QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP STUDIES ON SOME ANTI-HUMAN-IMMUNODEFICIENCY-VIRUS (ANTI-HIV) AGENTS – PROTEASE INHIBITORS

THESIS

Submitted in partial fulfilment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

BY

M. SURESH BABU

Under the Supervision of

Prof. S. P. GUPTA

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE
PILANI (RAJASTHAN)
INDIA
1999

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE
PILANI, RAJASTHAN – 333031

CERTIFICATE

This is to certify that the thesis entitled **"QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP STUDIES ON SOME ANTI-HUMAN-IMMUNO-DEFICIENCY-VIRUS (ANTI-HIV) AGENTS – PROTEASE INHIBITORS"** and submitted by **M. SURESH BABU**, ID. No. **1993PHXF012** for the award of **Ph.D. Degree** of the Institute, embodies the original work done by him under my supervision.

Date: 18.12.1999

(Dr. S. P. Gupta)
Professor
Chemistry Group
B.I.T.S., Pilani

Dedicated to

My Grand Parents

Late. Smt. & Shri. Valluri Seetharamaiah

PREFACE

In the present era, the acquired immunodeficiency syndrome (AIDS) is the most fatal disorder, caused by human immunodeficiency virus (HIV), for which no successful chemotherapy has been developed so far. The pandemic spread of this disease has promoted an unprecedented scientific and clinical effort to understand and combat it. Recognition of molecular events critical to viral replication has suggested several strategies for potential chemotherapeutic intervention. A large number of chemotherapeutics have been delineated, but only the inhibitors of reverse transcriptase and protease have met with clinical success.

Till now thirteen inhibitors of reverse transcriptase and protease have been approved by the U.S. Food and Drug administration (FDA) in recent years and are currently used alone or as a part a combination regimen for the treatment of HIV infection and AIDS patients. However, all these drugs have limited or transient clinical benefits in HIV-infected individuals due to rapid development of HIV resistance, adverse side effects, and/or toxicity. Therefore, many research approaches are still underway to discover diverse anti-HIV agents with novel structures or mechanism(s) of action.

Today biological or therapeutic activity is considered to be a function of physicochemical properties. The correlation of molecular structure with biological activity is at the heart of modern medicinal chemistry, being fundamental both to the understanding of how drugs act, and to the rational design of more effective analogues. This resulted in discovery, examination, and interpretation of SAR in a more systematic

way, which led to the introduction of quantitative structure-activity relationship (QSAR) studies.

Since HIV mutates at the level of protease to render resistance to antiviral drugs, targeting the enzyme, there is marked interest in developing structurally diverse and/or small molecules. This study can be very fruitfully facilitated by QSAR studies which investigate the physicochemical and structural properties of molecules that can increase the potency of the drugs. In the present thesis, therefore, an attempt has been made to discuss the QSAR studies of various HIV protease inhibitors.

The thesis consists of three chapters. Chapter 1 presents a brief introduction about HIV and its chemotherapy. Further it also describes briefly about QSAR and its applications and limitations. Chapter 2 discusses significance of different physico-chemical and structural parameters used in the correlation study and the methods of their calculations, and Chapter 3 embodies the results and discussion of our QSAR studies made on various protease inhibitors acting as anti-HIV-1 agents.

All the work, however, that the thesis embodies was done under the able guidance of my supervisor, Prof. S.P. Gupta, for which I express my deepest regards to him. I owe a special debt of gratitude to our Director Prof. S. Venkateswaran and Deputy Directors Prof. B.M. Mithal and Prof. L.K. Maheshwari for providing facilities, which are essential to expedite this work. I am also thankful to Prof. M. Ramachandran, Dean, Research and Consultancy Division for constantly monitoring the progress.

My thanks are also due to Prof. V.S.Rao, Dean, Practice School Division, who provided various other facilities needed for the work. I also wish to put on record the support and suggestions that I received from my Group Leader Prof. R.N.Saha.

During the work, I received the cooperation, encouragement, and the valuable suggestions from my various present and past colleagues namely Late Prof. P. Parimoo, Drs. V. Addepalli, Rajni Garg, R. Jagadeesh Babu, Alka Kurup, L. Rajendra, R. Mahesh, C.V.N. Prasad, K.R. Babu, P. Srinivas, P. Anitha, and M/s N. Sridhar, C.N.V.H.B. Gupta, Sajeev Chandran, S.V.S.A.K. Gupta, and Abby Abraham, for which I express my sincere thanks to them.

The special help that I received in the preparation of the thesis from Mr. Jitender Reddy and Mr. Vinay Kumar is duly acknowledged.

With deepest regards I place on the record my sincere gratitude to my parents. I am grateful to all my family members for supporting me in various ways. I have no words to express my gratitude to my wife, Subhashini, who exhibited immense tolerance during my work, and to my supervisor's wife, Mrs. Kanak Gupta, who always inspired me.

M. SURESH BABU

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1 INTRODUCTION

1.1 Human Immunodeficiency Virus (HIV): Discovery and Origin of HIV

The Human Immunodeficiency Virus (HIV) is the causative agent of acquired immunodeficiency syndrome (AIDS),[1,2] which is now spreading rapidly among many populations and has become a serious global threat to human health and life. The HIV has been identified as a retrovirus that belongs to *lentiviridae* family.[3,4]

AIDS is end stage manifestation of a prolonged infection with HIV. AIDS is a fatal multisystem disease that causes breakdown of a part of the body's immune system. As a result chemotaxis, antigen identification and the functioning of monocytes and macrophages are gradually diminished. The patient is vulnerable to a variety of unusual life threatening illness. The patients are susceptible to infections including unusual type of pneumonia and rare forms of skin cancer not usually found in a person with a normal ability to fight infection.

The disease AIDS, when first appeared, was recognized only as a severe, previously unseen, immunodeficiency syndrome of unknown cause. In 1981, initial reports in U.S.A.,[5] described the HIV epidemic as a clinical syndrome of immune deficiency well before research identified HIV and showed it to be the cause of the syndrome. First patient of AIDS in India was reported in May 1986 and the source of infection was traced to blood transfusion given in U.S.A..

The causative agent for this pandemic spread of the disease was isolated by Montagnier and co-workers[2] in 1983 and recognised as a new retrovirus from the lymph node of a man with persistent lymphadenopathy syndrome (LAS). In 1984, Gallo and co-workers[3,4,6,7] characterized another human retrovirus named HTLV-III (Human T-cell Leukemia Virus-III), isolated from peripheral blood mononuclear cells (PBMC) of adult and pediatric AIDS patients. In the same year, Levy and co-workers[8] isolated retrovirus from AIDS patients of different known risk groups and they named it as AIDS-associated retrovirus (ARV).

Within a short time, the three prototype viruses LAV, HTLV-III, and ARV were recognized, as antigenically indistinguishable, members of same group of retrovirus that belongs to *lentiviridae* family. In 1986, International Committee on Taxonomy of viruses recommended giving the AIDS virus a separate name, human immunodeficiency virus (HIV).[9]

Shortly after the identification of this virus (HIV), in 1985, a second AIDS virus was recognized from AIDS patients from West Africa with similar modes of transmission and associated clinical syndromes.[10] To distinguish these viruses, the initially discovered retrovirus associated with most of the world's HIV disease is designated as HIV-1 and the virus detected in the West Africa is designated as HIV-2. HIV-2 differed by more than 55% from previous HIV-1 strains isolated and was antigenically distinct. In common usage, 'HIV' usually indicates HIV-1, because HIV-2 is rare in most parts of the world due to less efficient transmission, and mortality rate from HIV-2 infection is only two-thirds than that from HIV-1.

HIV Transmission

HIV is transmitted in three ways.[11, 12]

(i) through unprotected vaginal or anal intercourse with an infected man or woman.[13, 14]

(ii) the infected blood entering the blood stream through sharing injection needles, transfusion of infected blood or blood products, intravenous drug users (those who take injections regularly for pleasure), needle injury, etc.[15]

(iii) from a woman with HIV to her baby either during pregnancy or during delivery.[16, 17]

There is absolutely no evidence that HIV is transmitted by casual contact or that the virus can be spread by insects such as a mosquito bite.

HIV Testing

HIV testing usually means the tests that determine whether or not infected with HIV, which causes AIDS. These tests look for "antibodies" to HIV. Antibodies are proteins produced by the immune system to fight a specific germ.

The standard screening test for HIV infection is the detection of anti-HIV antibodies using Enzyme Linked Immunosorbent Assay (ELISA). The test is highly sensitive (> 99.5%) and is quite specific. Before a positive ELISA test is reported, it is confirmed by Western Blot technique, which is a commonly used confirmatory test and detects antibodies to HIV antigens of specific molecular weights.[18]

Other approaches to detect infection in individuals without requiring blood samples include the examination of urine[19] and saliva,[20] for anti-HIV antibodies give upto 99%

sensitivity. However, measurement of antibodies in serum by ELISA technique is still the most sensitive routine method for the detection of HIV infection.

Other HIV tests, like 'Viral Load Test' and 'T-cell Test', are used when people already know that they are infected with HIV. These tests help measure how quickly the virus is multiplying (Viral Load Test) or the health of your immune system (T-cell Test). Healthy people have between 500 to 1,500 CD4+ (T-helper) cells in a milliliter of blood. HIV disease becomes AIDS when CD4+ cells count is less than 200 or on appearance of opportunistic infections.[21] Most common opportunistic infections are:

(i) Pneumocystis carini pneumonia (PCP), a lung infection

(ii) Kaposi's sarcoma (KS), a skin cancer

(iii) Cytomegalovirus (CMV), an infection that usually affects the eyes

(iv) Candida, a fungal infection that can cause thrush (a white film in mouth) or infections in throat or vagina.

HIV Infection

The infection that targets monocytes expressing surface CD4 receptors eventually produces profound defects in cell mediated immunity.[22] Overtime infection leads to serve depletion of CD4+ T-lymphocytes (T-cells), resulting in opportunistic infections, neurologic and neoplastic diseases, and ultimately death. Besides T-cells, other cells expressing CD4+ on their surface may also harbor HIV-1 and thereby act as a reservoir for the virus, thus extending the latency period associated with infection. These include macrophages, monocytes, and lymphoid cells.[23]

Events occuring in the days and weeks following infection are critical in determining the ultimate course of HIV disease.[24-28] These events include

(i) HIV spread to tissues and events that ultimately may represent hard to eradicate viral reservoirs.

(ii) Extensive damage to lymph node cellular architecture

(iii) Stimulation of an immune response against HIV

(iv) Loss of HIV specific CD4+

(v) Rapid HIV replication and mutation creating a more genetically diverse population of HIV genomes, some perhaps more virulent.

The extent, quality, and consequences of these events vary greatly among individuals, accounting in part for differences in subsequent rate of HIV disease progression.

Structural Components and Life Cycle of HIV-1

High resolution electron microscopy has illustrated that HIV-1 is an enveloped virus of about 100 nm diameter.[29] As shown in Figure (1.1), it contains an outer lipid layer, derived from the host cell during maturation, and consists of two major viral glycoproteins, the external *gp120* and the transmembrane *gp41* (*gp* stands for glycoprotein and the number refers to the mass of protein in thousands of Dalton.). Immediately beneath the outer envelope is a membrane-associated protein *p17*, which provides a matrix for the viral structure and is vital for the integrity of the virion. The matrix surrounds a characteristic dense, cylindrical nucleoid containing capsid protein *p24*. Inside this nucleoid are two

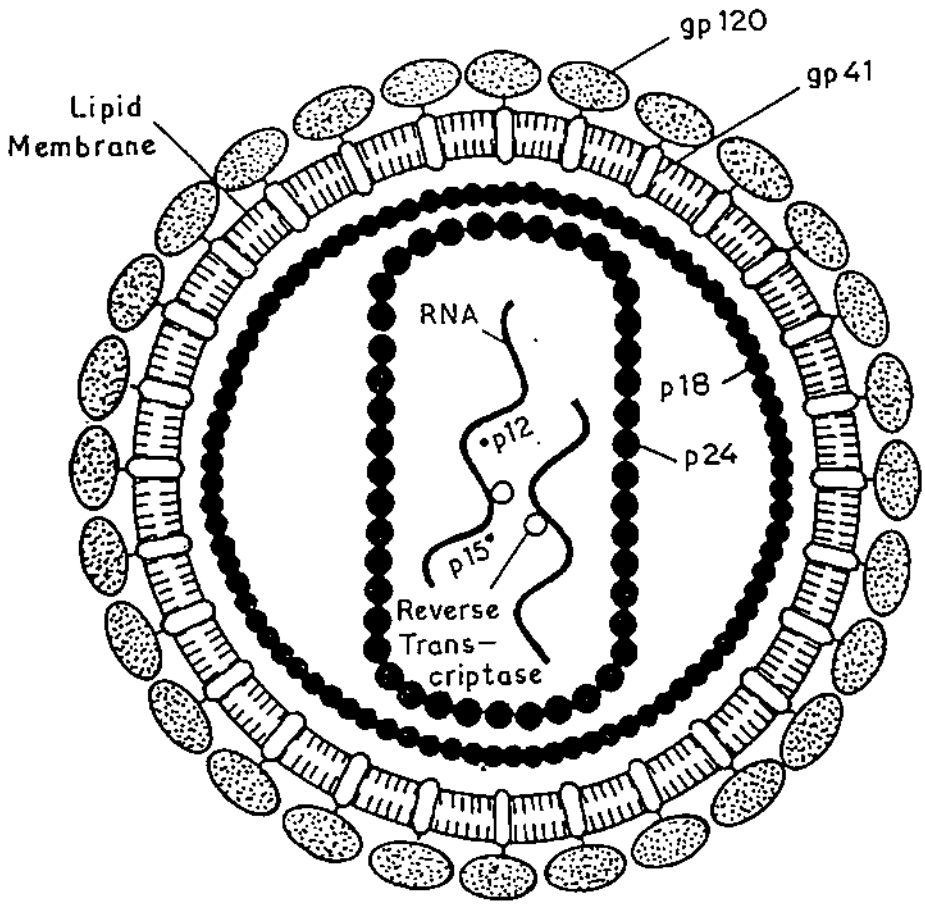


Figure (1.1): Structure of HIV-1[29]

identical RNA strands with which the viral RNA-dependent DNA polymerase (*pol*) *p66/p55*, called reverse transcriptase, is in association with nucleoprotein *p9*, integrase protein *p12*, and protease *p15* components.

The HIV life cycle begins with high affinity binding *gp120* envelope protein to its receptor CD4+ on the host cell surface (Figure (1.2)).[30] The CD4+ receptor is a protein molecule found predominantly on a subset of T-lymphocytes responsible for helper or inducer function in the immune response. Following binding, the fusion of virus with host cell membrane occurs via the *gp41* molecules and the HIV genomic RNA is uncoated and internalized. The enzyme reverses transcription of genomic RNA into double-stranded DNA. The DNA migrates to the nucleus to be integrated into the host cell chromosome through the action of virally encoded enzyme, integrase. The incorporation of this "provirus" into the cell genome is permanent. The provirus may remain transcriptionally inactive (latent) or manifest a high level of gene expression with active production of virus.

The activation of provirus (the gene expression) from the latent state by selective and constructive host transcription factors, notably the NF-Kβ family of DNA enhancer binding proteins, leads to the sequential production of various viral m-RNAs. These m-RNAs are translated into regulatory proteins- *Tat*, *Rev*, and *Nef*. The viral core is formed by the assembly of these proteins, enzymes, and genomic RNA at the plasma membrane of the cells. Budding of the progeny virion occurs through the host cell membrane, where the core acquires its external envelope. During the final budding process, the cleavage of *gag-pol* polyprotein precursor by HIV protease occurs, leading to morphological maturation of virions.

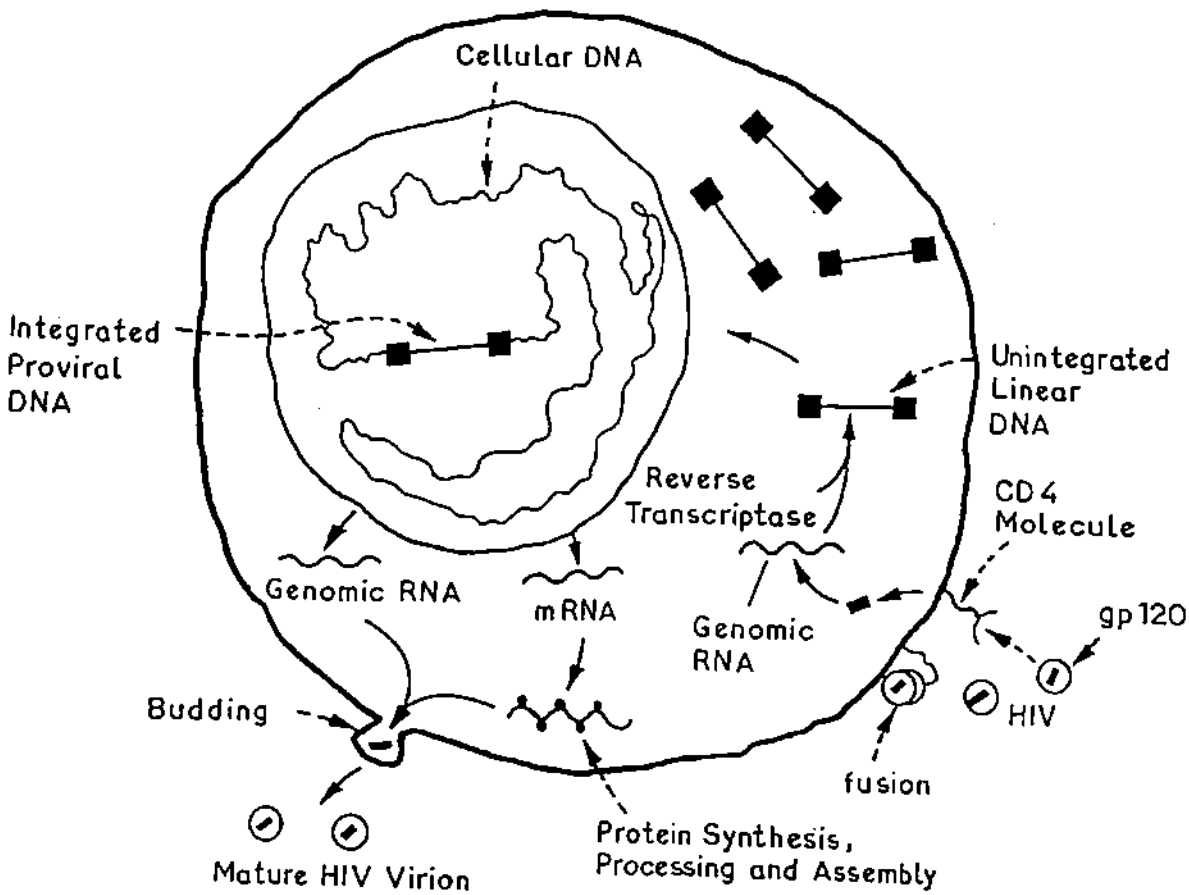


Figure (1.2): Life cycle of HIV-1.[30]

HIV-1 Genome

The integrated form of HIV-1, also known as the provirus, is approximately 9.8 kilobases in length.[31] Both the ends of the provirus are flanked by a repeated sequence known as the long terminal repeats (LTRs). The genes of HIV are located in the central region of the proviral DNA and encode at least nine proteins (Figure (1.3)).[32] These proteins are divided into three classes:

1. The major structural proteins : *Gag*, *Pol* and *Env*
2. The regulatory proteins : *Tat* and *Rev*
3. The accessory proteins : *Vpu*, *Vpr*, *Vif*, and *Nef*

The major difference in the genomes of HIV-1 and HIV-2 is the fact that the latter lacks the *Vpu* gene and instead has a *Vpx* gene not contained in HIV-1.

Anti-HIV-1 Drugs

Ideal Anti-HIV agent is expected to arrest the virulence and further infection of healthy cell without displaying toxicity towards normal cellular physiology. The inhibition of this virus forms the most fundamental aspect of the design and development of Anti-AIDS agent. A large number of chemotherapeutics have been delineated but only the inhibitors of reverse transcriptase and protease have met with clinical success. Till now thirteen inhibitors of reverse transcriptase and protease have been approved by the U.S. Food and Drug Administration (FDA) in recent years and are currently used alone or a part of a combination regimen for the treatment of HIV infection and AIDS patients. However,

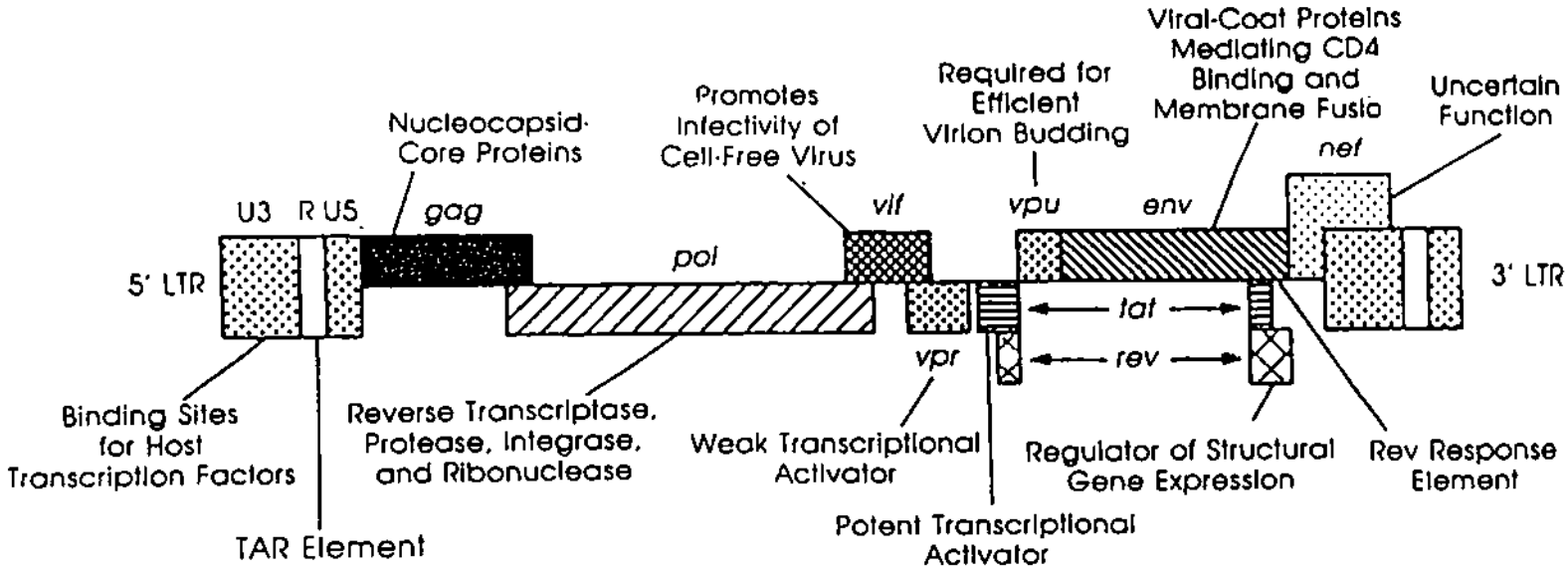


Figure (1.3): The genome of HIV-1.[32]

these drugs have a number of shortcomings, e.g., emergence of HIV-1 mutant strains having single or multiple resistance to the drugs used[34-39] and adverse side effects and/or toxicity. Therefore, it is essential to develop compounds with high potential therapeutic activity and less side effects.

Progress in HIV biology has provided detailed knowledge of molecular events in the replication cycle of HIV-1. Recognition of molecular events critical to virus replication has suggested several strategies for potential chemotherapeutic intervention.[33] Brief descriptions of these stages are given below.

A. Virus Binding Inhibitors

It was demonstrated that a truncated CD4 (sCD4) molecule was capable of inhibiting the binding of *gp120* to CD4 receptor and thus the viral replication in cell cultures.[40] However, further clinical studies of sCD4 with viral isolates were disappointing. The reasons were attributed to the insensitivity of the latter for the former and the difficulty in attaining sufficient therapeutic plasma levels due to short half-life of sCD4.

Some polyanionic compounds, whatever anion they are based upon, have also been found to inhibit the virus adsorption. Suramin, a hexasulfonatenapthylurea derivative, was the first compound to enter clinical trials as a possible chemotherapeutic agent against AIDS.[41] Polyanionic substances suffer from a number of pharmacokinetic and toxicological drawbacks, which seem to mar their clinical utility.[42]

B. Virus Cell Fusion Inhibitors

A number of compounds such as mannose-specific plant lecitins,[43,44] the polypeptidepolyphemusin,[45] negatively charged albumin,[46,47] and triterpene derivatives[48] have been postulated to interfere specifically with virus-cell fusion. The virus-cell fusion depends on the interaction of the envelope glycoproteins *gp120* and *gp41* with the cell membrane, but it is as yet not clear with which region(s) of *gp120* or *gp41* the fusion inhibitors actually interact. Further, it has been difficult so far to assess the clinical usefulness of the virus-cell fusion inhibitors, as the toxicological and pharmacokinetic profiles for most of these compounds remain to be established.

C. Virus Uncoating Inhibitors

The virus uncoating has been regarded as an appropriate target for antiviral agents. It has been speculated that HIV *p24* capsid protein can interact with the virus uncoating inhibitors.[49] At present, however, there is only one group of compounds, bicyclams, which have been found to inhibit the virus uncoating.

D. Reverse Transcriptase Inhibitors

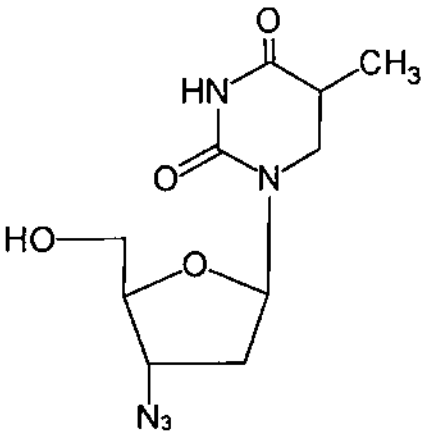
The process of reverse transcription of genomic RNA into double-stranded DNA by the enzyme reverse transcriptase (RT) is essential to the replication of HIV. Therefore, the inhibition of this key biochemical event in the viral life cycle provides the most attractive target for anti-HIV drug development. Most of the compounds approved so far by FDA in United States for the treatment of HIV infection are RT inhibitors.[50] Among

them zidovudine (AZT) (1), zalcitabine (DDC) (2), didanosine (DDI) (3), stavudine (D4T) (4), lamivudine (3TC) (5), and abacavir succinate (6) belong to the class of 2′,3′-dideoxynucleoside (ddN) analogues, while nevirapine (7), delaviridine (8), and efavirenz (9) belong to non-nucleoside class. Besides, several other non-nucleoside reverse transcriptase inhibitors (NNRTIs) have proceeded onto clinical development such as tivirapine (10) and HEPT derivative MKC-442 (11).[51] NNRTIs have recently gained an increasingly important role in the therapy of HIV infections.

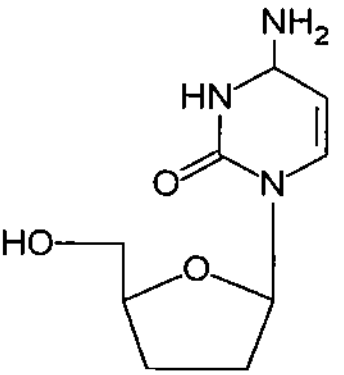
While ddN analogues, after being converted to corresponding triphosphates, compete with natural substrates to interact with the enzyme, nonnucleoside analogues have been found to interact non-competitively with an allosteric site, leading to the inactivation of the enzyme.[52,53]

E. Integrase Inhibitors

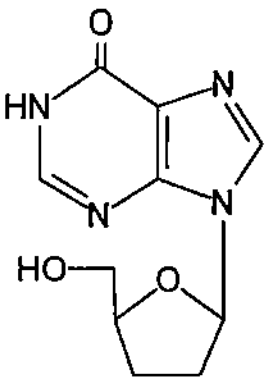
Incorporation of viral DNA into the host cell genome could be translated as the basis of life-long infection. Therefore, this biochemical event, catalyzed by the enzyme integrase, is a pivotal step in viral life cycle and thus worthy of being exploited to develop for anti-HIV chemotherapy. The enzyme integrase is produced by protease-mediated cleavage of the *gag-pol* precursor during virion maturation. A wide array of compounds has been speculated to act as integrase inhibitors. Several DNA binding agents were found to inhibit HIV-1 integrase, probably due to a nonspecific interaction with the DNA binding domain of the enzyme.[54] Catechol derivatives have also been found to act as integrase inhibitors, but they have been postulated to elicit their effects by interfering with the coordination of the metal ions that are required for the phosphoryl transfer.[55] However, the



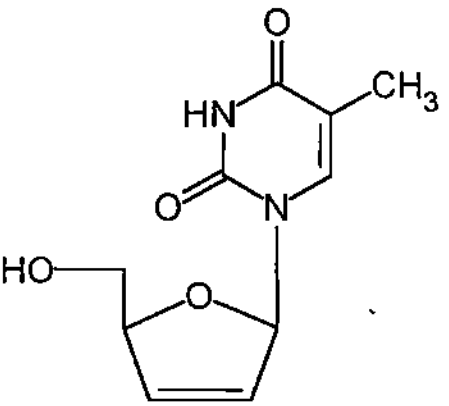
1
Zidovudine



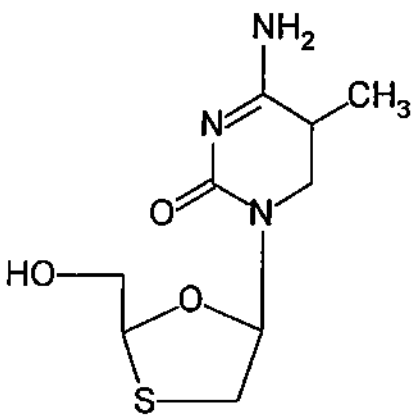
2
Zalcitabine



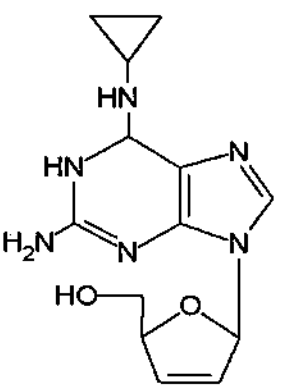
3
Didanosine



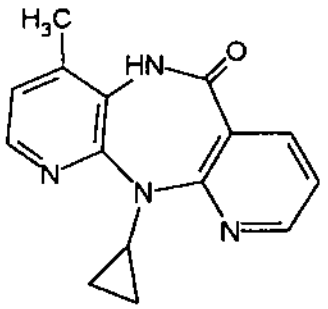
4
Stavudine



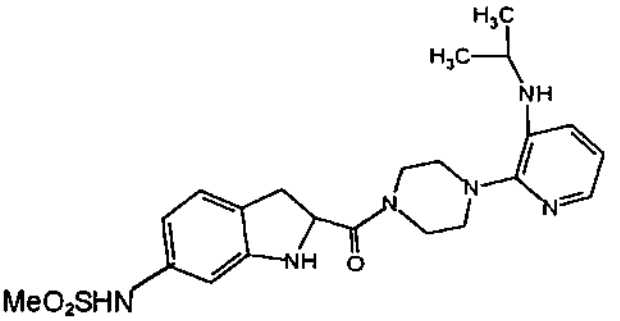
5
Lamivudine



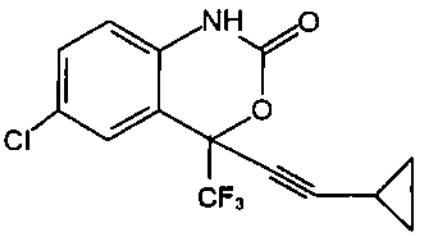
6
Abacavir succinate



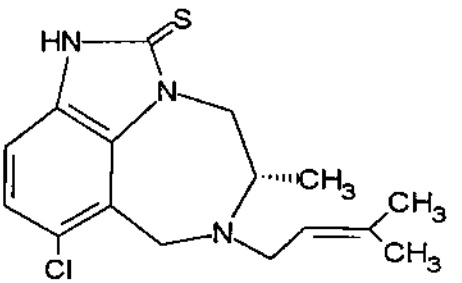
7
Nevirapine



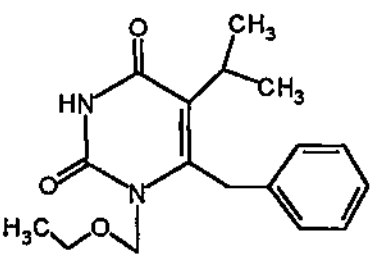
8
Delaviridine



9
Efavirenz



10
Tivirapine



11
MKC-442
HEPT derivative

catechol derivatives do not exhibit much antiviral specificity in cell culture and hence are no longer considered to be worth pursuing.[56]

F. Gene Expression Inhibitors

The viral integration into host cell genome becomes a sure cause of chronic infection, as the replicative machinery of the host cell will continue producing viral gene products (mRNA). If the translation of mRNA (gene expression) is inhibited, it may lead to the prevention of the spread of infection. Antisense oligonucleotides are generally thought to be the plausible inhibitors of this process due to their capacity to form stable duplexes with complimentary sequences of the viral mRNA. However, some pertinent problems related to the cost of synthesis, bioavailability, site specific delivery, and hybridization at a desired location will have to be addressed in order to fully realize the therapeutic utility of antisense oligonucleotides.[57]

G. Protease Inhibitors

HIV encodes an aspartic protease whose function is essential for proper virion assembly and maturation. Inactivation of this protease leads to the production of immature, non-infectious viral particles,[58-60] and thus to the prevention of further propagation of the virus. Consequently, the study of the inhibition of this enzyme has drawn the considerable interest of medicinal chemists for the development of anti-HIV chemotherapy.[61] Several protease inhibitors are approved for the treatment of HIV-1 infection. Details of these compounds and mechanism of protease inhibition are discussed later.

H. Glucosidase Inhibitors

The final step in the viral replication, leading to virion maturation, involves the processing of surface glycoproteins by the enzyme HIV glucosidase. This enzyme cleaves off glucose units from the oligosaccharide chain and thus helps the maturation of infectious virion. The inhibition of this enzyme, therefore, will lead to the inhibition of virion maturation. Polyhydroxylated compounds such as castenospermine and N-butyl-deoxynojirimycin have demonstrated inhibitory potential in preclinical evaluation.[62] However, selectivity of these compounds and their ability to distinguish between cellular and viral glycosylation has to be confirmed before wide scale use.

Thus, efforts have been made to exploit all the above mentioned intervention stages in the viral life cycle to develop anti-HIV chemotherapy.

1.2 HIV-1 Protease

HIV-1 Protease (HIVp) is an aspartic protease, which is an endopeptidase. HIVp is a C_2 symmetrical homodimer[63,64] with each monomer having 99 residues. The enzyme's C_2 axis lies between and perpendicular to the catalytic aspartates (Asp25 and Asp25′) in the active site. With standard nomenclature[65] the S_1 and S_1' (S_2 and S_2', etc) subsites are structurally identical. The two equivalent S_1 subsites are very hydrophobic, the S_2 subsites are mostly hydrophobic, except for the Asp29, Asp30, and Asp30′. The S_3 subsites are adjacent to the S_1 subsites and are mostly hydrophobic with the exception of Arg8 and Arg8′ (Figure (1.4)).

