molecules of the two kinds being otherwise so similar that there is no heat of mixing. They were able to show that under these conditions deviations from Raoult's law are to be anticipated. The

[6] R. H. Fowler and S. G. Rushbrooke, *Trans. Faraday Soc.*, **33**, 1272 (1937).

deviations are in such a direction as to increase the osmotic pressure. Powell, Clark, and Eyring [7] as well as Huggins [8] have derived equations for solutions of long-chain molecules in terms of the bond angles and freedom of orientation at each joint in the chain. They were able to

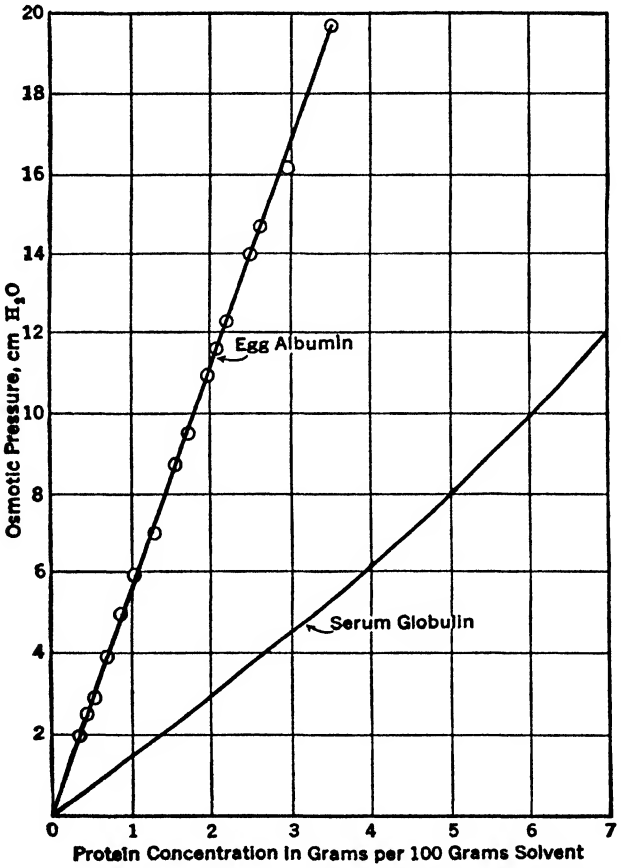


FIG. 4. Osmotic pressures of egg albumin (Bull) and of serum globulin (Burk and Greenberg) as functions of protein concentration at 25° C.

show that these effects would lead to an increase in the osmotic pressure of such solutions. It is doubtful, however, that there is very much flexibility in a protein molecule; too many polar groups are present. The extent to which these ideas are applicable to protein solutions is at present obscure.

[7] R. E. Powell, C. R. Clark, and H. Eyring, *J. Chem. Phys.*, **9**, 268 (1941).

[8] M. L. Huggins, *J. Phys. Chem.*, **46**, 151 (1942); *J. Am. Chem. Soc.*, **64**, 1712 (1942).

The author prefers to explore another approach to this problem. Protein hydration can account for the fact that osmotic pressure increases faster than the concentration. The hydrated protein is more concentrated than calculated, inasmuch as water is associated with the protein. Burk and Greenberg at first offered this explanation for the departure of the pressure-concentration relation from linearity which they observed in a number of their experiments. They calculated the true or effective protein concentration by the following relation on the assumption that the deviation from a straight line was entirely due to hydration

$$C = \frac{100C'}{100 - hC'} \qquad 27$$

where C' is the measured concentration in grams of protein per 100 grams of solvent, and h is the number of grams of solvent held per gram of protein required to give a straight line between concentration and osmotic pressure. They became embarrassed, however, when calculations showed that each gram of gelatin would have to have 4.68 cc. of water of hydration to account for the departure of the osmotic pressure-concentration curve from linearity. This was a great deal more than any one had considered possible. They subsequently abandoned this way of explaining the departure of experiment from theory. It was not realized that what was being measured was the amount of water in a system which was unable to act as a solvent for protein. This is quite a different matter from measuring the amount of water in a system which is incapable of dissolving a small molecule such as glucose. The author has gone through Burk's work and calculated the water associated with various native and denatured proteins on the assumption that the deviations from theory are entirely due to protein hydration. These results are shown in Table 1.

TABLE 1

SWELLING OF PROTEINS IN WATER AS CALCULATED FROM THE OSMOTIC-PRESSURE MEASUREMENTS OF BURK

Protein	Swelling in cc. per gram protein
Casein in urea	2.6
Hemoglobin (native)	1.6
Hemoglobin in urea (denatured)	2.8
Serum albumin (native)	Too little to detect
Serum albumin in urea (denatured)	6.3
Serum globulin (native)	2.0
Serum globulin in urea (denatured)	19.0

Let us look at this question of osmotic pressure and hydration a little more closely. From equation 2, the ratio of the activity coefficient of water in a solution of 1 gram of a protein of molecular weight 10,000 to that in pure solvent is about 0.999982. This lowering has been produced entirely by the colligative effect of the protein. Any water bound to the protein with an activity coefficient less than the above value will tend to increase the osmotic pressure. This conclusion is evident from equation 17. It is entirely conceivable that, if a protein molecule has a very loose structure, considerable water can enter in somewhat the same way as water enters a sponge. The activity coefficient of such water will not be greatly decreased; but it need not be greatly decreased to change the osmotic relations of the protein.

Donnan Equilibrium

In 1911 Donnan [9] published his well-known theory of membrane equilibrium. It has played a conspicuous role in the elucidation of many physical and biological problems. A Donnan equilibrium arises

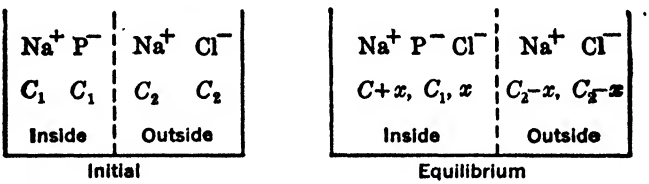


FIG. 5. Diagrammatic representation of the conditions leading to a Donnan equilibrium. The broken vertical line represents a membrane which is permeable to all components of the system except the protein ion, which is denoted by P^-.

whenever an ion or charged particle is constrained in its movements in any way. The classical constraint is a membrane through which the charged particle cannot pass. Consider the very simplified situation shown in Fig. 5.

In order to attain the equilibrium state (Fig. 5), NaCl has diffused into the protein compartment. This diffusion has resulted from the concentration gradient of Cl^-. In order to maintain electrical neutrality an equivalent amount of Na^+ must diffuse along with the Cl^-. This process continues until the concentration gradient of Na^+ outward equals the concentration gradient of Cl^- in the opposite direction. At this point an equilibrium state is reached. Let us turn now from this qualitative discussion to a quantitative one.

The free energy required to transfer one mole of Cl^- from the outside to the inside in a reversible manner is

[9] F. G. Donnan, *Z. Elektrochem.*, 17, 572 (1911).

$$\Delta F = RT \ln \frac{[Cl^-]_i}{[Cl^-]_o} \qquad 28$$

and similarly, the free energy required to transport one mole of Na^+ is

$$\Delta F = RT \ln \frac{[Na^+]_i}{[Na^+]_o} \qquad 29$$

where the subscripts i and o denote the inside and outside compartments, respectively. At equilibrium the total free energy change must be zero and

$$RT \ln \frac{[Cl^-]_i}{[Cl^-]_o} + RT \ln \frac{[Na^+]_i}{[Na^+]_o} = 0 \qquad 30$$

from which

$$\frac{[Cl^-]_o}{[Cl^-]_i} = \frac{[Na^+]_i}{[Na^+]_o} \qquad 31$$

The ratio of the concentration inside to that outside of any univalent anion must equal the ratio of the concentration outside to inside of any univalent cation. This statement is a perfectly general one and can be proved for any pair of univalent ions in precisely the same way as we have shown it to be true for Na^+ and Cl^-. If polyvalent cations and anions are introduced the relation has to be extended, and we find, in general, that for any pair of anions and cations

$$\left(\frac{[\text{Anions}]_i}{[\text{Anions}]_o}\right)^{1/Z_a} = \left(\frac{[\text{Cations}]_o}{[\text{Cations}]_i}\right)^{1/Z_c} \qquad 32$$

where Z_a is the valence of the anion and Z_c is the valence of the cation.

Returning to the simple situation depicted in Fig. 5, we have at equilibrium from equation 31

$$(C_1 + x)x = (C_2 - x)^2 \qquad 33$$

or

$$x = \frac{C_2^2}{C_1 + 2C_2} \qquad 34$$

and also from equation 31

$$\frac{[NaCl]_o}{[NaCl]_i} = \frac{[Cl^-]_o}{[Cl^-]_i} = \frac{C_2 - x}{x} \qquad 35$$

Substituting the value of x from equation 34 into equation 35 and rearranging there results

$$\frac{[NaCl]_o}{[NaCl]_i} = 1 + \frac{C_1}{C_2} \qquad 36$$

C_1 is evidently the equivalent concentration of protein, i.e., the number of charges per molecule multiplied by the molecular weight and divided by the weight concentration, and equation 36 can be rewritten

$$\frac{[NaCl]_o}{[NaCl]_i} = 1 + \frac{\text{Equivalent concentration of protein}}{\text{Initial salt concentration outside}} \qquad 37$$

We conclude from equation 37 that, the greater the equivalent concentration of protein (the greater the non-diffusible charges), the more uneven will be the final distribution of electrolytes inside and outside. On the other hand, if the initial salt concentration is large, the distribution of salt will approach unity at equilibrium and the Donnan distribution will be of no practical importance. For example, if the equivalent concentration of protein is 0.10 and we add varying concentrations of NaCl to the outside compartment, the ratio $[Cl^-]_o/[Cl^-]_i$ at equilibrium will be given in Table 2.

TABLE 2

RATIO OF $[Cl^-]_o$ TO $[Cl^-]_i$ AT EQUILIBRIUM FOR AN EQUIVALENT PROTEIN CONCENTRATION OF 0.10

See Fig. 1.

NaCl added to outside	$[Cl^-]_o/[Cl^-]_i$
0.01	11.05
0.05	3.00
0.10	2.0
1.00	1.1

Owing to the uneven distribution of ions in a Donnan equilibrium there must exist a potential difference between the inside and outside compartment. This potential, in the example we have been considering (Fig. 5), will be

$$E = \frac{RT}{F} \ln \frac{[Cl^-]_i}{[Cl^-]_o} \qquad 38$$

The source and meaning of this potential usually constitute a stumbling block to the uninitiated. The potential arises from the fact that the concentration of Cl^- outside is greater than that inside. The equalization of this ionic concentration difference is prevented by the requirement for electrical neutrality. If, however, reversible electrodes are inserted into the two compartments and the electrodes connected through an external circuit, a mechanism is provided whereby cations can flow

into the inside compartment and out of the outside compartment. This enables the Cl^- to migrate from the outside to the inside compartment in response to its concentration gradient, and a current flows through the external circuit; electrical neutrality is thereby maintained. It must not be thought that, if the current is allowed to flow and the electrodes removed, the same Donnan equilibrium will appear as was originally present. If this were true, an arrangement could be made which would function as a perpetual-motion machine. Remember that ions have flowed into and out of the two electrodes and the situation after current flow is quite different from the original one. A peculiarity is that the electrodes used to measure the potential must not be capable of producing an ion which is involved in the Donnan equilibrium. If such an electrode is used, the potential across the inside and outside compartments will always appear to be zero. Hydrogen electrodes, accordingly, cannot be used because water always contains hydrogen ions which distribute themselves according to the Donnan equilibrium. Suitable electrodes are calomel half-cells connected by KCl salt bridges to the compartments.