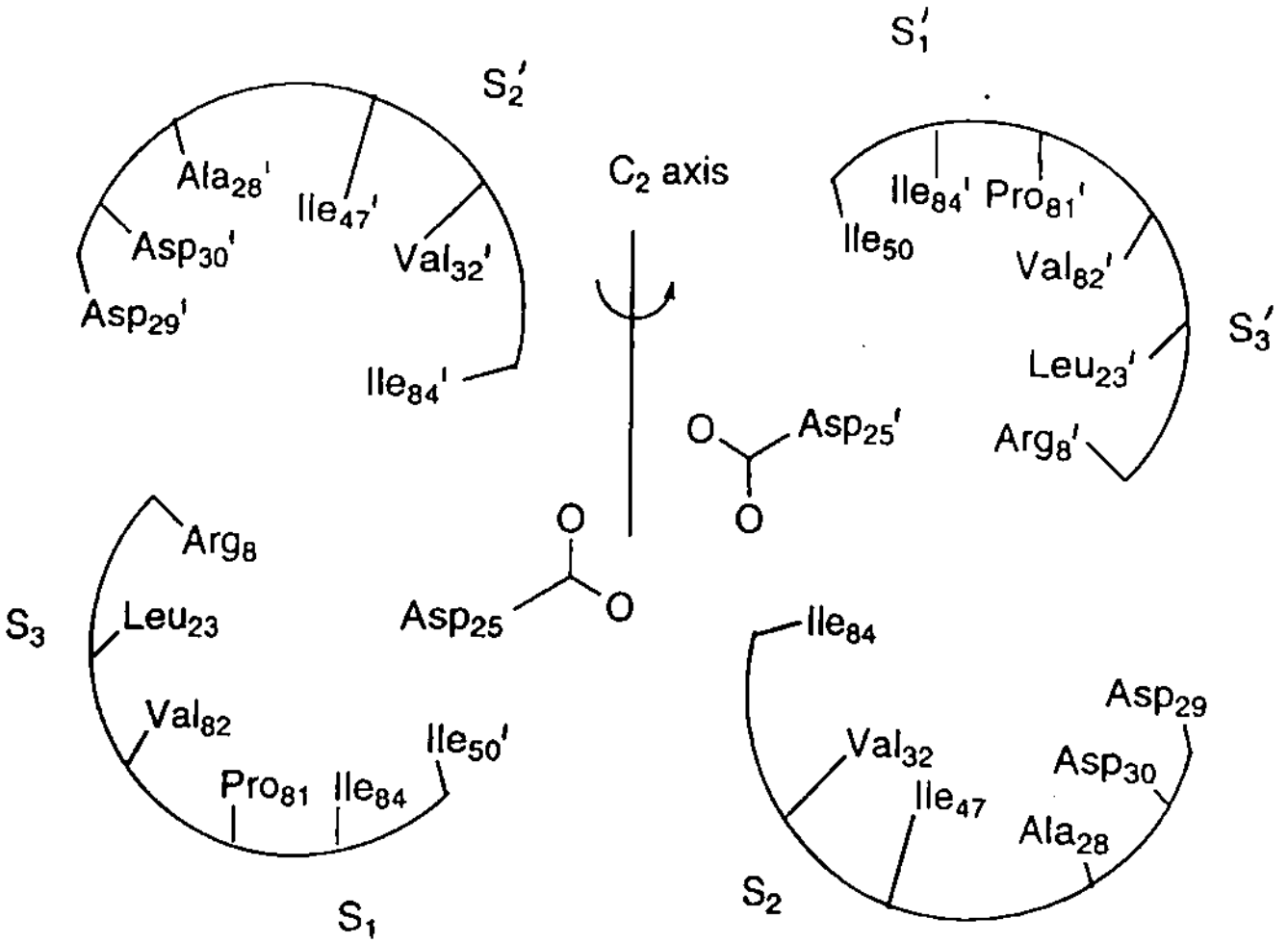


Figure (1.4): Amino acid residues which form binding sites of HIVp.[66]

The primary sequence of an aspartic protease has two different Asp-Thr-Gly sequence and the apostructure of it shows these two chains running in opposite directions with a water molecule bound between two aspartates. This water molecule is believed to be a nucleophile for the enzyme-catalysed amide hydrolysis of the substrate.[66] The substrate possesses a scissile bond (Figure (1.5)) which, in the substrate-enzyme interaction, is attached by the water molecule of the enzyme, and a few amino acid residues of the substrate interact with corresponding binding sites of the enzyme. This interaction is stabilized by several hydrogen bondings between the backbone of the substrate and the enzyme.[66,67] Since abundant structural informations are available on this enzyme,[68,69] it has become an attractive target for computer-aided drug design strategies and consequently a prime focus for the development of anti-HIV chemotherapy.[61]

Role of HIV Protease in Viral Replication

The HIV-1 protease is an aspartyl protease. Protease activity is required for cleavage of the *gag* and *gag-pol* polyprotein precursors. During viral maturation, the virally encoded protease cleaves the *pol* polypeptide away from *gag* and further digests it to separate the protease, reverse transcriptase, RNAse H and integrase functional proteins. Protease also cleaves *gag* polyprotein into smaller structural proteins. The proteolysis of the HIV *gag* and *pol* proteins occurs during the process of virus budding and the timing of this event is crucial to the successful generation of infectious virions.[69-71] Inhibition of this protease causes the delayed processing of polyprotein precursors that leads to the

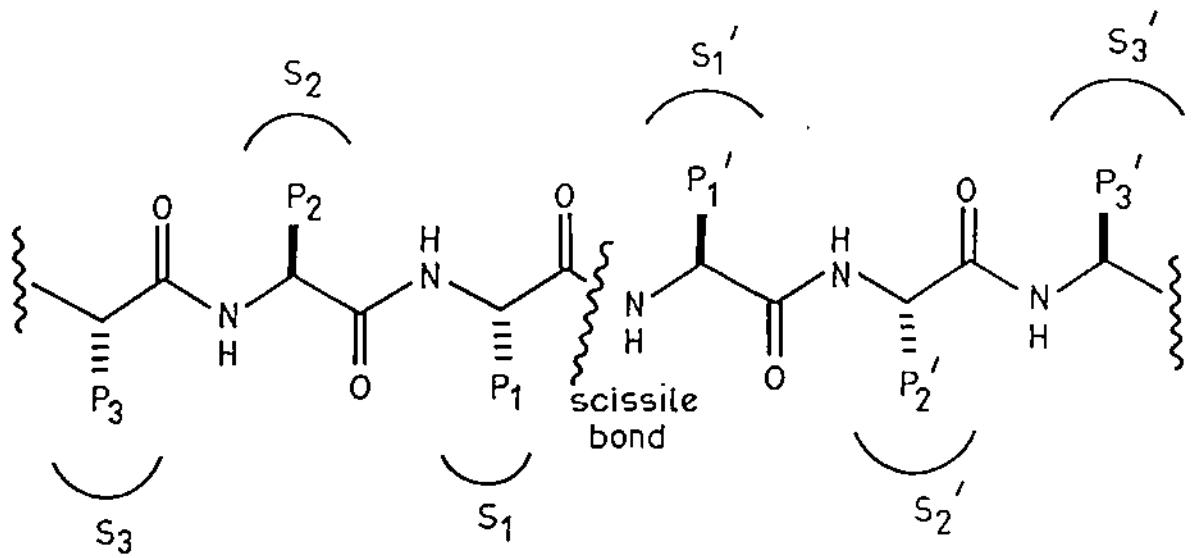


Figure (1.5): A peptidic substrate of aspartic proteases. The P_1, P_2,.....P_n and P_1', P_2',.....P_n' are amino acid residues, and S_1, S_2,....S_n and S_1', S_2',.....S_n' are the corresponding binding sites at the enzyme.[66]

These nomenclatures are according to the Schechter, I. and Burger, A. *Biochem. Biophys. Res. Commun.*, 1967, 27, 157.

production of non-infectious virions of altered morphology and thus to the prevention of further propagation of the virus.

1.3 Protease Inhibitors

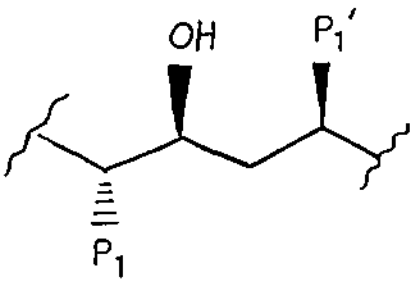
Protease inhibitors can be put into two categories: peptide based inhibitors and nonpeptidic inhibitors.

Peptide-Based Inhibitors

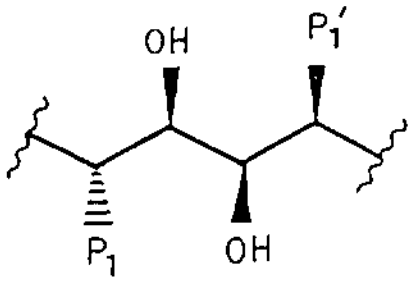
The discovery of peptide-based substrate mimicking HIV-1 protease inhibitors was based on the synthesis of substrate analogues in which the scissile bond (Figure (1.5)) was replaced by a non-cleavable isostere with tetrahedral geometry that could mimick the tetrahedral transition-state of the proteolytic action. A number of such isosteres that have been studied are shown in Figure (1.6).

Several inhibitors such as hydroxyethylamine[72](12), hydroxyethylene,[73,74] (R)-(hydroxyethyl)-urea,[75] norstatin,[76] and C_2 symmetric monoalcohols[77,78] and aminodiol[79] derivatives were prepared that could bind with the enzyme as shown in Figure (1.7). In the inhibitor enzyme interaction, the enzyme's water molecule makes hydrogen bonds with both the inhibitor and the enzyme with approximate tetrahedral geometry. This water molecule in the complex is known as "flap" water.

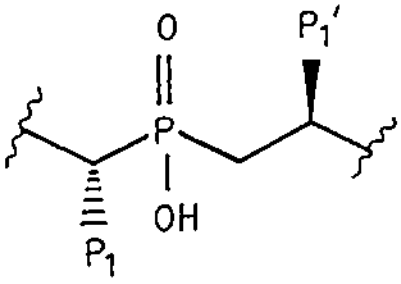
The crystal structures of enzyme-inhibitor complexes have provided deeper insight into the mechanism of protease inhibition.[80-83] Structures of the enzyme complexed with four structurally different peptide isostere exhibited that all four inhibitors were bound in



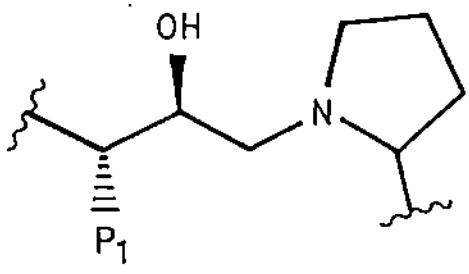
Hydroxyethylene



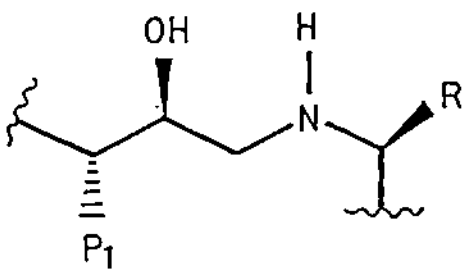
Dihydroxyethylene



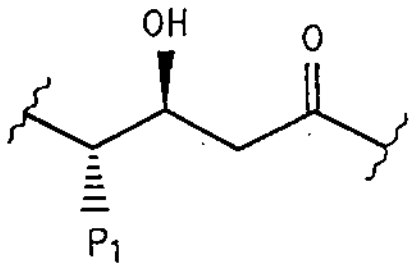
Phosphinate



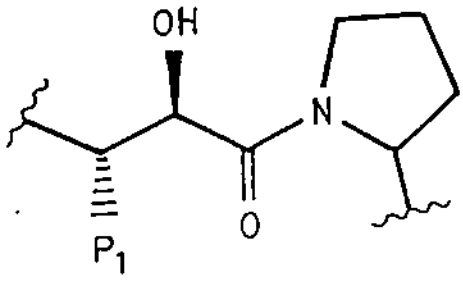
Hydroxyethylamine (P_1' cyclic)



Hydroxyethylamine (P_1' cyclic)



Statine



norstatine

Figure (1.6): Some nonhydrolyzable transition-state isosteres employed to replace P1-P1′ amide bond of the substrate for the design of HIV-1-Pr inhibitors.

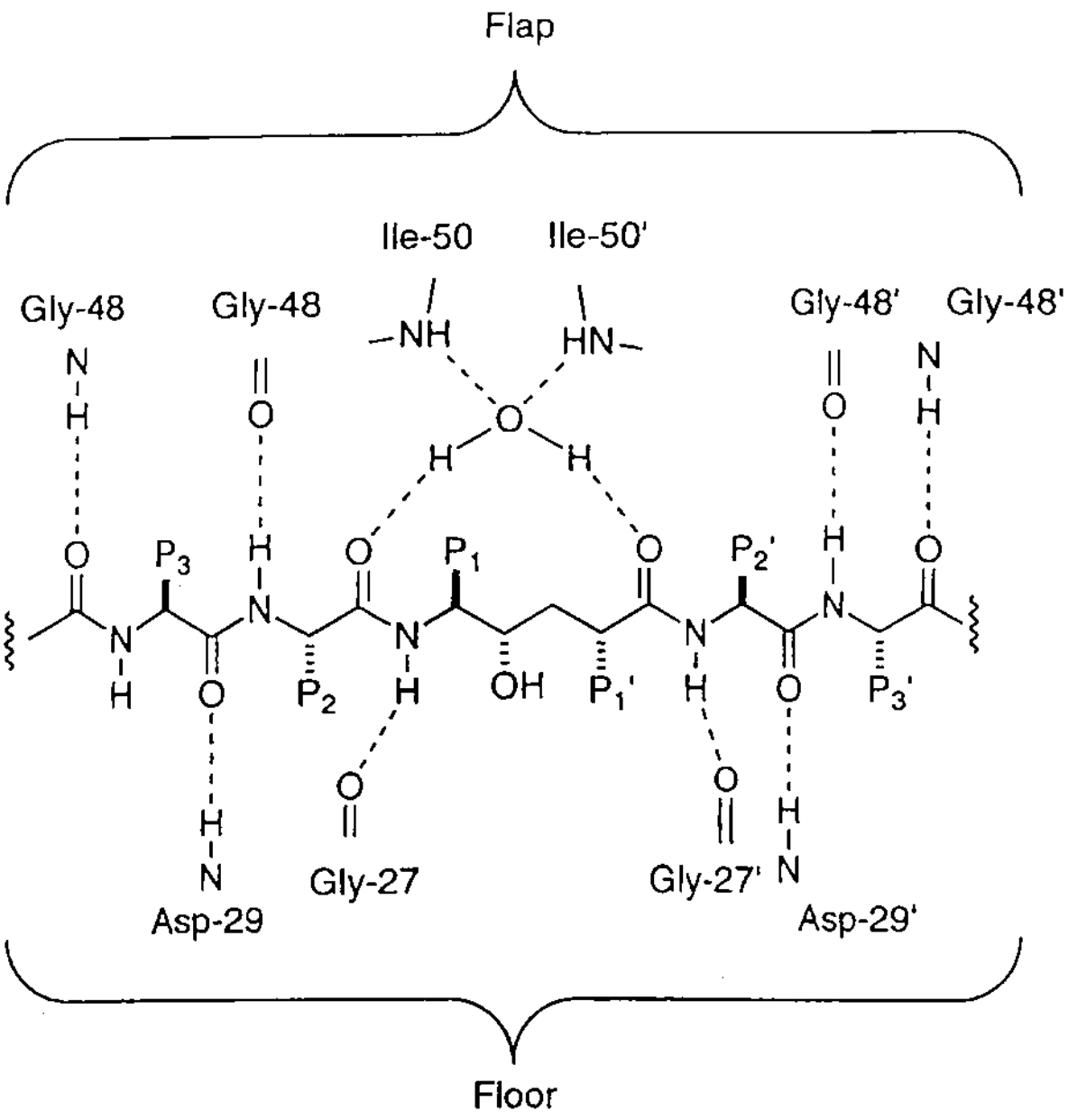


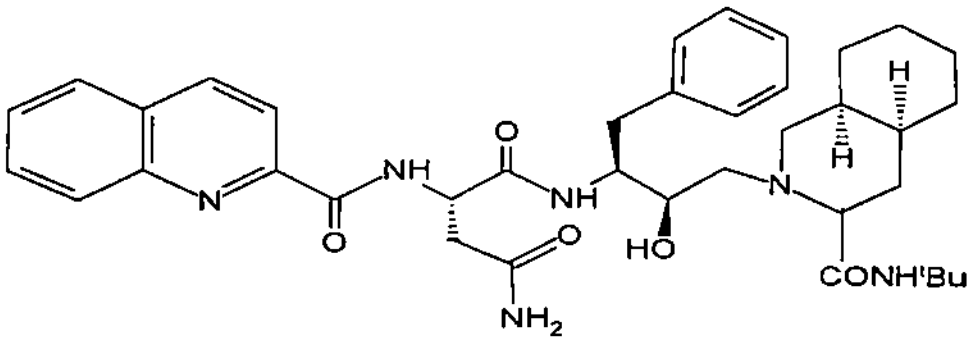
Figure (1.7): A model of binding of a substrate-based HIV-1-Pr inhibitor with with the enzyme.[66]

an extended conformation, spanning from P4 to P3′. An extensive network of hydrogen bonds could be illustrated between the enzyme and the polar atoms in the inhibitor. These postulated that hydrogen bonds are formed primarily with backbone atoms of the floor and flap regions of HIVp (Figure (1.7)). One striking feature of all four inhibitor complexes is that a tightly bound water molecule bridges the two enzyme flaps to the inhibitor through hydrogen bonds formed by the Ile50 and Ile50′ amide hydrogens and P_2 and P_1′ carbonyl oxygens of the inhibitors.

Majority of these transition-state peptidomimetics suffer from poor oral bioavailability due to their peptide-like structure.[84] Substantial progress in the identification of agents with high oral bioavailability has been made.[85-88] Currently, the FDA in United States has approved several HIV-1 protease inhibitors for the treatment of HIV, e.g., saquinavir (13), ritonavir (14), indinavir (15), and nelfinavir (16). All these compounds were developed on the C_2-symmetric structures of HIV-1 protease. Saquinavir was the first protease inhibitor that FDA approved for use in combination with nucleoside analogues AZT and DDC. Saquinavir is the least potent of the four FDA approved protease inhibitors. Indinavir perhaps the most potent of FDA approved protease inhibitors.

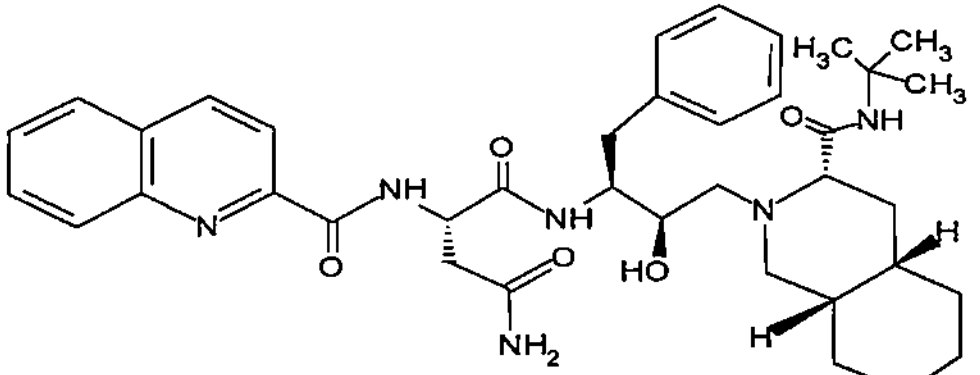
Nonpeptidic Inhibitors

Clinical development of peptide-derived compounds has been hindered by their poor pharmacokinetics, including low oral bioavailability and rapid excretion[84,89] and complex expensive synthesis.[90] On the basis of the knowledge of the X-ray crystal structure of the HIV protease dimer, the variety of nonpeptidic compounds were developed

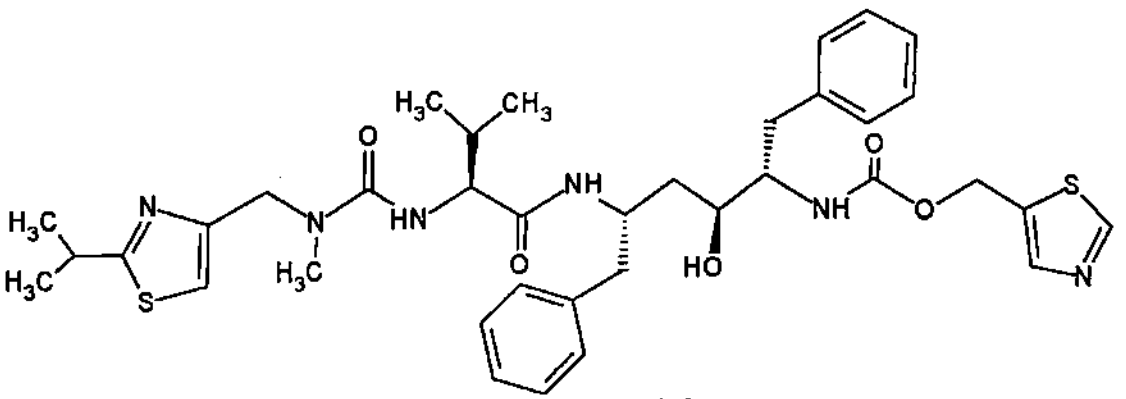


12

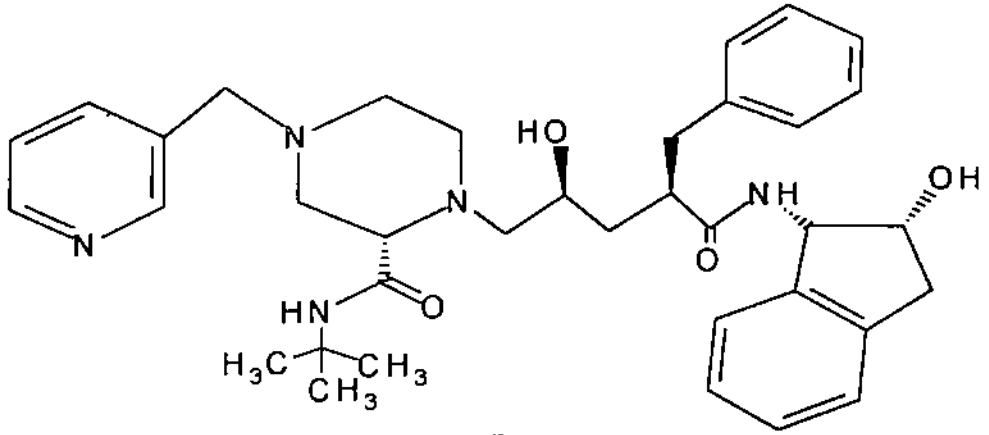
Ro 31-8959
Hydroxyethylamine isostere derivative



13
Saquinavir

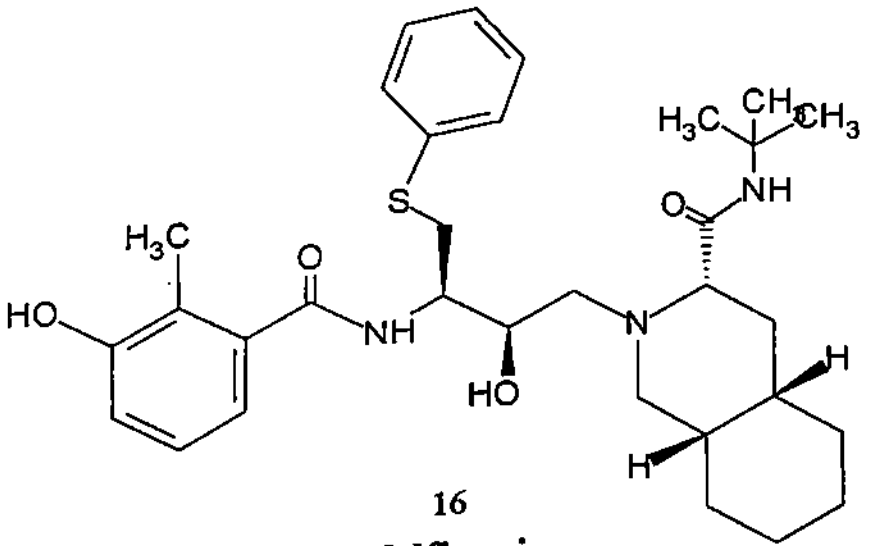


14
Ritonavir



15

Indinavir



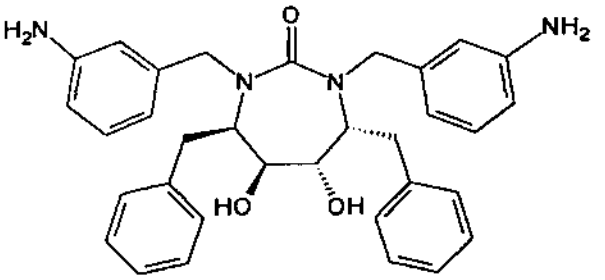
16

Nelfinavir

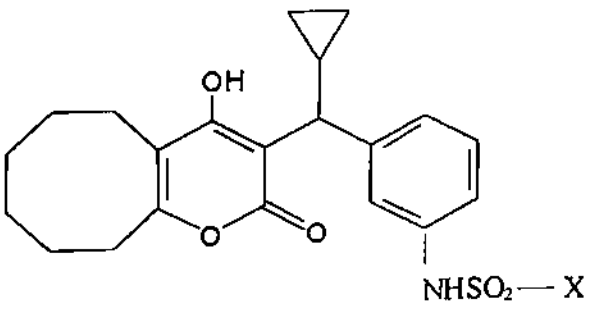
e.g., cyclic urea derivatives[91-93] (17), cycloalkylpyranones[94-97] (18), cyclic sulfolanes[98] (19), (4-hydroxy-6-phenyl-2-oxo-2H-pyran-3yl)thiomethanes[99] (20), and cyclic cyano-guanidines[100] (21). A common feature in all peptidomimetic protease inhibitors is the presence of flap water hydrogen binding to both the protein and the inhibitor. Except a few, the nonpeptidic protease inhibitors create an effective hydrogen bond network between the aspartyl residues and the flap region of the enzyme without intervention of water molecule. Cyclic urea derivative DMP450 (17) is presently in human clinical trails.[101] The development of anti-HIV chemotherapy based on protease inhibitors will always be an ongoing need because the virus has the ability to rapidly generate resultant mutants.[102-104]

1.4 HIV Resistance to Protease Inhibitors

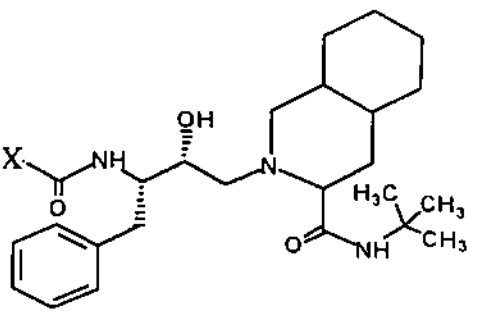
One notable problem with antiviral treatments has been the emergence of resistant virus strains. Mutations in the protease gene that confer resistance have been described for all currently available protease inhibitor drugs.[105,106] The flexibility of structure in the protease enzyme may make resistance more easily attainable than with other enzymes. Sustained concentrations of the protease inhibitors need to be maintained to avoid the emergence of resistance,[106] but the resistance occurs less frequently when the drugs are used in combination with other antiretroviral therapies. With some of the mutations, cross resistance to five structurally different protease inhibitors was found.[105] This rapid drawinian evolution occurring within the patients, as well as the toxicity of the drugs, almost certainly explain why the benefits of anti-HIV drugs are only temporary.



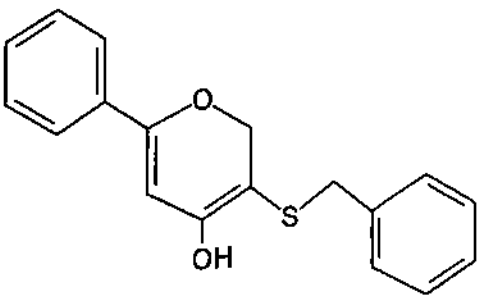
17



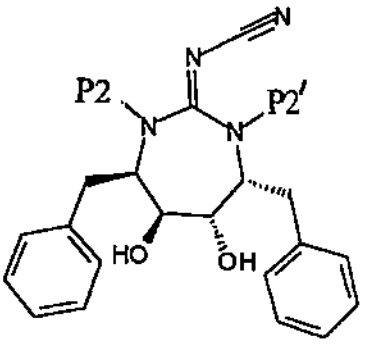
18



19



20



21

1.5 Quantitative Structure-Activity Relationship Studies

Over the past two decades, the intellectual focus of medicinal chemistry has shifted dramatically from, how to make a molecule, to what molecule to make. The lead molecule, as we know, is a prototype compound that has the desired biological or pharmacological activity, but may have undesirable characteristics such as toxicity, other biological activities, insolubility or metabolic problems. Early structure-activity relationship (SAR) studies (prior to 1960s), simply involved the synthesis of as many analogues as possible of the lead and then testing to determine the effects of structure on activity, based on the assumption that biological activity of a compound is a function of its chemical structure.

Today biological or therapeutic activity is considered to be a function of physicochemical properties. With this concept, structure-activity relationships are developed, when a set of physicochemical properties of a group of congeners is found to explain the variations in biological responses of these compounds. This has resulted in the discovery, examination, and interpretation of structure-activity relationships in a more systematic way, which has led to the introduction of quantitative structure-activity relationship (QSAR) studies.

Since its inception in 1962 by Corwin Hansch`s classic work, QSAR in biochemistry and biology[107] has progressed steadily.[108] However, with the advent of 3-D molecular graphics early in the 1980`s in Langridge`s Laboratory in the university of California, at Sanfrancisco, an explosive growth began to occur in methodology.[109-115] The attractive 3-D pictures of ligands bound to enzymes of established structure captured

researchers attention, but mechanism based on physical organic chemistry was forgotten. This is not meant to downgrade graphics. Indeed graphics can be of enormous help if it is based on QSAR and receptors whose structures are known.

The development of automated synthesis capability along with the formulation of combinatorial chemistry approach has enabled the rapid synthesis of large number of molecules. This large increase in synthetic capacity has been accompanied by the automation of *in vitro* bioassays affording high throughput screening systems capable of generating massive amounts of data in a relatively short period of time. The combinatorial possibilities of this strategy for even simple systems can be explosive. The alternative to this labor intensive approach to compound optimization is to develop a theory that quantitatively relates variations in the biological activity to changes in the molecular descriptors which can easily be obtained for each compound. A Quantitative Structure-Activity Relationship can then be utilized to help guide chemical synthesis.

The correlation of molecular structure with biological activity is at the heart of modern medicinal chemistry, being fundamental both to our understanding of how drugs act, and to the rational design of more effective analogues. Over the last few decades, considerable advances have been made in studies of structure-activity relationship, largely because of the trend towards expressing all aspects of "structure" in quantitative terms relative to some standard. The most significant contributions to this endeavour have been made by Hansch and co-workers.[116-118]

The quantitative approach to understanding drug action depends upon the ability to express structure by numerical values and then to relate these values to corresponding changes in activity. The response is going to be determined by the structure, i.e., by the physicochemical properties of the compound, and within a closely related or so-called congeneric series of compounds, changes in structure can be related to changes in biological activity.

The QSAR study tries to explain the reasons of observed variations in biological activities of a group of congeners in terms of molecular modifications or variations caused by the change of the substituents. QSAR studies generally have two important aspects: predictive aspect and diagnostic aspect. The predictive aspect, as the name implies, deals with the extrapolation and interpolation of a correlation study to identify synthesis of more active derivatives and to avoid the synthesis and testing of derivatives of same or equivalent activity, minimizing the time needed to find a better derivative. The diagnostic aspect, on the other hand, answers mechanistic aspects of the relation, i.e., it helps obtain the information about the type of binding forces involved and about the mode of actions of drugs. Results of both these aspects can lead to a tailor-made design of new drugs of better activity with lesser or no side effects.

Some important approaches used in QSAR studies are the nonparametric methods like Free-Wilson approach[119] or Fujita-Ban approach,[120] the parametric method developed by Hansch,[116-118] discriminant analysis,[121] and the pattern recognition technique.[122,123] Various factors such as quality of the biological data, number of compounds tested, degree

of variance in the results, and ratio of the time required for synthesis and biological testing dictate the choice of approach for the QSAR study.

The most popular and widely used approach continues to be the so called Hansch approach,[116-118] where the variance in biological effect (ΔBE) is explained by the variance in certain linear free-energy related substituent constants which describe the changes in lipophilic/hydrophilic (ΔL/ΔH), electronic (ΔE), steric (ΔS), and other properties of the parent molecule induced by the substituents. This model can be expressed as follows:

$$\Delta BE = f(\Delta L/\Delta H, \Delta E, \Delta S, \ldots..) \qquad (1.1)$$

The lipophilicity of a molecule can be described by the logarithm of partition coefficient P, measured in octanol-water system.[124] The change in lipophilicity or hydrophobicity due to a substituent is described by the lipophilic or hydrophobic constant π of the substituent defined as $\pi = \log P_X - \log P_H$, where X refers to the substituted derivative and H to the parent compound.[125] Lipophilicity can also be described by R_m values obtained from reverse-phase chromatography and by log k obtained from High Pressure Liquid Chromatography (HPLC). The change in electronic properties can be expressed by Hammett constant (σ),[126] charge densities, spectroscopic properties like chemical shift from IR or UV spectra, field constant (F), and resonance constant (R). The steric influence of the substituents can be described by the Taft steric constant (E_s),[127] molar volume (MV), and molar refractivity (MR).[128,129]

Besides, many a drug activities have been found to depend exclusively upon the molecular size,[130-138] which can be described by the van der Waals volume (V_w), and upon the molecular graph which is delineated by molecular connectivity index (χ).[139] In addition to these, Verloop's [140] width parameters B and length parameter L, evaluated by measuring the dimensions of substituents in a restricted number of directions with the aid of a computer program called STERIMOL, were also found to be useful in QSAR study. These parameters in their dimensional nature are indicative of the deviations of a substituent from spherical shape and their use might provide a better understanding of steric requirements in ligand-receptor interactions. In the present thesis, an extensive use has been made of these parameters.

In a stepwise linear multiple regression analysis, the biological activity (BA) can be related to various physicochemical, electronic, and steric parameters as:

$$BA = a\,\pi\ (\text{or } \log P) + b\,\pi^2\ (\text{or } [\log P]^2) + c\sigma + dE_s + k \qquad (1.2)$$

Where a, b, c and d are the regression coefficients and k the intercept obtained by least square method. Biological activity can be expressed by negative logarithm of the concentration of drug leading to a desired response. This reciprocal of the concentration used reflects the fact that greater potency is associated with a lower dose. Equation (1.2) shows a nonlinear, i.e., a parabolic dependence of activity on the hydrophobic character of molecules. Actually, Hansch had assumed a "random walk" of the molecules, where hydrophilic molecules tend to remain in aqueous phase, while hydrophobic molecules tend to go into lipid phase.

Only those molecules that have an optimal hydrophilic/hydrophobic balance tend to reach their goal in reasonable time and concentration. The nonlinear dependence of activity on π or log P value in an *in vivo* system is due to the nonlinear dependence of the rate constant of drug transport thorough aqueous and bio-organic phases on lipophilicity, whereas in *in vitro* systems, such as enzyme inhibition, such nonlinear relationships result from equilibrium distribution of the drug toward different areas at the enzyme surface, from limited binding space at the active site, or from limited solubility of more lipophilic congeners.

However, in many cases the relationships between activity and lipophilicity were found to be strictly linear [116-118] and although the parabolic model proved to be extremely useful for practical purposes, there was an inconsistency between it and the linear model. Although much less is known about the dependence of biological activities on lipophilic character beyond the point of optimal lipophilicity (log P_o or π_o), most often a linear relationship is observed with a negative slope beyond it. To overcome such inconsistencies between the linear and nonlinear models, a number of different models[141-147] were proposed, out of which Kubinyi's bilinear model was found, after Hansch's parabolic model, to be the most useful model[148-154] to describe the nonlinear relationships.

An alternative method, Free and Wilson analysis, is useful in systems in which the series of analogues are substituted at different positions. A Hansch treatment of such a series is complicated by the large number of physicochemical constants to be investigated.

The additive model of Free and Wilson gives the relationship between biological activity and the presence or absence of a substituent was then expressed by the following equation:

$$\text{Activity} = A + G_{ij}A_{ij} \qquad (1.3)$$

where A was defined as the average biological activity for the series, G_{ij} the contribution to activity of a functional group i in the jth position and X_{ij} the presence (1.0) or absence (0.0) of the functional group i in the jth position.

In the Free-Wilson treatment, hydrogen is treated as a substituent. The intercept of the resulting model does not represent the activity of the unsubstituted parent structure, but merely an average of biological activities of all compounds in the series.

In the Fujita and Ban analysis, a modified version of the Free-Wilson treatment, the H-substituent value at each position is set equal to zero. This is just a linear transformation of the Free-Wilson equation performed by subtracting the H-substituent value from each substituent constant, at that position, and adding the same value to the intercept. The intercept now represents the activity of the parent structure.

1.6 Applications of QSAR in Drug Design

After formulation of a statistically significant as well as physicochemically meaningful correlation equation for a given set of compounds, the informations contained in the equation can be used to design new compounds. According to the method of

utilization of the informations, examples could be classified into at least three categories:

1. Extrapolation of certain parameters toward directions enhancing the potency. As the correlation may or may not be linear, the best way this can be done is to gradually extend the extrapolation until the maximum potency is generated.

2. Identification of optimum structures with respect to certain parameters. If a parabolic dependence of the activity on certain parameters is revealed, the structure can be optimized by being modified so that the value of the parabolic parameter term is close to the maximum. This way, the best compound in the series can be identified and depending on this one could make a decision to continue or discontinue the synthetic program.

3. Transposition of QSAR informations to other series of compounds. The QSAR informations derived from a set of compounds $A\text{-}X_{1\sim n}$, prepared mainly on the basis of introduction or replacement of substituents, can be utilized to design new structures, $A\text{-}X_m$, where A is the basal skeletal structure that is kept unchanged and X means variable substituents or substructures.

A number of examples can be quoted where various combinations of the above points have been utilized to design compounds actually exhibiting the predicted activity.

1.7 Limitations of QSAR

Though QSAR studies can be successfully utilized to predict the activity of new analogues and discuss the mechanisms of drug-receptor interactions, they have some drawbacks and limitations as described below.[155]

The substituent effect on hydrophobicity is characterized by logP based on an octanol-water system; hence, even a very significant correlation can not represent a true model for hydrophobic interaction between a drug molecule and the receptor. The value of logP also depends on the electronic characters and the hydrogen bonding properties of the substituents.[156,157] Thus, if one gets a correlation with logP only, one cannot conclude that there is only hydrophobic interaction between drug and receptor and that no electronic interaction or hydrogen bonding takes place. Another factor that may influence logP values is steric effect that can prevent the access of water to a hydrophilic group.[158] Steric interactions are extremely difficult to extrapolate from system to system. The use of parameters like MR, MW, V_w, etc., do not give any idea in what way steric effects would effect the drug–receptor interaction. A more serious problem arises with the electronic parameters. The Hammett constants do not reflect which portion of the drug molecule would be actually involved in the interaction with the receptor. Quantum mechanical calculations can provide some help in this, but they are time consuming and expensive.

Although molecules are represented as rigid structures on paper, they may in fact, be quite different in solution and their dynamic nature should be recognized. There is considerable evidence that macromolecules, even in crystalline state, exhibit a wide spectrum of motions.[159-167] These motions may be involved in some molecular

conformational changes on substrate or group binding. Both drugs and biomolecules are three dimensional objects whose chemical features are related to their three dimensional structures. The interaction between them involves a complementarity or fit between the two objects. Even a successful QSAR study will provide only indirect information about the three-dimensional aspects of drug-biomolecule interaction

Many structural features that affect the activity but can not be parameterized by the usual variables like π, σ, E_s, etc., are accounted for by the use of indicator variables. These indicator variables are arbitrarily assigned two values: one to indicate the presence of the specific structural feature and other to indicate its absence. If the entire series of congeners is divided into two sets, one with and one without the specific structural feature, one would obtain two equations almost parallel, with a difference in their intercepts only. An indicator variable thus can be pictured simply as a constant that adjusts two parallel equations into one. If two sets are far apart in data space described by the usual parameters, one builds in a large amount of variance with the indicator variable leading to a much higher correlation coefficient (r).[168] Despite the better 'r', the new correlation may be a poorer one, and thus, one can be misled if other statistical parameters are not available.

Another serious problem in QSAR analysis is the problem of collinearity.[169] For example, π and MR most often turn out to be so collinear that it becomes impossible to tell whether one or both are involved in SAR. Over and above all, a QSAR study may be incorrectly interpreted if the biological property of interest is not correctly measured. A measured biological response maybe a complex result of several processes, and an *in vitro* model of drug-receptor interaction does not always represent the true *in vivo* model.

REFERENCES

1. Gallo, R.C., Sarin, P.S., Gelmann, E.P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V.S., Mann, D., Sidhu, G.D., Stahl, R.E., Zolla-Prazner, S., Leibowitch, J. and Popovic, M. *Science*, **220**, 865 (1983).
2. Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Grest, J., Dauguest, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W. and Montagnier, L. *Science*, **220**, 868 (1983).
3. Popovic, M., Sarngadharan, M.G., Read, E and Gallo, R.C. *Science*, **224**, 497 (1984).
4. Gallo, R.C., Salahuddin, S.Z., Popovic, M., Shearer, G.M., Kaplan, M., Haynes, B.F., Foster, P. and Markham, P.D. *Science*, **224**, 500 (1984)
5. Gottlieb, M.S., Schroff, R. and Schanker, H.M. *N. Engl. J. Med.*, **305**, 1425 (1981)
6. Sarngadharan, M.G., Popovic, M., Bruch, L., Schupbach, J. and Gallo, R.C. *Science*, **224**, 506 (1984).
7. Schupbach, J., Popovic, M. Gilden, R.V., Gonda, M.A., Sarngadharan, M.G. and Gallo, R.C. *Science*, **224**, 503 (1984).
8. Levy, J.A., Hoffman, A.D., Kramer, S.M., Landis, J.A., Shimabukuro, J.M. and Oshino, L.S. *Science*, **225**, 840 (1984).
9. Coffin, J., Haase, A., Levy, J.A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P. and Weiss, R. *Science*, **232**, 697 (1986).
10. Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M.-A., Santos-Ferreira, A. M.O., Laurent, A.G., Dauguet, C., Katlama, C., Rouzioux, C., Klatzmann, D., Champalimaud, J.L. and Montagnier, L. *Science*, **233**, 343 (1986).
11. Brookmeyer, R. *Science*, **253**, 37 (1991).
12. Quinn, T.C. *Lancet*, **348**, 99 (1996).
13. Jaffe, H.W., Bregman, D.J. and Selik, R.M. *J. Inf. Dis.*, **148**, 339 (1983).
14. Harris, C., Small, C.B., Klein, R.S., Friedland, G.H., Moll, B., Emeson, E.E., Spigland, I. and Steigbigel, N.H. *N. Eng. J. Med.*, **308**, 1181 (1983).

15. Masur, H., Michelis, M.A. and Greene, J.B. *N. Engl. J. Med.*, 305, 1431 (1981).
16. Amman, A.J., Cowan, M.J., Wara, D.W., Weintrub, P., Dritz, S., Goldman, H. and Perkins, H.A. *Lancet*, i: 956 (1983).
17. Dleske, J., Minnefor, A., Copper, Jr. R., Thomas, K., dela Cruz, A., Ahdieh, H., Guerrero, I., Joshi, V.V. and Desposito, F. *J. Am. Med. Assoc.*, **249**, 2345 (1983).
18. Schochetman, G., Epstein, J.S. and Zuck, T.F. *Annu. Rev. Microbiol.*, **43**, 629 (1989).
19. Gallo, D., George, J.R., Fitchen, J.H., Goldstein, A.S. and Hindahl, M.S. *J. Am. Med. Assoc.*, **277**, 254 (1997).
20. Berrios, D.C., Avins, A.L., Haynes-Sanstad, K., Eversley, R. and Woods, W.J. *Arch. Pathol. Lab. Med.*, **119**, 139 (1995).
21. Mildvan, D., Mathur, U. and Enlow, R.W. *Ann. Intern. Med.*, **96**, 700 (1982).
22. Bower, D.L., Lane, H.C. and Fauci, A.S. *Ann. Intern. Med.*, **103**, 704 (1985).
23. Embreston, J., Zupanic, M., Ribas, J.L., Burke, L., Racz, P., Tenner-Racz, K. and Haase, A.T. *Nature (London)*, **362**, 359 (1993).
24. Pedersen, C., Katzenstein, T. and Nielsen, C. *J. Acq. Imm. Def. Syn. Hum. Retrovirol.*, **16**, 93 (1997).
25. Mellors, J.W., Kingsley, L.A. and Rinaldo, Jr. C.R. *Ann. Intern. Med.*, **122**, 573 (1995).
26. Mellors, J.W., Rinaldo, Jr. C.R. and Gupta, P. *Science*, **272**, 1167 (1996)
27. Keet, I.P., Krijnen, P. and Koot, M. *AIDS*, 7, 51 (1993).
28. Schechter, M.T., Craib, K.J. and Le T.N. *AIDS*, **4**, 185 (1990).
29. (a) Isselbacher, K.J., Eugene, B., Wilson, J., Martin, J.B., Fauci, A.S. and Dennis, L.K. In: *Harrison's Principles of Internal Medicine*, Mc Graw-Hill, New York (1994), Vol. 1, 13th edition. (b) Gallo, R.C. *the human immunodeficiency virus (HIV)*, Scientific American, January, 1987, p. 46.
30. Fauci, A.S. *Science*, **239**, 617 (1998).
31. Muesing, M.A., Smith, D.H. and Cabradila, C.D. *Nature*, **313**, 450 (1985).
32. Greene, W.C. *N. Engl. J. Med.*, **324**, 308 (1991).

33. Schmit, J.C., Ruiz, L., Clotet, B., Raventos, A., Tor, J., Leonard, J., Desmyter, J. and De Clercq, E. *AIDS*, **10**, 995 (1996).

34. Gu, Z.X., Fletcher, R.S., Arts, E.J., Wainberg, M.A. and Parniak, M.A. *J. Biol. Chem.*, **269**, 28118 (1994).

35. Najera, I., Richman, D.D., Olivares, I., Rojas, J.M., Peinado, M.A., Perucho, M., Najera, R. and Lopez-Galindez, C. *AIDS Res. Hum. Retrov.*, **10**, 1479 (1994).

36. Stclair, M.H., Martin, J.L., Tudor-williams, G., Bach, M.C., Vavro, C.L., King, D.M., Kellam, P., Kemp, S.D. and Larder, B. *Science*, **253**, 1557 (1991).

37. Elfarrash, M.A., Kuroda, M.J., Kitazaki, T., Masuda, T., Kato, K., Hatanaka, M. and Harada, S. *J. Virol.*, **68**, 233 (1994).

38. Borman, A.M., Paulous, S. and Clavel, F. *J. Gen. Virol.*, **77**, 419 (1996).

39. De Clercq, E. *J. Med. Chem.*, **38**, 2491 (1995).

40. Hussey, R.E., Richardson, N.E., Kowalski, M., Brown, N.R., Chang, C.-H., Siliciano, R.F., Dorfman, T., Walker, B., Sodroski, J. and Reinherz, E.L. *Nature (London)*, **331**, 78 (1988).

41. Broder, S., Yarchoan, R. and Collins, J.M. *Lancet*, **2**, 627 (1985).

42. Lorentsen, K.J., Hendrix, C.W., Collins, J.M., Komhauser, D.M., Petty, B.G., Klecker, R.W., Flexner, C., Eckel, R.H. and Lietman, P.S. *Ann. Inter. Med.*, **111**, 561 (1989).

43. Balzarini, J., Neyts, J., Schols, D., Hosoya, M. Van Damme, E., Peumans, W. and De Clercq, E. *Antiviral Res.*, **18**, 191 (1992).

44. Balzarini, J., Schols, D., Neyts, J., Van Damme, E., Peumans, W. and De Clercq, E. *Antimicro. Agents Chemother.*, **35**, 410 (1991).

45. Nakashima, H., Masuda, M., Murakami, T., Koyanagi, Y., Matsumoto, A., Fuji, N. and Yamamoto, N. *Antimicrob. Agents Chemother.*, **36**, 1249 (1992).

46. Jansen, R.W., Molema, G., Pauwels, R., Schols, D., De Clercq, E., Meijr, D.K.F. *Mol. Pharmacol.*, **39**, 818 (1991).

47. Jansen, R.W., Schols, D., Pauwels, R., De Clercq, E. and Meijer, D.K.F. *Mol. Pharmacol.*, **44**, 1003 (1993).

48. Mayaux, J.F., Bousseau, A., Pauwels, R., Huet, T., Henin, Y. and Pecq, J.-B. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 3564 (1994).

49. Rossaman, M.G. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 4625 (1988).

50. Taken from the web site: http://www.niaid.nih.gov/daids/dtpdb/fdadrug.htm

51. De Clercq, E. *Il Farmaco*, **54**, 26 (1999).

52. De Clercq, E. *Med. Res. Rev.*, **13**, 229 (1993).

53. De Clercq, E. *Exp. Opin. Invest. Drugs*, **3**, 253 (1994).

54. Fesen, M.R., Kohn, K.W., Leteurtre, F., Pommier, Y. *Proc. Natl. Acad. Sci. U.S.A.*, 90, 2399 (1993).

55. Fesen, M.R., Pommier, Y., Leteurtre, F., Hiroguchi, S., Yung, J. and Kohn, D.W. *Biochem. Pharmacol.*, **48**, 595 (1994).

56. Mayaux, J.F., Huet, T., Pernelle, C., Becquart, J., Gueguen, J.C., Tahraui, L., Evers, M., Henin, Y., Bousseau, A., Dereu, N. *Abstracts of the NIH conference on Retroviral Integrase*, Bethesda, MD, January, 19-20 (1995).

57. Stein, C.A. and Cheng, Y.-C. *Science*, 261, 1004 (1993).

58. Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J. and Heimbach, J.C. *Proc. Natl. Acad. Sci. U.S.A.*, 85, 4186 (1988).

59. McQuade, T.J., Tomasselli, A.G., Liu, L., Karacostas, V. and Moss, B. *Science*, **247**, 454 (1990).

60. Seelmier, S., Schmidt, H., Turk, V. and von der Helm, K. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 6612 (1988).

61. Boehme, R.E., Borthwick, A.D. and Wyatt, P.G. *Annu. Rep. Med. Chem.*, **30**, 139 (1995).

62. Ratner, L. and Van der Heyden, N. *AIDS Res. Human Retrov.*, **9**, 291 (1993).

63. Toh, H., Ouo, M., Saigo, K. and Miyata, T. *Nature*, **315**, 691 (1985).

64. Miller, M., Jaskolski, M., Rao, J.K.M., Leis, J. and Wlodawer, A. *Nature*, **337**, 576 (1989).

65. Schechter, I. and Berger, A. *Biochem. Biophys. Res. Commun.*, **27**, 157 (1967).

66. Babine, R.E. and Bender, S.L. *Chem. Rev.*, **97**, 1359 (1997).

67. Wlodawer, A. and Ericson, J.W. *Annu. Rev. Biochem.*, **62**, 543 (1993).

68. Applet, K. *Perspec. Drug Dis. Des.*, 1, 23 (1993).

69. Kaplan, A.H. *Proc. Natl. Acad. Sci. U.S.A.*, 88, 4528 (1991).

70. Krausslich, H.G. *Proc. Natl. Acad. Sci. U.S.A.*, 88, 3213 (1991).

71. Krausslich, H.G. *J. Virol.*, 66, 567 (1992).

72. Roberts, N.A., Martin, J.A., Kinchington, D., Broadhurst, A.V., Craig, J.C., Duncan, I.B., Galpin, S.A., Handa, B.K., Kay, J., Krohn, I.B., Lambert, R.W., Merrett, J.H., Mills, J.S., Parkes, K.E.B., Redshan, S., Ritchi, A.J., Taylor, D.L., Thomas, G.J. and Machin, P.J. *Science*, 248, 358 (1990).

73. McQuade, T.J., Tomasselli, A.G., Liu, L., Karacostas, V., Moss, B., Sawyer, T.K., Heinrikson, R.L. and Tarpley, W.G. *Science*, 247, 454 (1990).

74. Rich, D.H., Sun, C.-Q., Vara Prasad, J.V.N., Pathiasseril, A., Toth, M.V., Marshall, G.R., Clare, M., Mueller, R.A. and Houseman, K. *J. Med. Chem.*, 34, 1225 (1991)

75. Getman, D.P., DeCrescenzo, G.A., Heintz, R.M., Reed, K.L., Talley, J.J., Bryant, M.L., Clare, M., Houseman, K.A., Marr, J.J., Muller, R.A., Vazquez, M.L., Shieh, H.-S., Stallings, W.C. and Stegeman, R.A. *J. Med. Chem.*, 36, 288 (1993).

76. Kageyama, S., Mimoto, T., Murakawa, Y., Nomizu, M., Ford, H. Jr., Shirasaka, T., Gulnik, S., Erickson, J., Takada, K., Hayashi, H., Broder, S., Kiso, Y. and Mitsuya, H. *Antimicrob. Agents Chemother.*, 37, 810 (1993).

77. Erickson, J., Neidhart, D.J., Vandrie, J., Kempf, D.J., Wang, X.C., Norbeck, D.W., Plattner, J.J., Rittenhouse, J.W., Turon, M., Wideburg, N., Kohilbrenner, J.W., Simmer, R., Helfrich, R., Paul, D.A. and Knigge, M. *Science*, 249, 527 (1990).

78. Kempf, D.J., Norbeck, D.W., Codacovi, L., Wang, X.C., Kohlbrenner, W.E., Wideburg, N.E., Paul, D.A., Knigge, M.F., Vasavanonda, S., Craig-Kennard, A., Saldivar, A., Rosen-brook, W. Jr., Clement, J.J., Platter, J.J. and Erickson, J. *J. Med. Chem.*, 33, 2687 (1990).

79. Barrish, J.C., Gordon, E., Alam, M., Lin, P.-F., Bisacchi, G.S., Chen, P., Cheng, P.T.W., Fritz, A.W., Greytok, J.A., Hermsmeier, M.A., Humphreys, W.G., Lis, K.A., Marella, M.A., Merchant, Z., Mitt, T., Morrison, R.A., Obermeier, M.T.,

Pluscec, J., Skoog, M., Slusarchyk, W.A., Spergel, S.H., Stevenson, J.M., Sun, C.-q., sundeen, J.E., Taunk, P., Tino, J.A., Warrack, B.M., Colonno, R.J. and Zahler, R. *J. Med. Chem.*, 37, 1758 (1994).

80. Miller, M., Schneider, J., Satyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H. and Woldawar, A. *Science*, 246, 1149 (1989).

81. Fitzgerald, P.M.D., McKeever, B.M., Van Middlesworth, J.F., Springer, J.P., Heimbach, J.C., Leu, C.-T., Herber, W.K., Dixon, R.A.F. and Darke, P.L. *J. Biol. Chem.*, 265, 14209 (1990).

82. Erickson, J., Neidhart, D.J., Van Drie, J., Kempf, D.J., Wang, X.C., Norbeck, D.W., Plattner, J.J., Rittenbouse, J.W., Turon, M., Wideburg, N., Kohlbrenner, W.E., Simmer, R., Helfrich, R., Kohlbrenner, W.E., Simmer, R., Helfrich, R., Paul, D.A. and Knigge, M. *Science*, 249, 52 (1990).

83. Swain, A.L., Miller, M.M., Green, J., Rich, D.H., Schneider, J., Kent, S.B.H. and Wlodawar, A. *Proc. Natl. Acad. Sci. U.S.A.*, 87, 8805 (1990).

84. Plattner, J.J. and Norbeck, D.W. In: *Drug Discovery Technologies*; Clark, R., Moos, W.H., Eds., Ellis Horwood Ltd., Chichester, pp 92 (1990).

85. Kempf, D.J., Marsh, K.C., Denissen, J.F. McDonald, E., Vasavananda, S., Flentge, C.A., Green, B.E., Fino, L., Park, C.H., Kong, X.-P., Wideburg, N.E., Saldivar, A., Ruiz, L., Kati, W.M., Sham, H.L., Robinn, T., Stewart, K.D., Hau, A., Plattner, J.J., Leonard, J.M. and Norbeck, D.W. *Proc. Natl. Acad. Sci. U.S.A.*, 92, 2484 (1995).

86. Vacca, J.P., Dorsey, B.D., Schleif, W.A., Levin, R.B., McDaniel, S.L., Darke, P.L., Zugae, J., Quintero, J.C., Blahy, O.M., Roth, E., Sardana, V.V., Schlabach, A.J., Graham, P.I., Condra, J.H., Gotlib, L., Holloway, M.K., Lin, J., Chen, L.-W., Vastag, K., Ostovic, D., Anderson, P.S., Eimini, E.A. and Huff, J.R. *Proc. Natl. Acad. Sci. U.S.A.*, 91, 4096 (1994).

87. Kalish, V., Kaldor, S., Shetty, B., Tatlock, J., Davies, J., Hammend, M., Dressman, B., Fritz, J., Applet, K., Reich, S., Musick, L., Wu, B.-W. and Su, K. *Eur. J. Med. Chem.*, 30, 201 (1995)

88. Kim, E.E., Baker, C.T., Dwyer, M.D., Murcko, M.A., Rao, B.G., Tung, R.D. and Navia, M.A. *J. Am. Chem. Soc.*, 117, 1181 (1995).

89. Olson, G.L., Bolin, D.R., Bonner, M.P., Bos, M., Cook, C.M., Fry, D.C., Graves, B.J., Hatada, M., Hill, D.E., Kahn, M., Madison, V.S., Rusiecki, V.K., Sarabu, R., Sepinwall, J., Vincent, G.P. and Voss, M.E. *J. Med. Chem.*, 36, 3039 (1993).

90. Maligres, P.E., Upadhyay, V., Rossen, K., Cianciosi, S.J., Purick, R.M., Eng, K.K., Reamer, R.A., Askin, D., Volante, R.P. and Reider, P.J. *Tetrahedron Lett.*, 36, 2195 (1995).

91. Lam, P.Y.S., Jadhav, P.K., Eyermann, C.J., Hodge, C.N., Ru, Y., Bacheler, L.T., Meek, J.L., Otto, M.J., Raynner, M.M., Wong, N.Y., Chang, C.-H., Wever, P.C., Jackson, D.A., Sharpe, T.R. and Erickson-Viitanen, S. *Science*, 263, 380 (1994).

92. Nugiel, D.A., Jacobs, K., Worley, T., Patel, M., Kalten-batch III, R.F., Meyer, D.T., Jadhav, P.K., DeLucca, G.V., Smyser, T.E., Klabe, R.M., Bacheler, L.T., Rayner, M.M. and Seith, S.P. *J. Med. Chem.*, 39, 2156 (1996).

93. Lam, P.Y.S., Ru, Y., Jadhav, P.K., Aldrich, P.E., DeLucca, G.V., Eyerman, C.J., Chang, C.-H., Emmett, G., Holler, E.R., Danekar, W.F., Li, L., Confalone, P.N., Mc Hugh, R.J., Han, Q., Li, R., Markwader, J.A., Seitz, S.P., Sharpe, T.R., Bacheler, L.T., Rayner, M.M., Klabe, R.M., Shum, L., Winslow, D.L., Kornhauser, D.M., Jackson, D.A., Erickson-Viitanen, S. and Hodge, C.N. *J. Med. Chem.*, 39, 3514 (1996).

94. Romines, K.R., Watenpaugh, K.D., Tomich, P.K., Howe, W.J., Morris, J.K., Lovasz, K.D., Mulichak, A.M., Finzel, B.C., Lynn, J.C., Horng, M.-M, Schwende, F.J., Ruwart, M.J., Zipp, G.L., Chong, K.-T., Dolak, L.A., Toth, L.N., Howard, G.M., Rush, B.D., Wilkinson, K.F., Possert, P.L., Dalga, R.J. and Hinshaw, R.R. *J. Med. Chem.*, 38, 1884 (1995).

95. Romines, K.R., Watenpaugh, K.D., Howe, W.J., Tomich, P.K., Lovasz, K.D., Morris, J.K., Janakiraman, M.N., Lynn, J.C., Horng, M.-M., Chong, K.-T., Hinshaw, R.R. and Dolak, L.A. *J. Med. Chem.*, 38, 4463 (1995).

96. Romines, K.T., Morris, J.K., Howe, W.J., Tomich, P.K., Horng, M.-M., Chong, K.-T., Hinshaw, R.R., Anderson, D.J., Strohbach, J.W., Turner, S.R. and Mizsak, S.A. *J. Med. Chem.*, 39, 4125 (1996).

97. Skulnick, H.I., Johnson, P.D., Aristoff, P.A., Morris, J.K., Lovaszk, K.D., Howe, W.J., Watenpaugh, K.D., Janakiraman, M.N., Anderson, D.J., Reischer, R.J., Schwartz, T.M., Banitt, L.S., Tomich, P.K., Lynn, J.C., Horng, M.-M., Chong, K.-T., Hinshaw, R.R., Dolak, L.A., Seest, E.P., Schwende, F.J., Rush, B.D., Howard, G.M., Toth, L.N., Wilkinson, K.R., Kakuk, T.J., Johnson, C.W., Cole, S.L., Zaya, R.M., Zipp, G.L., Possert, P.L., Dalga, R.J., Zhong, W.-Z., Williams, M.G. and Romines, K.R. *J. Med. Chem.*, 40, 1149 (1997).

98. Ghosh, A.K., Lee, H.Y., Thomson, W.J., Culberson, C., Holloway, K., McKee, S.P., Munson, P.M., Duong, T.T., Smith, A.M., Darke, P.L., Zugay, J.A., Emini, E.A., Schleif, W.A., Huff, J.R. and Anderson, P.S. *J. Med. Chem.*, 37, 1177 (1994).

99. Vara Prasad, J.V.N., Para, K.S., Tummino, P.J., Ferguson, D., McQuade, T.J., Lunney, E.A., Rapundalo, S.T., Batley, L.B., Hingorani, G., Domalguda, J.M., Gracheck, S.J., Bhat, T.N., Liu, B., Baldwin, E.T., Erickson, J.W. and Sawyer, T.K. *J. Med. Chem.*, 38, 898 (1995).

100. Hodge, C.N., Aldrich, P.E., Bacheler, L.T., Chang, C.-H, Eyerman, C.J., Gribb, M.F., Jackson, D.A., Jadhav, P.K., Korant, B., Lam, P.Y.S., Maurin, M.B., Meek, J.L., Otto, M.J., Rayner, M.M., Sharpe, T.R., Shum, L., Winslow, D.L. and Erickson-Viitanen, S. *Chem. Biol.*, 3, 301 (1996).

101. Jadhav, P.K., Woerner, F.J., Lam, P.Y.S., Hodge, C.N., Eyermann, C.J., Man, C.J., Man, H.-W., Danekar, W.F., Bachelor, L.-T., Rayner, M.M., Meek, J.L., Erickson-Viitanen, S., Jackson, D.A., Calabrese, J.C., Schadt, M. and Chang, C.-H. *J. Med. Chem.*, 41, 1446 (1998).

102. Ala, P.J., Huston, E.E., Klabe, R.M., McCabe, D.D., Duke, J.L., Rizzo, C.J., Korant, B.D., DeLoskey, R.J., Lam, P.Y.S., Hodge, C.N. and Chang, C.-H. *Biochemistry*, 7, 1573 (1997).

103. Jacobson, H., Yasargil, K., Winslow, D.L., Craig, J.C., Krohn, A., Duncan, I.B. and Mous, J. *Virol.*, 206, 527 (1995).

104. Ridky, T. and Leis, J. *J. Biol. Chem.*, 270, 29621 (1995).

105. Condra, J.H., Schleif, W.A., Blahy, O.M., Gabryelski, L.J., Graham, D.J., Quintero, J.C., Rhodes, A., Robbins, H.L., Roth, E., Shivaprakash, M., Titus, D., Yang, T., Teppler, H., Squires, K.E., Deutsch, P.J. and Emini, E.A. *Nature*, 374, 569 (1995).

106. Deeks, S.G., Smith, M., Holodnity, M. and Kahn, J.O. *J. Am. Med. Assoc.*, 277, 145 (1997).

107. Hansch, C., Maloney, P.P., Fujita, T. and Muir, R.M. *Nature*, 194, 178 (1962).

108. Hansch, C. and Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington, D.C (1995).

109. Hansch, C. and Klein, T. *Acc. Chem. Res.*, 19, 392 (1986).

110. Blaney, J.M. and Hansch, C. Comprehensive Medicinal Chemistry, Vol. 4, Application of Molecular Graphics to the Analysis of Macromolecular Structures. Pergamon Press (1990).

111. Blaney, J.M., Hansch, C. Silipo, C. and Vittoria, A. *Chem. Rev.*, 84, 333 (1984).

112. Hansch, C. and Blaney, J.M. In, *Drug Design: Fact or Fantasy?*, G. Jolles and Wooldridge, Eds., Academic Press, pp.185.

113. Hansch, C. and Klein, T. *Methods in Enzymology*, 202, 512 (1991).

114. Carrieri, A., Altomare, C., Barreca, M.L., Contento, A., Carotti, A. and Hansch, C. *Farmaco*, 49, 573 (1994).

115. Selassie, C.D. and Klein, T.E. In: *3-D QSAR and Drug Design Theory Methods and Applications*, H. Kubinyi, Ed., ESCOM Publishers, Leiden, pp. 257.

116. Hansch, C. *Acc. Chem. Res.*, 2, 232 (1969).

117. Hansch, C. and Fujita, T. *J. Am. Chem. Soc.*, 86, 1616 (1969).

118. Hansch, C. In: *Drug Design*, Ed. Arien's E.J. vol. 1, Academic Press, New York, p. 271 (1971).

119. Free, S.M. Jr. and Wilson, J.W. *J. Med. Chem.*, 7, 395 (1964).

120. Fujita, T. and Ban, T. *J. Med. Chem.*, 14, 232 (1971).

121. Martin, Y.C., Holland, J.B., Jaorbe, C.H. and Plotnikoff, N. *J. Med. Chem.*, **17**, 409 (1974).
122. Kowalski, B.R. and Bonder, C.F. *J. Am. Chem. Soc.*, **94**, 5632 (1972).
123. Chu, K.C. *Analyt. Chem.*, **46**, 1181 (1974).
124. Levitan, H. and Berker, J.L. *Science*, **176**, 1423 (1974).
125. Leo, A., Hansch, C. and Elkins, D. *Chem. Rev.*, **71**, 525.
126. Hammet, L.P. *Physical Organic Chemistry*, McGraw Hill, New York (1940).
127. Taft, R.W. In: *Steric Effects in Organic Chemistry*, Ed. Newman, M.S., Wiley, New York, p. 556 (1956).
128. Hansch, C., Leo, A., Unger, S.H., Kim, K.H., Nikaitani, D. and Lein, E.J. *J. Med. Chem.*, **16**, 1207 (1973).
129. Dunn, W.J. *Eur. J. Med. Chem.*, **12**, 109 (1977).
130. Moriguchi, J. and Kanada, Y. *Chem. Pharm. Bull.*, **25**, 926 (1977).
131. Gupta, S.P, Bindal, M.C. Singh, P. and Bhatnagar, R.P. *Res. Commun. Chem. Pharmacol.*, **25**, 111 (1979).
132. Bindal, M.C., Singh, P., Bhatnagar, R.P. and Gupta, S.P. *Arzneim Forsch.*, **30**, 924 (1980).
133. Kumar, K., Bindal, M.C., Singh, P. and Gupta, S.P. *Int. J. Quant. Chem.*, **20**, 123 (1981).
134. Gupta, S.P., Singh, P. and Bindal, M.C. *Eur. J. Med. Chem.*, **16**, 446 (1981).
135. Gupta, S.P., Singh, P. and Bindal, M.C. *Indian J. Biochem. Biophys.*, **19**, 143 (1982).
136. Handa, A., Bindal, M.C., Prabhakar, Y.S. and Gupta, S.P. *Indian J. Biochem. Biophys.*, **20**, 318 (1983).
137. Verloop, A., Hoogenstraaten, W. and Tipker, J. In: *Drug Design*, vol. 7, Ed. Arien`s E.J., Academic Press, New York (1976), p. 165.
138. Gupta, S.P., Prabhakar, Y.S. and Handa, A. *Res. Commun. Chem. Pathol. Pharmacol.*, **42**, 455 (1983).
139. Gupta, S.P., Prabhakar, Y.S., Handa, A. and Bindal, M.C. *Arzneim Forsch.*, **34**, 147 (1984).

140. Kier, L.B. and Hall, L.H. *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York (1976).

141. Macfarland, J.W. *J. Med. Chem.*, **13**, 1192 (1970).

142. Penniston, J.T., Beckett, L., Bentley, D.L. and Hansch, C. *Mol. Pharmacol.*, **5**, 333 (1969).

143. Higuchi, T. and David, S.S. *J. Pharm. Sci.*, **59**, 1376 (1970).

144. Hyde, R.M. *J. Med. Chem.*, **18**, 231 (1975).

145. Franke, R. and Schmidt, W. *Acta. Biol. Med. Ger.*, **31**, 273 (1973).

146. Yalkowsky, S.H. and Flynn, G.L. *J. Pharm. Sci.*, **62**, 210 (1973).

147. Kubinyi, H. *Arzneim. Forsch.*, **26**, 1991 (1976).

148. Kubinyi, H. *Prog. Drug Res.*, **23**, 97 (1979).

149. Kubinyi, H. *Arzneim. Forsch.*, **29**, 1067 (1979).

150. Seydel, J.K. In, *Strategy in Drug Research*, Ed. Buisman, J.A.K. (Ed.), Elsevier, Amsterdam (1982), p. 179.

151. Seydel, J.K. and Schaper, K.J. *Pharmacol. Ther.*, **15**, 131 (1982).

152. Kubinyi, H. *J. Med. Chem.*, **20**, 625 (1977).

153. Li, R.L., Hansch, C., Mathews, D., Blaney, J.M., Langridge, R., Delcamp, T.J., Susten, S.S. and Freisheim, J.H. *Quant. Struct. Act. Relat. Pharmacol. Chem. Bio.*, **1**, 1 (1982).

154. Kubinyi, H. In: *QSAR in Design of Bioactive Compounds*, Ed. Kuchar, M., J.R. Prous Science, Barcelona, Spain (1984), p. 321.

155. Martin, Y.C. *J. Chem.*, **24**, 229, (1981).

156. Leo, A. *J. Chem. Soc. Perkin Trans.*, **2**, 825 (1983).

157. Fujita, T. *J. Pharm. Sci.*, **72**, 285 (1983).

158. Martin, Y.C. In: *Quantitative Drug Design*, Marcel Dekker, New York (1978), p. 74.

159. McCammon, J.A., Gelin, B.R., Karplus, M. and Wolynes, P.G. *Nature*, **262**, 325 (1976).

160. McCammon, J.A., Gelin, B.R., Karplus, M. and Wolynes, P.G. *Nature*, **267**, 585 (1977).

161. McCammon, J.A. and Karplus, M. *Nature*, **268**, 765 (1977).

162. Karplus, M. and Ann. N.Y. *Acad. Sci.*, **367**, 407 (1981).

163. Karplus, M. and Ann. N.Y. *Acad. Sci.*, **439**, 107 (1985).

164. Maston, G.B. *J. Chem. Phys.*, **65**, 4147 (1976).

165. Wuthrich, K., Wagner, G. and Demarco, A. *Experimentia.*, **32**, 815 (1976).

166. Hallenga, K. and Koenig, S.H. *Biochemistry*, **15**, 4255 (1976).

167. Baldwin, R.L. *Annu. Rev. Biochem.*, **44**, 453 (1975).

168. Hansch, C. and Calef, D.E. *J. Org. Chem.*, **41**, 1240 (1976).

169. Hansch, C. and Leo, A.J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*, John Wiley, New York (1979).

CHAPTER 2

PARAMETERS USED AND THEIR CALCULATIONS

2 PARAMETERS USED AND THEIR CALCULATIONS

This chapter discusses the methodology of calculation of various distinct parameters, on which most of the biological activities are found to depend and which are very useful in QSAR studies.