We have not discussed the situation which arises when acids or bases are added to the outside compartment instead of neutral salt. The introduction of acids or of bases will evidently complicate the distribution because they will alter the charge on the protein and change its equivalent concentration. The situation can be analyzed, however, and it will be found at equilibrium that

$$\frac{[H^+]_i}{[H^+]_o} = \frac{[OH^-]_o}{[OH^-]_i} = \frac{[Cl^-]_o}{[Cl^-]_i} = \frac{[Na^+]_i}{[Na^+]_o} \qquad 39$$

If the protein is acid to its isoelectric point, the outside solution will be more acid than the inside; and if the protein is basic to its isoelectric point, the outside will be more basic than the inside.

The presence of Donnan equilibrium has important consequences for the osmotic pressure of a system. In general, the osmotic-pressure difference between the inside and outside compartments would be expected to be

$$P_{0p} = RT(\Sigma C_i - \Sigma C_o) \qquad 40$$

where ΣC_i is the total molal concentration of dissolved material inside and ΣC_o that of outside. It is quite clear that a Donnan equilibrium will always increase the osmotic pressure of a protein solution above that expected from the protein alone. In osmotic studies of proteins away from their isoelectric points, it is necessary to add sufficient electrolytes to depress the Donnan equilibrium in accord with equation 37. In

Chapter XVIII, we shall show how it is possible to use the Donnan equilibrium to explain the swelling of protein gels as the gels are shifted from their isoelectric points. The swelling is due to an increase of the osmotic pressure inside the gel.

It appears from experimental work that the osmotic-pressure difference between the inside and outside compartments is not exactly calculable with equation 40. The osmotic-pressure difference is always less than that calculated. This discrepancy, first described by Hammarsten,[10] is known as the "Hammarsten effect." He believed that the presence of a large colloidal ion acted in some way to decrease the activity coefficient of the small ions. This effect has been investigated by several other workers. A careful series of experiments on the osmotic relations of gum arabic has been reported by Briggs.[11] The cause of the Hammarsten effect is still obscure.

Osmotic Relations in Living Cells

It has been pointed out in Chapter II that living cells behave as more or less complete osmotic systems, that is, the living cell is in osmotic equilibrium with its environment. That this is true is shown by the work of Lucke and McCutcheon [12] and of Lucke [13] as well as by others. Some living cells are so perfect in their osmotic behavior that molecular weights of small organic molecules can be determined by a quantitative study of the shrinkage in the volume of the cell produced by an external solution of the compound.

The Donnan equilibrium has been very helpful in understanding the ionic exchange of red blood cells during the respiratory cycle. The red cell, unlike most other body cells, is not only in osmotic equilibrium with its environment but also very nearly in ionic equilibrium as well, so that the equations of the Donnan equilibrium apply directly to this cell. The general application of the Donnan principles to body cells is very questionable, unless it can be shown in each specific case that the application is permissible. The metabolism of red cells is very low compared with that of other body cells; accordingly, equilibrium is more nearly approached. The ionic exchanges of the red cell are discussed at length by Henderson [14] and also by Van Slyke.[15]

[10] E. Hammarsten, *Biochem. Z.*, **144**, 383 (1924).

[11] D. R. Briggs, *Cold Spring Harbor Symposia Quant. Biol.*, I, 152 (1933).

[12] B. Lucke and M. McCutcheon, *Physiol. Revs.*, **12**, 68 (1932).

[13] B. Lucke, *Cold Spring Harbor Symposia Quant. Biol.*, **8**, 123 (1940).

[14] L. J. Henderson, *Blood*, Yale University Press, New Haven, 1928.

[15] D. D. Van Slyke, *Factors Affecting the Distribution of Electrolytes, Water and Gases in the Animal Body*, J. B. Lippincott Co., Philadelphia, Pa., 1926.

Chapter XVII

MEMBRANES AND CELL PENETRATION

No problem is more fundamental to biology than that of the passage of materials into and out of living cells. Although it is true that artificial membranes shed only a limited light on this subject, they are by their own right worthy of as much attention as we can give them. It is proposed to discuss artificial membranes first and then to summarize our knowledge about living cell membranes.

Artificial membranes are of two types. The sieve membrane is made of such substances as parchment paper, collodion, or cellophane. Such membranes are essentially gels having water as the continuous phase. The other type is the homogeneous membrane, which involves the separation of two water phases by a third, non-aqueous phase. An example of such membranes is an oil layer between two aqueous solutions. Although we have made the conventional distinction between these two types of artificial membranes, it should be added that in the limit, as the pore sizes of the sieve membrane are made progressively smaller, the distinction tends to disappear; the sieve membrane becomes a homogeneous membrane. See, however, Sollner and Carr.[1]

Sieve Membranes

Elford[2] has given directions for preparing what he calls gradocol membranes. Such membranes have reproducible characteristics and are, within limits, as dense or open as may be desired. The principle of their preparation depends upon the addition of the proper solvent to the collodion previous to the formation of the membrane. Elford found that, in general, the addition of good solvents for nitrocellulose (ordinary collodion is a solution of nitrocellulose in an ether-alcohol mixture) caused a decrease in membrane porosity, and non-solvents or precipitating agents an increase in porosity. It was possible to compose mixtures of ether, ethyl alcohol, amyl alcohol, and acetone to which were added small quantities of other reagents for the preparation of membranes covering a wide range of porosities (2 μ to 2 mμ). Fine adjustments in porosities were made by changing the evaporation time of the solvents.

[1] K. Sollner and C. W. Carr, *J. Gen. Physiol.*, **26**, 17 (1942).

[2] W. J. Elford, *Trans. Faraday Soc.*, **33**, 1094 (1937); *J. Path. Bact.*, **34**, 505 (1931).

The membranes were formed by pouring a thin layer of the collodion solution on a carefully prepared glass plate and allowing evaporation of the solvents to proceed for a specified time. The film was then covered with water and the solvents washed out. Collodion sacks can be formed inside test tubes; the sacks, after the solvents have evaporated, are treated with water and carefully withdrawn from the test tube. Sacks, however, are unsuited for work requiring uniform and reproducible membranes; disk membranes must be used in such work.

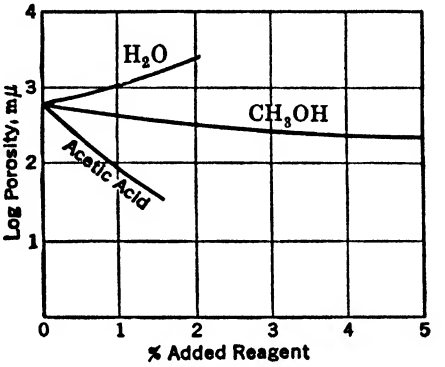


FIG. 1. Porosities of collodion membranes as a function of the composition of the solvent. (Elford.)

Figure 1 shows the range of porosities of collodion membranes having various solvents incorporated in the parent collodion.

McBain and Steuwer [3] found that zinc chloride increased the porosities of cellophane membranes. Cellophane sheets (cellophane is pure cellulose) on the market have a very low porosity, and Seymour,[4] stimulated by the work of McBain and Steuwer, has studied the effect of zinc chloride solutions on the porosity of cellophane membranes. He has outlined the conditions necessary to obtain cellophane membranes of the desired porosity within certain limits.

Pore Diameter

The term pore diameter as applied to such membranes as those of collodion and cellophane is certainly misleading. In all probability, what one deals with are holes and cracks in the membrane of every conceivable degree of irregularity as to shape, size, and pathway through the membrane. (A good physical picture, much enlarged, is presented by a wad of cotton.) Nevertheless the "average pore diameter" (abbre-

[3] J. W. McBain and R. Steuwer, *J. Phys. Chem.*, **40**, 1157 (1936).
[4] W. B. Seymour, *J. Biol. Chem.*, **134**, 701 (1940).

viated APD) of membranes is determined. Although these determinations serve to characterize a membrane, it would be most dangerous to base any physical calculations on such measurements.

The APD calculations involve Poiseuille's law of flow through circular capillaries. As we have seen in Chapter XIII, the integrated form of this law can be expressed by the equation

$$V = \frac{\pi r^4 P}{8\eta l} \qquad 1$$

where V is the volume of liquid in cubic centimeters flowing through the capillary in 1 second, P is the pressure in dynes per square centimeter, r is the radius of the capillary, η is the coefficient of viscosity of the liquid, and l is the length of the capillary. If we have N such capillaries with an average pore radius of r_a, evidently the total volume of liquid, V_t, flowing through all the capillaries is

$$V_t = \frac{\pi r_a^4 PN}{8\eta l} \qquad 2$$

The total volume of all the capillaries is

$$V_c = N\pi r_a^2 l \qquad 3$$

Combining equations 2 and 3 and rearranging, there results

$$\text{APD} = 4l\sqrt{\frac{2V_t\eta}{PV_c}} \qquad 4$$

The value of l is obtained by measuring the thickness of the membrane directly and assuming that the thickness represents the length of the capillaries. This is probably far from correct in view of the tortuous course of the capillaries through the membrane. V_c is obtained by drying the wet membrane and determining the amount of water expelled. Curiously enough, for collodion membranes, V_c is independent of pore size from about 20 mμ to over 1 μ, and the proportion of free space averages about 87 per cent. As can be readily appreciated, the values of APD obtained by the above method should not be taken too seriously; they should be used for comparative purposes only.

A method similar to the above, which, however, utilizes the rate of flow as well as the electrical conductance through the membrane to evaluate APD, was devised by Bull and Moyer.[5]

[5] H. B. Bull and L. S. Moyer, *Science*, **83**, 242 (1936).

The maximum pore diameter can be estimated by measuring the pressure necessary to force air through the wet membrane. The maximum pore diameter is then calculated by the relation (Cantor's law)

$$d = \frac{4\sigma}{P} \qquad 5$$

where σ is the surface tension of the air-water surface in dynes per centimeter, and P is the air pressure in dynes per square centimeter required to force visible bubbles of air through the membrane. This relation is based upon the assumption of circular pores, of which probably none are present.

Erbe[6] has described a method for determining the pore-size distribution of a membrane. The method depends upon the measurement of the pressure required to replace a liquid in a membrane by another, immiscible liquid. A convenient pair of liquids is isobutyl alcohol and water, the two liquids being mutually saturated. The interfacial tension of this combination is 1.8 dynes per centimeter at 25° C. The membrane is thoroughly wet with isobutyl alcohol, and water is forced through the membrane. The pressure is gradually increased, and the rate of flow through the membrane is measured. The rate of flow is then plotted against the pressure, a curve being obtained like that shown schematically in Fig. 2.

Pressure at P_1 evidently corresponds to the largest pore size (see equation 5); that at P_9 is a measure of the smallest pore size. The rate of flow between pressures A and B is divided into ΔS segments, each of which is given by the relation

$$\Delta S = \frac{NPr^4\pi}{8\eta l} \qquad 6$$

where N is the number of capillaries which have been opened up by the displacement of isobutyl alcohol by water in the capillaries at the designated pressure. Since, as we have seen (equation 5),

$$r = \frac{2\sigma}{P} \qquad 7$$

we can substitute relation 7 in equation 6 and rearrange to obtain

$$N = \frac{\eta l P^3 \Delta S}{2\pi\sigma^4} \qquad 8$$

[6] F. Erbe, *Kolloid-Z.*, **62**, 277 (1933).

We then have a distribution of pore sizes as follows:

$$N_1 = \frac{\eta l P_2{}^3 \Delta S_1}{2\pi\sigma^4} \quad \text{with capillaries of radius } r_1$$

$$N_2 = \frac{\eta l P_3{}^3 \Delta S_2}{2\pi\sigma^4} \quad \text{with capillaries of radius } r_2$$

$$N_3 = \frac{\eta l P_4{}^3 \Delta S_3}{2\pi\sigma^4} \quad \text{with capillaries of radius } r_3$$

$$N_n = \frac{\eta l P_n{}^3 \Delta S_{n-1}}{2\pi\sigma^4} \quad \text{with capillaries of radius } r_n$$

If l is unknown, only the relative number of pores is obtained. All studies on the distribution of pore sizes show the membranes to have a

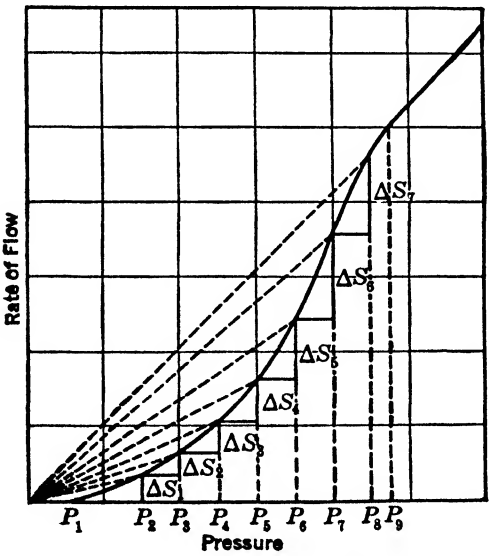


FIG. 2. Rate of flow as a function of pressure. Water is forced through the membrane, displacing isobutyl alcohol.

fairly wide distribution of pore diameters; typically the distribution follows that of a probability curve, the maximum number having a size close to the average pore diameter. The Erbe method involves all the uncertainties of the APD method as well as some of its own. For example, the liquid which is being displaced has a certain viscosity; accordingly, the time required for its displacement from a given capillary, even after the critical pressure has been exceeded, may be appreciable.

Therefore, time must be allowed at each pressure increment for a steady flow to be attained.

From the above discussion, it should be evident that there is no unambiguous method for the measurement of pore sizes in a membrane. Indeed, the concept of pore size itself is a fuzzy one.

For an excellent discussion of the general problem of the flow of liquids through porous solids see the review paper by Sullivant and Hertel.[7]

Electrical Properties of Sieve Membranes

Beginning with the early work of Loeb and others much attention has been centered on the electrical properties of sieve membranes. Michaelis'[8] contributions were especially significant. He noted that well-dried collodion membranes (membranes of very low porosity) showed a differential permeability for cations and anions, the cations penetrating more rapidly than the anions. Accordingly, when KCl solutions of different concentrations were placed on two sides of the membrane, a potential difference was observed across the membrane. With the more dense membranes, at 25° C., this potential approached the theoretical limit of 59 millivolts for a concentration gradient of 1 to 10 across the membrane. (See Chapter VI for a discussion.) These membrane potentials were maintained for long periods of time (several months). Sollner and co-workers,[9] who have reinvestigated this problem, conclude that the differential permeability to ions shown by the dense collodion membranes is due to acid groups in the collodion. It was quite impossible for them to realize such potentials with the purer grades of collodion on the market. If, however, the collodion was oxidized with NaBrO, it acquired the electrical properties of which we speak. Sollner further concluded that the electrical behavior of such membranes could not be ascribed to ion adsorption but must be due to carboxyl groups in the collodion which attract and allow the cations to pass through the membrane and repel the anions. The negatively charged carboxyl group thus stands as a sentinel at the entrance of pores and allows only cations to pass.

Differential ion permeability gives rise to what is known as anomalous osmosis. It has been found that under certain conditions water will

[7] R. R. Sullivant and K. L. Hertel, *Advances in Colloid Science*, Interscience Publishers, Inc., New York, 1942.

[8] L. Michaelis, *J. Gen. Physiol.*, **8**, 33 (1925); **10**, 575, 671, 685 (1927); **11**, 147 (1928); **12**, 55, 221, 473, 487 (1929); *Colloid Symposium Monograph*, **5**, 135 (1928).

[9] K. Sollner et al., *J. Gen. Physiol.*, **24**, 1 (1940); **24**, 467 (1941); **25**, 7 (1941); **25**, 411 (1942).

flow from a concentrated to dilute salt solution. The two solutions are separated by a porous membrane. The flow may continue for several hours and build up a considerable pressure. After a time, however, the pressure tending to force the water from the concentrated to the dilute solution falls to zero and the electrolyte concentration on both sides of the membrane becomes identical. Anomalous osmosis is a kinetic and not an equilibrium condition. If the water flows from the dilute to the

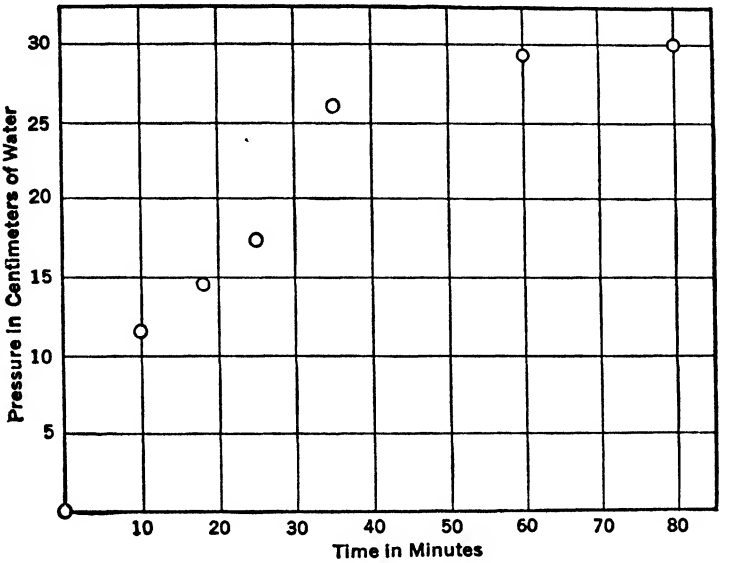


FIG. 3. Negative anomalous osmosis across a porous magnesium silicate membrane. The membrane separated water from a LiCl solution. (Data of Grollman and Sollner.)

concentrated salt solution, the process is called positive anomalous osmosis; if the direction of flow is from the concentrated to the dilute solution, it is called negative anomalous osmosis.

Figure 3 shows the course of negative anomalous osmosis occurring between water and a solution of LiCl separated by a porous membrane of maghesium silicate.[10] It is reported by Grollman and Sollner that the pressure of 300 mm. of water remained unchanged for about 1 hour, after which an appreciable drop occurred, and several hours later the pressure difference had entirely disappeared.

Anomalous osmosis is clearly the result of the membrane potentials which we have discussed above. Such potentials arise, as we have seen, from a differential permeability of the membrane to anions and cations.

[10] A. Grollman and K. Sollner, *Trans. Electrochem. Soc.*, **61**, 477 (1932).

Since the potentials produce an electroosmotic flow of water across the membrane, the direction of flow depends upon the sign of the diffusion potential, which in turn depends on the relative mobility of the anions and cations through the membrane.

The conditions for negative anomalous osmosis may be outlined as follows: If the ion whose sign of charge is the same as that of the membrane has the greater mobility through the membrane, negative anomalous osmosis will occur. In order for this condition to be fulfilled, the pores of the membrane must be fairly large and an electrolyte must be used whose ions show considerable difference in mobility, the greater mobility belonging to the ion whose sign of charge is the same as that of the membrane. A specific example of anomalous negative osmosis is presented by a solution of LiCl or $LiNO_3$ and a magnesium silicate membrane. The mobility of Li^+ ions is much less than the mobilities of either NO_3^- or Cl^- ions. ($V_{Li^+} = 33.4$, $U_{Cl^-} = 65.5$, and $U_{NO_3} = 61.7$.) In spite of the negative charge on the magnesium silicate, the pores of this membrane are so large that the mobility of the anion through it is not significantly diminished and, accordingly, the anions move through the membrane faster than the cations. Consequently the water side of the membrane becomes negatively charged in respect to the salt side. The water being positively charged in the negative pores is drawn from the salt side to the water side by electroosmosis, and negative anomalous osmosis results.

On the other hand, positive anomalous osmosis will occur if the pores of the membrane are very small. In such a situation the ions whose sign of charge is opposite to that of the membrane pores will penetrate faster. This will give rise to a charge on the salt side of the membrane which has the same sign as that on the membrane and opposite to that of the water contained in the pores of the membrane. This will result in an electroosmotic flow of water from the dilute to the concentrated side of the membrane, i.e., positive anomalous osmosis. Naturally, the pore size of a negatively charged membrane need not be small if the cation initially has a greater mobility than the anion; all acids fulfill this condition.

Abrams and Sollner [11] report that, if a collodion membrane is activated with NaBrO and interposed between a 0.01 *M* solution of potassium citrate and pure water, positive anomalous osmosis will result. If counter pressure is not allowed to develop, a considerable transfer of water occurs which may be as high as 90 cc. per 100 sq. cm. of membrane per hour. The transfer is only slightly reduced if water is replaced by a uni-univalent salt or by a sugar solution of the same osmotic activity as that of the citrate solution opposing it.

[11] I. Abrams and K. Sollner, *Am. J. Physiol.*, **133**, 189 (1941).

Ultrafiltration

An ultrafilter is defined as a filter whose pores or interstices are of colloidal or molecular dimensions. Filtration through such a membrane is called ultrafiltration. In general, colloid particles cannot pass through such a membrane, and, accordingly, the filtrate, which is known as the ultrafiltrate, is colloid free. An apparatus which can be used for ultrafiltration is of simple design (see Fig. 4); more elaborate apparatus is required for careful work. Membranes in the form of disks must be used, and they must be well supported to prevent distortion of the membrane resulting from the applied pressure.