2.1 Hydrophobic Parameter [log P]

The fragment method suggested by Hansch and Leo[1] for calculating log P, where P is the partition coefficient of the solute in octanol-water system, is known as constructionist or synthetic approach. Experimentally determined log P values can often be reproduced or approached theoretically with the help of this approach. The basic assumption of this approach is: the log P of a solute can be expressed as a linear sum of fundamental structural constants known as fragments (*f*) and factors (*F*) that affect the partitioning equilibrium.

$$\text{LogP} = \sum_{1}^{n} a_n f_n + \sum_{1}^{m} b_m F_m \qquad (2.1)$$

Carefully conducted partitioning experiment and statistical survey of the then available partition data have been used in assigning values to the fragments and factors. The working principle is summarized in the following paragraphs.

In this approach carbon atoms are divided into two categories: isolating carbons (IC) and nonisolating carbons (NIC). ICs are those having either four single bonds (at least two of which are to non heteroatoms) or else are multiply bonded to other carbon atoms.

NIC atoms are carbon atoms multiply bonded to hetero atoms. For example –C= in $CH_2=CH_2$ is an IC but not in $H_2C=O$. Fragments are of two types: (i) fundamental fragments defined as fragments whose free valency will lead to isolating carbons; (ii) derived fragments, a derivative of fundamental fragments (e.g. CF_3). A fundamental fragment can be either a single atom or a group of multiple atoms (e.g., –C=O, -C≡N, etc.). A single atom fundamental fragment can be either an isolating carbon atom or a hydrogen or a hetero atom all of which are bonded to ICs. Depending on its nature a fragment will come under one of the following classes:

(i) **Non-polar fragments**: These are simple ICs and hydrogens attached to ICs.

(ii) **H-polar fragments**: A fragment that can be expected to form H-bonds either by accepting or donating an electron pair (e.g. -OH, -COOH, $-NH_2$ etc.).

(iii) **S-polar fragments**: A fragment that is strongly electron withdrawing with little tendency to form H-bonds (e.g. halogens). In expressing fragments, the structural formulae (or WLN code) of the respective fragments will be written as subscripts of "*f*" for example as *f*-NH-CO-NH for expressing the fragment –NH-CO-NH- present in $CH_3NHCONHCH_3$. Various factors (*F*) are designed to account for the intramolecular forces and factors that affect the partitioning equilibrium of the solute. All these F_s are identified with the help of different subscripts and superscripts. The subscripts are mentioned in the factors table (Table (2.2)). The superscripts are applicable also to fragments.

They are listed as:

1. None = aliphatic structural attachment
2. ϕ = attachment to aromatic ring; if bivalent the attachment is from left as written
3. $1/\phi$ = as 2 but attachment from right as written
4. $\phi\phi$ = two aromatic attachments
5. X = aromatic attachment, value enhanced by second, electron–withdrawing substituent ($\sigma_I \geq \pm 0.35$) and
6. IR = benzyl attachment.

Underlining any symbol means it is present in a ring system. Whenever halogens and H-polar fragments are separated by only one ICs an additional factor will come into operation.

In calculating the logP of any compound, the first step is dividing that compound into "well defined" fragments based on the above discussion and then searching for different factors operating in between the fragments within the structure of the molecule. Now the sum of all these fragments and factors will give the calculated logP of that compound. It is always safe to break any compound, especially compound containing hetero atoms, into fundamental fragment rather than into derived fragment. Some important fragment values and factor values are listed in Tables (2.1) and (2.2), respectively. A simple example for logP calculation is shown below:

Table (2.1): Some common fragment constants*

Without Carbon	f	f^{ϕ}	$f^{\phi\phi}$	With Carbon	f	f^{ϕ}	$f^{\phi\phi}$
-Br	0.20	1.09		C	0.20	0.20	
- Cl	0.06	0.94		$-CF_3$[a]		1.11	
- F	-0.38	0.37		-CN	-1.27	0.34	
- I	0.59	1.35		-CON	-3.04	-2.80	-1.93
- N	-2.18	-0.93	-1.13	-C(O)-	-1.90	-1.09	-0.50
$- NO_2$	-1.16	-0.03		$-CO_2-$	-1.49	-0.56	-0.09
- O-	-1.82[b]	-0.61	0.53	$-CO_2$	-5.19	-4.13	
- H	0.23	0.23		-COH	-1.10	-0.42	
- NH-	-2.15	-1.03	-0.09	$-CO_2H$	-1.11	-0.03	
$- NH_2$	-1.54	-1.00		$-CONH_2$	-2.18	-1.26	
- OH	-1.64	-0.44		-CONH-	-2.71	-1.81	-1.06
- SH	-0.23	0.62		-NHCONH-	-2.18	-1.57	-0.82

Fused in Aromatic Ring

Without Carbon	f^{ϕ}	Without Carbon	f^{ϕ}	With Carbon	f^{ϕ}	With Carbon	f^{ϕ}
- N =	-1.12	-N = N-	-2.14	C	0.13	-CH-	0.35
- N	-1.60	-O-	-0.08	C	0.25[c]	-C(O)-	-0.59
$- N^{\phi}$	-0.56	-NH-	-0.65	C*	0.44[d]	-OC(O)-	-1.40

*Taken from ref. 1, [a]Derived fragment, [b]For methyl ethers and ethylene oxide, use –1.54, [c]For ring fusion carbon, [d]For ring fusion – hetero

Table (2.2): *List of some factors**

	Involving Bonds			
	Unsaturation			Geometric
	Double	Triple	Proportional to Length: x(n-1)	Short Chains: 1-time
Normal	F(=) = -0.55	F(≡) = -1.42	Chain: F_b = -0.12	Alkane Chain: F_{cBr} = 0.13
Conjugate to φ	F^{φ} (=) = -0.42		Ring[a] : F_b = -0.09	H-polar Fragment: F_{gBr} = -0.22
Conjugate to 2 φ	$F^{\varphi\varphi}$ (=) = 0.0	$F^{\varphi\varphi}$ (≡) = 0.0	Branching: F_{bYN} = -0.20[b] F_{bYP} = -0.20[c]	Ring Cluster: F_{rCl} = -0.45

	Involving Multiple Halogenation[d]		
	On same Carbon (geminal) F_{mhCm}	(n=2) = 0.30 (n=3) = 0.53 (n=4) = 0.72	On adjacent Carbon (vicinal) F_{mhVn} : 0.28 (n-1)

	Involving H-polar Proximity				
Chain	$F_p^1 = -0.42\,\Sigma f_1 + f_2$ $F_p^2 = -0.26\,\Sigma f_1 + f_2$ $F_p^3 = -0.10\,\Sigma f_1 + f_2$	Aliphatic: Ring:	$F_p^1 = -0.32\,\Sigma f_1 + f_2$ $F_p^2 = -0.20\,\Sigma f_1 + f_2$	Aromatic :	$F^{\varphi}{}_p^1 = -0.16\,\Sigma f_1 + f_2$ $F^{\varphi}{}_p^2 = -0.08\,\Sigma f_1 + f_2$

Involving Intramolecular H-bond	
F_{HBN} = 0.60 for Nitrogen	F_{HBO} = 1.0 for Oxygen

*Taken from Ref.1. [a]Aromatic rings are excluded. [b]For amine
[c]For Phosphorus esters. [d]Value per halogen atom

Example, Toluene ($C_6H_5CH_3$): This can be treated as a compound comprising six aromatic carbons, one aliphatic carbon and eight hydrogens.

The fragments can be expressed as:

$$6f^{\circ}\ c + fc + 8f_H = \log P\ (\text{Toluene})$$

$$6(0.13) + 0.20 + 8(0.23) = 2.82\ (\text{calcd.}),\ 2.80\ (\text{obsd.})$$

Since aromatic ring is excluded from bond factor there is no F_b term in the above equation. And here aliphatic chain length is one ($-CH_3$), so $(n-1)\ F_b$ is equal to zero (C-H bonds are excluded from factors). The logP of this compound can also be calculated from two derived fragments as:

$$f^{\circ}_{\underline{C6H5}} + f CH_3 = \log P\ (\text{toluene})$$

$$1.9 + 0.89 = 2.79\ (\text{calcd.})$$

Sometimes calculated logP values of compounds deviate very much from the experimentally determined values. For example, observed logP of 1,2-methylene-dioxybenzene is 2.08, but the calculated value comes out to be 1.34 only. This large difference may be due to factors beyond the control of this method. However, since it is an additive model, it will serve the purpose of drug design when used in a congeneric series of compounds. Further details are given in the literature.[1]

2.2 Hydrophobic Constant (π) of Substituents

Although logP can be used as a measure of the hydrophobicity of a whole molecule, one often works with a set of derivatives of a parent compound in which a

large portion of the structure remains constant. In such a case, knowing the relative hydrophobicity of substituents can be sufficient for correlation analysis. Sometimes it has been found that only substituents in certain positions interact hydrophobically with a given biosystem.[2,3] To enable one to work with the relative hydrophobicity of substituents and in this way separate hydrophobic character from electronic and steric effects of substituents, the parameter π has been defined as

$$\pi_x = \log P_x - \log P_H \qquad (2.2)$$

In this expression, Px is the partition coefficient of a derivative and P_H that of the parent compound, for example:

$$\pi_{Cl} = \log P_{C_6H_5Cl} - \log P_{C_6H_6} \qquad (2.3)$$

$$0.71 = 2.84 - 2.13$$

A positive value for π means that relative to H the substituent favours the octanol phase. A negative value indicates its hydrophilic character relative to H. The value of π varies somewhat from system to system. Certain π values are given in Table (2.3).

2.3 Electronic Parameter (σ)

The development of electronic parameter is one of the most important breakthrough for mechanistic organic chemistry which came in 1935 when L.P. Hammett[4] proposed the following equation to define an electronic parameter σ.

$$\sigma = \log K_X - \log K_H \qquad (2.4)$$

Table (2.3): Data on Physicochemical parameters for some important substituents*

No.	Substituent	π	σ_m	σ_p	MR
1	H	0.00	0.00	0.00	1.03
2	CH_3	0.56	-0.07	-0.17	5.65
3	C_2H_5	1.02	-0.07	-0.15	10.30
4	C_3H_7	1.05	-0.07	-0.13	14.96
5	$i-C_3H_7$	1.53	-0.07	-0.15	14.96
6	$n-C_4H_9$	2.13	-0.08	-0.16	19.61
7	F	0.14	0.34	0.06	0.92
8	Cl	0.71	0.37	0.23	6.03
9	Br	0.86	0.39	0.23	8.88
10	I	1.12	0.35	0.18	13.94
11	OCH_3	-0.02	0.12	-0.27	7.87
12	NH_2	-1.23	-0.16	-0.66	5.42
13	OH	-0.67	0.12	-0.37	2.85
14	COOH	-0.32	0.37	0.45	6.93
15	$COOCH_3$	-0.01	0.37	0.45	12.87
16	CF_3	0.88	0.43	0.54	5.02
17	NO_2	-0.28	0.71	0.78	7.36
18	CHO	-0.65	0.35	0.42	6.88
19	C_6H_5	1.96	0.06	-0.01	25.36
20	CN	-0.57	0.56	0.66	6.33
21	N_3	0.46	0.27	0.15	10.20
22	NHOH	-1.34	-0.04	-0.34	7.22
23	$CH=CH_2$	0.82	0.05	-0.02	10.99
24	$COCH_3$	-0.55	0.38	0.50	11.18

Contd...

No.	Substituent	π	σ_m	σ_p	MR
25	$COOC_2H_5$	0.51	0.37	0.45	17.47
26	$COOC_3H_7$	1.07	0.37	0.45	22.17
27	CH_2OH	-1.03	0.00	0.00	7.19
28	$CHOHCH_3$	-0.86	0.00	-0.07	11.82
29	CH_2OCH_3	-0.78	0.02	0.03	12.07
30	SCH_3	0.61	0.15	0.00	13.82
31	NHCHO	-0.98	0.19	0.00	10.31
32	$OCOCH_3$	-0.64	0.39	0.31	12.47
33	$OCH(CH_3)_2$	0.85	0.10	-0.45	17.06
34	OC_3H_7	1.05	0.10	-0.25	17.06
35	$N(CH_3)_2$	0.18	-0.15	-0.83	15.55

* Taken from ref. 1

In equation (2.4), K_H is the ionisation constant of benzoic acid in water at 25°C and K_X is the ionisation constant for the meta or para derivative under the same experimental conditions. Positive values of σ represent the electron-withdrawing and the negative ones electron-donating character of substituents in the aromatic ring. For certain substituents, σ values are given in Table (2.3).

2.4 Molar Refractivity (MR)

In various organic reactions, dispersion forces play an important role and these could be modeled by the molar refractivity (MR) of substituents. Experimentally, MR is usually obtained via the Lorentz-Lorenz equation.

$$MR = \frac{n^2 - 1}{n^2 + 2} \cdot \frac{MW}{d} \qquad (2.5)$$

where n is the index of refraction, d is the density, and MW is the molecular weight of a compound. Since MR is an additive constituent property of molecules, fragment values have been calculated for many common groups of atoms. It has generally been assumed that a positive coefficient with an MR term in a correlation equation suggests a binding action via dispersion forces. Such binding could produce a concomitant conformational change in a macromolecular binding site. If the conformational change favoured the process under study, one would certainly expect a positive coefficient with the MR term, however, if the conformational change were detrimental, a negative coefficient could result for the MR term. Negative coefficient with MR have also been assumed to reflect

steric hindrance of one kind or another. Some MR values used are tabulated in Table (2.3).

2.5 van der Waals Volume (V_w)

The van der Waals volume (V_w) has been found to be one of the most fundamental characteristics of the drug structure controlling biological activity. This determines the molecular size and shape of the compounds which are very important in the aspect of drug-receptor interactions.

To calculate V_w of molecules, spherical shapes are assumed for all atoms according to Bondi[5] because of the absence of generally accepted pear shapes.

The values of the van der Waals radii used and calculated volume of atoms are listed in Table (2.4). Since van der Walls radii are greater than covalent radii, a correction for sphere overlapping due to covalent bonding between atoms is needed for the calculation of V_w of polyatomic molecules. The covalent bond lengths and correction values are tabulated in Table (2.5). A correction for branching in the molecule is also included in the V_w calculation. Such correction is also mentioned in the Table (2.5). All these values have been taken from the literature.[6]

2.6 Molecular Connectivity Index (χ)

Kier and Hall[7] introduced this additive topological parameter to drug design. Here the molecular connectivity index, χ, signifies the degree of branching or connectivity in a molecule.

Table (2.4): van der Waals radius and volume of atoms*

Atom		Radius (Å)	Sphere Volume (10^2Å^3)
C		1.7	0.206
H		1.1	0.056
N		1.5	0.141
O		1.4	0.115
S		1.8	0.244
F		1.4	0.115
Cl	Aliphatic	1.7	0.206
	Aromatic	1.8	0.244
Br	Aliphatic	1.8	0.244
	Aromatic	1.9	0.287
I	Aliphatic	2.0	0.335
	aromatic	2.1	0.388
B		2.1	0.388
He		1.2	0.072
Ne		1.6	0.171
Ar		1.9	0.287
Kr		2.0	0.335
Xe		2.2	0.446

*Taken from ref. 6

Table (2.5): Correction values of van der Waals volume, for sphere overlapping due to covalent bonding and branching*

Bond	Bond length (Å)	Correction value (10^2 Å^3)
C-C	1.5	-0.078
C-H	1.1	-0.043
C-N	1.4	-0.060
C-O	1.4	-0.056
C-S	1.8	-0.066
C-F	1.4	-0.056
C-Cl (aliphatic)	1.8	-0.058
C-Cl (aromatic)	1.8	-0.066
C-Br (aliphatic)	1.9	-0.060
C-Br (aromatic)	1.9	-0.068
C-I (aliphatic)	2.1	-0.063
C-I (aromatic)	2.1	-0.072
C-B	1.6	-0.113
H-H	0.7	-0.030
N-H	1.0	-0.038
N-N	1.4	-0.050
N-O	1.4	-0.042
N-S	1.6	-0.061
O-H	1.0	-0.034
O-B	1.5	-0.079
S-H	1.3	-0.040
S-S	2.0	-0.062

Contd...

Bond	Bond length (Å)	Correction value (10^2 Å^3)
S-F	1.6	-0.052
C=C	1.3	-0.094
C=N	1.3	-0.072
C=O	1.2	-0.068
C=S	1.6	-0.081
N=N	1.2	-0.061
N=O	1.2	-0.053
S=O	1.5	-0.057
C≡C	1.2	-0.101
C≡N	1.2	-0.079
C≡C (aromatic)	1.4	-0.086
Branching for saturated bond except bonding with H		-0.050

* Taken from ref. 6

Different versions of χ are calculated from the hydrogen-suppressed graph of the molecule. For this purpose the hydrogen-suppressed graph will be decomposed, depending on the χ considered, into uniform parts called as subgraph(s). Here two types of connectivity indices, simple molecular connectivity index ($^{m}\chi$) and valance molecular connectivity index ($^{m}\chi^{v}$) are discussed. The superscript 'm' is known as order of the connectivity index and is numerically equal to the number of non hydrogenic sigma bonds present in the subgraph of the particular χ.

A simple version of simple molecular connectivity index is first-order simple molecular connectivity index, $^{1}\chi$, and it is computed by

$$^{1}\chi = \Sigma C_{ij} = \Sigma (\delta_i \ \delta_j)^{-1/2} \qquad (2.6)$$

where the summation extends to all connection or edges (C_{ij}) of the hydrogen-suppressed graph and δ_i and δ_j are integers assigned to each atom indicating the number of atoms adjacent or connected to atoms i and j which are formally bonded. Here, in deriving this index, only the number of non-hydrogenic adjacent atoms are considered but not the nature of the atoms and the unsaturation in the molecule.

The valence molecular connectivity index, in contrast to the simple molecular connectivity index, takes into account the nature of the atoms as well as the unsaturation present in the molecules. Here the connectivity term, δ^{v} is defined as:

$$\delta^{v} = Z^{v}_{i} - N_H \qquad (2.7)$$

in which Z^v_i is the number of valence electrons present in atom i and N_H is the number of hydrogens attached to it. A simple version of valence molecular connectivity index is first-order valence molecular connectivity index, $^1\chi^v$, and is formulated as:

$$^1\chi^v = \Sigma C_{ij} = \Sigma (\delta^v_i \ \delta^v_j)^{-1/2} \qquad (2.8)$$

The application of equation (2.7) for atoms beyond the second row in the periodic table leads to the same δ^v value for each family member, for the example, seven for each halogen and six for each chalcogen. Consideration of valence electrons (Z^v_i) of atom i together with its atomic number (Z_i) and number of hydrogen atoms (h_i) attached to that atom will give appropriate δ^v value for atoms beyond second row in the periodic table.[8] The mathematical expression for this is:

$$\delta^v = (Z^v_i - h_i) / (Z^v_i - h_i - 1) \qquad (2.9)$$

According to this equation $\delta_{Cl} = 0.78$ and $\delta_{Br} = 0.26$. The δ^v value for some heteroatoms including halogens are listed in Table (2.6).

Only the above discussed connectivity indices are used in our studies. Higher order connectivity indices are discussed by Kier and Hall in their monograph.[7]

2.7 Steric Parameter (E_s)

Steric effect of substituents in organic reactions are very important. The first generally successful numerical definition of steric effects in organic reactions was

Table (2.6): Valence delta (δ^v) values for heteroatoms*

Group	δ^v	Group	δ^v
NH_2	3	OH	5
NH	4	O	6
N	5	C=O	6
C≡N	5	Furan O	6
C=NH	4	O=NO	6
Pyridine N	5	H_2O	4
Nitro N	6	H_3O^+	3
NH_3	2	F	7
NH_4^+	1	Cl	0.78[a]
$\rangle N\langle^+$	6	Br	0.26[a]
$= NH_2^+$	3	I	0.16[a]
		S	0.67[a]

*Taken from ref. 7, [a] Obtained from equation (2.9)

proposed by Taft.[9, 10] Following a suggestion of Ingold, Taft defined the steric constant E_s as:

$$E_s = \log (K_X/K_H) \qquad (2.10)$$

where K refers to the rate constant for the acid hydrolysis (denoted by A) of esters of type $X\text{-}CH_2COOR$. The size of X will affect attainment of the transition state, which is an essential step for acid hydrolysis by water.

2.8 Verloop's Steric Parameters (L and B)

Verloop's steric parameters[11] L and B referring to length and breadth of the substituents are calculated by a computer program called STERIMOL. There is only one length parameter L but there are five width parameters $B_1 - B_5$. All are calculated from standard bond angles, bond lengths, van der Waals radii, and user-determined reasonable confirmations. The width parameters are measured perpendicular to the bond axis and describe the positions, relative to the point of attachment and the bond axis, of five planes which closely surround the group. In contrast to E_s values which, because of the reaction on which they are based, cannot be determined for many substituents, the Verloop's parameters are available for any substituent. Table (2.7) lists the Verloop's parameters for some important substituents.

For QSAR studies in this thesis, standard values for different parameters for various substituents have been taken from literature.[1]

Table (2.7): Verloop's parameters for some important substituents*

No.	Substituent	L	B_1	B_2	B_3	B_4
1	H	2.06	1.00	1.00	1.00	1.00
2	F	2.65	1.35	1.35	1.35	1.35
3	Cl	3.52	1.80	1.80	1.80	1.80
4	Br	3.83	1.95	1.95	1.95	1.95
5	I	4.23	2.15	2.15	2.15	2.15
6	CH_3	3.00	1.52	1.90	1.90	2.04
7	C_2H_5	4.11	1.52	1.90	1.90	2.97
8	n-C_3H_7	5.05	1.52	1.90	1.90	3.49
9	i-C_3H_7	4.11	2.04	2.76	3.16	3.16
10	c-C_3H_7	4.14	1.98	2.24	2.29	2.88
11	$CH_2C_6H_5$	3.63	1.52	3.11	3.11	6.02
12	CF_3	3.30	1.98	2.44	2.44	2.61
13	COOH	3.91	1.60	1.60	2.36	2.66
14	$COOCH_3$	4.85	1.90	1.90	2.36	3.36
15	$CONH_2$	4.06	1.60	1.60	2.42	3.07
16	CN	4.23	1.60	1.60	1.60	1.60
17	C_6H_5	6.28	1.70	1.70	3.11	3.11
18	p-Cl C_6H_4	7.74	1.80	1.80	3.11	3.11
19	OH	2.74	1.35	1.35	1.35	1.93
20	OCH_3	3.98	1.35	1.90	1.90	2.87
21	O C_2H_5	4.92	1.35	1.90	1.90	3.36
22	$OCH_2C_6H_5$	8.20	1.35	3.03	3.11	3.11
23	NH_2	2.93	1.50	1.50	1.84	1.84
24	$NHCH_3$	3.53	1.50	1.90	1.90	3.08
25	NO_2	3.44	1.70	1.70	2.44	2.44

* Taken from ref. 11

REFERENCES

1. Hansch, C. and Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*, John Wiley, New York (1970)
2. Hansch, C., Deutsch, F.W. and Smith, R.N. *J.Am. Chem. Soc., 87, 2738 (1965)*
3. Yoshimoto, Y. and Hansch, *C. J. Med. Chem., 19, 71 (1976)*
4. Hammet, L.P. *Chem. Rev., 17, 125 (1935)*
5. Bondi, A. *J. Phys. Chem., 68, 441 (1964)*
6. Moriguchi, I., Kanada, Y. and Komatsu, K. *Chem. Pharm. Bull., 24, 1799 (1976)*
7. Kier, L.B. and Hall, L.H *Molecular connectivity in Chemistry and Drug Research*, Academic Press, New York (1976)
8. Kier, L.B. and. Hall L.H. *J. Pharm. Sci., 72, 1170 (1983)*
9. Taft, R.W. *J. Am Chem. Soc., 71, 3129 (1952)*
10. Taft, R.W. In, *Steric Effects in Organic Chemistry*, Ed., Newman, M.S., John Wiley, New York (1956), p. 556
11. Verloop, A., Hoogenstraaten, W. and Tipker, J. In: *Drug Design*, Ed. Ariens, E.J., Academic Press, New York (1976), Vol. 7, p.165

CHAPTER 3

QSAR STUDIES

RESULTS AND DISCUSSION

3 QSAR STUDIES: RESULTS AND DISCUSSION

The aspartyl protease (PR) of HIV-1 is a homodimeric enzyme that cleaves the polyprotein products of the *gag* and *pol* viral genes, yielding structural proteins and enzymes that are essential to the life cycle of the virus. Inhibition of this protease leads to the production of non-infectious viral particles[1,2] and thus the prevention of further propagation of the virus. Since abundant informations are available on this enzyme, it has become an attractive target for computer-aided drug design strategies[3,4] and consequently a prime focus for the development of anti-HIV chemotherapy.[5]

A number of peptide-derived compounds have been identified as HIV-PR inhibitors[6] but their clinical development has been hindered by their poor pharmacokinetics including low oral bioavailability and rapid excretion,[7] and complex and expensive synthesis.[8] Therefore, attention has been focussed to investigate nonpeptidic or modified peptide inhibitors of low molecular weight. Since HIV mutants at the level of protease to render resistance to antiviral drugs, targeting the enzyme,[9] there is a marked interest in developing structurally diverse and/or small molecules. This can be fruitfully facilitated by quantitative structure-activity relationship (QSAR) studies.

Several nonpeptide inhibitors were synthesised in the recent years. A quantitative analysis of the biological activity and the physicochemical properties of these compounds will precisely determine the extent of the role played by different physicochemical and structural properties of the molecules that can lead them to strongly bind with the

receptor. Further, the correlation equations (Eqs.) may be exploited to design a better compound.

In this chapter, we present the detailed QSAR studies on the following categories of HIV-Pr inhibitors.

Cyclic urea derivatives[11,12]

Cyclic cyanoguanidines[13]

Cycloalkylpyranones[16-19]

Cyclic sulfolanes[24]

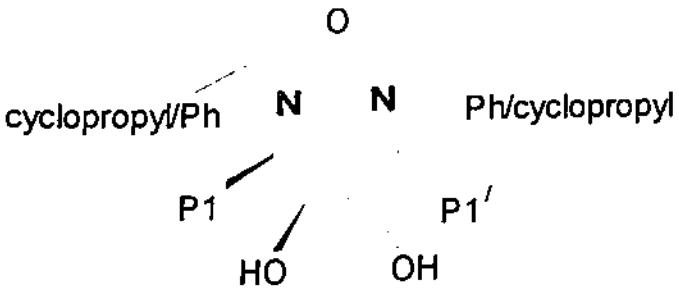
Arylthiomethanes[25]

Ritonavir analogues[32]

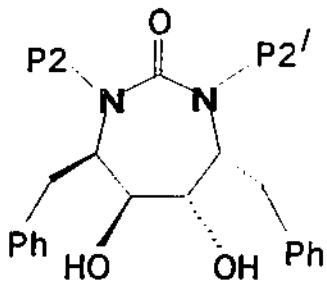
All the compounds have been taken from the various sources as indicated by the references. The results on them are discussed one by one.

3.1 Cyclic Urea Derivatives

In the development of non-peptidic inhibitors some authors have recently paid attention towards seven-membered cyclic urea scaffold (1),[10-12] which creates an effective hydrogen bond network between the aspartic residues and flap region of the enzyme with out intervention of a water molecule commonly found in linear inhibitors. Two large series of cyclic urea derivatives (1 and 2) studied for their HIV-1 protease (HIV-1-PR) inhibition activity and antiviral potency by Nugeal et al.[11] and by Lam et al.[12] were



1



2

compiled and listed in Tables (3.1) and (3.2). The enzyme inhibition activity was reported in terms of K_i , the inhibition constant, and the antiviral potency in terms of IC_{90}, the molar concentration of the compound required to reduce the concentration of HIV viral RNA by 90% from the level measured in an infected culture. A Hansch analysis was performed on these activities of the compounds to find out the correlation between them and the physicochemical properties of the substituents. A multiple regression using least square method was adopted to derive QSAR equations.

For the series of compounds belonging to 1 (Table (3.1)) the first 26 compounds, the P2/P2′ substituent is the benzyl group and, for the remaining compounds (27-38), it is cyclopropylmethyl group. Thus, the series can be divided into two groups: Group A, containing compounds 1-26, and Group B, containing compounds 27-38. When a regression analysis was performed initially for Group A, a significant correlation was obtained, (Eq. (3.1)), between the enzyme inhibition activity and the van der Waals volume of the P1/P1′-substituent and two indicator parameters I_a and I_o. The indicator

Table (3.1): Cyclic urea derivatives (1) and their HIV-I protease inhibition activity and antiviral potency studied by Nugiel et al.[11] along with physico-chemical parameters used.

No.	P1/P1′	V_w	I_p	I_o	log(1/K_i) Obsd.	log(1/K_i) Calcd. Eq. (3.2)	log(1/IC_{95}) Obsd.
	P2,P2′ – dibenzyl derivatives						
1	benzyl	0.94	1	0	8.41	8.34	6.10
2	methyl	0.25	0	0	5.30	5.03	-
3	4-isopropylbenzyl	1.40	1	0	8.96	8.19	-
4	4-(methylthio)benzyl	1.28	1	0	8.47	8.33	5.89
5	isobutyl	0.71	0	0	5.77	6.41	-
6	2-(methylthio)ethyl	0.73	0	0	5.96	6.45	-
7[a]	3-indolylmethyl	1.12	1	0	6.24	8.37	-
8	cyclohexylmethyl	1.07	0	0	7.55	6.78	-
9[a]	phenethyl	1.08	1	0	6.50	8.37	-
10	2-napthylmethyl	1.36	1	0	8.01	8.25	5.48
11	3-furanylmethyl	0.76	1	0	8.08	8.12	5.11
12	3-(methylthio)benzyl	1.28	1	0	8.00	8.33	5.38
13	4-(methylthiosulfonyl)benzyl	1.40	1	0	8.60	8.20	6.33
14	2-methoxybenzyl	1.17	1	1	7.22	7.29	5.06
15	2-hydroxybenzyl	1.01	1	1	7.46	7.28	5.19
16	3-methoxybenzyl	1.17	1	0	8.33	8.39	6.46
17	4-methoxybenzyl	1.17	1	0	8.07	8.39	6.22
18	4-hydroxybenzyl	1.01	1	0	8.96	8.38	6.73
19	3-aminobenzyl	1.04	1	0	8.55	8.39	5.89
20	3-(dimethylamino)benzyl	1.36	1	0	8.37	8.24	5.92
21	4-aminobenzyl	1.04	1	0	8.07	8.39	5.85
22	4-aminobenzyl.2HCl	1.04	1	0	8.15	8.39	5.89
23	4-(dimethylamino)benzyl	1.36	1	0	7.34	8.24	5.57
24	4-pyridylmethyl	0.90	1	0	7.65	8.30	5.24
25	3-(2,5-dimethylpyrolyl)benzyl	1.80	1	0	6.80	7.26	-
26	3,4-(methylenedioxy)benzyl	1.15	1	0	8.89	8.40	6.30
	P2,P2′-bis(cyclopropylmethyl)derivatives						
27	benzyl	0.94	1	0	8.72	8.34	5.74
28	isobutyl	0.71	0	0	7.07	6.41	-
29	isopropyl	0.55	0	0	6.60	6.06	-
30	2-(methylthio)ethyl	0.73	0	0	5.60	6.45	-
31	4-fluorobenzyl	0.99	1	0	8.24	8.37	5.50
32	2-methoxybenzyl	1.17	1	1	7.19	7.29	-
33	3-methoxybenzyl	1.17	1	0	9.06	8.39	6.19
34	3-hydroxybenzyl	1.01	1	0	7.89	8.38	5.59
35	4-methoxybenzyl	1.17	1	0	8.54	8.39	6.50
36	2-napthylmethyl	1.36	1	0	8.37	8.25	5.47
37	3,5-dimethoxybenzyl	1.41	1	0	8.57	8.18	6.42
38	2-thienylmethyl	0.87	1	0	8.04	8.27	5.11

[a] Not included in the derivation of Eq. (3.2).

$$\log(1/K_i) = 5.741\ (\pm 3.221)V_w - 2.592\ (\pm 1.414)V_w^2 + 1.542\ (\pm 0.708)I_a$$
$$- 1.010\ (\pm 0.774)\ I_o + 3.646$$
$$n = 24, \quad r = 0.89, \quad s = 0.48, \quad F_{4,19} = 18.58\ (4.50) \qquad (3.1)$$

parameter I_a has been used, with a value of unity, for a substituent containing aromatic moiety, and I_o has been used, with a value of unity, for an ortho-substituent in the latter. In the equation, n is the number of data points, r is the correlation coefficient, s is the standard deviation, F is the F-ratio between the variances of calculated and observed activities. The value of F given in the parenthesis is of 99% level. All these parameters show that the correlation is statistically quite significant, suggesting that the molecular size of the P1/P1′-substituent would be an important factor in the protease inhibition and that an aromatic substituent will have an added advantage but an ortho-substituent in it will have a detrimental effect.

In the derivation of Eq. (3.1), however, compounds 7 and 9 were not included, as they exhibited an aberrant behaviour. The activity values of these compounds as predicted by Eq. (3.1) were much higher than their corresponding observed values (8.37 vs 6.24 and 8.37 vs 6.50). The reason for the low observed activities of these two compounds may be due to some steric effects produced by the substituents or the misorientation of the substituents towards the active site of the receptor.

Equation (3.1) beautifully absorbs all the 12 compounds of Group B, maintaining essentially all its statistical and analytical characteristics (Eq. (3.2)).

$$\log(1/K_i) = 5.567\ (\pm 2.984)V_w - 2.491\ (\pm 1.308)V_w^2 + 1.527\ (\pm 0.606)I_a$$
$$- 1.101\ (\pm 0.602)I_o + 3.767$$
$$n = 36, \quad r = 0.89, \quad s = 0.48, \quad F_{4,31} = 28.79\ (4.00) \qquad (3.2)$$

$$\log(1/K_i) = 5.465\ (\pm 3.024)V_w - 2.431\ (\pm 1.331)V_w^2 + 1.538\ (\pm 0.612)I_a$$
$$- 1.100\ (\pm 0.608)I_o - 0.127\ (\pm 0.357)\ D + 3.878$$
$$n = 36, \quad r = 0.89, \quad s = 0.48, \quad F_{5,30} = 22.79\ (3.70) \qquad (3.3)$$

Equation (3.3) was derived to see the effect of an alteration in the P2/P2′ substituent, using a dummy variable D with a value of 1 for Group A, where this substituent is a benzyl group, and with a value of zero for Group B, where this substituent is a cyclopropylmethyl group. As is obvious from this equation, the D parameter was totally insignificant, suggesting that a variation in this substituent was of no consequence.

We were, however, unable to correlate so significantly the antiviral activity (IC_{90}) of these compounds with any physicochemical parameters. However, for the series of compounds belonging to 2 (Table (3.2)), where the variation is in the P2/P2′-substituent, the IC_{90} data were found to be significantly correlated with the hydrophobic property of the substituents and some indicator variables (Eq. (3.4)). The indicator variable $I_H = 1$

$$\log(1/IC_{90}) = 2.531\ (\pm 0.899)\ \pi - 0.534\ (\pm 0.212)\ \pi^2 + 2.010\ (\pm 0.422)I_H$$
$$- 0.749\ (\pm 0.493)I_o + 2.774$$
$$n = 40, \quad r = 0.90, \quad s = 0.40, \quad F_{4,35} = 35.66\ (3.91) \qquad (3.4)$$

Table (3.2): Cyclic urea derivatives (2)and their HIV-1 protease inhibition activity and antiviral potency studied by Lam et al.[12] along with physicochemical parameters used.