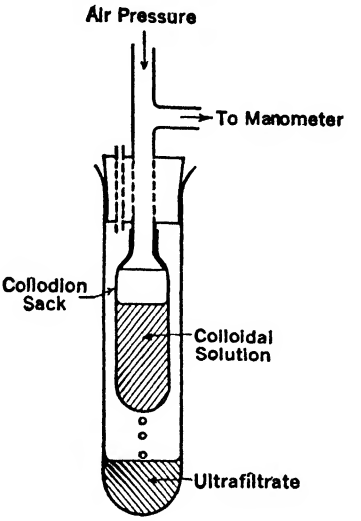


FIG. 4. Arrangement for ultrafiltration.

Ultrafiltration has various laboratory uses, among which may be mentioned the collection of an ultrafiltrate for analytical purposes. The ultrafiltrate should be employed with caution for analysis as it is subject to certain errors. Flexner [12] has considered some of the theoretical and experimental aspects of ultrafiltration. It is evident from his work that it is only at low pressures and moderate permeabilities that the concentration of the salts in the ultrafiltrate becomes substantially identical with that of the intermicellary fluid. A Donnan equilibrium effect must also be absent.

In forcing a given volume of filtrate through a membrane, the total amount of work done, in ergs, is equal to the product of the pressure in dynes per square centimeter and the volume of the filtrate in cubic centimeters. The amount of osmotic work gained in calories is (for dilute solutions where the activities are substantially equal to the concentrations)

$$\text{Work} = RT\left(n_a \ln \frac{A_1}{A_2} + n_b \ln \frac{B_1}{B_2} + \dots\right) \qquad 9$$

where n_a and n_b are the number of moles of substance A and B transferred by ultrafiltration in a liter of solution, A_1 and B_1 are the concentrations of A and B in the mother liquid, and A_2 and B_2 are the concentrations in the ultrafiltrate. (This equation is strictly valid only in the

[12] L. B. Flexner, *J. Biol. Chem.*, **121**, 615 (1937).

absence of Donnan equilibrium.) Flexner, in his investigation, found a very low efficiency; i.e., the osmotic work gained was only a small fraction of the work done.

Borsook and Winegarden [13] have considered the work done in the secretion of urine by the kidney. As will be recalled, the glomerular filtrate is essentially an ultrafiltrate of the blood. The only work in this step is that involved in overcoming the protein osmotic pressure of the plasma plus the mechanical work of transferring a given volume of fluid from the blood to the urine. Both these work terms are small. As the ultrafiltrate passes through the kidney tubules, considerable water is

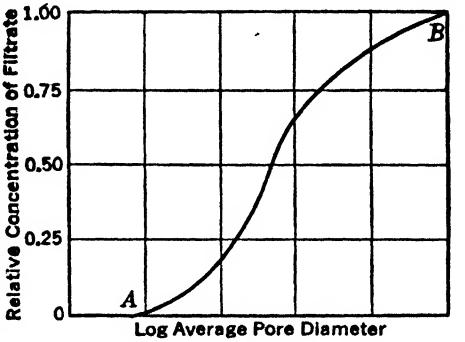


FIG. 5. A Schematic Filterability Curve.

reabsorbed into the blood stream along with such molecules as those of glucose, salts, and amino acids. This work involves the expenditure of about 700 calories per liter of urine finally excreted. Borsook and Winegarden point out that this is only a small fraction of the energy consumed by the kidney and that the kidney is only 1 or 2 per cent efficient.

Dense membranes will, of course, completely prevent the passage of colloid particles into the ultrafiltrate, whereas very open membranes will not hinder the passage of such particles in any way. Between these limits are porosities which will permit some of the colloid to pass but its concentration in the ultrafiltrate will be smaller than in the original solution. A quantitative measure of the filterability of a given dispersed phase is obtained by filtering the solution through a series of membranes of different porosities. The maximum relative concentration of the colloid in the filtrate is determined for each experiment. The relative concentrations are then plotted against the logarithm of the average pore diameter. The porosity of the most porous of the membranes retaining

[13] H. Borsook and H. M. Winegarden, *Proc. Natl. Acad. Sci. U. S.*, **17**, 313 (1931).

the colloid completely corresponds to what is called "the filtration end point." Figure 5 is a schematic representation of such a filterability curve. Point *A* represents the filtration end point, and point *B* that of total filtration.

Grabar [14] has used such filterability curves to estimate the sizes of a number of protein molecules. Curiously enough a membrane has to be "lubricated" with such substances as Na-oleate or Na-glycocholate before the protein molecules will pass through the membrane. The "lubricating" agent does not change the pore size of the membrane but probably acts by decreasing the adsorption of the protein on the membrane.

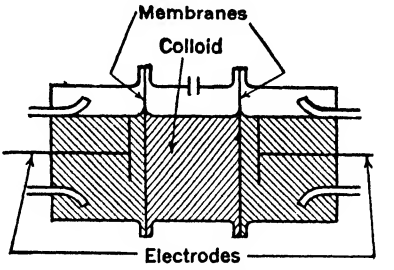


FIG. 6. Electrodialysis apparatus. (Pauli.)

Dialysis and electrodialysis are two very convenient techniques for purifying a colloidal solution. In dialysis, a colloidal solution is placed in a collodion or cellophane sack and the sack is immersed in pure water. Artificial sausage casings make excellent dialysis sacks. The small molecules such as salts and sugars diffuse out of the sack. The outside water is frequently changed, and the colloid solution is thus progressively purified of the smaller, diffusible molecules. Electrodialysis employs somewhat the same arrangement, but the removal of the electrolytes is hastened by passing an electrical current through the colloid solution. A scheme for rapid dialysis has been published by Lauffer.[15] Figure 6 shows an apparatus for electrodialysis (used by Wo. Pauli).

Homogeneous Artificial Membranes

We have made a distinction between sieve membranes and homogeneous membranes. A homogeneous membrane consists of a single non-aqueous phase separating two aqueous phases. Actually, as the pore sizes of the sieve membrane become progressively smaller, such a membrane approaches the condition of a homogeneous membrane. We might properly speak of a substance which penetrates such a membrane as being "soluble" in the membrane, the solute passing between the "molecules" of the sieve membrane in much the same sense as it does between the molecules of a homogeneous membrane.

Osterhout and Stanley [16] undertook to devise an artificial homogeneous

[14] P. Grabar, *Cold Spring Harbor Symposia Quant. Biol.*, **6**, 252 (1938).
[15] M. A. Lauffer, *Science*, **95**, 363 (1942).
[16] W. J. V. Osterhout and W. M. Stanley, *J. Gen. Physiol.*, **15**, 667 (1932).

membrane which would simulate some of the properties of living membranes. They found that a mixture of 70 per cent guaiacol and 30 per cent p-cresol fulfilled their expectations to a surprising degree. It has been reported, for example, that the giant plant cell *Valonia* accumulates potassium and sodium from sea water, the potassium being in greater concentration than the sodium. The homogeneous membrane of Osterhout also accumulates potassium and sodium. The arrangement of the model membrane is shown in Fig. 7.

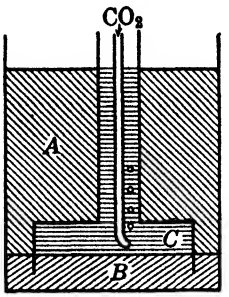


FIG. 7. Osterhout-Stanley model for a living membrane. *A* is a 0.04 *M* KOH solution. *B* is a non-aqueous layer of P-cresol and guaiacol. *C* is water into which CO_2 gas is bubbled.

The KOH in *A* reacts with the guaiacol to give the potassium salt. This potassium salt then diffuses until it comes into contact with *C* and potassium bicarbonate is produced and is released into compartment *C*. This leads to an accumulation of potassium in compartment *C*, the accumulation being brought about through the action of CO_2. The CO_2 is supposed to represent the end products of metabolism. Figure 8 shows a graph of some of their results using 0.05 *N* NaOH and 0.05 *M* KOH in compartment *A*.

Although the Osterhout model does not reproduce all the ramifications of electrolyte penetration into living cells, it is useful in demonstrating one of the outstanding features of such penetration, namely, the dependence of electrolyte accumulation upon an active mechanism which is represented in the model by a continuous supply of CO_2.

Since there are many examples of living membranes, the selection of a "typical" living membrane is indeed difficult. Physiologists have sought as diligently for such a membrane as Diogenes did for an honest man and with about the same success. The red blood cell membrane has caught the attention of some investigators. Red cells have the advantage that they are readily available and their study is comparatively simple. The rate of metabolism of these cells, however, is very low, and, as a consequence of this and other factors, the permeability of red cells to electrolytes and small organic molecules is not typical of most body cells. The red cell is as nearly in equilibrium with its environment as it is possible for a living cell to be.

There are many membranes in the animal body across which molecules are transported. The gastric mucosa secretes into the stomach acid which is of the order of four million times more concentrated than is the hydrogen-ion concentration of blood. The parietal cells of the stomach mucosa do a minimum of about 1,700 calories of work in the secretion

of a liter of gastric juice. It is evident that the gastric mucosa is a highly specialized membrane.

The membrane which separates the yolk from the white in a hen's egg does an astonishing job. The *p*H of the white of an egg is about 8.5 (the fresher the egg the higher the *p*H), while that of the yolk is about 5.5. Thus a thousandfold difference in hydrogen-ion concentration is

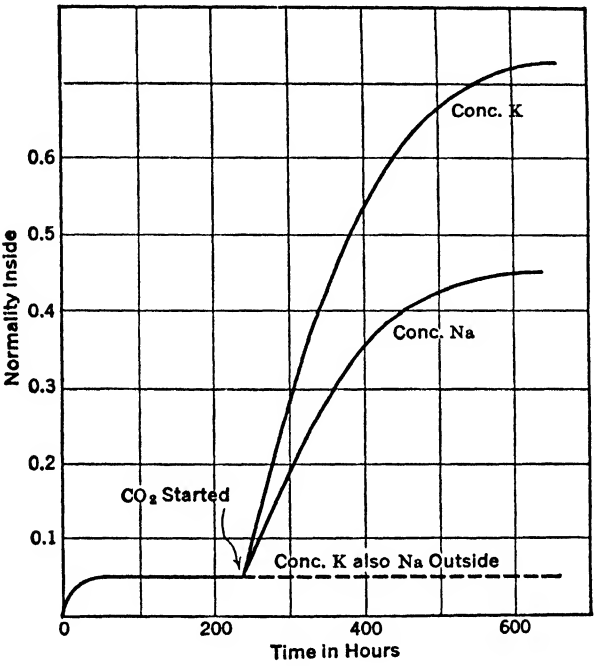


FIG. 8. Graph showing accumulation of sodium and potassium in compartment *C* (see Fig. 7) due to the presence of CO_2. (Osterhout and Stanley.)

maintained between the white and the yolk. This membrane is not typical of anything except the hen's egg.

Some physiologists work with plant root membranes; others prefer such material as slices of potatoes.

The large marine plant cells of *Valonia* and of *Halicystis* have proved very useful for certain types of study. These cells grow in warm clear sea water on protected parts of the coast. Single cells are frequently as large as grapes. In these cells the protoplasm forms a thin layer having an inner, non-aqueous surface, an aqueous layer, and an outer non-aqueous surface. The protoplasm is protected by a cellulose wall which is apparently so permeable that it can be neglected in the study of the permeability of the cell membrane. The inside of the cell is filled with

sap which makes up by far the largest part of the volume of the cell and is essentially a dilute salt solution. Osterhout and his co-workers have studied these cells with much profit. Their papers can be found for the most part in the *Journal of General Physiology.*

Penetration through Living Membranes

This part of the discussion can be conveniently divided into three parts: (1) water penetration, (2) electrolyte penetration, and (3) penetration of unionized organic molecules.

All living cells are permeable to water, and, in general, no active mechanism is involved; water is transferred into and out of cells in response to differences in osmotic pressure. The subject has been ably reviewed by Lucke and McCutcheon [17] and more recently by Lucke.

Not only do living cells swell or shrink in response to a decrease or increase of the osmotic pressure of the external medium but the relation is, in general, a quantitative one:

$$P_e(V_e - b) = P_0(V_0 - b) \qquad 10$$

where P_e is the known osmotic pressure of the external medium, V_e is the expected equilibrium volume of the cell, P_0 is the original osmotic pressure of the external medium, V_0 is the original volume of the cell, b is a constant which is in the nature of a correction for the volume occupied by the osmotically inactive materials in the cell. The value of b varies considerably with different varieties of cells. Thus for sea-urchin eggs it is about 12 per cent of the initial total volume of the cell while for mammalian red blood cells it is much larger—about 45 per cent.

The rate of uptake of water by living cells in response to osmotic changes varies from one type of cell to another but is usually surprisingly rapid, complete adjustment frequently requiring only a few minutes. The osmotic pressure of the external medium can be controlled by electrolyte addition or dilution and the osmotic response of the cell is still in accord with equation 10. This indicates that the rate of water penetration through cell membranes greatly exceeds that of electrolyte penetration.

Much has been written about the penetration of electrolytes through cell membranes. One of the outstanding features of living cells is their high electrical resistance, which indicates that ions enter and leave living cells with difficulty. The summary of the problem of electrolyte penetration, as outlined by Stewart,[18] appeals very strongly to the

[17] B. Lucke and M. McCutcheon, *Physiol. Revs.*, **12**, 68 (1932).
B. Lucke, *Cold Spring Harbor Symposia Quant. Biol.*, **8**, 123 (1940).
[18] F. C. Stewart, *Trans. Faraday Soc.*, **33**, 1006 (1937).

author. Stewart points out that the accumulation of electrolytes by living cells is not, in general, an example of simple diffusion. The concentration within the cell may be a thousandfold greater than that of the external medium and yet accumulation of electrolytes will proceed. The accumulation is closely related to the growth of cells. It is also closely related to the metabolism of cells. If the oxygen supply is decreased below a certain level, the accumulation ceases. In fact, if the oxygen supply is decreased sufficiently, salts are lost by the cell. There is, however, no simple relation between salt accumulation and respiration, but rather a general parallelism. Active cells produce CO_2 equal to many times the chemical equivalent of salts absorbed. Salt accumulation is related to the total acrobic respiration but the energy expanded on salt absorption is only a small fraction of that represented by the total metabolism of tissue. For example, the energy used in salt accumulation by potato slices is only about 0.2 per cent of the total energy of respiration of the slices.

The metabolism of organic acids seems to be involved. When anions not accompanied by cations are absorbed, the tissue metabolizes organic acids and the loss of organic anions balances the discrepancy between anion and cation uptake. The evidence indicates that equal or unequal uptake of anion and cation is not merely a property of the ions concerned but is determined also by the metabolism of the tissue.

Stewart goes ahead to point out that the ion series which represents the ease of uptake of salts with a common anion $K^+ > Na^+ > Ca^{++} > Mg^{++} > Sr^{++} > Li^+$ or with a common cation $NO_3^- > Cl^- > HCO_3^- > HPO_4^{--} > SO_4^{--}$, and which are commonly ascribed to differences in ionic mobility, need to be re-examined as to their effects on cell metabolism.

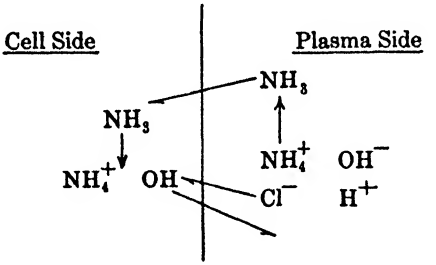
The test of whether a given accumulation or secretion of electrolytes is a spontaneous process or one which requires work must, of course, be the sign of the free energy involved. The free energy of any such transfer is

$$\Delta F = n_1 RT \ln \frac{A_2}{A_1} + n_2 RT \ln \frac{B_2}{B_1} + n_3 RT \ln \frac{C_2}{C_1} \qquad 11$$

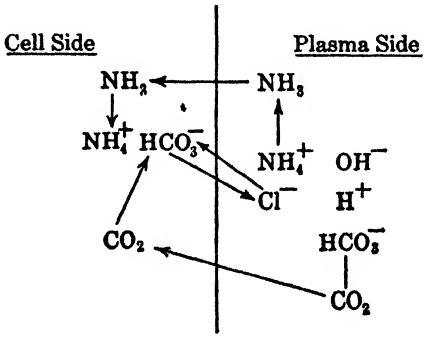
where n_1, n_2, n_3, etc., are the number of mole ions of A, B, C, etc., respectively transferred in a given volume of the solution; A_1, B_1, C_1 represent the concentration of the ions in their original situation; and A_2, B_2, C_2, that in their final situation. Exact calculations require A, B, C to be expressed in activities, but the accuracy of the measurements seldom justifies this refinement and concentrations can be used on most occasions. If ΔF is zero, the system is at equilibrium (the situa-

tion for the Donnan equilibrium). If it is negative, the transfer of ions is spontaneous and requires no work on the part of the cell. If, on the other hand, it is positive, it means that the cell has had to do work to bring about the transfer, the magnitude of which is given by ΔF in calories. In general, such a process is not very efficient, and several times the work calculated is actually required. One of the important unsolved problems of physiology is how the cell does work—work of any kind. The problem of heat production has been extensively investigated and is understood; but how the body transforms chemical energy into mechanical or into osmotic work is a complete blank.

Jacobs [19] has called attention to a mechanism by which the transport of ions across membranes can be catalyzed by the presence of CO_2. It is well known that cell membranes are much more permeable to undissociated weak acids and bases such as NH_3, CO_2, and H_2S than they are to ions. It has been found, for example, that the rate of hemolysis of mammalian erythrocytes in solutions of ammonium chloride can be greatly increased—50 times or more—by the addition of a small amount of bicarbonate. The mechanism for the entrance of NH_4Cl into the red cell can be pictured as follows:



If a little NH_4HCO_3 is added to the plasma, we have



[19] M. H. Jacobs, *Cold Spring Harbor Symposia Quant. Biol.*, **8**, 30 (1940). M. H. Jacobs and D. R. Stewart, *J. Gen. Physiol.*, **25**, 539 (1942).

Jacobs points out that the mass law requires that at equilibrium the product (NH_4^+) inside $\times$ (HCO_3^-) inside must equal (NH_4^+) outside $\times$ (HCO_3^-) outside. Since (NH_4^+) outside is initially much greater than (NH_4^+) inside, there will be a tendency to force (HCO_3^-) inside above (HCO_3^-) outside. Since the red cell is anion permeable, this would lead to an exchange of HCO_3^- for Cl^-. As a final result of these two processes, NH_4Cl has entered the cell and HCO_3^- is back again on the outside ready to repeat the cycle. Jacobs' theory may have rather wide application.

Permeability of Cells to Non-Electrolytes

An excellent reference to this topic is an article by Collander,[20] who investigated the penetration of 45 non-electrolytes into cells of the alga *Chara ceratophylla*. All experimental data seem to indicate that non-electrolytes penetrate by a simple diffusion process in which the only significant concentration gradient is that across the protoplasm itself. Thus Ficks' law is directly applicable in the form

$$P = \frac{V}{qt} \ln \frac{C}{C - C_1} \qquad 12$$

where V is the volume of the cell in cubic centimeters, q is the surface of the cell in square centimeters, C_1 is the internal concentration of the penetrating substance at time t, C is the equilibrium internal concentration (almost identical with its external concentration), and P is a constant which measures permeability and which is, therefore, termed the permeability constant.

Collander also determined the distribution of the various non-electrolytes between water and olive oil. This distribution was accepted as a provisional measure of their relative lipid solubility. The molar index of refraction was used as a measure of the molecular volume of non-electrolytes. The results are plotted in Fig. 9.

There is clearly a somewhat close correlation between oil-solubility of the substance on one hand, and their permeability constants on the other. This is not merely a general concordance but, at least approximately, a direct proportionality. On the other hand, the smallest molecules obviously penetrate faster than would be expected on account of their oil-solubility alone. It seems, therefore, that whereas the medium-sized and large molecules penetrate the plasma membrane only when dissolved in lipids, the smallest molecules can penetrate it in some other way also. Thus, the plasma membrane seems to act both as a selective

[20] R. Collander, *Trans. Faraday Soc.*, **33**, 985 (1937).

solvent and as a molecular sieve. The author suggests that there need not be a separate mechanism for the penetration of the two sizes, but that the problem can be understood on the basis of a combination of a sieve and lipid (non-aqueous) membrane in series, the oil being the limiting factor for the small molecules and the sieve and the oil being the factors governing the penetration of large molecules. It is possible that cellulose constitutes the sieve for plant cells and molecular layers

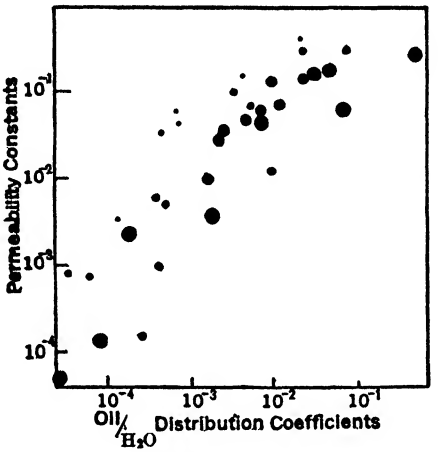


FIG. 9. Permeability of the cells of *Chara ceratophylla* to various non-electrolytes. Ordinate indicates the permeability constants (centimeters per hour). The abscissa gives the oil-water distribution of the substances. The relative sizes of the circles indicate the relative sizes of the molecules. (Collander.)

of proteins that for animal cells. Sometimes, of course, non-electrolyte penetration does involve an active mechanism. For example, the uptake of glucose from the intestines of mammals seems to depend on such an active mechanism.

Bellows and Gutmann [21] were able to increase considerably the penetration of sulfonamides through the corneal membrane of the eye by local application of the sulfonamide in the presence of a wetting agent. The wetting agents tried included among others Aerosol OS, Aerosol OT, Tergitol, and bile salts. The application of this work to the treatment of infections of the eye is obvious.