No.	P2/P2′	π	I_H	I_o	I_e	log(1/K_i) Obsd.	log(1/K_i) Calcd. Eq.(3.6)	log(1/K_i) Calcd. Eq.(3.7)	log(1/IC_{90}) Obsd.	log(1/IC_{90}) Calcd. Eq.(3.5)
1[a]	methyl	0.65	0	0	0	5.24	6.63	-	-	-
2	n-ethyl	1.18	0	0	0	7.00	7.61	7.81	-	-
3[b]	n-propyl	1.71	0	0	0	8.10	8.24	8.34	4.27	-
4	n-butyl	2.24	0	0	0	8.85	8.53	8.53	6.17	5.85
5	n-pentyl	2.77	0	0	0	8.80	8.48	8.38	5.82	5.73
6	n-hexyl	3.30	0	0	0	8.34	8.08	7.89	-	-
7	n-heptyl	3.83	0	0	0	6.59	7.34	7.06	-	-
8	$CH_2CH_2OCH_3$	0.88	0	0	1	6.10	5.82	5.67	-	-
9	$CH_2CH_2OCH_2CH_3$	1.41	0	0	1	5.96	6.65	6.40	-	-
10	$CH_2CH_2OCH_2CH_2OCH_3$	0.95	0	0	1	5.11	5.94	5.78	-	-
11[a]	i-butyl	2.11	0	0	0	7.31	8.49	-	-	-
12	i-pentyl	2.64	0	0	0	7.92	8.52	8.46	5.50	5.79
13	i-hexyl	3.17	0	0	0	8.16	8.21	8.04	5.09	5.42
14	i-heptyl	3.70	0	0	0	7.52	7.55	7.30	-	-
15	i-octyl	4.23	0	0	0	6.96	6.55	6.22	-	-
16	neohexyl	3.04	0	0	0	7.44	8.32	8.17	-	-
17	allyl	1.23	0	0	0	8.28	7.68	7.86	5.33	5.16
18	2-methylpropen-3-yl	1.63	0	0	0	8.14	8.16	8.27	5.12	5.58
19	isoprenyl	2.54	0	0	0	8.75	8.54	8.48	6.06	5.82
20	$CH_2CH_2OCH=CH_2$	1.40	0	0	1	7.22	6.64	6.39	-	-
21	3-propynyl	0.75	0	0	0	7.66	6.85	7.13	4.38	4.42
22	cyclopropylmethyl	1.63	0	0	0	8.68	8.16	8.27	5.75	5.58
23	cyclobutylmethyl	2.18	0	0	0	8.89	8.52	8.52	6.00	5.85
24	cyclopentylmethyl	3.02	0	0	0	8.38	8.33	8.19	5.77	5.56
25	cyclohexylmethyl	3.31	0	0	0	7.43	8.07	7.88	-	-
26	N-morpholino-2-ethyl	0.87	0	0	1	5.40	5.80	5.65	-	-
27	benzyl	2.27	0	0	0	8.52	8.54	8.53	6.08	5.86
28	2-picolyl	0.78	0	0	0	6.84	6.90	7.17	4.31	4.47
29[a]	3-picolyl	0.78	0	0	0	8.01	6.90	-	5.06	4.47
30	4-picolyl	0.78	0	0	0	7.05	6.90	7.17	4.01	4.47
31	α-napthylmethyl	3.45	0	0	0	7.07	7.90	7.69	4.80	5.08
32[a]	β-napthylmethyl	3.45	0	0	0	9.51	7.90	-	5.41	5.08
33	o-flurobenzyl	2.42	0	1	0	7.47	7.11	7.26	5.26	5.04
34	m-fluorobenzyl	2.42	0	0	0	8.52	8.55	8.51	6.15	5.85
35	p-fluorobenzyl	2.42	0	0	0	8.85	8.55	8.51	6.22	5.85
36	o-chlorobenzyl	2.99	0	1	0	6.62	6.91	6.97	4.95	4.77
37	m-chlorobenzyl	2.99	0	0	0	9.05	8.35	8.22	5.89	5.58
38	p-chlorobenzyl	2.99	0	0	0	8.28	8.35	8.22	5.35	5.58

Contd.....

No.	P2/P2'	π	I_H	I_n	I_e	log(1/K_i) Obsd.	log(1/K_i) Calcd. Eq.(3.6)	log(1/IC_{90}) Calcd. Eq.(3.7)	log(1/IC_{90}) Obsd.	log(1/IC_{90}) Calcd. Eq.(3.5)
39	m-bromobenzyl	3.14	0	0	0	8.85	8.23	8.07	5.92	5.45
40	p-bromobenzyl	3.14	0	0	0	7.57	8.23	8.07	5.09	5.45
41	m-methylbenzyl	2.80	0	0	0	8.16	8.46	8.36	5.62	5.71
42	p-methylbenzyl	2.80	0	0	0	8.25	8.46	8.36	5.37	5.71
43	m-(trifluoromethyl)benzyl	3.16	0	0	0	7.66	8.22	8.05	5.11	5.51
44	p-(trifluoromethyl)benzyl	3.16	0	0	0	7.29	8.22	8.05	5.14	5.51
45	o-methoxybenzyl	2.19	0	1	1	5.73	5.80	5.59	4.64	5.04
46[a]	m-methoxybenzyl	2.19	0	0	1	8.80	7.24	-	5.89	5.85
47	p-methoxybenzyl	2.19	0	0	1	6.81	7.24	6.84	5.12	5.85
48	m-nitrobenzyl	2.02	0	0	0	8.55	8.45	8.49	6.01	5.81
49[a]	m-iodobenzyl	3.40	0	0	0	9.38	7.96	-	5.52	5.15
50	p-(hydroxymethyl)benzyl	1.24	1	0	0	9.47	9.62	9.64	7.22	7.25
51	m-(hydroxymethyl)benzyl	1.24	1	0	0	9.85	9.62	9.64	7.42	7.25
52	p-hydroxybenzyl	1.61	1	0	0	9.92	10.08	10.02	7.50	7.48
53	m-hydroxybenzyl	1.61	1	0	0	9.92	10.08	10.02	7.27	7.48
54	m-(aminobenzyl).2CH_3SO_2H	1.05	1	0	0	9.55	9.33	9.38	6.89	6.83

[a]Not used in the derivation of Eq. (3.7). [b]Not used in the derivation of Eq. (3.5).

is meant for the last 5 compounds (50-54) which differ from the others in that they have an OH or NH_2 group in their P2/P2′-substituent. The variable I_o = 1 is meant for a benzyl substituent bearing a group at the ortho-position.

In the derivation of Eq. (3.4) compound 3 was, however, found to be a misfit. The exclusion of this compound led relatively to a much better correlation (Eq. (3.5)).

$$\log(1/IC_{90}) = 2.732\ (\pm 0.759)\pi - 0.592\ (\pm 0.180)\pi^2 + 1.914\ (\pm 0.357)I_H - 0.810\ (\pm 0.414)I_o + 2.705$$

$$n = 39,\quad r = 0.93,\quad s = 0.33,\quad F_{4,34} = 50.80\ (3.93) \qquad (3.5)$$

So far as the enzyme inhibition activity was concerned, the most relevant Eq. obtained for that was:

$$\log(1/K_i) = 2.966\ (\pm 1.002)\ \pi - 0.612\ (\pm 0.224)\ \pi^2 + 1.929\ (\pm 0.690)I_H - 1.440\ (\pm 0.832)I_o - 1.279\ (\pm 0.572)I_e + 4.963,$$

$$n = 54,\quad r = 0.85,\quad s = 0.67,\quad F_{5,48} = 23.92\ (3.42) \qquad (3.6)$$

In this equation, I_e is an additional parameter, which has been used for ethereal substituents including those having a methoxy group. The negative coefficient of this parameter suggests that such substituents are not conductive to good activity.

Although Eq. (3.6) predicts for certain compounds very high or very low activity as compared correspondingly to their observed activity, giving a difference of 1 log unit

(see Table (3.2)) high activity for compounds I and II and low activity for compounds 29, 32, 46 and 49), it expresses statistically a quite significant correlation, accounting overall for 72% of the variance in the activity ($r^2 = 0.72$). If these six compounds, which are misfit in the correlation are excluded, a highly improved correlation as expressed by Eq. (3.7) is obtained.

$$\log(1/K_i) = 2.751\ (\pm 0.771)\pi - 0.606\ (\pm 0.166)\pi^2 + 1.766\ (\pm 0.486)I_H$$
$$- 1.251\ (\pm 0.603)I_o - 1.680\ (\pm 0.436)I_c + 5.400$$
$$n = 48, \quad r = 0.93, \quad s = 0.46, \quad F_{5,42} = 51.78\ (3.49) \qquad (3.7)$$

Thus Eqs. (3.4) to (3.7) suggest that both the antiviral and the enzyme inhibition activities of the compounds are governed by the hydrophobic property of the P2/P2′-substituent with an almost equal optimum value of π, 2.31 for the former (Eq. (3.5)) and 2.27 for the latter (Eq. (3.7)), and that in both the cases, an OH- or NH_2- containing substituent would be of more advantage. This leads us to assume that P2/P2′-substituents may have predominantly a hydrophobic interaction with the receptor and that the presence of an OH or NH_2 group in them may be involved in effective hydrogen bonding with the receptor.

In both the cases, it is however also indicated that an ortho-substituted benzyl group will produce an adverse effect. This adverse effect can be attributed to some steric role of the ortho substituent. In the case of the enzyme inhibition, an ethereal substituent is also found to produce a negative effect, which can be attributed to the repulsive effects of lone pairs of electrons at the oxygen.

The similarity of the effects in this series on the antiviral and enzyme inhibition activities is due to the fact that both the activities are mutually well correlated (n = 40, r = 0.82).

The main interaction of cyclic urea inhibitors with HIV-PR, however, involves multiple hydrogen bonds. The high resolution structural studies on the complexes of HIV-1-PR with peptidomimetic inhibitors have revealed the presence of a structural water molecule which is hydrogen bonded to both the mobile flaps of enzyme and the two carbonyls flanking the transition-state mimic of the inhibitors.[12] Cyclic ureas incorporate this structural water and preorganize the side chain residues into optimum binding conformations. They undergo reasonably symmetrical binding with the enzyme. The urea oxygen accepts two hydrogen bonds from the backbone of NH of Ile50/50′ and the diols form multiple hydrogen bonds with catalytic Asp25/25′. The OH or NH_2 present in the P2/P2′-substituent is supposed to be involved in hydrogen bonding with the backbone NH of Asp29/29′ and Asp30/30′. A Schematic representation of the bindings is given in Figure (3.1), where P2/P2′-substituents project symmetrically into S2/S2′ pockets of the enzyme for hydrophobic interaction, and the P1/P1′-substituents into the S1/S1′ pockets, most likely for dispersion interaction.

3.2 Cyclic Cyanoguanidines

Jadhav et al.[13] synthesised a series of cyclic cyanoguanidines (3) as structurally diverse class of seven membered cyclic nonpeptide inhibitors and studied their enzyme inhibition activity. Tables (3.3) and (3.4) give series of analogues belonging to cyclic

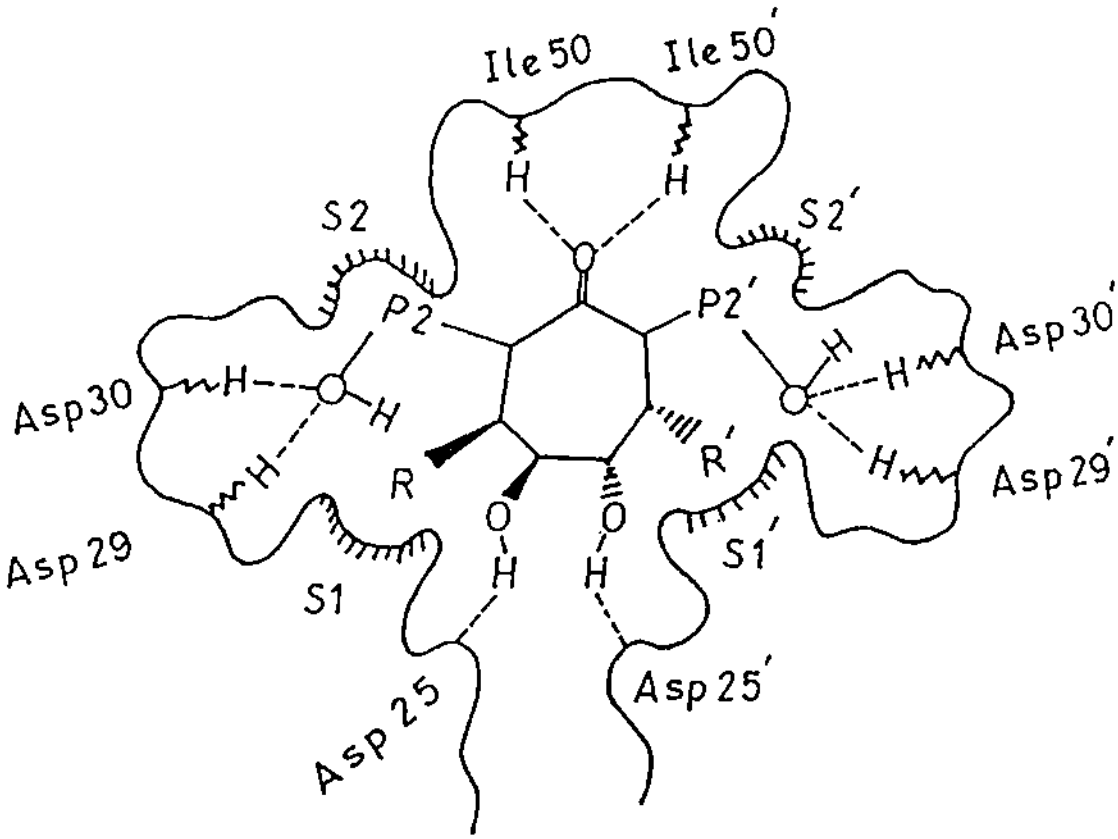
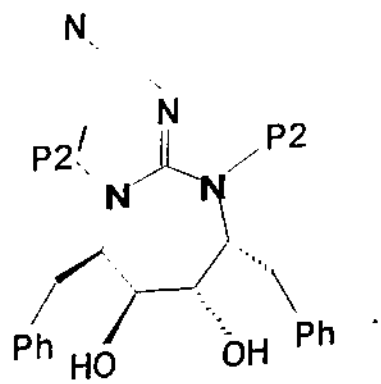


Figure (3.1): Schematic representation of the interaction of cyclic ureas with HIV-1 proteases.

cyanoguanidines and cyclic urea derivatives with various P2/P2′-substituents, respectively. The activity parameter IC_{90} is a measure of antiviral potency and refers to the molar concentration of the compound required to reduce the concentration of HIV viral RNA by 90% from the level measured in an infected culture. The inhibition of HIV-1-PR was measured by assaying the cleavage of fluorescent peptide substrate using high performance liquid chromatography.



3

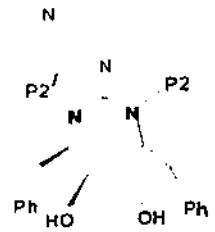
The antiviral potency of the compounds Table (3.3) (cyclic cyanoguanidines) was found to have a very good parabolic correlation with hydrophobic constant π of P2/P2′ substituents of the compounds and with two indicator parameters I_H and I_a (Eq. (3.8)).

$$\log(1/IC_{90}) = 3.219(\pm1.411)\pi - 0.604(\pm0.280)\pi^2 + 1.954(\pm0.600)I_H - 0.565(\pm0.438)I_a + 1.670 \quad (3.8)$$

$$n = 21, \quad r = 0.88, \quad s = 0.35, \quad F_{4,16} = 13.68\,(4.77), \quad \pi_o = 2.66$$

The indicator parameter I_H has been used with a value of unity for a P2/P2′ substituent containing OH/NH_2 group, and I_a has been used with a value of unity for an aromatic P2/P2′-substituent. The π_o refers to the optimum value of π.

Table (3.3): Cyclic cyanoguanidines (3) and their antiviral potency and HIV-1 protease inhibition activity studied by Jadhav et al.[13] along with physicochemical parameters used.



No.	P2/P2'	π	I_H	I_a	I_c	log (1/IC$_{90}$) Obsd.	log (1/IC$_{90}$) Calcd. Eq (3.8)	log (1/K$_i$) Obsd.	log (1/K$_i$) Calcd. Eq (3.9)
1	H	0.00	1	0	0	---[a]	3.62	---[b]	7.05
2	Allyl	1.00	0	0	0	4.29	4.29	7.43	7.12
3	n-Propyl	1.71	0	0	0	5.13	5.41	7.85	7.94
4	n-Butyl	2.24	0	0	0	5.59	5.85	8.56	8.32
5	3-3-Dimethylallyl	2.15	0	0	0	5.33	5.80	7.52	8.27
6	3-Methylbutyl	2.64	0	0	0	5.96	5.96	8.42	8.47
7	Cyclopropylmethyl	1.63	0	0	0	5.30	5.31	7.66	7.87
8	Cyclobutylmethyl	2.18	0	0	0	6.08	5.82	8.70	8.28
9	Cyclopentylmethyl	2.74	0	0	0	6.46	5.96	8.82	8.48
10	Cyclohexylmethyl	3.31	0	0	0	5.96	5.71	8.24	8.46
11	Benzyl	2.27	0	1	0	5.42	5.30	7.70	7.51
12	3-Nitrobenzyl	2.02	0	1	0	4.75	5.14	7.05	7.36
13	4-Nitrobenzyl	2.02	0	1	0	4.71	5.14	7.17	7.36
14	3-Aminobenzyl	1.05	1	1	0	6.30	5.78	8.13	8.06
15	4-Aminobenzyl	1.05	1	1	0	5.64	5.78	7.60	8.06
16	3-Cyanobenzyl	1.71	0	1	0	5.51	4.85	7.59	7.12
17	4-Cyanobenzyl	1.71	0	1	0	5.11	4.85	6.89	7.12
18	3-Hydroxybenzyl	1.61	1	1	0	6.89	6.68	9.14	8.72
19	4-Hydroxybenzyl	1.61	1	1	0	6.60	6.68	8.58	8.72
20	3-(Benzyloxy)benzyl	3.96	0	1	1	---[a]	4.38	5.86	5.96
21	3-(Benzyloxy)benzyl	3.96	0	1	1	4.17	4.38	6.05	5.96
22	3-(Hydroxymethyl)benzyl	1.24	1	1	0	6.23	6.12	8.77	8.31
23	4-(Hydroxymethyl)benzyl	1.24	1	1	0	5.49	6.12	7.96	8.31
24	2-Napthylmethyl	3.45	0	1	0	---[a]	5.02	7.66	7.59

[a] Not used in the derivation of Eq. (3.8). [b] Not used in the derivation of Eq. (3.9)

A similar correlation with an additional parameter I_e was obtained for enzyme inhibition activity of these compounds (Eq. (3.9)).

$$\log(1/K_i) = 2.124(\pm1.609)\pi - 0.359(\pm0.348)\pi^2 + 1.694(\pm0.574)I_H - 0.824(\pm0.411)I_a - 1.367(\pm1.214)I_e + 5.357$$

$$n = 23, \quad r = 0.92, \quad s = 0.32, \quad F_{5,16} = 19.28\ (4.44), \quad \pi_o = 2.96 \qquad (3.9)$$

I_e is equal to 1 for a P2/P2' substituent containing an ethereal moiety (C-O-C) and zero for others. The negative coefficient of it suggests that such a substituent would not be favourable to the enzyme inhibition activity of these compounds. Similarly, the negative coefficient I_a in both Eqs. (3.8) and (3.9) indicates that the presence of an aromatic substituent would be detrimental to both antiviral and enzyme inhibition activities of these compounds. Since the correlation is parabolic with hydrophobic constant π, the receptor site may have the limited bulk tolerance, as unlike in *in vivo*, there is no membrane-like lipid-water barrier in the *in vitro* system to optimize the lipophilic effect. Thus, Eqs. (3.8) and (3.9) suggest that both the antiviral potency and the enzyme inhibition activity of cyclic cyanoguanidines are governed by the hydrophobic property of the P2/P2' substituents with an optimal value of π equal to 2.66 and 2.96, respectively, which are essentially the same. However, in both the cases the positive coefficient of I_H suggests that an –OH or $-NH_2$ containing substituent would be more beneficial than any other substituent. This is probably because OH or NH_2 moiety may form the hydrogen bond with the receptor.

In deriving Eq. (3.8) compounds 1, 20, and 24, and in deriving Eq. (3.9) compound 1 were not included because the corresponding activity data for them are not reported. Equation (3.8), however, predicts the antiviral potency of 1, 20 and 24 as 3.62, 4.38 and 5.02, respectively, and Eq. (3.9) predicts enzyme inhibition activity of 1 as 7.05.

In the case of cyclic urea derivatives (Table (3.4)), too, both antiviral potency and enzyme inhibition activity of the compounds were found to be significantly correlated with the hydrophobic property of the substituents and indicator variable I_H (Eqs. (3.10) and (3.11)).

$$\log(1/IC_{90}) = 1.587(\pm1.366)\pi - 0.340(\pm0.307)\pi^2 + 1.780(\pm0.428)I_H + 3.962$$

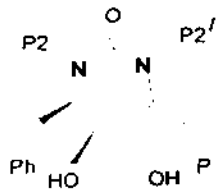
$$n = 19, \quad r = 0.94, \quad s = 0.27, \quad F_{3,15} = 37.53\ (5.42), \quad \pi_o = 2.33 \qquad (3.10)$$

$$\log(1/K_i) = 2.986(\pm0.870)\pi - 0.688(\pm0.175)\,\pi^2 + 1.637(\pm0.581)I_H + 5.241$$

$$n = 23, \quad r = 0.92, \quad s = 0.42, \quad F_{3,19} = 34.49\ (5.01), \quad \pi_o = 2.17 \qquad (3.11)$$

In deriving Eq. (3.10), compounds 1, 20 and 21 (Table (3.4)) were not included because their activity data are not reported. Equation (3.10), however, predicts their activities as 5.74, 4.92 and 4.92, respectively. Compounds 3 and 10 were also not included, as they exhibited an aberrant behaviour. The activities of these compounds as predicted by Eq. (3.10) were much higher than their corresponding observed values (5.68 versus 4.27 and 5.49 versus 4.02, respectively). The reason for the low observed activities of these compounds may be due to some steric effects produced by the substituents, or misorientation of the substituents towards the active site of the receptor.

Table (3.4) : Cyclic urea derivatives and their antiviral potency and HIV-1 protease inhibition activity studied by Jadhav et al.[13] and physicochemical parameters



No.	P2/P2'	π	I_H	I_a	I_c	log (1/IC$_{50}$) Obsd.	log (1/IC$_{50}$) Calcd. Eq.(3.10)	log (1/K$_i$) Obsd.	log (1/K$_i$) Calcd. Eq (3.11)
1	H	0.00	1	0	0	---[a]	5.74	6.57	6.88
2	Allyl	1.00	0	0	0	5.33	5.21	8.28	7.54
3	n-Propyl	1.71	0	0	0	4.27[a]	5.68	8.10	8.34
4	n-Butyl	2.24	0	0	0	6.16	5.81	8.85	8.48
5	3-3-Dimethylallyl	2.15	0	0	0	6.07	5.80	8.80	8.48
6	3-Methylbutyl	2.64	0	0	0	5.37	5.78	7.92	8.33
7	Cyclopropylmethyl	1.63	0	0	0	5.74	5.64	8.68	8.28
8	Cyclobutylmethyl	2.18	0	0	0	6.00	5.81	8.89	8.48
9	Cyclopentylmethyl	2.74	0	0	0	5.77	5.75	8.37	8.26
10	Cyclohexylmethyl	3.31	0	0	0	4.02[a]	5.49	7.43	7.59
11	Benzyl	2.27	0	1	0	6.08	5.81	8.52	8.49
12	3-Nitrobenzyl	2.02	0	1	0	6.01	5.78	8.55	8.47
13	4-Nitrobenzyl	2.02	0	1	0	5.08	5.78	7.49	8.47
14	3-Aminobenzyl	1.05	1	1	0	6.89	7.03	9.55	9.26
15	4-Aminobenzyl	1.05	1	1	0	6.96	7.03	8.96	9.26
16	3-Cyanobenzyl	1.71	0	1	0	5.66	5.68	8.52	8.34
17	4-Cyanobenzyl	1.71	0	1	0	5.24	5.68	7.28	8.34
18	3-Hydroxybenzyl	1.61	1	1	0	7.27	7.42	9.92	9.90
19	4-Hydroxybenzyl	1.61	1	1	0	7.49	7.42	9.92	9.90
20	3-(Benzyloxy)benzyl	3.96	0	1	1	---[a]	4.92	6.47	6.28
21	3-(Benzyloxy)benzyl	3.96	0	1	1	---[a]	4.92	6.27	6.28
22	3-(Hydroxymethyl)benzyl	1.24	1	1	0	7.42	7.19	9.85	9.52
23	4-(Hydroxymethyl)benzyl	1.24	1	1	0	7.42	7.19	9.85	9.52
24	2-Napthylmethyl	3.45	0	1	0	5.41	5.38	9.51[b]	7.35

[a] Not used in the derivation of Eq. (3.10). [b] Not used in the derivation of Eq (3.11)

Also in deriving Eq. (3.11) compound **24** has not been included, as it behaves as an outlier. Its predicted activity value (7.35) is quite low as compared to its observed activity (9.51). The high observed activity may be due to a very good π-stacking interaction of β-naphthyl moiety with S2/S2′ pockets of the enzyme.

Now it is to be noted that in both the series the P2/P2′ substituents are the same, but if we compare Eqs. (3.8) and (3.9) with Eqs. (3.10) and (3.11) correspondingly, we find that the aromatic substituents in cyclic cyanoguanidine series produce an adverse effect on both antiviral potency as well as enzyme inhibition activity but not in cyclic urea series. Further, an ethereal moiety in the substituent has an additional negative effect on the enzyme inhibition activity of cyclic cyanoguanidines.

Such an effect of an ethereal moiety on the enzyme inhibition activity of cyclic ureas was, however, also observed when a large series of these compounds, studied by the same group of authors[12] who reported the compounds of Table (3.2), were subjected to QSAR (Eq. (3.7)). In that series, a few of the compounds of Table (3.4) (**2-12** and **24**) were common, but the correlation obtained was:

$$\log(1/K_i) = 2.751(\pm 0.771)\pi - 0.606(\pm 0.166)\pi^2 + 1.766(\pm 0.486)I_H - 1.251(\pm 0.603)I_o - 1.680(\pm 0.436)I_e + 5.400$$

$$n = 48, \quad r = 0.93, \quad s = 0.46, \quad F_{5,42} = 51.78\ (3.49), \quad \pi_o = 2.27 \qquad (3.7)$$

The absence of parameter I_e in Eq. (3.11) may be due to the facts that there are only two compounds (**20** and **21**) in the series (Table (3.4)) for which $I_e = 1$ and both the

compounds have very low activity which can be attributed to a more dominant factor than an ethereal moiety. It may be that the bulky (benzyloxy)benzyl substituent in these compounds produces a dominant steric effect, leaving no scope for the ethereal moiety to play any role. The negative effect of the ethereal moiety (C-O-C fraction), whatsoever, can be assumed to be due to the repulsion between the electron pairs at its oxygen and an anionic site at the receptor. Compounds **20** and **21** of Table (3.4) were not present in the series (Table (3.2)) for which Eq. (3.7) was obtained.

The additional parameter I_o in Eq. (3.7) stands for an aromatic P2/P2′ substituent bearing an ortho group. No such substituents are present in the present series (Table (3.4)), hence I_o did not appear in Eq. (3.11). An ortho group may create some steric hindrance in the interaction of the molecule with the receptor, hence the negative effect.

Leaving I_e and I_o parameters apart, there seems an excellent similarity between Eqs. (3.11) and (3.7) with respect to the coefficients of variables as well as the values of statistical parameters r, s and π_o. This similarity between two equations, where Eq. (3.11) has been obtained for a smaller group and Eq. (3.7) relatively for a larger group, shows the validity of the correlation.

For the cyclic urea derivatives of Table (3.2) an ortho group in P2/P2′ aromatic substituent was found to affect the antiviral potency (Eq. (3.5)). Since there are no aromatic P2/P2′ substituents bearing an ortho group in the present case (Table (3.4)), the

parameter I_o did not appear in Eq. (3.10). Otherwise, Eqs. (3.10) and (3.5) are also quite similar and equally significant validating the correlation for antiviral potency, too.

$$\log(1/IC_{90}) = 2.732(\pm0.759)\pi - 0.592(\pm0.180)\pi^2 + 1.914(\pm0.357)I_H - 0.810(\pm0.414)I_o + 2.705$$

$$n = 39, \quad r = 0.93, \quad s = 0.33, \quad F_{4,34} = 50.80\ (3.93), \quad \pi_o = 2.31 \qquad (3.5)$$

To make a comparative study of the activities of cyclic cyanoguanidines and cyclic urea derivatives, compounds of Table (3.3) and (3.4) were merged and Eqs. (3.12) and (3.13) were obtained for the combine, where the parameter I_g is equal to one for cyclic cyanoguanidines and zero for cyclic urea derivatives. The negative coefficient of I_g in both Eqs. (3.12) and (3.13) indicates that cyclic cyanoguanidines would be less active than cyclic urea derivatives both as an antiviral and as a PR-inhibitor. In deriving Eqs. (3.12) and (3.13), all those compounds were excluded which were not taken in the derivation of Eqs. (3.8) and (3.10) (for antiviral potency) and Eqs. (3.9) and (3.11) (for enzyme inhibition activity).

$$\log(1/IC_{90}) = 2.416(\pm1.139)\pi - 0.474(\pm0.236)\pi^2 + 1.696(\pm0.411)I_H - 0.532(\pm0.282)I_g + 2.957$$

$$n = 40, \quad r = 0.861, \quad s = 0.41, \quad F_{4,35} = 25.12\ (3.95) \qquad (3.12)$$

$$\log(1/K_i) = 2.942(\pm0.878)\pi - 0.646(\pm0.171)\pi^2 + 1.392(\pm0.491)I_H - 0.625(\pm0.332)I_g + 5.208$$

$$n = 46, \quad r = 0.847, \quad s = 0.53, \quad F_{4,41} = 26.07\ (3.82) \qquad (3.13)$$

The comparison of X-ray crystal structures of cyclic cyanoguanidines with those of cyclic urea derivatives reveal that there is an almost perfect overlapping between the seven membered rings of cyclic cyanoguanidines and cyclic urea derivatives except some difference in the orientation of their P2/P2′ substituents.[13] This difference in the orientation of their P2/P2′ substituents may be due to the steric interaction in cyclic cyanoguanidines of the methylene moiety with the cyano group, which is rotated out of the plane formed by the –N–C≡N atoms. As in cyclic urea derivatives, the structural water molecule is displaced by the exo-cyclic nitrogen of the cyclic cyanoguanidines, but the cyano group present on the exocyclic nitrogen causes movement of one of the flaps of the enzyme from the active site which presumably results in the enlargement of the S_2 pocket of the enzyme.[13] Therefore, in cyclic cyanoguanidines, large alkyl groups in P2 can be more readily accommodated in the active site. Our theoretical study also confirms this, but more potent cyclic cyanoguanidines may be obtained with large alkyl group with an optimum π value of 2.66 for antiviral potency and 2.96 for enzyme inhibition activity. A hydrogen bond forming group, like OH or NH_2, will have an added advantage. A schematic representation of the bindings of substituents with the enzyme is shown in Figure (3.2), where a P2 substituent projects into the S_2 pocket of the enzyme for hydrophobic interaction and P1/P1′ substituents into S1/S1′ pockets, most probably for dispersion interaction as in cyclic urea derivatives. Since P2′ substituents in cyclic cyanoguanidines are not assumed to interact with the enzyme, it may be the reason for these compounds to be less active than cyclic urea derivatives where both P2/P2′ substituents are assumed to interact with the enzyme (Figure (3.1)).

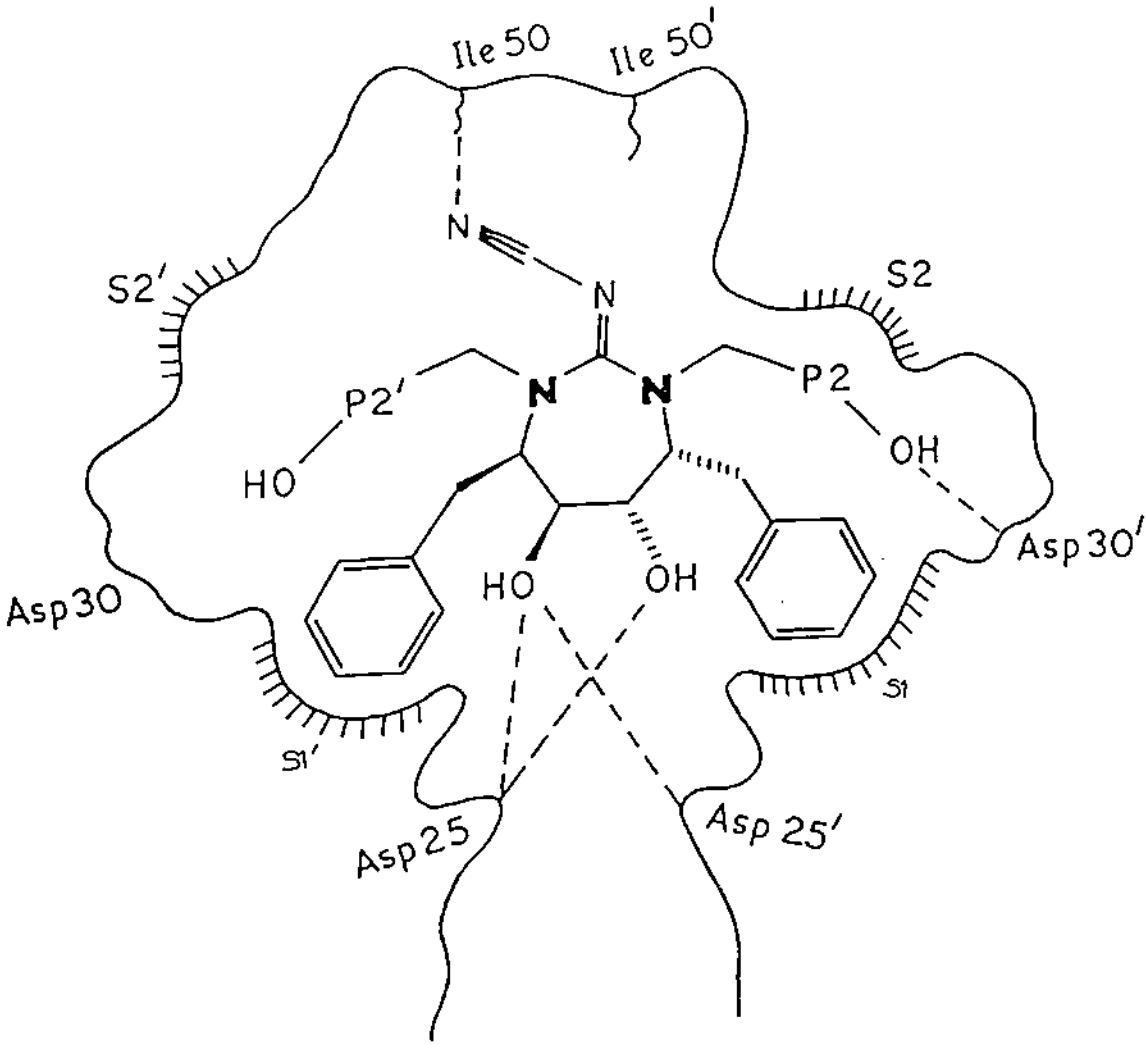


Figure (3.2): A proposed model for the interaction of cyclic cyanoguanidines with HIV-1 protease.

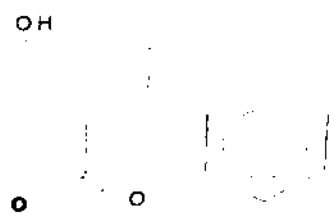
As resistant variants may emerge more quickly against symmetric inhibitors,[18] asymmetric cyclic urea derivatives may have better affinity with higher efficacy than cyclic cyanoguanidines.