[21] J. G. Bellows and M. Gutmann, private communication.

Chapter XVIII

COLLOIDAL STRUCTURES

In the chapter on colloidal solutions we discussed systems in which there was little or no interaction between colloidal particles. In the present chapter we deal with systems in which this interaction is the most conspicuous feature. Such interaction leads characteristically to the formation of a coagulum or to gelation.

Coacervates

Although coacervates have little or no structure, it is convenient to discuss them at this point. Coacervation was first described as such by Bungenberg de Jong and Kruyt.[1] Most of the work on coacervation has been done by Bungenberg de Jong and his co-workers; for reviews on the subject see papers by Bungenberg de Jong [2] and by Koets.[3] The word coacervation, derived from the Latin "acervus" (heap or swarm) combined with the prefix "co" (together), means, literally, a swarming together. We have seen in our discussion of colloidal solutions that, in general, hydration and electrostatic charges are responsible for the stability of hydrophilic colloids. If two hydrophilic colloids of opposite sign are mixed, there is a marked tendency for the colloidal particles to precipitate each other. On the other hand, the water of hydration may prevent a coagulation of the colloids. The result is a weak union between the oppositely charged particles. The coacervate particles separate first into fine droplets which gradually coalesce to form a separate phase. There will, however, be a series of equilibrium values for the coacervate as the ratio of the components of the coacervate is changed; and, for every variation of the equilibrium liquid, there is a corresponding variation in the coacervate. Frequently, a small amount of the colloid remains suspended in equilibrium with the coacervate.

As we have noted, coacervates can be formed by mixing two hydrophilic colloids of opposite sign. An example of such a coacervate is that which forms between gelatin and gum arabic at the proper *p*H. The

[1] H. G. Bungenberg de Jong and H. Kruyt, *Proc. Acad. Sci. Amsterdam*, **32**, 849 (1929).

[2] H. G. Bungenberg de Jong, *Protoplasma*, **15**, 110 (1932).

[3] P. Koets, *J. Phys. Chem.*, **40**, 119 (1936).

gum arabic is negatively charged at all *p*H values while the gelatin is positively charged below its isoelectric point, which is at *p*H 4.8; accordingly, below a *p*H 4.8 a coacervate will form between gelatin and gum arabic whereas while above a *p*H 4.8 the solution will remain perfectly clear.

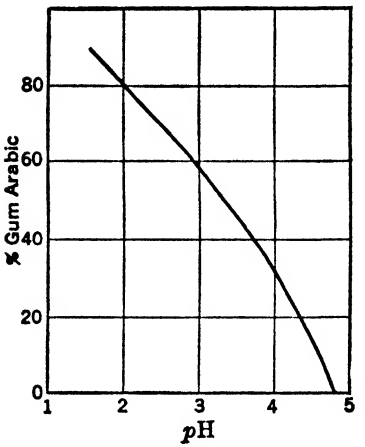


FIG. 1. Optimal ratio of gum arabic to gelatin as a function of *p*H. (Bungenberg de Jong.)

The mixing of oppositely charged colloids results in a partial or complete neutralization of the charge on the coacervate. Electrophoretic experiments thus show that the coacervate may be positive or negative; but, for optimal coacervation, the resulting droplets are uncharged. Figure 1 shows the ratio of gelatin to gum arabic which gives optimal coacervation at various *p*H values.

The formation of a coacervate is easily followed by viscosity measurements. The viscosity of the mixture of gelatin and gum arabic is determined in a region where no coacervate forms, *p*H 5. Then the viscosity of the mixture of the same composition is determined in the *p*H region where coacervation takes place. It is found that the viscosity of the coacervate

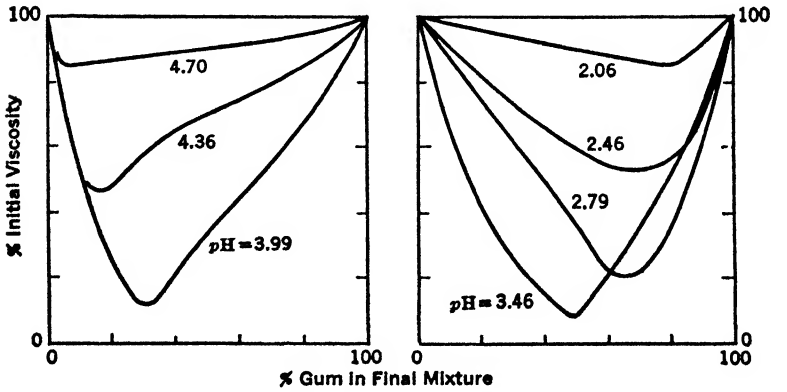


FIG. 2. Viscosity of gum arabic-gelatin mixtures as a function of *p*H. The greatest decrease in viscosity represents the optimum coacervate formation. (Bungenberg de Jong.)

is always less than that of the corresponding mixture in which no coacervate was able to form, and, in general, the decrease of the viscosity is proportional to the extent of coacervate formation. These relations are

shown in Fig. 2. The decrease in viscosity upon coacervate formation results from the loss of hydration of the hydrophilic colloids incidental to their union, i.e., water is squeezed out between the particles.

Coacervate formation is reversible. A coacervate can be dispersed by the addition of electrolytes which act, as we have seen, to decrease the electrostatic influence and, accordingly, decrease the attraction of the oppositely charged particles for each other. The higher the valence, the more effective is the ion in dispersing the coacervate.

Coacervates are classified by Bungenberg de Jong into simple and complex coacervates. Simple coacervates result from a dehydration of the colloid by means of a molecularly dispersed substance or by a second hydrophilic colloid of the same sign, whereas complex coacervates result from a dehydration brought about by a second hydrophilic colloid of opposite sign (gelatin and gum arabic) or by the addition of a polyvalent ion. An example of this last is the addition of a hexo salt of cobalt to gum arabic.

Bungenberg de Jong points out that there are a number of analogies between coacervates and protoplasm. For example:

1. The coacervate is a colloid-rich liquid which is not misicible with the equilibrium liquid.
2. Vacuoles appear in the coacervate in response to a rise in temperature or when an external electrical field is applied.
3. Oil drops are taken up by coacervates just as living cells engulf such drops.

On the other hand, there are many obvious differences between coacervates and protoplasm. Some of these differences are:

1. Coacervates are true liquids in that the rate of flow is proportional to the applied force. This is not, in general, true of protoplasm.
2. Coacervates show no preferential permeability.

Tactoids

Large boat- or spindle-shaped particles have been observed to form in concentrated solutions of iron oxide and of vanadium pentoxide, and also in solutions of several dyes. Such aggregations of particles were called tactoids by Zocher,[4] who first described them. Tactoids are optically anisotropic as revealed by the polarizing microscope. This indicates that they are composed of oriented particles, and the direction of orientation coincides with the long axis of the tactoid.

Bernal and Fankuchen [5] report that tobacco mosaic virus protein exhibits very well-defined tactoids. The larger the virus tactoids, the

[4] H. Zocher, *Z. anorg. Chem.*, **147**, 91 (1925).
[5] J. D. Bernal and I. Fankuchen, *J. Gen. Physiol.*, **25**, 111 (1941).

more nearly they approach a spherical form; the smaller they are, the more cylindrical they become. Bernal and Fankuchen have outlined a theory of their formation.

A phenomenon related to tactoid formation is the formation of Schiller planes.[6] In aged iron oxide or tungsten oxide solutions, it is found that the particles tend to occur in horizontal planes which show a vertical periodicity of 0.2 to 0.4 μ. The effect is so pronounced that they frequently show interference colors. Schiller planes can also be produced at will if iron chloride solutions of 0.3 to 1 per cent are slowly hydrolyzed. The forces which bring about the formation of Schiller planes are undoubedly electrostatic in nature and are of the kind described by Levine and Langmuir (see Chapter XII). The particles are in energy troughs produced by the minimum in the resultant of the electrostatic repulsive and attractive forces. In platy particles, these equilibrium distances can reach 5,000 Å or more and are quite sufficient to account for the formation of Schiller planes.

Gels

Gels are usually classified as elastic or swelling gels and as inelastic or non-swelling gels. An example of the inelastic type is silicic acid gel. In this discussion we shall confine ourselves to swelling gels.

The classic example of a swelling gel is presented by gelatin in water. Such a gel is typically formed by dissolving the gelatin in hot water and allowing the solution to cool. There are many examples of swelling gels. Almost any protein can be made to form a gel in water if the proper technique is used. A gel has a definite structure which resists mechanical deformation. If the displacement produced by a mechanical force is not too great, the gel will usually restore itself to its original form when the stress is removed.

It is evident that, in order for a gel to form, there has to be a marked interaction between the colloidal particles present. Numerous theories of gel structure have been advanced, such as the "brush-heap" entanglement of highly elongated particles, the solvation theory of large particles, and the formation of network structures which extend throughout the volume, immobilizing the entrapped liquid. Bernal and Fankuchen believe that the formation of tobacco mosaic virus gels is due to the fusion of tactoids which produces a protein network throughout such a gel. All these theories have one point in common; they all stress the importance of large colloidal particles.

[6] H. Freundlich, *Kapillarchemie*, Bd. II, *Academische Verlagsgesellschaft*, Leipzig, 1932.

Flory [7] has proposed a theory of gelation and has defined certain conditions which are necessary for gelation. He points out that polymerizations which are propagated through the intermolecular reactions of "bivalent" or bifunctional molecules lead to soluble, diffusible products, whereas the incorporation of units of higher functionality permits formation of gelled or insoluble products.

Consider, as Flory has done, a substance containing groups A and B. $A—A$ and $B—B$ represent bifunctional units while $A—\begin{matrix}A\\A\end{matrix}$ represents the trifunctional branch unit. AB or BA is the product of the condensation of two functional groups. We then have the situation pictured in Fig. 3.

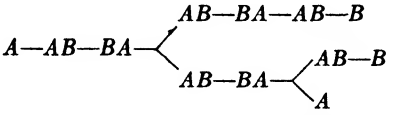


FIG. 3. Trifunctionally branched polymers. (Flory.)

Let α be the probability that a chain leading from a trifunctional unit eventually leads to another branched unit. If $\alpha < \frac{1}{2}$, the indefinite extension of a network is impossible since a given chain has less than an even chance of reproducing two new chains. On the other hand, if $\alpha > \frac{1}{2}$, branching of successive chains may continue the network indefinitely. The condition $\alpha = \frac{1}{2}$ is, therefore, critical for the formation of infinite structures or for the formation of a gel. Flory gives the equation for the general case, which turns out to be

$$\alpha_c = \frac{1}{(f-1)} \qquad 1$$

where f is the functionality of the branching units interpolymerized at random with bifunctional units. α_c is the critical probability above which a gel may form and below which no gel can form. As proteins, in general, contain many reactive groups, the possibility of gel formation is large.

Ordinarily when a solution changes to a gel there is little change in the gross physical and chemical properties of the system aside from the obvious appearance of mechanical structure. The electrical conductance through a gel in the presence of electrolytes is substantially the same as it is in the liquid condition. On the other hand, if the electrolyte content of a gelatin solution is quite small, the solution has a slightly higher

[7] P. J. Flory, *J. Phys. Chem.*, **46**, 132 (1942).

conductance than the gel.[8] Freundlich and Abramson [9] found that the electrophoretic mobility of quartz particles is the same through a gelatin gel as through the gelatin solution. They believed there was a local liquefaction of the gel in the immediate neighborhood of the particles which permitted their unhindered movement through the gel.

Freundlich,[10] on the basis of Heyman's [11] studies on the volume changes on gelation, was able to classify gels into three classes. Thixotropic gels show no volume change when gelation or solution occurs at a fixed temperature. The second type, exemplified by gelatin or agar, displays a small decrease of volume with setting; heat is evolved in the process. The third type which is represented by aqueous solutions of methylcellulose shows a small increase of volume and an absorption of heat during gelation. Thus, it is found that gels of the second type tend to set upon cooling, whereas those of the third type liquefy as the temperature falls.

Swelling of Gels

A very characteristic feature of proteins is their tendency to imbibe water and to swell. Dry isoelectric gelatin when placed in water will imbibe water and swell to a definite limit. Some other proteins, such as egg albumin, when placed in water will swell indefinitely and will eventually disperse completely. The reason why both the proteins swell in water is that polar groups on the protein molecule are hydrated and water is bound through hydrogen bonds (see Chapter XII). The reason gelatin swells to a definite limit is that intermolecular bonds between the gelatin molecules prevent further extension. These bonds, however, are comparatively weak and are fairly easily broken. If the temperature of the water in contact with the gelatin is increased above about 45° C., the gelatin disperses completely; the intermolecular bonds are broken at this low temperature. The intermolecular bonds re-form when the temperature is lowered. The dispersion of a gel at higher temperatures is roughly analogous to the melting of a crystal, although the solution temperature is far from sharp and the solution temperature is always appreciably higher than the gelation temperature.

If the *p*H of the completely swollen isoelectric gelatin gel is shifted either to the acid or basic side of the isoelectric point, additional swelling is noted. In Fig. 4 are shown some of the results of Jordan Lloyd and

[8] D. M. Greenberg and M. A. Mackey, *J. Gen. Physiol.*, **15**, 161 (1931).

[9] H. Freundlich and H. A. Abramson, *Z. physik. Chem.*, **133**, 51 (1928); *J. Gen. Physiol.*, **11**, 743 (1928).

[10] H. Freundlich, *J. Phys. Chem.*, **41**, 901 (1937).

[11] E. Heyman, *Trans. Faraday Soc.*, **31**, 846 (1935); **32**, 462 (1936).

Pleass [12] on the influence of *p*H on the swelling of gelatin. Also included is the effect of 0.1 *M* NaCl.

In the past there was an enormous amount of controversy on the cause of protein swelling. The correct view seems to be that the initial isoelectric swelling is simply hydration of the polar groups of the protein plus the osmotic pressure due to the protein alone. The additional

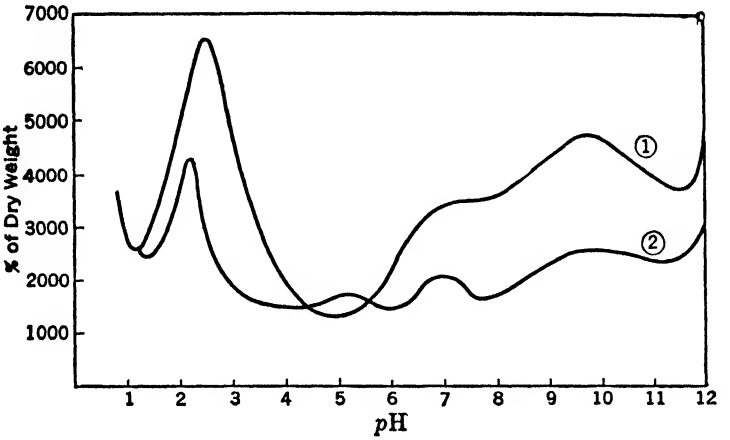


FIG. 4. Swelling of gelatin in response to *p*H changes. Curve 1 is in the absence of electrolytes other than HCl or NaOH. Curve 2 is in the presence of 0.1 *M* NaCl. Both curves are for 18° C. (Jordan Lloyd and Pleass.)

swelling resulting from a change of *p*H can be adequately explained on the basis of the original theory of Proctor and Wilson.[13] When the *p*H of the gel is shifted from its isoelectric point, a Donnan equilibrium is established between the interior of the gel and the acid or base on the outside of the gel. This leads to electrolyte accumulation by the gel with an increase in the osmotic pressure inside the gel, and water flows into the gel until the total swelling pressure is exactly equal to the elastic strength of the gel.

Let us look at the swelling problem in another way. Consider again the relation between the vapor-pressure lowering and the osmotic pressure as was discussed in Chapter XVI. On the basis of the maximum work principle we obtain

$$P_h \bar{V}_1 = RT \ln \frac{P_0}{P} \qquad 2$$

where P_h is the hydrostatic pressure of a solution whose vapor pressure is

[12] D. Jordan Lloyd and W. B. Pleass, *Biochem. J.*, **21**, 1352 (1927).

[13] H. R. Proctor and J. A. Wilson, *J. Chem. Soc.*, **109**, 307 (1916).

P in equilibrium with the pure solvent whose vapor pressure is P_0. $\overline{V}_1$ is the partial molar volume of the solvent. We can use this equation to calculate the swelling pressure of a gel since we can set P_h equal to the swelling pressure. The true swelling pressure of a gel is evidently equal to the swelling pressure calculated by means of equation 2 minus

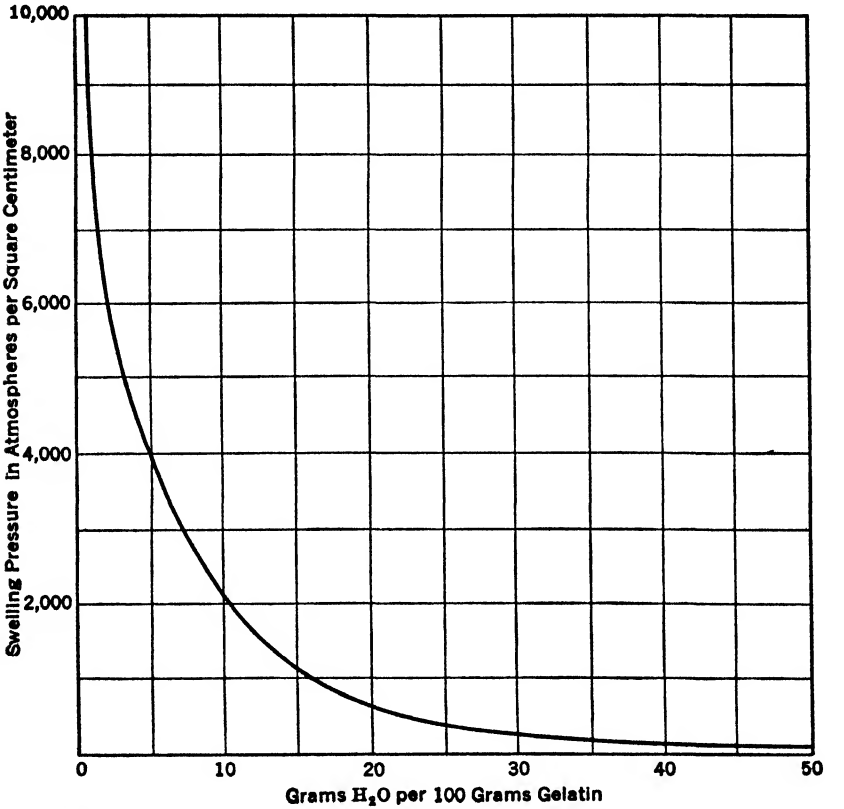


FIG. 5. Swelling pressure of isoelectric gelatin at 25° C. in atmospheres as calculated from aqueous vapor-pressure studies.

the pressure exerted by the elastic properties of the gel. The elastic pressure is ordinarily small and is of importance only at and near the saturation point of the gel with solvent.

The swelling of a gel can thus be described in terms of the vapor-pressure lowering of the solvent irrespective of what factors are responsible for this vapor-pressure lowering. Figure 5 shows the swelling pressure of isoelectric salt-free gelatin calculated from aqueous vapor-pressure studies.[14] In the application of equation 2 there is an uncer-

[14] H. B. Bull, unpublished results.

tainty regarding the value of the partial molar volume, $\overline{V}_1$. Upon the addition of water to a dry protein, there is always a volume contraction. That is, the sum of the volumes of the water and protein separately is always more than the volume of the two when mixed. The value of $\overline{V}_1$ is, accordingly, less than the molar volume of water. See for example Svedberg,[15] Neurath and Bull,[16] and Friedman and Brown.[17]

Gel Structure

A detailed discussion of the structure of protein gels would have to involve a discussion of protein structure and would require a great deal more information than we now have available. A gelatin gel is completely optically isotropic, indicating that there is no preferred orientation of particles. Such gels, when subjected to a mechanical stress, however, become anisotropic and exhibit birefringence.[18]

A gelatin gel acquires a certain internal structure, and, if the gelatin is dried and allowed to reimbibe water, it will tend to return to the original water content. For example, Gortner and Hoffman [19] studied the rehydration of gelatin. The original gels contained 10, 25, and 35 per cent gelatin. The gels, after drying, were powdered and passed through a 2-mm. sieve, and those particles retained in a 1-mm. sieve were used. It was found that the dried particles from the 10 per cent gel imbibed the most water, those from the 25 per cent gel imbibed less, and the particles from the 35 per cent gel imbibed the least. The influence of the past history of a gel on its present behavior is called hysteresis.

Some substances will form a gel at astonishingly low concentrations. Thus Gortner and Hoffman [20] were able to prepare an aqueous gel which contained as little as 0.1 per cent dibenzoyl-*l*-cystine. The ordinary jellyfish contains about 96 per cent water and yet has a definite structure.

Bernal and Fankuchen [5] in a most interesting paper conclude that tobacco mosaic virus protein gels have a highly organized structure. The virus protein particles are very elongated rods;[21] in a gel, these rods lie parallel to one another. There is, however, no order in the third dimension, so that the gel can be looked upon as a two-dimensional crystal.

[15] T. Svedberg, *J. Am. Chem. Soc.*, **46**, 2672 (1924).
[16] H. Neurath and H. B. Bull, *J. Biol. Chem.*, **115**, 519 (1936).
[17] L. Friedman and B. Brown, *J. Am. Chem. Soc.*, **57**, 508 (1935).
[18] M. Kunitz, *J. Gen. Physiol.*, **13**, 565 (1929–30).
[19] R. A. Gortner and W. F. Hoffman, *Proc. Soc. Exptl. Biol. Med.*, **19**, 257 (1922); *J. Phys. Chem.*, **31**, 464 (1927).
[20] R. A. Gortner and W. F. Hoffman, *J. Am. Chem. Soc.*, **43**, 2199 (1921).
[21] M. A. Lauffer, *J. Phys. Chem.*, **42**, 935 (1938).