An overall conclusion that can be drawn from the present study is that both antiviral and enzyme inhibition activities of cyclic cyanoguanidines as well as cyclic ureas are the function of hydrophobic property of P2/P2′ substituents. These substituents can be of further advantage if they contain OH– or NH_2– like hydrogen bond donor groups. However, a cyclic urea derivative is found to have better antiviral potency or PR-inhibition activity than corresponding cyclic cyanoguanidine derivative. This difference is assumed to be due to the binding of both P2 and P2′ substituents in cyclic ureas with the receptor, whereas in cyanoguanidines only P2 substituent is assumed to bind. Also an aromatic P2/P2′ substituent appears to have an adverse effect on the activities of cyanoguanidines.

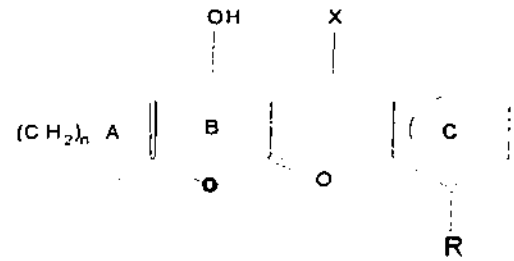
3.3 Cycloalkylpyranones

The investigation of nonpeptidic inhibitors of low molecular weight, resulted within a short time in the identification of a lead structure, 4-hydroxy coumarin (4).[15] This lead structure led to the development of several series of cycloalkyl-Pyranones(5),[16-19] which exhibited better enzyme inhibition activity than (4).

All the cycloalkylpyranones series listed in Tables (3.5) – (3.10) and their HIV-1-PR binding affinity (K_i) data have been taken from the reports of Romines et al..[17-19]



4

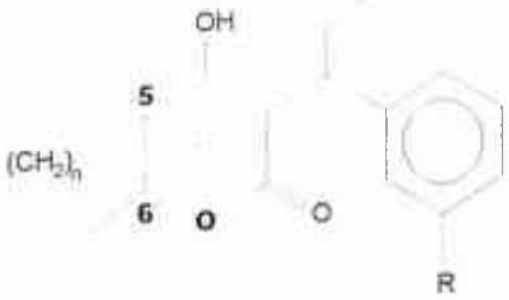


5

The analysis on this data has been performed using either the simple parametric method developed by Hansch[20] or nonparametric method developed by Fujita and Ban.[21]

The cycloalkylhydropyranones listed in Table (3.5) have been studied[18] for the effects on their activity of changes in the size of the alkyl ring A and of the various modifications in the substituent at the 3-position. In Table (3.5), compounds do not have much variations in the 3-position substituents, but vary in the size of the alkyl ring from 5-membered to 8-membered with saturation or unsaturation of the 5,6-bond. The effects of these changes could be easily specified by using only indicator variables I_1, I_2, and I_3 with a value of 1 each for n = 4, unsaturation of the 5,6-bond, and R = 4-cyano-2-pyridinesulfonamide, respectively. From the excellent correlation obtained (Eq. (3.14)), the maximum weight was found to be attached with I_3 variable, signifying a predominant role of 4-cyano-2-pyridinesulfonamide group at the meta position (3′) of the phenyl ring in the 3-substituent. The other two variables are also seen to produce positive effects on the activity of the compounds, but while I_1, signifying the effect of cycloalkyl ring size, is

Table (3.5) : Cycloalkylpyranones studied by Romines et al.[18] and their HIV-1-PR binding affinity.



No.	R	sat/un sat. at 5-6	n	I_1	I_2	I_3	log(1/K_i) Obsd.	log(1/K_i) Calcd. Eq.(3.14)
1	H	unsat.	1	0	1	0	6.16	6.24
2	H	unsat.	2	0	1	0	5.96	6.24
3	H	unsat.	3	0	1	0	6.32	6.24
4	H	unsat.	4	1	1	0	7.13	6.78
5	H	unsat.	6	0	1	0	-[a]	6.24
6	$NHSO_2$-C_5H_5N-CN	unsat.	1	0	1	1	10.13	10.27
7	$NHSO_2$-C_5H_5N-CN	unsat.	2	0	1	1	10.19	10.27
8	$NHSO_2$-C_5H_5N-CN	unsat.	3	0	1	1	10.74	10.27
9	$NHSO_2$-C_5H_5N-CN	unsat.	4	1	1	1	11.16	10.82
10	$NHSO_2$-C_5H_5N-CN	unsat.	6	0	1	1	9.60	10.27
11	H	sat.	1	0	0	0	-[a]	5.81
12	H	sat.	2	0	0	0	6.17	5.81
13	H	sat.	3	0	0	0	-[a]	5.81
14	H	sat.	4	1	0	0	5.92	6.36
15	$NHSO_2$-C_5H_5N-CN	sat.	1	0	0	1	9.83	9.84
16	$NHSO_2$-C_5H_5N-CN	sat.	2	0	0	1	10.30	9.84
17	$NHSO_2$-C_5H_5N-CN	sat.	3	0	0	1	9.72	9.84
18	$NHSO_2$-C_5H_5N-CN	sat.	4	1	0	1	10.13	10.39

[a]Not used in the derivation of Eq.(3.14)

statistically significant at 95% confidence level, I_2, delineating the effect of unsaturation of 5,6 bond, is not. However, since the coefficient of I_2 is positive, the unsaturation of the 5,6-bond would be preferred to its saturation.

$$\log(1/K_i) = 0.546(\pm0.504)I_1 + 0.429(\pm0.454)I_2 + 4.032(\pm0.454)I_3 + 5.809$$

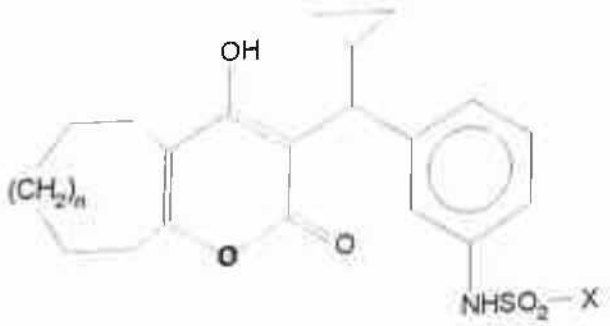
$$n = 15, \quad r = 0.986, \quad s = 0.39, \quad F_{3,11} = 127.20 \qquad (3.14)$$

The other values of n were also taken for I_1, but the best correlation was obtained with $n = 4$ only. n was also used as variable with all its values 1– 6, but it was found to be statistically insignificant. Thus a cyclooctyl ring was found to be optimal. This fact was also verified in Table (3.6), where primarily the variation in the 3'-sulfonamide group was studied.[19] A highly significant correlation (Eq. (3.15)) was obtained between the binding constant of the compounds and the calculated hydrophobic constant (π) of the X moiety at the sulfonamide group and an indicator variable I = 1 for $n = 2$ in the cycloalkyl ring. The correlation shows that $n = 2$ (a cyclooctyl ring) will give 10-fold higher activity than $n = 1$ (a cycloheptyl ring), for which I = 0. The correlation of activity with π_x suggests that the X moiety of the sulfonamide may have hydrophobic interaction with the receptor, but since the correlation is parabolic and π_x attains an optimum value equal to 2.79, the receptor site may be assumed to have limited bulk tolerance, since in *in vitro* system there is no membrane-like lipid–water transport barrier to optimize the role of hydrophobicity.

$$\log(1/K_i) = 0.762(\pm0.407)\pi_x - 0.136(\pm0.085)\pi_x^2 + 0.870(\pm0.293)I + 6.551$$

$$n = 10, \quad r = 0.979, \quad s = 0.13, \quad F_{3,6} = 47.21(9.78) \qquad (3.15)$$

Table (3.6) : Cycloalkylpyranones studied by Skulnick et al [19] and their HIV-1-PR binding affinity.



No.	n	X	π_X	I	log(1/K_i) Obsd.	log(1/K_i) Calcd. Eq. (3.15)
1	1	CH_3	0.56	0	6.92	6.93
2	1	CH_2CH_3	1.02	0	7.20	7.19
3	1	$CH{=}CH_2$	0.82	0	-[a]	7.08
4	2	CH_3	0.56	1	7.96	7.80
5	2	CH_2CH_3	1.02	1	8.00	8.06
6	2	$CH{=}CH_2$	0.82	1	7.75	7.95
7	2	n-Pr	1.55	1	8.40	8.27
8	2	i-Pr	1.53	1	7.48[a]	8.27
9	2	n-Bu	2.13	1	8.34	8.42
10	2	n-Octyl	4.20	1	8.22	8.21
11	2	C_6H_5	1.96	1	8.50	8.39
12	2	Cyclohexyl	2.76	1	8.44	8.48

[a]Not included in the derivation of Eq.(3.15)

In deriving Eq. (3.15), however, compound 8 (Table (3.6)) has not been included, as it behaved as an outlier. Its observed activity is much lower than that predicted by Eq. (3.15) (Table (3.6)). This low activity of the compound can be attributed to the branched character of the X moiety, which may produce some steric effects. Compound 3 also was not included because of its uncertain activity date.

Also in deriving Eq. (3.14) compounds 5, 11, and 13 (Table (3.5)) were not included because of their uncertain activity data. The equation, however, predicts their activity as 6.24, 5.81, and 5.81, respectively.

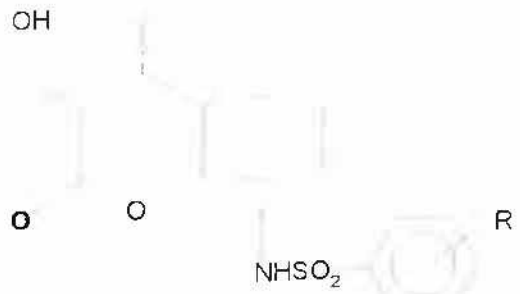
Table (3.7) contains a series of cyclooctylpyranones, where a study has been made[19] on the effect of substitution at the aryl portion of the 3′-arylsulfonamide group. For this series, the best Eq. that we could obtain was

$$\log(1/K_i) = 9.948 - 0.564(\pm 0.194)B5_2 - 0.118(\pm 0.073)B5_4 - 0.774(\pm 0.239)B5_5 + 0.662(\pm 0.328)D_{CN}$$

$$n = 43, \quad r = 0.856, \quad s = 0.21, \quad F_{4,38} = 26.07 \quad (3.86) \qquad (3.16)$$

where B5 is Verloop's STERIMOL parameter defining the maximum width of the substituents,[22] and D_{CN} is an indicator variable used with a value of unity for a CN group present at the 3- or 4-position. If this indicator variable is used for 2-CN also, a slightly poorer correlation is obtained (Eq. (3.17)), exhibiting a reduced effect of the 2-CN ring substitution. A positive coefficient of this variable in both Eqs. (3.16) and (3.17), thus,

Table (3.7): Cyclooctylpyranones studied by Skulnick et al.[19] and their HIV-1-PR binding affinity and physicochemical parameters



No.	R	$B5_2$	$B5_4$	$B5_5$	D_{CS}	log(1/Ki) Obsd.	log(1/Ki) Calcd. Eq. (3.16)
1	H	1.00	1.00	1.00	0	8.50	8.53
2	2-Me	2.04	1.00	1.00	0	8.03	7.95
3	2-F	1.35	1.00	1.00	0	7.96	8.34
4	2-Cl	1.80	1.00	1.00	0	8.03	8.08
5	2-CF_3	2.61	1.00	1.00	0	7.68	7.62
6	2-CN	1.60	1.00	1.00	0	8.13	8.19
7	3-Me	1.00	1.00	1.00	0	8.28	8.53
8	3-Cl	1.00	1.00	1.00	0	8.57	8.53
9	3-Br	1.00	1.00	1.00	0	8.59	8.53
10	3-CF_3	1.00	1.00	1.00	0	8.52	8.53
11	3-NO_2	1.00	1.00	1.00	0	8.33	8.53
12	3-COOH	1.00	1.00	1.00	0	8.35	8.53
13	3-CO_2Me	1.00	1.00	1.00	0	8.85	8.53
14	3-NH_2	1.00	1.00	1.00	0	8.64	8.53
15	3-CN	1.00	1.00	1.00	1	9.22	9.19
16	4-Me	1.00	2.04	1.00	0	8.48	8.41
17	4-Et	1.00	3.17	1.00	0	8.30	8.28
18	4-n-Pr	1.00	3.49	1.00	0	7.96	8.24
19	4-i-Pr	1.00	3.17	1.00	0	7.89	8.28
20	4-t-Bu	1.00	3.17	1.00	0	7.50[a]	8.28
21	4-F	1.00	1.35	1.00	0	8.51	8.49
22	4-Cl	1.00	1.80	1.00	0	8.60	8.44
23	4-Br	1.00	1.95	1.00	0	8.68	8.42
24	4-I	1.00	2.15	1.00	0	8.51	8.40
25	4-CF_3	1.00	2.61	1.00	0	8.21	8.34
26	4-CN	1.00	1.60	1.00	1	9.10	9.13
27	4-NO_2	1.00	2.44	1.00	0	8.57	8.36
28	4-CO_2H	1.00	2.66	1.00	0	7.96	8.34

Continued...

No.	R	$B5_2$	$B5_4$	$B5_6$	D_{CN}	log(1/K_i) Obsd.	log(1/K_i) Calcd. Eq.(3.16)
29	4-$CONH_2$	1.00	3.07	1.00	0	8.72	8.29
30	4-OMe	1.00	3.07	1.00	0	8.41	8.29
31	4-O-n-Bu	1.00	4.79	1.00	0	7.75	8.09
32	4-OCF_3	1.00	3.61	1.00	0	8.40	8.23
33	4-NH_2	1.00	1.97	1.00	0	8.16	8.42
34	4-NMe_2	1.00	1.35	1.00	0	8.55	8.49
35	4-N_3	1.00	4.18	1.00	0	8.59	8.16
36	2,6-diMe	2.04	1.00	1.00	0	< 6.70[a]	7.95
37	2,6-diCl	1.80	1.00	1.00	0	< 6.70[a]	8.08
38	2,5-diCl	1.80	1.00	1.80	0	7.13	7.46
39	2,4-diCl	1.80	1.80	1.00	0	8.21	7.99
40	2,4-diF	1.35	1.35	1.00	0	8.39	8.29
41	2,3-diCl	1.80	1.00	1.00	0	8.12	8.08
42	3,5-diCl	1.00	1.00	1.80	0	8.16	7.91
43	3,5-diCF_3	1.00	1.00	2.61	0	7.33	7.29
44	3,4-diCl	1.00	1.80	1.00	0	8.39	8.44
45	3,4-diOCH_3	1.00	3.07	1.00	0	8.17	8.29
46	2,3,4-triCl	1.80	1.80	1.00	0	8.06	7.99

[a]Not included in the derivation of Eq.(3.16)

indicates that a CN substitution at the ring would be beneficial with a dominant effect at the 3- or 4-position.

$$\log(1/K_i) = 10.059 - 0.612(\pm 0.209)B5_2 - 0.122(\pm 0.080)B5_4 - 0.778(\pm 0.261)B5_5 + 0.416(\pm 0.295)D_{CN}$$

$$n = 43, \quad r = 0.827, \quad s = 0.22, \quad F_{4,38} = 20.51 \quad (3.86) \qquad (3.17)$$

In both the equations, the negative coefficients of width parameter used for 2-, 4-, and 5-substituents suggest that all these substituents would produce steric effects. No physicochemical parameters for the 3-substituents were found to be significant, hence the 3-substituents were assumed to play no role in the binding of the compounds with the receptor.

The high coefficients of $B5_2$ and $B5_5$ as compared to that of $B5_4$ suggest that the 2- and 5-substituents have more detrimental effect than the 4-substituents. The beneficial effect of CN present at any position can be attributed to its highly polar nature and completely linear shape, so that it can be easily accommodated even in a narrowest site and have a strong dipole–dipole or any other kind of electronic interaction.

In the derivation of Eqs. (3.16) and (3.17), however, compounds **36** and **37** (Table (3.7)), due to having uncertain activity data, were not included. Another compound **20** was not included because it exhibited an aberrant behaviour. Its observed activity (7.50) is shown to be much less than predicted by Eq. (3.16) (8.28). This difference can be

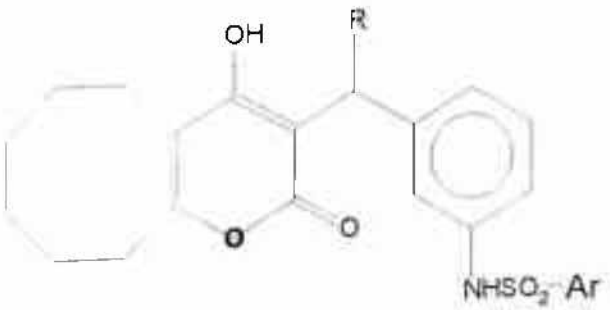
attributed to the bulk of 4-*t*-Bu substituent, whose effect probably cannot be accounted for by only a width parameter but by its total molar volume.

The fairly low activity of compounds **36** and **37** (Table (3.7)) can be due to the presence of substituents at both the ortho positions, which together might be producing strong steric effects.

In another series (Table (3.8)) too, the variation in the aryl moiety of the sulfonamide group was studied along with the variation at C-3α of the 3-alkyl substituent.[19] A Fujita–Ban analysis was performed to find out the activity contribution of each different aryl moiety and of each different alkyl group at C-3α. Among the aryl groups, p-Cl-Ph and 3-pyridyl were found to make, though statistically insignificant, negative contributions and of the remaining four, which were found to make positive contributions, the most dominant effect was found to be associated with the N-Me-imidazole group and the next with the p-CN-Ph group. The contributions of the other two, i.e., p-F-Ph and 8-quinoline, though positive, were found to be statistically insignificant.

Among the alkyl groups at C-3α, a statistically significant but negative contribution was found to be associated with the isopropyl group and a positive contribution, though only marginally significant statistically but more than those of others, with c-propyl. Thus, the two aryl groups, N-Me-imidazole and p-CN-Ph, and the two C-3α substitutes, i-pr and c-propyl, were found to dominate the binding affinity,

Table (3.8) Cyclooctylpyranones studied by Skulnick et al.[19] and their HIV-1-PR binding affinity.



No.	Ar	R	I_1	I_2	I_3	I_4	log(1/K_i) Obsd.	Calcd. Eq. (3.18)	Calcd. Eq. (3.19)
1	p-Cl-Ph	c-Pr	0	0	0	1	8.60	8.63	8.52
2	p-Cl-Ph	Et	0	0	0	0	8.46	8.47	8.52
3	p-Cl-Ph	n-Pr	0	0	0	0	8.40	8.47	8.52
4	p-Cl-Ph	i-Pr	0	0	1	0	7.75	7.93	7.93
5	p-CN-Ph	c-Pr	0	1	0	1	9.10	8.93	8.80
6	p-CN-Ph	Et	0	1	0	0	8.51	8.76	8.80
7	p-CN-Ph	n-Pr	0	1	0	0	8.85	8.76	8.80
8	p-CN-Ph	n-Bu	0	1	0	0	8.68	8.76	8.80
9	p-CN-Ph	i-Pr	0	1	1	0	8.23	8.22	8.21
10	p-CN-Ph	CH_2-i-Pr	0	1	0	0	8.82	8.76	8.80
11	p-F-Ph	c-Pr	0	0	0	1	8.51	8.63	8.52
12	p-F-Ph	n-Pr	0	0	0	0	8.68	8.47	8.52
13	p-F-Ph	n-Bu	0	0	0	0	8.22	8.47	8.52
14	p-F-Ph	i-Pr	0	0	1	0	7.96	7.93	7.93
15	p-F-Ph	CH_2-i-Pr	0	0	0	0	8.55	8.47	8.52
16	N-Me-imidazole	c-Pr	1	0	0	1	9.95	10.08	9.98
17	N-Me-imidazole	n-Pr	1	0	0	0	10.00	9.92	9.98
18	N-Me-imidazole	i-Pr	1	0	1	0	9.30	9.38	9.39
19	N-Me-imidazole	CH_2-i-Pr	1	0	0	0	10.07	9.92	9.98
20	8-Quinoline	c-Pr	0	0	0	1	8.72	8.63	8.52
21	8-Quinoline	n-Pr	0	0	0	0	8.68	8.47	8.52
22	8-Quinoline	i-Pr	0	0	1	0	8.38	7.93	7.93
23	8-Quinoline	CH_2-i-Pr	0	0	0	0	8.48	8.47	8.52
24	3-Pyridyl	c-Pr	0	0	0	1	8.66	8.63	8.52
25	3-Pyridyl	n-Pr	0	0	0	0	8.66	8.47	8.52
26	3-Pyridyl	i-Pr	0	0	1	0	7.70	7.93	7.93
27	3-Pyridyl	CH_2-i-Pr	0	0	0	0	8.03	8.47	8.52

accounting for 91% of the variance in it (Eq. (3.18), $r^2 = 0.91$). Their effects have been described by indicator variables I_1, I_2, I_3, and I_4, respectively, with a value of 1 each. If the variable I_4, used for cyclopropyl, is dropped, as in comparison to others it is statistically less significant, the correlation remains almost unaffected (Eq. (3.19)), still accounting for 89% of the variance in the binding affinity ($r^2 = 0.89$).

$$\log(1/K_i) = 1.457(\pm 0.229)I_1 + 0.295(\pm 0.197)I_2 - 0.539(\pm 0.200)I_3 + 0.165(\pm 0.200)I_4 + 8.466$$

$$n = 27, \quad r = 0.953, \quad s = 0.20, \quad F_{4,22} = 54.20(4.31) \qquad (3.18)$$

$$\log(1/K_i) = 1.461(\pm 0.238)I_1 + 0.281(\pm 0.204)I_2 - 0.588(\pm 0.199)I_3 + 8.516$$

$$n = 27, \quad r = 0.946, \quad s = 0.21, \quad F_{3,23} = 65.74(4.76) \qquad (3.19)$$

However, both Eqs. (3.18) and (3.19) show that the activity contributions of the N-Me-imidazole group would be about 15 times greater than that of the p-CN-Ph group. This difference can be attributed to the presence in the former of a heterocyclic nitrogen ortho to the sulfonamide that can participate in the hydrogen bonding interaction with the NH of Asp29. This has been pointed out by Skulnick et at.[19] based on X-ray crystal structures of related sulfonamides bonded to HIV protease.[23] This theory also explains the predominant effect of 4-cyano-2-pyridinesulfonamide group, pointed out by the variable I_3 in Eq. (3.14) obtained for the compounds of Table (3.5).

Although Eq. (3.18) does not show any significant role of c-propyl at C-3α among the various other substituents tried at this position, this group was found to be the

best. In Table (3.9) its effect has been reported against Et. It was also tried as a replacement of the phenyl group, the second substituent at C-3α (R^2). A relative assessment of it, both as R^1 and R^2, was made using an indicator variable I_1 equal to 1 for R^1 = c-propyl and zero for R^1 = Et, and another variable I_2 equal to 1 for R^2 = c-propyl and zero for R^2 = Ph, and then obtaining the equation,

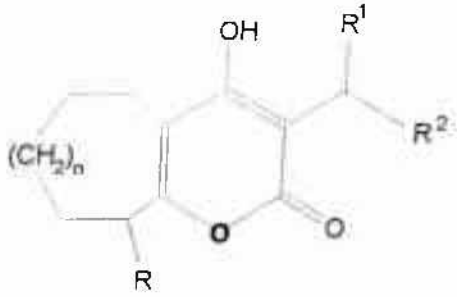
$$\log(1/K_i) = 0.879\ (\pm 0.294)I_1 - 0.827(\pm 0.356)I_2 - 0.594(\pm 0.301)I_R + 7.160$$

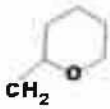
$$n = 15, \quad r = 0.919, \quad s = 0.197, \quad F_{3,11} = 19.79 \qquad (3.20)$$

which showed that R^1 = c-propyl would be preferred to R^1 = Et and that R^2 = c-propyl would be less advantageous than R^2 = Ph. The additional parameter I_R in Eq. (3.20) is to account for the effect of the R substituent of the cycloalkyl ring. For any R substituent it is equal to 1 and zero for no substituent. From its negative coefficient it is obvious that no R substituent at the cycloalkyl ring would be advantageous.

In deriving Eq. (3.20), however, compound **14** (Table (3.9)) has not been included. It behaves as an outlier. This compound is similar to compound 6, except that it is an heptyl ring instead of an octyl ring. Its low activity as compared to that of compound 6 should be attributed to this difference, but when a parameter I_n is used to account for this difference with a value of 1 for octyl ring and zero for heptyl ring, no statistical significance of it is obtained nor is there obtained any significant improvement in the correlation (Eq. (3.21)).

Table (3.9) Cycloalkylpyranones studied by Romines et al.[17] and their HIV-1-PR binding affinity.



No.	n	R^1	R^2	R	I_1	I_2	I_R	I_n	log(1/K_i) Obsd.	log(1/K_i) Calcd. Eq.(3.20)
1	2	Et	Ph	H	0	0	0	1	7.23	7.19
2	2	Et	Ph	Me	0	0	1	1	6.48	6.47
3	2	Et	Ph	CH_2-Ph	0	0	1	1	6.56	6.63
4	2	c-Pr	cPr	H	1	1	0	1	7.24[a]	7.19
5	2	c-Pr	cPr	CH_2-Ph	1	1	1	1	6.59	6.64
6	2	c-Pr	Ph	H	1	0	0	1	7.93	7.65
7	2	c-Pr	Ph	Et	1	0	1	1	7.48	7.50
8	2	c-Pr	Ph	n-Pr	1	0	1	1	7.59	7.70
9	2	c-Pr	Ph	n-Bu	1	0	1	1	7.51	7.59
10	2	c-Pr	Ph	CH_2-i-Pr	1	0	1	1	7.52	7.66
11	2	c-Pr	Ph	CH_2-c-Pr	1	0	1	1	7.62	7.58
12	2	c-Pr	Ph	$(CH_2)_2$-i-Pr	1	0	1	1	7.30	7.32
13	2	c-Pr	Ph		1	0	1	1	6.92	6.87
14	1	c-Pr	Ph	H	1	0	0	0	7.02[a]	7.40
15	1	c-Pr	Ph	$(CH_2)_2OCH_3$	1	0	1	0	7.55	7.46
16	1	c-Pr	Ph	CH_2.c-Pr	1	0	1	0	7.62	7.33

[a]Not included in the derivation of Eq.(3.20)

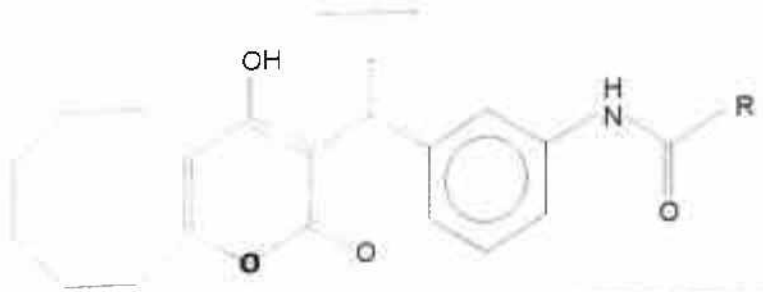
$$\log(1/K_i) = 0.847(\pm 0.299)I_1 - 0.789(\pm 0.380)I_2 - 0.611(\pm 0.302)I_R$$
$$- 0.179(\pm 0.343)I_n + 7.249$$
$$n = 16, \quad r = 0.929, \quad s = 0.194, \quad F_{4,11} = 15.64\ (5.67) \qquad (3.21)$$

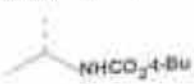
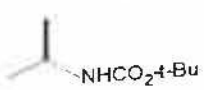
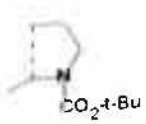
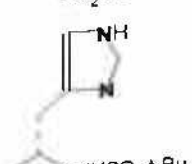
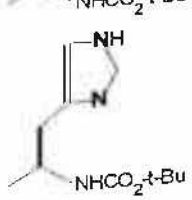
In addition to sulfonamide derivatives, a series of carboxamide derivatives (Table (3.10)) was also studied,[17] in which the R-moiety of carboxamide group was varied. We found a good correlation between the calculated π values of various R-moieties and the activity of the compounds as shown by

$$\log(1/K_i) = 8.480 - 0.750(\pm 0.397)\pi_R$$
$$n = 17, \quad r = 0.721, \quad s = 0.50, \quad F_{1,15} = 16.20(8.86) \qquad (3.22)$$

This correlation suggests that a lipophilic R-moiety will not be beneficial to the activity. The correlation, however, accounts for only 52% of the variance in the activity. A significant improvement in it was achieved (Eq. (3.23)) when an indicator variable I equal to 1 was used for compounds **12-15** (Table (3.10)). In these compounds, the R-moieties provide a hydrogen-bond donor or acceptor group just adjacent to the amide group (NH-CO), and the negative coefficient of I indicates that such moieties would not be favourable to the binding. It can be assumed that these moieties can form intramolecular hydrogen bonds with NH or CO of the amide group and thus hinder their interaction with the receptor. The correlation now, however, accounts for 75% variance in the binding affinity and is further improved significantly to account for 88% of the variance when compounds **5** and **14** are excluded from the regression (Eq. (3.24)). For both the compounds, Eq. (3.24) predicts a fairly low activity as compared to their

Table (3.10) : Cyclooctylpyranones studied by Romines et al.[17] and their HIV-1-PR binding affinity.



No.	R	π	I	log(1/Ki) Obsd.	log(1/Ki) Calcd. Eq.(3.24)
1	Ph	1.49	0	7.38	7.51
2	p-F-Ph	1.83	0	7.26	7.27
3	$CH_2CH=CHPh$	2.65	0	6.33	6.72
4	$CH_2CH_2NHCO_2$-t-Bu	0.91	0	8.26	7.99
5	$CH_2CH_2CH_2NHCO_2$-t-Bu	1.24	0	8.40[a]	7.68
6	$CH_2CH_2CH_2CH_2NHCO_2$-t-Bu	0.91	0	7.85	7.90
7		1.13	0	8.16	7.75
8		1.22	0	7.49	7.69
9		2.09	0	7.37	7.10
10		0.23	0	8.52	8.36
11		0.23	0	7.96	8.36
12	NHEt	0.47	1	7.14	7.24
13	NHPh	1.86	1	6.35	6.30
14	OEt	1.13	1	7.72[a]	6.79
15	OPh	2.11	1	6.18	6.13
16	CH_2Et	1.05	0	7.62	7.80
17	CH_2Ph	1.76	0	7.48	7.32

[a]Not included in the derivation of Eq (3.24)

corresponding observed activity (Table (3.10)). The reason of this difference seems hard to explain.

$$\log(1/K_i) = 8.622 - 0.718(\pm 0.297)\pi_R - 0.780(\pm 0.459)I$$

$$n = 17, \quad r = 0.868, \quad s = 0.37, \quad F_{2,14} = 21.36 \ (6.51) \qquad (3.23)$$

$$\log(1/K_i) = 8.518 - 0.680(\pm 0.215)\pi_R - 0.961(\pm 0.374)I$$

$$n = 15, \quad r = 0.938, \quad s = 0.27, \quad F_{2,12} = 43.89 \ (6.93) \qquad (3.24)$$

The use of any steric parameter in place of π_R was not found to be successful. Hence the negative dependence of the activity on π_R cannot be attributed to any kind of steric effect of the R substituent but purely to its hydrophobic nature. Therefore, to have the positive effect one should take the hydrophilic (polar) substituents, which might have a polar interaction with the receptor.

On the basis of these studies, we derive the following conclusions.

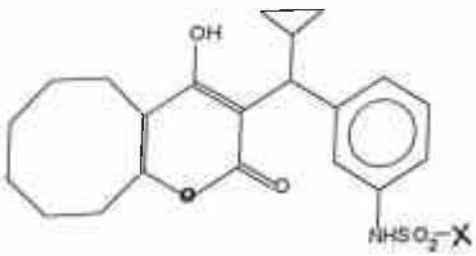
1. In cycloalkyl rings, a cyclooctyl ring with unsaturation in the 5,6-bond appears to be optimal.
2. At the 3-position of the pyranone ring, a cyclopropyl- and phenyl- substituted methyl group would be preferred.
3. Substitution at the meta position of the phenyl ring of the 3-substituent is found to enhance the activity and a 4-cyano-2-pyridinesulfonamide group is found to be the best.

4. A hydrophobic substituent on the sulfonamide would increase the binding affinity, provided it is not very bulky. It is assumed to have a hydrophobic interaction with the receptor. From this point of view an aryl substituent is found to be better than an alkyl substituent.

5. In the arylsulfonamides, if the aryl group is substituted, only a cyanogroup at the 3- or 4-position is found to be advantageous. However, in place of a 3- or 4-cyanophenyl, an N-Me-imidazole group is found to be better.

6. If, in place of a sulfonamide group, a carboxamide group is used, the presence of a polar substituent can increase the activity. However, carboxamide derivatives were found to be less active than sulfonamide derivatives.

On the basis of the above finding, the following compounds seem to have a bright future.



6 X= 4-CN-2-pyridine

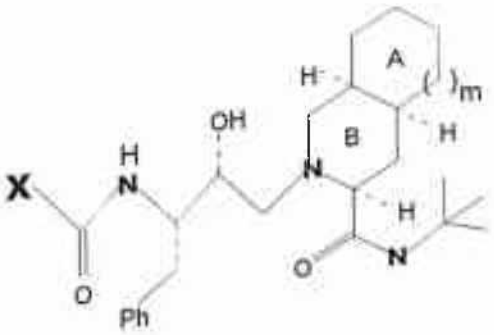
7 X= N-Me-imidazole

8 X= 3- or 4-CN-Ph

HIV proteases offer a number of possibilities for hydrogen bonding and in sulfonamides both NH and SO_2 moieties are assumed to form hydrogen bonds with the receptor.[16] We have already discussed how the above mentioned X substituents can further strengthen the binding of the molecule with the enzyme. In all X substituents the aromatic ring is assumed to have π-stacking interaction with the Arg8 residue of the protease.[19]

3.4 Cyclic Sulfolanes

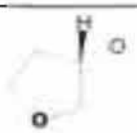
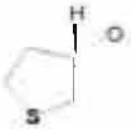
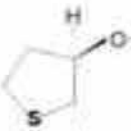
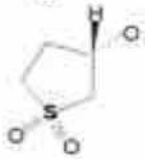
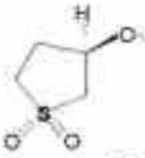
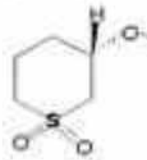
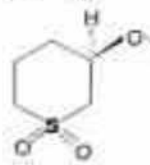
In search of small nonpeptidic HIV-1-PR inhibitors, Ghosh et al.[24] synthesized a series of sulfolanes (9) and studied their inhibition activity. The activity parameter IC_{50} refers to the molar concentration of the compound leading to 50% inhibition of the enzyme activity.



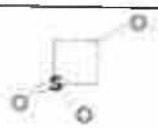
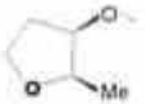
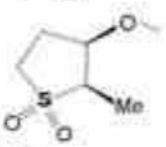
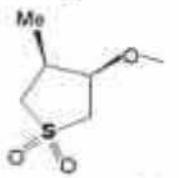
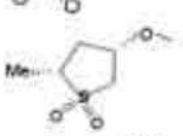
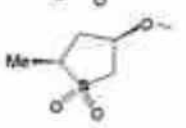
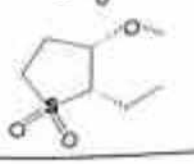
9

The inhibition activity of compounds of Table (3.11) was found to have a good parabolic correlation with the hydrophobic constant π of 2-substituent in the X group of the compounds (Eq. (3.25)).

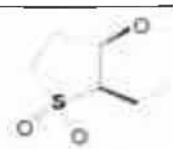
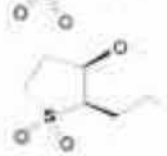
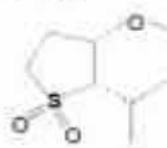
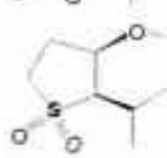
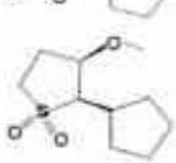
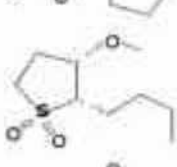
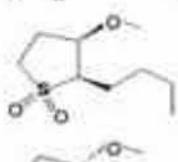
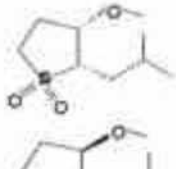
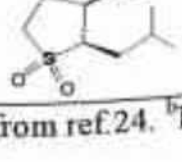
Table (3.11): The HIV-1-PR inhibition activity of some sulfolanes (9)and their physicochemical parameters used in the regression.