Each protein rod molecule is surrounded by six other protein rod molecules all at equidistance. The distance between the rods depends upon the water content. Above a certain water content, the protein solution separates into two phases; the top phase is a dilute solution of the protein whereas in the bottom phase the particles are almost completely oriented. The distance between these particles depends upon the electrolyte concentration.

Thixotropy

Some gels when mechanically agitated undergo a reversible, isothermal gel-sol transformation; the gel re-forms when the agitation is stopped. The word thixotropy is derived from the Greek "thixis" meaning a touch and "trope" meaning a turn, to change.

Almost any gel will show a certain degree of thixotropy, but some gels show this property to a particularly striking degree. Among these may be mentioned vanadium pentoxide, benzopurpurin, dibenzoyl-cystine, barium malonate, iron oxide, and certain clays such as bentonite. The dispersion medium for all these substances is water. The concentration of the suspended material is important. If it is too concentrated, the gel cannot be liquefied by shaking; if it is too dilute, no gel will form. The electrolyte concentration is also critical.

The clay bentonite is very convenient for the demonstration of thixotropy. The particles are very asymmetric, being long, thin plates. Bentonite is derived from volcanic dust, and its main constituent is the mineral montmorillonite. It is one of the few inorganic substances which swell in water. To obtain a thixotropic gel of bentonite, water is mixed with the clay until the desired consistency is obtained. The amount of water added determines the setting time of the gel. If the clay suspension is quite concentrated, it is possible to hear the movement of the liquefied suspension in response to a vigorous shaking of the gel in a test tube; but the gelling time may be so short that as soon as the shaking has stopped the gel sets and the liquid state is not observed.

Electrolytes added to the proper sol will produce a thixotropic system, the effectiveness of electrolytes in regard to valence, etc., exactly paralleling the influence of electrolytes on coagulation of sols. Thixotropy may be regarded in a sense as the first step in the direction of coagulation. Increasing temperature decreases the time required for a thixotropic sol to set to a gel just as increasing temperature increases the rate of coagulation of a sol.

There is a marked correlation between the asymmetry of particles and their tendency toward thixotropy. The more asymmetric the particles, the more likely the substance is to exhibit marked thixotropy. Freund-

lich,[22] however, reports an exception to this rule. He finds that the Solnhofen slate which has symmetrically shaped particles will form thixotropic systems in water. Water is not necessary for thixotropy as some suspensions in a non-aqueous medium are thixotropic. For example, mercaptobenzothiazol in such liquids as benzene, toluene, and gasoline will form thixotropic systems. The size of the particles appears to be unimportant as long as it is below about 5 μ. Particles larger than 5 μ will not form thixotropic systems.

The formulation of a theory of thixotropy has occupied the attention of a number of workers. One necessity for understanding thixotropy is the conception of long-range forces. These forces must be strong enough to hold the particles in place, but at the same time they cannot be so strong as to prevent rupture upon the application of comparatively weak mechanical force. They must also be of a non-specific character; otherwise the speed of reaction of particles of such size would be very slow. The most satisfactory theory of these forces is given by Levine and Langmuir, previously discussed in Chapter XII. The characteristic feature about such forces is the great distance through which they act. As we have pointed out, a minimum energy point is found at some distance from the particle. If the depth of this minimum is great enough, particles will be held at this distance apart. Such minima are of particular importance for asymmetric particles because here the ratio of the kinetic energy of the particles to the depth of the energy minima is small. One can, accordingly, understand the close parallelism between thixotropic properties and asymmetry of particles. However, this theory has a certain element of incompleteness. No account is taken of the influence of hydration or solvation. Solvation would appear to play an important role in thixotropy in preventing the too close approach of the particles. In short, solvation prevents the coagulation process from going to completion. Another point which disturbs the author is that increasing temperature favors gel formation and decreases the time required for a sol to revert to a gel. One would perhaps expect that with increasing temperature the particles would acquire sufficient kinetic energy to leave their energy troughs, and either complete coagulation or complete dispersal of the gel would result from the temperature increase.

Several examples of thixotropic behavior are found in nature. For example, Freundlich found that quicksand is a thixotropic system, and he seems to feel that protoplasm may have thixotropic properties. Myosin gels show well-defined thixotropic behavior, and Freundlich points out that thixotropic transformation is one of the very few colloidal processes which is reversible and rapid enough to function in a

[22] H. Freundlich, *Thixotropy*, Hermann et Cie, Paris, 1935.

muscle mechanism. The transformation in muscle must be due to a very rapid process; otherwise the high frequency of the vibrations of insects' wings would be unintelligible.

Methods of Investigation

A great many types of physical measurements can be made on colloidal and biological structures. Optical measurements give a great deal of information as to the degree of orientation of asymmetric particles. Orientation of asymmetric particles always gives rise to birefringence.

Probably the two most important tools for the investigation of structures are X-ray diffraction and the electron microscope. The electron microscope especially promises to be useful. Unfortunately, both these instruments are very expensive, the electron microscope being almost prohibitive in cost. The biologist and biochemist have been very timid in applying these techniques to their problems, and the professional physicist has been too busy with his own projects. Katz, Sponsler, Corey, Astbury,[23] Fankuchen,[24] and Schmitt [25] are to be congratulated on their pioneer work on the application of X-ray diffraction to biochemical and biological structures.

X-Ray Diffraction

X-rays are produced by the action of high-speed electrons on matter. They arise, as was pointed out in Chapter I, from an excitation of the electrons in the innermost shells of the target atoms. The *K*-shell gives rise to the hardest or shortest-wavelength X-rays; the *L*-shell electrons give off longer-wavelength rays, and the *M*-shell electrons produce still longer-wavelength X-rays. The wavelengths of visible light lie between 4,000 and 8,000 Å; those of X-rays are very much shorter, between 0.1 and 100 Å. To produce X-rays, a stream of electrons is made to strike the anode target, whereupon the atoms of the anode give off X-rays. The anode is usually made of copper, of chromium, or of iron. X-rays derived from these anodes are not monochromatic but show definite intensity maxima at several wavelengths. For accurate work, monochromatic X-rays must be used although the $K\alpha$ radiation from copper can be used for most work without purification. There are devices for producing a monochromatic beam.

The beam of X-rays from the tube is passed through a pinhole in a metal plate, and the tiny pencil of X-rays emerging from the pinhole is

[23] W. T. Astbury, *Ann. Rev. Biochem.*, **8**, 113 (1939).
[24] I. Fankuchen, *Ann. N. Y. Acad. Sci.*, **41**, 157 (1941).
[25] F. O. Schmitt, *Physiol. Revs.*, **19**, 270 (1939).

allowed to impinge on the specimen under investigation. If there is any regularity of structure in the specimen, diffraction of X-rays from this regular structure will result; and X-rays emerging from the specimen will be in phase or out of phase depending upon the extent of diffraction. Rays that are in phase are reinforced in intensity; those out of phase are decreased in intensity. A photographic plate placed behind the specimen will register these interferences. In practice, a lead pellet is placed over the photographic plate at the point where the undeviated beam of X-rays strikes the plate; otherwise the plate would be badly fogged by this intense beam. It is the diffracted rays which are important.

The angle of deflection of the X-rays is related to the wavelength of the rays and to the distance between the repeating units in the structure under investigation by Bragg's law which is

$$n\lambda = 2d \sin \alpha \qquad 3$$

where n is 1, 2, 3, or any integer; λ is the wavelength of the X-ray; d is the distance between the repeating units; and α is the angle through which the X-ray of a definite wavelength is deflected. A series of angles α_1, α_2, etc., will give deflections corresponding to values equal to 1, 2, etc. These are called the first-order spectrum, the second-order spectrum, etc. It is thus possible to determine the repeating distance in any material by means of X-rays, provided that repeating distance is commensurate with the wavelength of the X-rays used. It is seldom, indeed, that no structure at all is revealed; almost always there is some diffraction of the X-rays. Even water gives a diffraction picture.

Considerable information can be obtained from the quality of the diffraction. A perfect crystal such as sodium chloride gives sharp spots on the photographic plate; for a fiber, the spots tend to spread into an arc of a circle. The better the orientation of the fiber molecules, the narrower will be the arcs. A completely disoriented fiber with repeating units in it, such as displayed by a disoriented polypeptide chain, spreads the arcs out so much that they appear as complete concentric circles with the undeviated beam as a common center. A beautiful example of the transformation from a completely disoriented fiber to an oriented one is shown by the stretching of rubber. Upon stretching, the disoriented isoprene chains orient themselves in the direction of the stretch, as X-ray diffraction pictures show.[26]

The molecular spacings in fibers can be calculated by means of Bragg's law. It is to be noted that the wide angle deflections correspond to the shorter molecular spacings, and the spots, arcs, or rings closest to the

[26] E. A. Hauser and H. Mark, *Kolloid.-Beihefte*, **22**, 63 (1926).

central beam arise from the longer spacings in the specimen. In order to resolve these longer spacings, it is necessary to have the photographic plate a considerable distance from the specimen. In fiber work this distance should be 40 cm. or more.

The application of X-rays has a great many experimental and theoretical ramifications, far more than the simple account given above would indicate. It is seldom that X-ray diffraction studies alone give a completely unambiguous answer to the structure of any material. The X-ray data must be supplemented by optical and chemical data. A book on X-rays which is easily understood is one by Clark.[27]

Electron Microscope

The electron microscope involves the same optical principles as an ordinary microscope except that magnetic fields are used to focus the beam instead of glass lenses. The specimens show differential transparency to the electrons, and a true picture of the structure is obtained. The electron beam after passing through the specimen is greatly magnified and focused on a photographic plate or on a fluorescent screen. The resolving power of a microscope is limited by the wavelength of the light used. The wavelength of the electron beam is very small in comparison with that of visible light. The resolving power of an electron microscope is, accordingly, much greater than that of an ordinary microscope. Magnifications in excess of 25,000 have been achieved, and with such magnifications it is possible to "see" tobacco mosaic virus protein molecules.[28]

Schmitt, Hall, and Jakus[29] have published an electron microscopic study of collagen fibers. They find such fibers to be striated with dense bands which are about 440 Å long in the direction of the fiber axis and the light bands are about 200 Å in the direction of the fiber. Such fibers show long-range elasticity, but it was not possible to interpret their findings in terms of the molecular structure of collagen.

A recent review on the electron microscope is one by Anderson.[30]

[27] G. L. Clark, *Applied X-rays*, 3rd edition, McGraw-Hill Book Co., Inc., New York, 1940.

[28] W. M. Stanley and T. F. Anderson, *J. Biol. Chem.*, **139**, 325 (1941).

[29] F. O. Schmitt, C. E. Hall, and M. A. Jakus, *J. Cellular Comp. Physiol.*, **20**, 11 (1942).

[30] T. F. Anderson, *Advances in Colloid Science*, p. 353, Interscience Publishers, Inc., New York, 1942.

INDEX