No.	X	m	π_2	I_3	I_4	I_m	I_s	log(1/IC$_{50}$) Obsd.[a]	Calcd. Eq.(3.29)	Calcd. Eq.(3.30)
1		1	0.00	0	0	0	1	6.88	6.77	6.76
2		0	0.00	0	0	1	1	7.22	7.29	7.24
3		1	0.00	0	0	0	1	6.88	6.77	6.76
4		1	0.00	0	0	0	0	6.66	6.41	6.47
5		1	0.00	1	0	0	1	7.12	7.07	7.18
6		0	0.00	1	0	1	1	7.95	7.59	7.66
7		1	0.00	1	0	0	0	6.85	6.71	6.90
8		1	0.00	0	0	0	0	5.99	6.41	6.47
9		1	0.00	0	1	0	1	5.76	5.90	5.86
10		1	0.00	0	1	0	0	5.68	5.54	5.58

Contd....

No.	X	m	π_2	I_5	I_6	I_m	I_s	log(1/IC$_{50}$) Obsd.[a]	Calcd. Eq.(3.29)	Calcd. Eq.(3.30)
11		1	0.00	0	0	0	0	6.98	6.41	6.47
12		1	0.52	0	0	0	1	7.28	7.49	7.39
13		1	0.52	0	0	0	0	6.77	7.13	7.11
14		1	0.52	1	0	0	1	7.94	7.79	7.82
15		0	0.52	1	0	1	1	7.89	8.31	8.30
16		1	0.52	1	0	0	0	7.65	7.43	7.53
17		1	0.00	1	0	0	1	7.16	7.07	7.18
18		1	0.00	1	0	0	0	6.66	6.71	6.90
19		1	0.00	1	0	0	1	7.09	7.07	7.18
20		1	0.00	1	0	0	0	5.80[b]	6.71	6.90
21		1	1.05	1	0	0	1	8.27	8.18	8.15

Contd....

No.	X	m	π_2	I_s	I_o	I_m	I_p	log(1/IC$_{50}$) Obsd.[a]	Calcd. Eq.(3.29)	Calcd. Eq.(3.30)
22		1	1.05	1	0	0	0	7.88	7.81	7.86
23		1	1.58	1	0	0	1	8.21	8.20	8.16
24		1	1.58	1	0	0	0	7.70	7.84	7.87
25		1	1.45	1	0	0	1	7.96	8.23	8.19
26		0	1.45	1	0	1	0	8.52	8.39	8.38
27		1	1.45	1	0	0	0	8.46	7.87	7.90
28		1	1.99	1	0	0	1	8.21	7.98	7.94
29		1	1.99	1	0	0	0	7.65	7.61	7.65
30		1	2.11	1	0	0	1	7.83	7.88	7.85
31		1	2.11	1	0	0	0	7.34	7.52	7.56
32		1	1.98	1	0	0	1	7.92	7.99	7.95
33		1	1.98	1	0	0	0	7.52	7.63	7.67

[a] Taken from ref.24. [b] Not used in the derivation of Eq.(3.30).

$$\log(1/IC_{50}) = 2.130(\pm0.901)\pi_2 - 0.793(\pm0.460)\pi_2^2 + 6.699$$

$$n = 33, \quad r = 0.775, \quad s = 0.48, \quad F_{2,30} = 22.49 \ (3.32) \qquad (3.25)$$

This equation accounts for 60% of the variance in the activity ($r^2 = 0.60$) and thus the hydrophobic property of 2-substituents seems to play a major role in the inhibition potency of the compounds.

The correlation was found to be significantly improved when some indicators variables were used. A variable I_m was used to account for the effect of size of A ring. It was given a value of 1 for a five-membered ring (m=0) and zero for a six-membered ring (m=1). Another variable I_5 was used for a five-membered sulfolane ring in X-substituent. For such a ring it was given a value of 1 and for others 0. A third variable I_6 equal to 1 was defined for a six-membered ring in X-substituent. For others it was taken equal to zero. A fourth variable I_s = 1 was taken to describe the effect of a 3(S) configuration of the ring in X relative to a 3(*R*) configuration. All these variables were successively found to improve the correlation, giving finally a highly significant correlation as expressed by Eq. (3.29).

$$\log(1/IC_{50}) = 2.014(\pm0.809)\pi_2 - 0.715(\pm0.415)\pi_2^2 + 0.722(\pm0.502)I_m + 6.603$$

$$n = 33, \quad r = 0.832, \quad s = 0.43, \quad F_{3,29} = 21.72 \ (4.54) \qquad (3.26)$$

$$\log(1/IC_{50}) = 1.771(\pm0.768)\pi_2 - 0.662(\pm0.383)\pi_2^2 + 0.662(\pm0.463)I_m + 0.475(\pm0.382)I_5 + 6.395$$

$$n = 33, \quad r = 0.866, \quad s = 0.38, \quad F_{4,28} = 21.01 \ (4.07) \qquad (3.27)$$

$$\log(1/IC_{50}) = 1.667(\pm0.687)\,\pi_2 - 0.616(\pm0.342)\pi_2^2 + 0.614(\pm0.413)I_m + 0.324(\pm0.355)I_5 - 0.861(\pm0.603)I_6 + 6.581$$

$$n = 33, \quad r = 0.900, \quad s = 0.33, \quad F_{5,27} = 23.07\ (3.79) \qquad (3.28)$$

$$\log(1/IC_{50}) = 1.717(\pm0.594)\pi_2 - 0.634(\pm0.296)\pi_2^2 + 0.520(\pm0.362)I_m + 0.302(\pm0.307)I_5 - 0.867(\pm0.521)I_6 + 0.362(\pm0.231)I_s + 6.406$$

$$n = 33, \quad r = 0.930, \quad s = 0.28, \quad F_{6,26} = 27.59\ (3.59) \qquad (3.29)$$

However, even in Eq. (3.29), compound **20** was found to be misfit. Equation (3.29) predicts very high activity for it as compared to its observed activity (Table (3.11)) Therefore, when this compound was excluded, a further improved correlation was obtained (Eq. (3.30)), accounting for more than 90% of the variance in the activity (r^2 = 0.904). The exclusion of this compound makes the I_5 parameter also statistically significant at 95% confidence level, which is only marginally significant in Eq. (3.29).

$$\log(1/IC_{50}) = 1.524(\pm0.489)\pi_2 - 0.575(\pm0.240)\,\pi_2^2 + 0.479(\pm0.292)I_m + 0.426(\pm0.256)I_5 - 0.893(\pm0.420)I_6 + 0.287(\pm0.190)I_s + 6.470$$

$$n = 32, \quad r = 0.951, \quad s = 0.23, \quad F_{6,25} = 39.16\ (3.63) \qquad (3.30)$$

The fairly low activity of **20** can be attributed to the presence in sulfolane ring of a 5-Me group in R-configuration that might produce some steric effects. This gives a structure which is exactly opposite sterio-isomerically to **19**, which has a much better activity.

Now from Eq. (3.30), which exhibits a highly significant correlation, we can conclude that a five-membered A ring and a five-membered sulfolane ring in 3(S)-configuration and with a lipophilic 2-substituent (cis to 3-substituent) will be highly beneficial to the activity. Only a six membered ring will have the negative effect, which is probably due to some steric effects produced by it. The potency of compounds can be controlled by the lipophilicity of the 2-substituent. The correlation has been parabolic in π_2 and the optimum value of π_2, $(\pi_2)_{opt}$, as obtained from Eq. (3.30) is 1.32, which is essentially the same as can be obtained from Eq. (3.25) (1.34). Thus, a compound having 2-substituent of π = 1.32 and having all other positive factors can be the most potent compound with a predicted activity of 8.67. This compound can be obtained by changing the 2-substituent of **25** and its six-membered A ring to a five-membered one. Thus slightly a more potent compound than **26**, the most potent compound listed in Table (3.11), can be expected.

The discovery of substrate mimicking HIV-1-PR inhibitors was based on the synthesis of substrate analogues in which the scissile bond was replaced by a noncleavable isostere with tetrahedral geometry that could mimic the tetrahedral transition state of the proteolytic reaction. Thus, several inhibitors with hydroxyethylene or hydroxyethylamine isostere replacement were prepared, which could bind with the enzyme as shown in Figure (1.7). In the inhibitor–enzyme interaction, the enzyme's water molecule makes hydrogen bonds with both the inhibitor and the enzyme with approximately tetrahedral geometry. This water molecule in the complex is known as 'flap' water. A sulfolane can bind with this flap water as shown in Figure (3.3), which

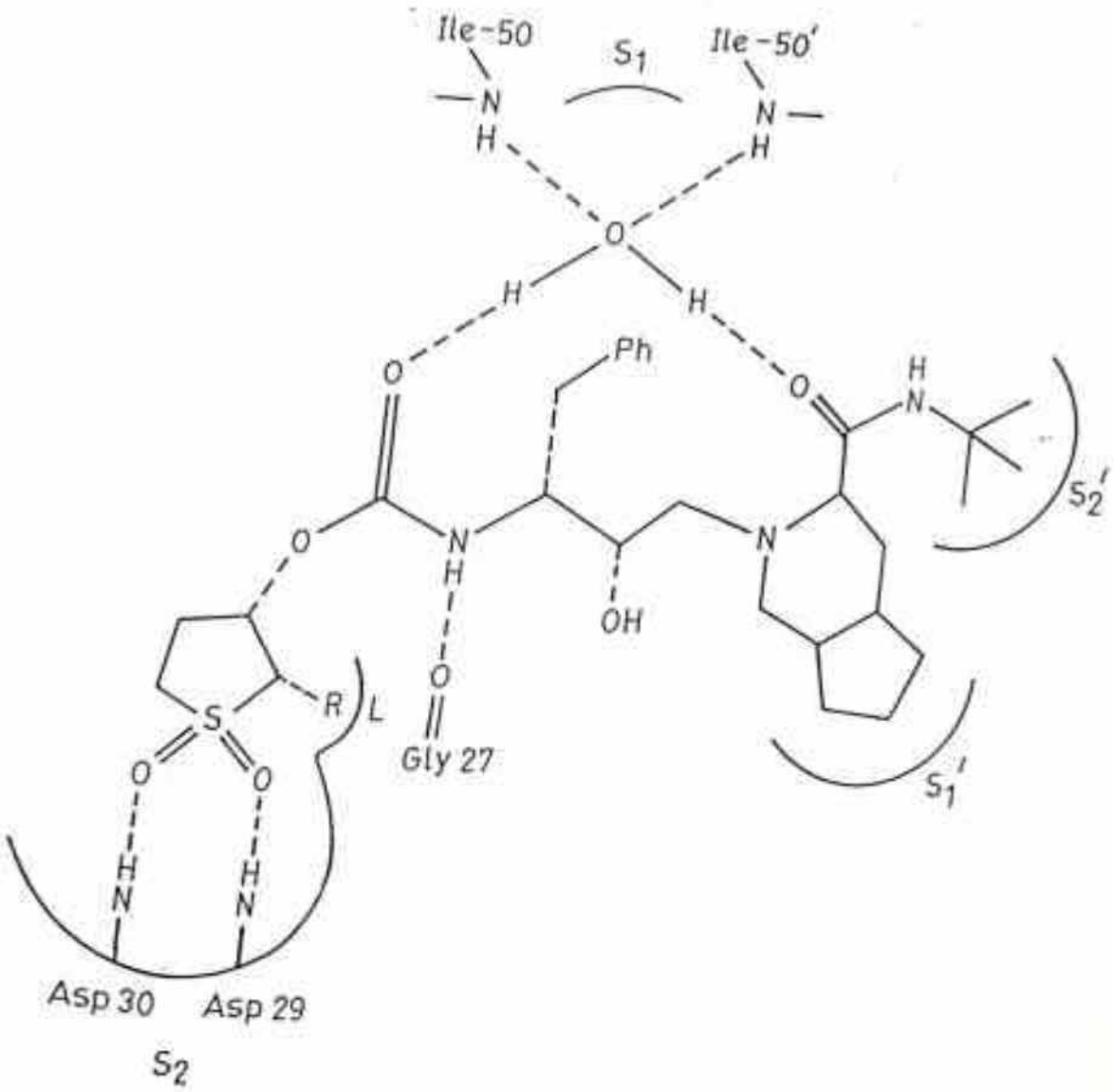


Figure (3.3): *A model proposed for the binding of a sulfolane with HIV-1-PR. S_1', S_2' can be hydrophobic sites in the enzyme. S_2 site is shown to participate in the hydrogen bondings but also to contain a small cavity L which can accommodate the hydrophobic 2-substituent.*

also exhibits how other portions of the inhibitor can interact with the enzyme leading to its inhibition. The sulfone oxygens can have the hydrogen bondings with Asp 29 and Asp 30 present in S_2 binding domain of the HIV-1 protease, and the hydrophobic 2-substituent can have the hydrophobic interaction with a hydrophobic cavity present in the S_2 region.

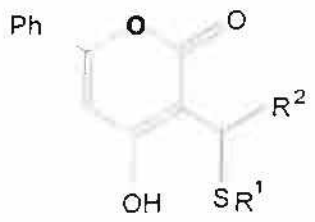
The independent variables used in Eq. (3.30), finally obtained for Table (3.11), were found to have no mutual correlations (Table (3.12)).

Table (3.12): Mutual correlations (r-values) among the variables used in Eq. (3.30)

	π_2	I_m	I_5	I_6	I_s
π_2	1.00	0.126	0.548	0.243	0.095
I_m		1.00	0.051	0.098	0.166
I_5			1.00	0.383	0.042
I_6				1.00	0.016
I_s					1.00

3.5 Arylthiomethanes

Vara Prasad et al.[25] synthesised a series of (4-hydroxy-6-phenyl-2-oxo-2H-pyran-3-yl)thiomethanes (**10**) and studied their inhibition activity. The parameter IC_{50} refers to the molar concentration of the compound leading to 50% inhibition of the enzyme activity. In the case of arylthiomethanes (Table (3.13)), too, the hydrophobic nature of



10

substituents is found to play a major role. A significant parabolic correlation is obtained between the enzyme inhibition activity of the compounds and the hydrophobic parameter of R^2-substituents (Eq. (3.31)), suggesting that these substituents may be involved in strong hydrophobic interaction with the enzyme.

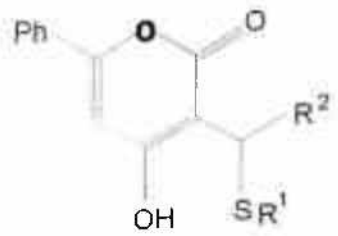
$$\log(1/IC_{50}) = 3.263(\pm 1.247)\pi_{R2} - 1.042(\pm 0.418)(\pi_{R2})^2 + 4.048$$

$$n = 19, \quad r = 0.811, \quad s = 0.43, \quad F_{2,16} = 15.39\ (6.23) \qquad (3.31)$$

However, since the correlation is parabolic, the receptor site may have the limited bulk tolerance, as unlike in in vivo there is no membrane-like lipid-water barrier in *in vitro* system to optimize the lipophilic effect. The optimum value of π_{R2} as obtained from Eq. (3.31) is 1.57, suggesting that only a moderately lipophilic or essentially only a moderately bulky R^2-substituent can be advantageous.

The correlation expressed by Eq. (3.31) was found to be further improved significantly when two indicator variables were introduced (Eq. (3.32)). A variable $I_{R2} = 1$ was used for all those R^2-substituents which were attached to C-3α through CH_2

Table (3.13) : The HIV-1-PR inhibition activity of some arylthiomethanes (10) and their physicochemical parameters used in the regression.



No.	R^1	R^2	π_{R2}	$V_{w,R2}$	I_{R2}	I_{R1}	log(1/IC$_{50}$) Obsd.[a]	Calcd. (Eq.3.32)	Calcd. (Eq.3.33)
1	C_6H_5	H	0.00	0.056	0	0	4.07	4.15	4.16
2	C_6H_5	C_6H_5	1.35	0.785	0	0	6.11	6.07	6.01
3	C_6H_5	2-naphthyl	2.52	1.205	0	0	5.11	5.33	4.95
4	C_6H_5	cyclohexyl	2.43	0.833	0	0	5.61	5.47	5.97
5	C_6H_5	$CH_2CH(CH_3)_2$	1.77	0.707	1	0	6.39	6.66	6.57
6	C_6H_5	$CH_2CH_2CH(CH_3)_2$	2.30	0.861	1	0	6.41	6.24	6.47
7	2-naphthyl	C_6H_5	1.35	0.785	0	0	5.61	6.07	6.01
8	$CH_2C_6H_5$	C_6H_5	1.35	0.785	0	0	6.32	6.07	6.01
9	$CH_2C_6H_5$	$CH_2CH(CH_3)_2$	1.77	0.707	1	0	6.59	6.66	6.57
10	$CH_2C_6H_5$	CH_2cyclopropyl	1.28	0.603	1	0	7.08	6.63	6.52
11	cyclohexyl	C_6H_5	1.35	0.785	0	0	6.32	6.07	6.01
12	cyclohexyl	$CH_2CH(CH_3)_2$	1.77	0.707	1	0	6.50	6.66	6.57
13	cyclohexyl	CH_2cyclopropyl	1.28	0.603	1	0	6.83	6.63	6.52
14	cyclohexyl	CH_2cyclopentyl	2.40	0.898	1	0	6.27	6.09	6.42
15	cyclohexyl	$CH_2C(CH_3)_3$	2.17	0.861	1	0	6.52	6.39	6.47
16	cyclohexylmethyl	$CH_2CH(CH_3)_2$	1.77	0.707	1	0	6.07	6.66	6.57
17	cyclopentyl	cyclopentyl	1.87	0.757	0	1	6.65	6.59	6.67
18	cyclopentyl	$CH_2CH(CH_3)_2$	1.77	0.707	1	1	7.24	7.24	7.22
19	cyclopentyl	CH_2cyclopropyl	1.28	0.603	1	1	7.16	7.21	7.17

[a] Taken from ref. 25

bridge group and the other variable I_{R1} = 1 was used for a cyclopentyl group in R^1-substituents.

$$\log(1/IC_{50}) = 2.516(\pm0.844)\pi_{R2} - 0.813(\pm 0.279)(\pi_{R2})^2 + 0.591(\pm0.309)I_{R2} + 0.586(\pm0.403)\, I_{R1} + 4.150$$

$$n = 19, \quad r = 0.940, \quad s = 0.25, \quad F_{4,14} = 26.67\ (5.03), \ (\pi_{R2})_{opt} = 1.55 \qquad (3.32)$$

The statistically quite significant positive coefficients of these variables indicate that such substituents, for which these variables stand, are crucial for the activity. In R^2-substituent, a CH_2 bridge can provide conformational flexibility to the substituent, because of which the substituent may be able to have the desired hydrophobic interaction with the receptor site. The advantageous role of cyclopentyl group in R^1-substituents can be assumed to be due to its ability to make a complete steric fit with the receptor site.

The use of van der Waals volume (V_w) in place of π for R^2-substituents gave equally significant correlation (Eq. (3.33)), supporting the suggestion that very bulky R^2-substituent will not be tolerated at the receptor site. However, there exists a very good correlation between $V_{w,R2}$ and π_{R2} ($r = 0.865$), hence it is difficult to say whether it is the hydrophobic interaction or the dispersion interaction which really takes place between the R^2-substituent and the receptor.

$$\log(1/IC_{50}) = 6.238(\pm2.186)V_{w,R2} - 4.404(\pm1.726)(V_{w,R2})^2 + 0.535(\pm0.330)I_{R2} + 0.644(\pm0.415)I_{R1} + 3.828$$

$$n = 19, \quad r = 0.937, \quad s = 0.26, \quad F_{4,14} = 25.16\ (5.03), \ (V_{w,R2})_{opt} = 0.708 \qquad (3.33)$$

Since it has been observed that protease has four hydrophobic pockets near its active site and that the favourable hydrophobic interactions with these pockets are desirable for an inhibitor to achieve nanomolar potency,[26] it is plausible to assume, in the present case, that it is only the hydrophobic interaction which is responsible for the binding of the R^2-substituent, and even the R^1-substituent, with the enzyme. Wang et al.[27] also observed that at least two additional factors are important in the binding of a compound to HIV-1 protease. The first is the conformational flexibility of the inhibitor molecule and the second is the hydrophobic interactions between an inhibitor and the enzyme. According to Vara Prasad et al.[25] the R^1- and R^2-substituents interact with S_1' and S_2' sites of the enzyme (Figure (3.4)). Our theoretical study fully conforms to it and predicts the same compound **18** (Table (3.12)) to possess the highest activity as the one observed by experiment.

The independent variables used in the Eq. (3.32), finally obtained for Table (3.13), were found to have no mutual correlations (Table (3.14)).

Table (3.14): Mutual correlations (r-values) among the variables used in Eq. (3.32)

	π_{R2}	I_{R2}	I_{R1}
π_{R2}	1.00	0.297	0.025
I_{R2}		1.00	0.208
I_{R1}			1.00

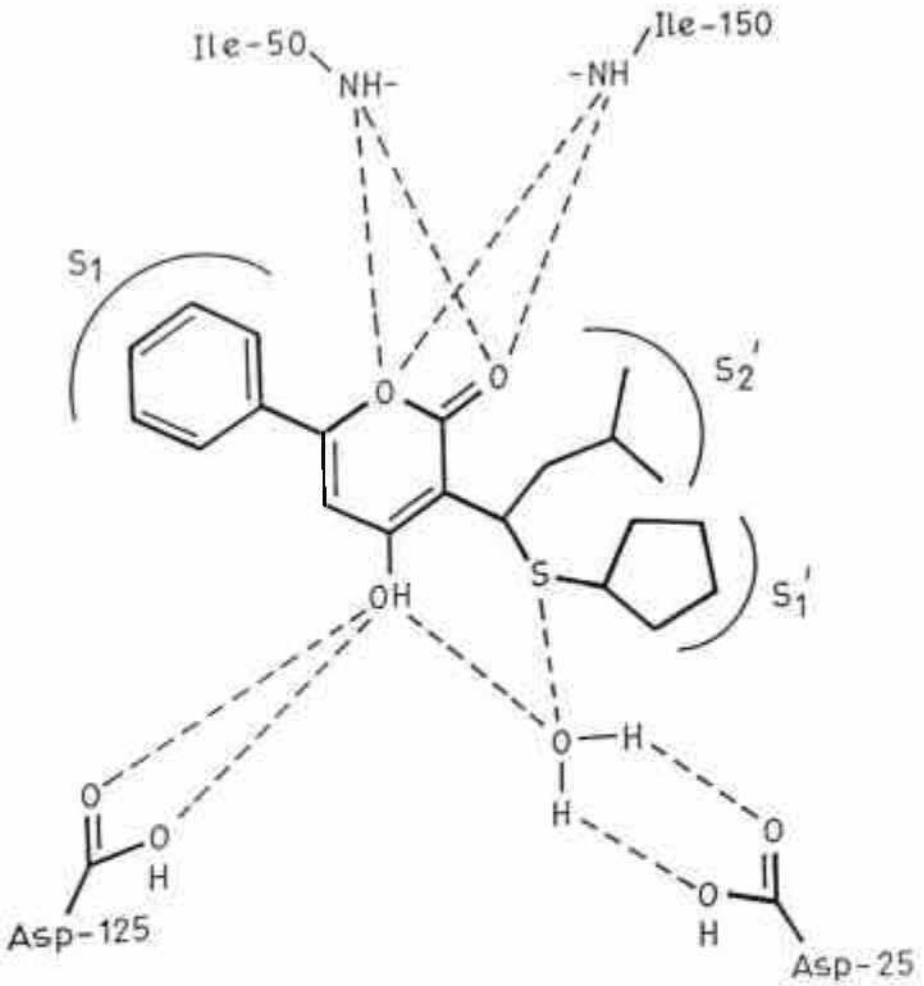
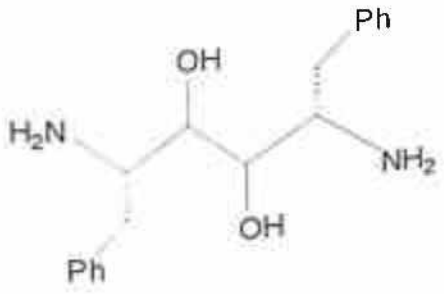


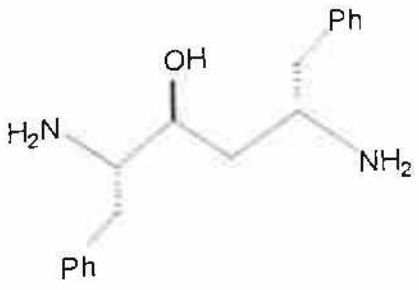
Figure (3.4): A model of binding of arylthiomethanes with HIV-1-PR based on X-ray crystallographic studies.[25] S_1, S_1', S_2' all can be hydrophobic sites.

3.6 Ritonavir Analogues

Since HIV protease exists as a C_2-symmetric homodimer, a variety of peptidic inhibitors that were developed were based on this C_2-symmetric structure. But their clinical development has been hindered by their poor pharmacokinetics, including low oral bioavailability and rapid excretion. Therefore, attempts have been made to develop modified peptidic inhibitors and substantial progress in the identification of agents with high oral bioavailability has been made. Kempf et al.[28,29] reported series of inhibitors based on a symmetric core diamine **11** and a pseudo symmetric core diamine **12**.

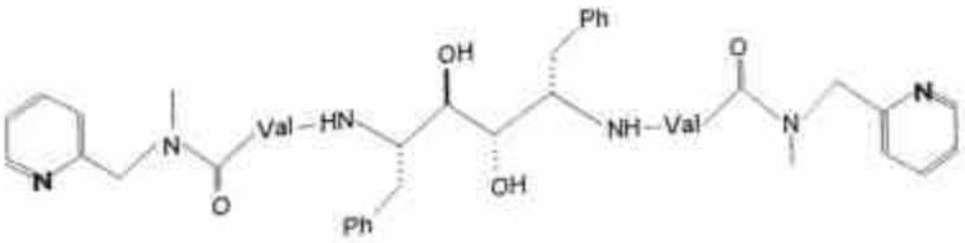


11

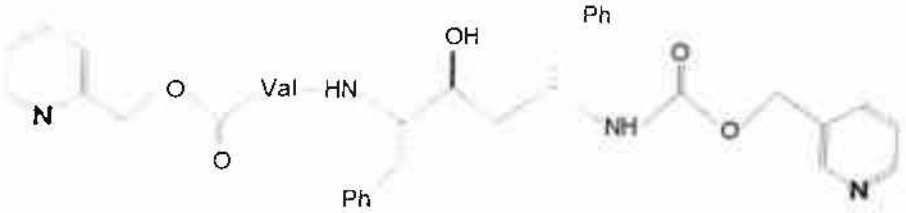


12

Structure-activity studies on the derivatives of **11** and **12** led Kempf et al.[29,30] to identify two compounds **13** (A-77003) and **14** (A-80987), possessing adequate anti-HIV activity. Out of these two, A-80987 was found to possess better oral bioavailability in both animal

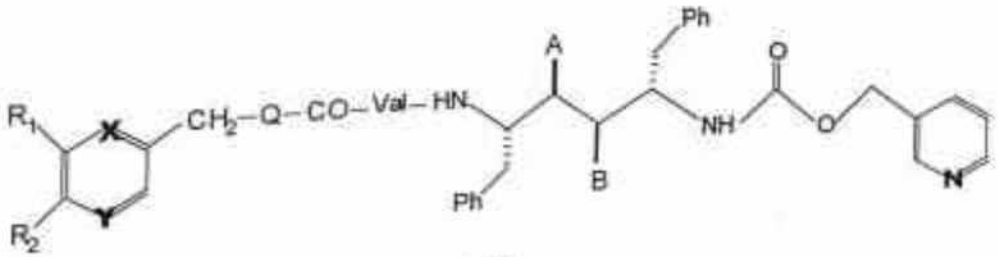


13 (A-77003)

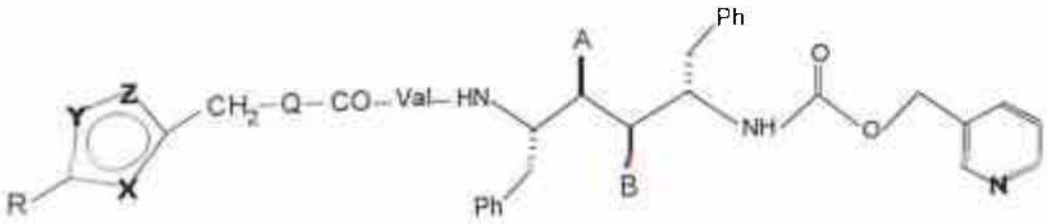


14 (A-80987)

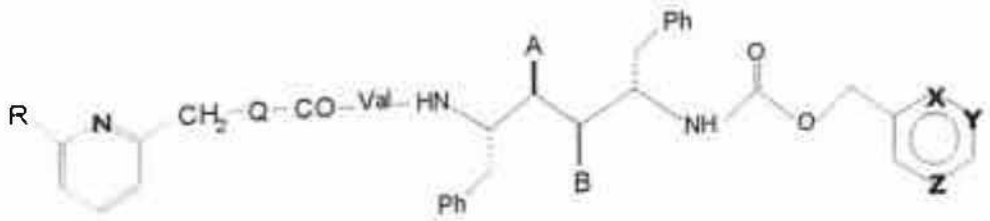
models and in humans.[31] Further studies on the analogues of A-80987 (**15-17**) led to the development of ritonavir (ABT-538) (**18**) which possesses high oral bioavailability in both animals and humans and substantially reduced rate of metabolism.[31,32] For further investigation, the analogues of ritonavir (**19**) were also studied.[32] Systematic and quantitative study was performed on the structure-activity of various analogues of A-80987 (**15-17**) and those of ritonavir (**19**) (Tables (3.15) – (3.18)) inorder to investigate the possibility of designing still better analogues. Analogues that are listed in Tables (3.15) – (3.18) have been taken from Kempf et al..[32] Kempf et al.[32] evaluated the anti-HIV activity of these compounds in terms of the ability of the compound to block the spread of HIV-1 in the immortalized human T-cell line MT4 by measuring the cytopathic effect of the virus in those cells by uptake of a tetrazolium dye. The EC_{50} values listed in



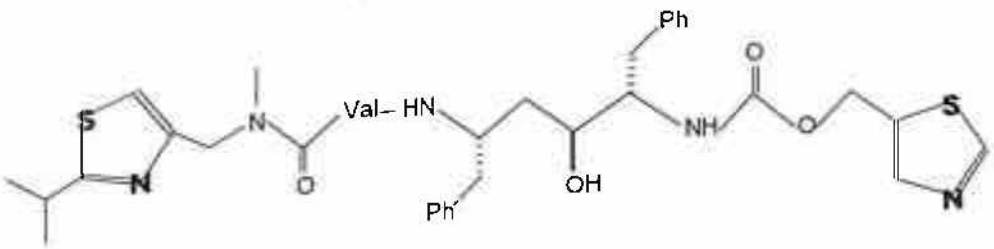
15



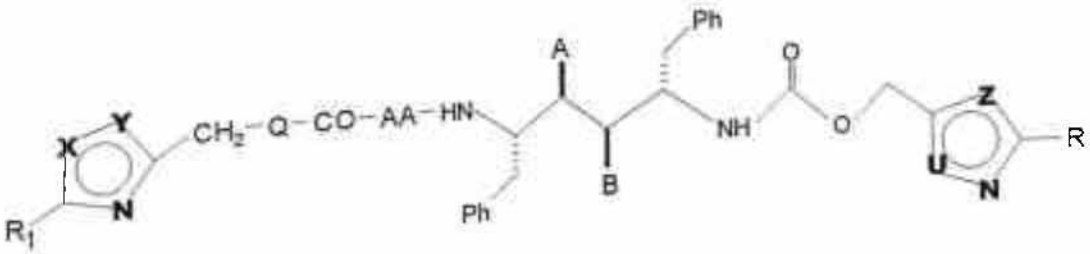
16



17



18 (ABT-538)



19

the tables refers to the molar concentration of the compounds producing 50% effect. Kempf et al.[32] had also studied the cytotoxic effects of the compounds in terms of $CCIC_{50}$, the concentration of the compound required to reduce by 50% the number of mock-infected MT4 cells.

We analyzed the structure-activity relationships of these compounds, using Fujita-Ban method,[21] (Eq. (3.34)). In Eq. (3.34), where the activity contribution of each substituent or moiety can be obtained. In this method the total activity of a compound is given by

$$\text{Activity} = \sum_i \alpha_i \chi_i + \mu \qquad (3.34)$$

α_i is the activity contribution of i^{th} substituent relative to H or the substituent defined in the parent structure, and χ_i is a parameter which takes a value of 1 or 0 depending on the presence or absence of i^{th} substituent in the molecule. The constant μ is the activity of the parent structure of the molecule.

In the present case, the substitutions have occurred within the rings and in the bridge groups also. Therefore, in each case the parent structure need to be specified:

15: $R_1 = R_2 = H$, (X = N, Y= CH), Q = O, (A = OH, B = H)

16: R = H, (X = N, Y = S, Z = CH), Q = O, (A = OH, B = H)

17: R = H, Q = O, (X–Y = S, Z = N), (A = OH, B= H)

19: R_1 = H, (X = S, Y = CH), Q = O, (Z = S, U = CH), AA = Val, (A = OH, B = H)

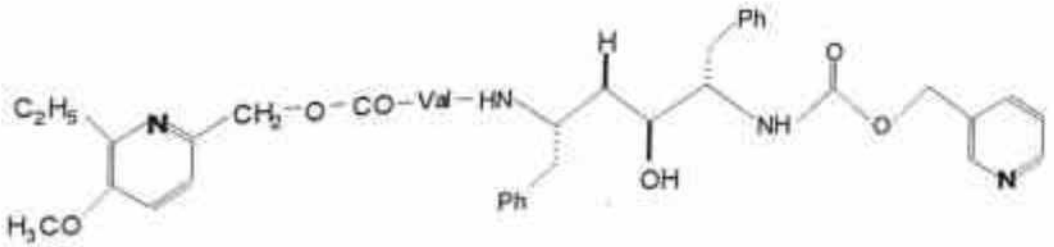
Deviations from such parent structures have been parameterized. For the substituents at the rings, referred to by R, R_1, or R_2, the values of χ as defined in Eq. (3.34) have been used. For the remaining we have defined the values of χ as follows.

1. χ is equal to 1 in all structures for A = H, B = OH and 0 for A = OH, B = H as defined in the parent structures.
2. χ is equal to 1 in all structures if Q is NH, N – alkyl, CH_2, or CH_2O. It is zero for Q = O (the parent structure).
3. For the ring of **15**, χ = 1 if X = CH, Y = N or if X = Y = N it is zero if X = N, Y = CH (parent structure).
4. For the ring of **16**, χ = 1 if X = N, Y = O, Z = CH, or if X = N, Y = CH, Z = S, or if X = S, Y = N, Z = CH or if X = CH, Y = O, Z = N, or if X = O, Y = N, Z = CH, or if X = CH Y = N, Z = $C(CH_3)$. It will be zero for X = N, Y = S, Z = CH (parent structure).
5. For the ring of **17**, χ = 1 if X–Y = CH, Z = O, or if X–Y = S, Z = N. It is zero for X–Y = CH–N, Z = N (parent structure).
6. For the rings of **19**, left hand side ring: χ = 1 if X = O, Y = CH, or if X = CH, Y = S. It is zero, if X = S, Y = CH (parent structure). For right hand side ring: χ = 1 if Z = O, U = CH, or if Z = CH, U = O. It is zero for Z = S, U = CH (parent structure).
7. In **19**, χ = 1 for AA = Ala in the bridge group. It is zero for AA = val (parent structure).

For the compounds of Table (3.15) the Fujita-Ban analysis revealed the contribution of substituents to anti-HIV activity of the compounds as shown in Table (3.19). The figures within parentheses are 95% confidence intervals. The table also reports the number of compounds (n) used in the analysis and the values obtained for the correlation coefficient (r), the standard deviation (s), and the F-statistics (F). Certain compounds as indicated in the table (Table (3.15)) were not included in the analysis as they were exhibiting aberrant behaviour.

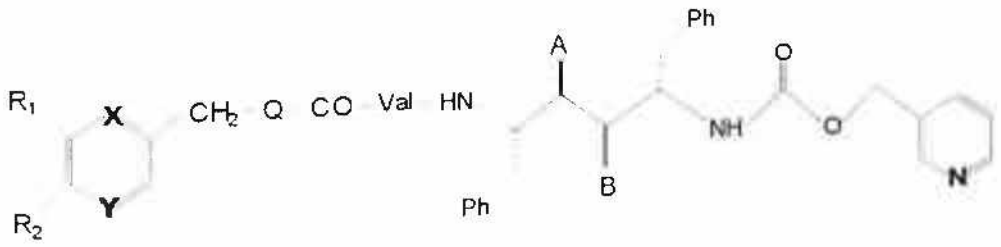
Table (3.19) shows that the activity contributions of certain substituents (indicated by asterisk) are statistically insignificant at 95% confidence intervals. Hence ignoring them, the activity contributions were reanalyzed and the results were those as shown in the next set of data in Table (3.19). These results do not appear much different from the the previous ones. The statistical parameters remain almost unchanged.

Both sets of results reveal the highest activity contributions associated with $R_1 = C_2H_5$, $R_2 = OCH_3$, and A = H, B = OH, and thus the most potent compound is predicted to be 20 with log $(1/EC_{50})$ equal to 8.30 (from set 1) or 8.18 (from set 2).



20 (pEC_{50} = 8.18)

Table (3.15) : P3 Pyridine analogues (15), their HIV-1 antiviral potency and cytotoxicity studied by Kempf et al.[12]



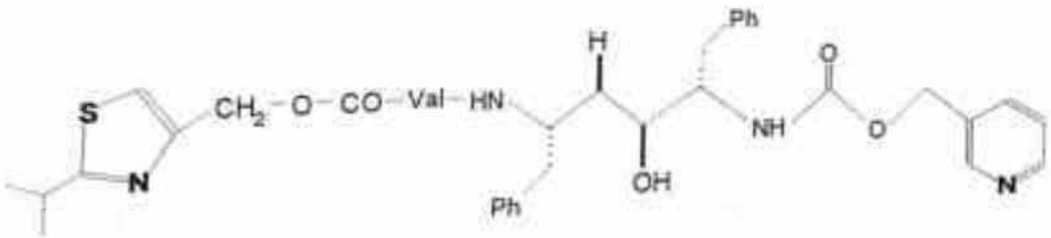
No.	R_1	R_2	X	Y	Z	A	B	log (1/EC_{50}) Obsd	log (1/EC_{50}) Calcd[a]	log (1/EC_{50}) Calcd[b]	log (1/$CCIC_{50}$) Obsd	log (1/$CCIC_{50}$) Calcd[c]
1	H	H	N	CH	O	OH	H	6.69	6.72	6.88	-	-
2	H	H	N	CH	NCH_3	H	OH	6.24	6.37	6.54	-	-
3	CH_3	H	N	CH	O	OH	H	7.00	7.00	7.04	-	-
4	CH_3	H	N	CH	O	H	OH	7.29	7.18	7.21	-	-
5	CH_3	H	N	CH	NCH_3	OH	H	6.48	6.47	6.53	4.09	4.16
6	CH_3	H	N	CH	NCH_3	H	OH	6.59	6.65	6.70	4.23	4.16
7	CH_3	H	N	CH	NH	OH	H	6.52	6.34	6.39	4.13	4.16
8	CH_3	H	N	CH	NH	H	OH	6.47	6.52	6.56	-	-
9	Et	H	N	CH	O	OH	H	7.72	7.46	7.45	4.09	4.16
10	Et	H	N	CH	O	H	OH	7.64	7.64	7.62	-	-
11	Et	H	N	CH	NCH_3	OH	H	7.17	6.93	6.94	4.24	4.16
12	Et	H	N	CH	NCH_3	H	OH	6.60	7.11	7.12	4.23	4.16
13	*i*-Pr	H	N	CH	O	OH	H	7.31	7.27	7.26	-	-
14	*i*-Pr	H	N	CH	O	H	OH	7.57	7.45	7.43	4.70	4.71
15	*i*-Pr	H	N	CH	NCH_3	OH	H	6.75	6.74	6.76	4.72	4.71
16	*i*-Pr	H	N	CH	NCH_3	H	OH	6.75	6.92	6.93	4.72	4.71
17	*t*-Bu	H	N	CH	O	OH	H	7.19	6.97	6.88	4.72	4.71
18	*t*-Bu	H	N	CH	O	H	OH	7.26	7.15	7.05	4.69	4.71
19	*t*-Bu	H	N	CH	NCH_3	OH	H	6.22	6.44	6.37	4.72	4.71
20	*t*-Bu	H	N	CH	NCH_3	H	OH	6.51	6.62	6.54	4.72	4.71
21	CH3	H	CH	N	O	OH	H	7.04	7.12	7.04	-	-
22	CH3	H	CH	N	O	H	OH	7.29	7.30	7.21	-	-
23	CH3	H	CH	N	NCH_3	OH	H	6.32	6.59	6.53	-	-
24	CH3	H	CH	N	NCH_3	H	OH	6.92	6.77	6.70	-	-
25	H	CH_3	CH	N	O	OH	H	7.06	6.99	6.88	-	-
26	H	CH_3	CH	N	O	H	OH	7.24	7.17	7.05	-	-
27	H	CH_3	CH	N	NCH_3	OH	H	6.39	6.46	6.37	-	-

Contd.....

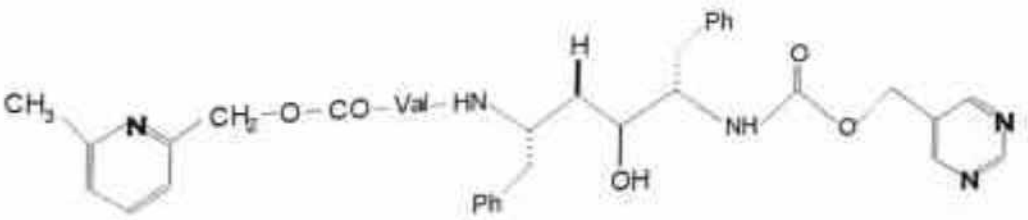
No.	R_1	R_2	X	Y	Z	A	B	log (1/EC_{50}) Obsd	log (1/EC_{50}) Calcd[a]	log (1/EC_{50}) Calcd[b]	log (1/$CCIC_{50}$) Obsd	log (1/$CCIC_{50}$) Calcd[c]
28	H	CH_3	CH	N	NCH_3	H	OH	6.89	6.64	6.54	-	-
29	H	CH_3	CH	N	NH	OH	H	6.20	6.33	6.24	-	-
30	H	OCH_3	N	CH	O	OH	H	7.41	7.39	7.44	4.19	4.16
31	H	OCH_3	N	CH	O	H	OH	7.31	7.57	7.61	4.13	4.16
32	H	OCH_3	N	CH	NCH_3	OH	H	7.00	6.86	6.94	4.22	4.16
33	H	OCH_3	N	CH	NCH_3	H	OH	7.12	7.04	7.11	4.06	4.16
34	OCH_3	H	CH	N	O	OH	H	7.22	7.16	7.15	-	-
35	OCH_3	H	CH	N	O	H	OH	7.34	7.34	7.32	-	-
36	OCH_3	H	CH	N	NCH_3	OH	H	6.57	6.63	6.64	-	-
37	OCH_3	H	CH	N	NCH_3	H	OH	6.80	6.81	6.82	-	-
38	H	OCH_3	CH	N	O	OH	H	7.27	7.51	7.44	-	-
39	H	OCH_3	CH	N	O	H	OH	7.72	7.69	7.61	4.26	4.16
40	H	OCH_3	CH	N	NCH_3	OH	H	7.20	6.98	6.94	4.24	4.16
41	H	OCH_3	CH	N	NCH_3	H	OH	7.17	7.16	7.11	4.22	4.16
42	H	NH_2	N	CH	O	OH	H	6.07	6.38	6.37	-	-
43	H	NH_2	N	CH	O	H	OH	6.16	6.56	6.54	-	-
44	H	NH_2	N	CH	NCH_3	OH	H	6.11	5.85	5.86	-	-
45	H	NH_2	N	CH	NCH_3	H	OH	6.47	6.03	6.04	-	-
46	NH_2	H	CH	N	O	OH	H	6.46	6.36	6.36	-	-
47	NH_2	H	CH	N	O	H	OH	6.43	6.53	6.53	-	-
48	H	NH_2	CH	N	O	OH	H	5.12[d]	6.78	6.37	4.05	4.16
49	H	NH_2	CH	N	O	H	OH	5.02[d]	6.68	6.54	-	-
50	H	H	N	N	O	OH	H	6.70	6.68	6.88	-	-
51	H	H	N	N	O	H	OH	6.96	6.85	7.05	-	-
52	CH_3	H	N	N	O	OH	H	7.01	6.95	7.04	-	-
53	CH_3	H	N	N	O	H	OH	7.15	7.13	7.21	4.23	4.16
54	CH_3	H	N	N	NCH_3	OH	H	5.84[d]	6.42	6.53	-	-
55	CH_3	H	N	N	NCH_3	H	OH	6.59	6.60	6.70	-	-
56	H	CH_3	N	N	O	OH	H	6.66	6.82	6.88	-	-
57	H	CH_3	N	N	O	H	OH	7.00	7.00	7.05	4.02	4.16
58	CH_3	CH_3	N	N	O	OH	H	7.15	7.10	7.04	-	-
59	CH_3	CH_3	N	N	O	H	OH	7.41	7.28	7.21	-	-
60	CH_3	CH_3	N	N	O	OH	H	6.18	6.57	6.53	-	-
61	CH_3	CH_3	N	N	O	H	OH	6.92	6.75	6.70	-	-

[a]Using the set 1 activity contributions of substituents as given in Table (3.19). [b]Using the set 2 activity contributions of substituents as given in Table (3.19). [c]Using activity contributions of Table (3.23). [d]Not used in derivation of set 1 and set 2 values of Table (3.19).

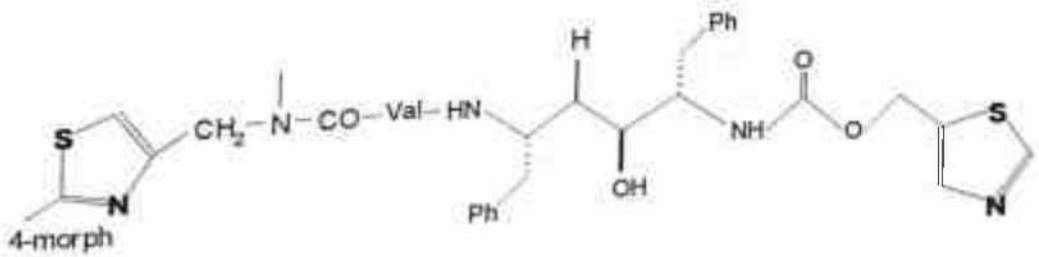
In a similar way, the activity contributions of substituents were obtained for the compounds of Tables (3.16) – (3.18) and reported in Tables (3.20) – (3.22), respectively. In each case, the most favourable substituents are bold faced. Thus in the series of Table (3.16) the most active compound is predicted to be **21**, in the series of Table (3.17), **22**, and in the series of Table (3.18), **23**. The prediction in each case has been based on the activity contributions listed in set 2 in each of Tables (3.20) – (3.22).



21 (pEC50 = 7.90)

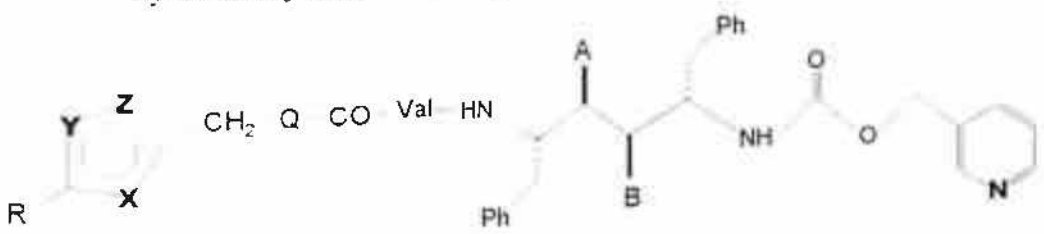


22 (pEC_{50} = 7.40)



23 (pEC50 = 8.22)

Table (3.16) : P3 Five-membered heterocyclic analogues (16), their antiviral potency and cytotoxicity data studied by Kempf et al.[32]



No.	R	X	Y	Z	Q	A	B	log (1/EC50) Obsd	log (1/EC50) Calcd[a]	log (1/EC50) Calcd[b]	log (1/CCIC50) Obsd	log (1/CCIC50) Calcd[c]
1	H	N	S	CH	NCH_3	OH	H	6.13	6.14	6.48	4.08	4.13
2	H	N	S	CH	NCH_3	H	OH	6.31	6.30	6.67	-	-
3	CH_3	N	S	CH	NCH_3	OH	H	6.68	6.63	6.48	-	-
4	CH_3	N	S	CH	NCH_3	H	OH	6.74	6.79	6.67	-	-
5	Et	N	S	CH	O	OH	H	7.19	7.17	7.09	4.23	4.26
6	Et	N	S	CH	O	H	OH	7.38	7.34	7.27	4.19	4.26
7	Et	N	S	CH	NCH_3	OH	H	6.54	6.59	6.48	4.22	4.13
8	Et	N	S	CH	NCH_3	H	OH	6.59	6.76	6.67	4.22	4.13
9	*i*-Pr	N	S	CH	O	OH	H	7.96	7.76	7.72	-	-
10	*i*-Pr	N	S	CH	O	H	OH	8.00	7.92	7.90	4.72	4.60
11	*i*-Pr	N	S	CH	NCH_3	OH	H	6.54[d]	7.45	7.11	4.49	4.47
12	*i*-Pr	N	S	CH	NCH_3	H	OH	7.54	7.35	7.29	4.33	4.47
13	*t*-Bu	N	S	CH	NCH_3	OH	H	6.43	6.43	6.48	4.72	4.75
14	*t*-Bu	N	S	CH	NCH_3	H	OH	7.54[d]	6.86	6.66	4.77	4.75
15	*i*-Pr	N	O	CH	O	OH	H	7.75	7.74	7.72	-	-
16	*i*-Pr	N	O	CH	O	H	OH	7.70	7.91	7.90	4.28	4.32
17	*i*-Pr	N	O	CH	NCH_3	OH	H	7.48	7.17	7.11	4.21	4.19
18	*i*-Pr	N	O	CH	NCH_3	H	OH	7.22	7.33	7.29	4.20	4.19
19	*i*-Pr	N	S	CH	CH_2	OH	H	6.80	6.90	6.89	4.36	4.36
20	*i*-Pr	N	S	CH	CH_2	H	OH	7.17	7.07	7.08	4.00[e]	4.34
21	$MeOCH_2$	N	S	CH	O	OH	H	7.02	7.22	7.09	-	-
22	$MeOCH_2$	N	S	CH	O	H	OH	7.12	7.38	7.27	-	-
23	$MeOCH_2$	N	S	CH	NCH_3	OH	H	6.48	6.64	6.48	4.22	4.13
24	$MeOCH_2$	N	S	CH	NCH_3	H	OH	7.12	6.80	6.67	4.19	4.13
25	$(Me)_2N$	N	S	CH	O	OH	H	7.05	7.05	7.09	4.28	4.26
26	4-morph	N	S	CH	O	OH	H	7.48	7.40	7.09	4.54[e]	4.26

Contd.....

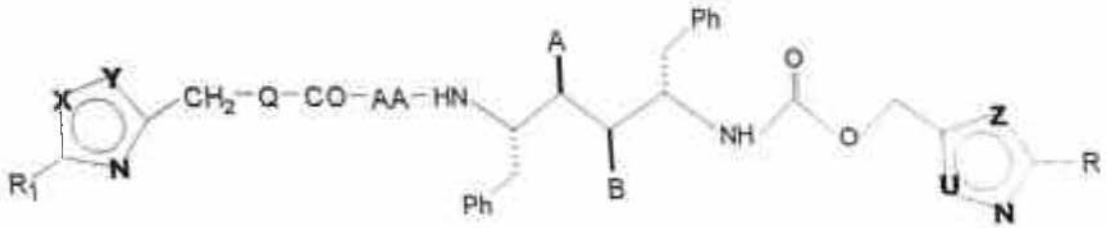
No.	R	X	Y	Z	Q	A	B	log (1/EC_{50})			log (1/$CCIC_{50}$)	
								Obsd	Calcd[a]	Calcd[b]	Obsd	Calcd[c]
27	4-morph	N	S	CH	O	H	OH	7.48	7.56	7.27	4.22	4.26
28	H	N	CH	S	O	OH	H	7.43	7.08	7.09	-	-
29	H	N	CH	S	O	H	OH	6.89	7.24	7.27	-	-
30	Et	S	N	CH	O	OH	H	6.54	6.63	6.70	4.18	4.26
31	Et	S	N	CH	O	H	OH	7.04	6.79	6.88	-	-
32	*i*-Pr	S	N	CH	O	OH	H	7.00	7.22	7.33	-	-
33	*i*-Pr	S	N	CH	O	H	OH	7.12	7.38	7.51	-	-
34	H	O	N	CH	O	OH	H	6.27	6.39	6.37	-	-
35	H	O	N	CH	O	H	OH	6.66	6.55	6.56	-	-
36	$MeOCH_2$	S	N	CH	O	OH	H	6.75	6.67	6.70	-	-
37	$MeOCH_2$	S	N	CH	O	H	OH	7.07	6.84	6.88	-	-
38	H	S	N	$C(CH_3)$	CH_2O	OH	H	6.96	6.94	6.70	-	-
39	H	S	N	$C(CH_3)$	CH_2O	H	OH	7.09	7.11	6.88	4.22	4.26
40	*t*-Bu	CH	O	N	O	OH	H	7.17	6.91	7.09	-	-
41	*t*-Bu	CH	O	N	O	H	OH	7.14	7.07	7.27	4.72	4.50
42	*t*-Bu	CH	O	N	NCH_3	OH	H	6.00	6.33	6.48	4.25	4.37
43	*t*-Bu	CH	O	N	NCH_3	H	OH	7.32[d]	6.76	6.67	4.26	4.37
44	CH_3O	CH	N	O	O	OH	H	6.66	6.79	7.09	4.22	4.26
45	CH_3O	CH	N	O	O	H	OH	7.08	6.95	7.27	-	-

[a]Using the set 1 activity contributions of substituents as given in Table (3.20). [b]Using the set 2 activity contributions of substituents as given in Table (3.20). [c]Using activity contributions of Table (3.24). [d]Not used in derivation of set 1 and set 2 values of Table (3.20). [e]Not used in the derivation of activity contributions of Table (3.24).

No.	R	Q	X–Y	Z	A	B	Obsd	Calcd[a]	Obsd (μm)
1	H	O	CH–N	N	OH	H	6.60	6.59	>100
2	H	NCH_3	CH–N	N	OH	H	6.17	6.27	>100
3	H	NCH_3	CH–N	N	H	OH	6.35	6.27	>100
4	H	O	CH	O	OH	H	6.64[b]	6.28	>100
5	H	NCH_3	CH	O	OH	H	5.97	5.96	58
6	CH3	O	CH	O	OH	H	7.05	7.09	>100
7	CH3	O	CH	O	H	OH	7.13	7.09	>100
8	H	O	S	N	OH	H	6.28	6.28	>100
9	H	O	S	N	H	OH	6.26	6.28	>100

[a]Using the set 2 activity contributions of substituents as given in Table (3.21).
[b]Not included in derivation of activity contributions of Table (3.21)

Table (3.18) : Ritonavir analgoues (19), their antiviral potency and cytotoxicity studied by Kempf et al.[32]



No.	R_1	X	Y	Q	AA	Z	U	R_2	A	B	log (1/EC_{50})			log (1/$CCIC_{50}$)	
											Obsd	Calcd[a]	Calcd[b]	Obsd	Calcd[c]
1	*i*-Pr	S	CH	O	Val	S	CH	H	OH	H	8.00[d]	6.04	6.13	4.00[e]	4.55
2	*i*-Pr	S	CH	O	Val	S	CH	H	H	OH	8.31[d]	6.80	6.91	4.29	4.55
3	*i*-Pr	S	CH	NCH_3	Val	S	CH	H	OH	H	6.72	6.49	6.60	4.25	4.41
4	*i*-Pr	S	CH	NCH_3	Val	S	CH	H	H	OH	7.60	7.25	7.39	4.24	4.41
5	*i*-Pr	O	CH	NCH_3	Val	S	CH	H	OH	H	6.01	6.43	6.60	4.24	4.42
6	*i*-Pr	O	CH	NCH_3	Val	S	CH	H	H	OH	7.37	7.20	7.39	4.22	4.42
7	*i*-Pr	S	CH	NCH_3	Val	O	CH	H	H	OH	7.89	7.89	7.39	4.22	4.41
8	*i*-Pr	O	CH	NCH_3	Val	O	CH	H	H	OH	<	-	-	-	-
9	CH3	S	CH	NCH_3	Val	S	CH	H	H	OH	6.10[d]	7.17	7.39	4.21	4.23
10	CH3	O	CH	NCH_3	Val	S	CH	H	H	OH	7.11	7.11	7.39	-	-
11	Et	S	CH	NCH_3	Val	S	CH	H	OH	H	6.55	6.70	6.60	4.20	4.23
12	Et	S	CH	NCH_3	Val	S	CH	H	H	OH	7.28	7.47	7.39	4.23	4.23
13	Et	O	CH	NCH_3	Val	S	CH	H	H	OH	7.75	7.41	7.39	4.00	4.03

Contd.....

No.	R_1	X	Y	Q	AA	Z	U	R_2
14	3-pent	S	CH	NCH_3	Val	S	CH	H
15	*c*-Pr	S	CH	NCH_3	Val	S	CH	H
16	*c*-Pr	S	CH	NCH_3	Val	S	CH	H
17	*c*-Bu	S	CH	NCH_3	Val	S	CH	H
18	*c*-Bu	S	CH	NCH_3	Val	S	CH	H
19	4-morph	S	CH	O	Val	S	CH	H
20	4-morph	S	CH	O	Val	S	CH	H
21	*i*-Pr	CH	S	O	Val	S	CH	H
22	*i*-Pr	CH	S	O	Val	S	CH	H
23	*i*-Pr	S	CH	NEt	Val	S	CH	H
24	*i*-Pr	S	CH	NEt	Val	S	CH	H
25	*i*-Pr	S	CH	N*c*Pr	Val	S	CH	H
26	*i*-Pr	S	CH	N*c*Pr	Val	S	CH	H
27	*i*-Pr	S	CH	O	Ala	S	CH	H
28	*i*-Pr	S	CH	O	Ala	S	CH	H
29	*i*-Pr	S	CH	NCH_3	Ala	S	CH	
30	*i*-Pr	S	CH	NCH_3	Ala	S	CH	
31	*i*-Pr	O	CH	NCH_3	Ala	S	CH	

A	B	log (1/EC_{50})			log (1/$CCIC_{50}$)	
		Obsd	Calcd[a]	Calcd[b]	Obsd	Calcd[c]
H	OH	6.85	6.85	7.39	4.72[c]	4.23
OH	H	6.82	6.69	6.60	4.35	4.23
H	OH	7.33	7.46	7.39	4.32	4.23
OH	H	6.64	6.64	6.60	4.70	4.65
H	OH	7.41	7.41	7.39	4.59	4.65
OH	H	7.10	6.97	6.96	-	-
H	OH	7.60	7.73	7.74	4.22	4.37
OH	H	5.96	6.02	6.13	4.44	4.55
H	OH	6.85	6.79	6.91	4.66	4.55
OH	H	5.85	6.42	6.13	4.85	4.55
H	OH	6.60	7.19	6.91	4.62	4.55
H	OH	6.55	6.26	6.26	4.72	4.55
OH	H	5.57	5.50	5.49	4.70	4.55
OH	H	5.77	5.97	6.13	4.20	4.34
H	OH	7.00	6.73	6.91	4.25	4.34
OH	H	5.87	6.42	6.60	4.22	4.20
H	OH	7.41	7.18	7.39	4.28	4.20
H	OH	7.23	7.12	7.39	4.00	4.00

Contd.....

No.	R_1	X	Y	Q	AA	Z	U	R_2	A	B	log (1/EC_{50})			log (1/$CCIC_{50}$)	
											Obsd	Calcd[a]	Calcd[b]	Obsd	Calcd[c]
32	*I*-Pr	S	CH	NEt	Ala	S	CH	H	OH	H	7.05	6.35	6.13	-	-
33	*i*-Pr	S	CH	NEt	Ala	S	CH	H	H	OH	7.86	7.11	6.91	4.28	4.34
34	*i*-Pr	S	CH	NPr	Ala	S	CH	H	H	OH	6.85	6.99	6.91	4.72	4.60
35	*i*-Pr	S	CH	N*i*Bu	Ala	S	CH	H	H	OH	6.75	6.90	6.91	4.77	4.81
36	*i*-Pr	S	CH	N*n*Bu	Ala	S	CH	H	H	OH	6.54	6.54	6.91	4.77	4.77
37	*i*-Pr	S	CH	NcPr	Ala	S	CH	H	OH	H	4.85	5.42	5.49	4.35	4.34
38	*i*-Pr	S	CH	NcPr	Ala	S	CH	H	H	OH	6.26	6.18	6.26	4.39	4.34
39	*i*-Pr	S	CH	NCH_3	β-Ala	S	CH	H	OH	H	5.66	5.36	5.41	4.24	4.19
40	*i*-Pr	S	CH	NCH_3	β-Ala	S	CH	H	H	OH	5.70	6.12	6.20	4.26	4.19
41	*i*-Pr	S	CH	NcPr	β-Ala	S	CH	H	OH	H	4.87	4.36	4.29	4.23	4.33
42	*i*-Pr	S	CH	NcPr	β-Ala	S	CH	H	H	OH	4.75	5.13	5.08	4.32	4.33
43	*i*-Pr	S	CH	NEt	Gly	S	CH	H	H	OH	5.68	5.98	5.87	4.39	4.31
44	*i*-Pr	S	CH	NPr	Gly	S	CH	H	H	OH	6.00	5.86	5.87	4.46	4.58
45	*i*-Pr	S	CH	N*i*Bu	Gly	S	CH	H	H	OH	5.92	5.77	5.87	4.82	4.78
46	*i*-Pr	S	CH	CH_2O	Val	S	CH	H	H	OH	7.00	7.00	6.91	-	-
47	*i*-Pr	S	CH	CH_2O	Val	S	CH	H	OH	H	7.77[d]	6.24	6.13	4.60	4.55
48	*i*-Pr	S	CH		Val	S	CH	H	OH	H	6.09	6.04	6.13	-	-
49	*i*-Pr	S	CH		Val	S	CH	H	H	OH	6.68	6.80	6.91	-	-
50	*i*-Pr	S	CH	CH_2NMe	Val	S	CH	H	H	OH	6.74	6.95	6.91	4.60	4.55

Contd.....

No.	R_1	X	Y	Q	AA	Z	U	R_2	A	B	log (1/EC_{50})			log (1/$CCIC_{50}$)	
											Obsd	Calcd[a]	Calcd[b]	Obsd	Calcd[c]
51	*i*-Pr	S	CH	CH_2NEt	Val	S	CH	H	H	OH	6.26	6.02	5.98	4.75	4.79
52	*i*-Pr	S	CH	CH_2NPr	Val	S	CH	H	H	OH	5.96	5.79	5.75	4.75	4.86
53	*i*-Pr	S	CH	CH_2NMe	Ala	S	CH	H	H	OH	7.09	6.88	6.91	4.23	4.34
54	*i*-Pr	S	CH	CH_2NEt	Ala	S	CH	H	H	OH	5.70	5.94	5.98	4.62	4.58
55	*i*-Pr	S	CH	CH_2NPr	Ala	S	CH	H	H	OH	5.54	5.71	5.75	4.75	4.64
56	*i*-Pr	S	CH	NCH_3	Val	S	CH	CH_3	OH	H	7.13	6.66	6.60	4.66	4.71
57	*i*-Pr	S	CH	NCH_3	Val	S	CH	CH_3	H	OH	6.96	7.43	7.39	4.75	4.71
58	*i*-Pr	S	CH	NCH_3	Val	S	CH	*i*-Pr	OH	H	5.82	5.88	5.87	4.72	4.72
59	*i*-Pr	S	CH	NCH_3	Val	S	CH	*i*-Pr	H	OH	6.70	6.64	6.65	4.72	4.72
60	*i*-Pr	S	CH	NCH_3	Val	CH	O	H	H	OH	8.05	7.85	7.39	4.55	4.41
61	*i*-Pr	O	CH	NCH_3	Val	CH	O	H	H	OH	7.92	7.79	7.39	4.24	4.22
62	*i*-Pr	O	CH	NCH_3	Ala	CH	O	H	H	OH	7.38	7.71	7.39	-	-

[a]Using the set 1 activity contributions of substituents as given in Table (3.22). [b]Using the set 2 activity contributions of substituents as given in Table (3.22). [c]Using activity contributions of substituents as given in Table (3.25). [d]Not used in the derivation of set 1 and set 2 values of Table (3.22). [e]Not included in the derivation of activity contributions of Table (3.25).

Table (3.19): Activity contributions of substituents of analogues of **15** (Table (3.15)).
The parent structure is defined as:

$R_1 = R_2 = H$, (X = N, Y = CH), Q = O, (A = OH, B = H)

R_1	R_2	X , Y	Q	A and B	μ
CH_3 = 0.279(±0.180)	CH_3* = 0.146(±0.196)	(X=CH, Y=N)* = 0.120(±0.182)	NCH_3 = −0.531(±0.118)	(A=H, B=OH) = 0.179(±0.109)	6.724
C_2H_5 = 0.735(±0.284)	OCH_3 = 0.667(±0.227)	(X=N, Y=N)* = −0.049(±0.215)	NH = −0.661(±0.276)		
i-Pr = 0.547(±0.284)	NH_2 = −0.345(±0.284)				
t-Bu* = 0.247(±0.284)					
NH_2 = −0.488(±0.354)					
OCH_3 = 0.315(±0.286)					
			n = 58, r = 0.914, s = 0.178, F = 15.54		
Activity contributions ignoring statistically insignificant substituents					
CH_3 = 0.157(±0.153)	OCH_3 = 0.563(±0.189)		NCH_3 = −0.506(±0.119)	(A=H, B=OH) = 0.171(±0.113)	6.879
C_2H_5 = 0.571(±0.242)	NH_2 = −0.509(±0.241)		NH = −0.645(±0.266)		
i-Pr = 0.383(±0.242)					
NH_2 = −0.520(±0.326)					
OCH_3 = 0.271(±0.242)					
			n = 58, r = 0.897, s = 0.193, F = 19.425		

* Statistically insignificant substituents

R	X, Y, Z	Q	A, B	μ
C_2H_5 = 0.456(±0.414)†	(X = S, Y = N, Z = CH) = –0.548(±0.263)	NCH_3 = –0.579(±0.228)	(A = H, B = OH) = 0.166(±0.152)	6.984
i-Pr = 1.043(±0.434)	(X = O, Y = N, Z = CH) = –0.334(±0.528)	CH_2 = –0.856(±0.419)		
t-Bu* = 0.293(±0.588)	(X = S, Y = N, Z = $C(CH_3)_2$ = 0.772(±0.549)†			
$MeOCH_2$ = 0.503(±0.414)†	(X = CH, Y = N, Z = O)* = 0.071(±0.528)			
4-morpho = 0.681(±0.528)†	(X = N, Y = CH, Z = S)* = 0.361(±0.528)			
$N(CH_3)_2$* = 0.334(±0.631)	(X = N, Y = O, Z = CH)* = –0.015(±0.335)			
CH_3 = 0.490(±0.476)†	(X = CH, Y = O, Z = N)* = –0.101(±0.572)	n = 42, r = 0.926, s = 0.176, F = 8.526		
Activity contributions ignoring statistically insignificant substituents				
i-Pr = 0.628(±0.200)	(X = S, Y = N, Z = CH) = –0.392(±0.224)	NCH_3 = –0.607(±0.192)	(A = H, B = OH) = 0.182(±0.162)	7.090
	(X = O, Y = N, Z = CH) = –0.716(±0.393)	CH_2 = –0.824(±0.419)		
		n = 42, r = 0.862, s = 0.237, F = 16.842		

* Statistically insignificant substituents. † Consistently not sifnificant.

Table (3.21) : Activity contributions of substituents of analogues of 17 (Table (3.17)).
The parent structure is defined as:

R = H, (X–Y = S, Z = N), Q = O, (A = OH, B = H)

R	X–Y, Z	Q	A, B	μ
CH_3 = 0.700(±0.583)	(X–Y = CH–N, Z = N) = 0.370(±0.393)	NCH_3 = –0.380(±0.393)	(A= H, B = OH) = 0.080(±0.248)	6.230
	(X–Y = CH, Z = O)* = 0.120(±0.583)	n = 8, r = 0.996, s = 0.038, F = 48.528		
Activity contributions ignoring statistically insignificant substituents				
CH_3 = 0.813(±0.189)	(X–Y = CH–N, Z = N) = 0.310(±0.171)	NCH_3 = –0.320(±0.171)		6.277
		n = 8, r = 0.992, s = 0.054, F = 79.673		

* Statistically insignificant

Table (3.22) : Activity contributions of substituents of analogues of 19 (Table (3.18)).
The parent structure is defined as:

$R_1 = H$, (X = S, Y = CH), Q = O, (A = OH, B = H), (Z = S, U=CH), AA = Val

R_1	X, Y	AA	Q	Z, U	R_2	A, B	μ
i-Pr = 0.403(±0.925)*	(X=CH, Y=S) = −0.018(±0.739)*	Ala = −0.075(±0.315)*	NCH_3 = 0.449(±0.543)	(Z = CH,U = O) † = 0.594(±0.591)	CH_3= 0.173(±0.698)*	(A = H, B = OH) = 0.763(±0.273)	5.638
C_2H_5 = 0.617(±0.981)*	(X=O, Y=CH) = −0.058(±0.470)*	β-Ala = −1.129(±0.522)	NEt = 0.381(±0.573)*	(Z = O, U = CH)* = 0.637(±0.925)	i-Pr = −0.612(±0.698)		
c-Pr = 0.607(±1.029)*		Gly = −1.210(±0.671)	NcPr = −0.546(±0.563)				
c-Bu = 0.557(±1.029)*			NPr = 0.264(±0.856)*				
CH_3= 0.318(±1.269)*			N*i*Bu=0.174(±0.801)*				
4-morpho = 1.331(±1.164)			N*n*Bu=−0.189(±0.953)*				
			CH_2O=0.196(±0.956)*				
			CH_2NMe = 0.148(±0.734)*				
			CH_2NEt = −0.787(±0.734)				
			CH_2NPr = −1.017(±0.734)	n = 57, r = 0.929, s = 0.299, F = 7.215			

...Contd.,

R_1	X, Y	AA	Q	Z, U	R_2	A, B	μ
Activity contributions ignoring statistically insignificant substituents							
4-morpho = 0.830(±0.581)		β-Ala = −1.188(±0.435) Gly = −1.046(±0.492)	NCH_3 = 0.475(±252) NcPr = −0.649(±0.399) CH_2NEt = −0.933(±0.584) CH_2NPr = −1.163(±0.584)		i-Pr = −0.735(0.571)	(A = H, B = OH) = 0.786(±0.231)	6.127
				n = 57, r = 0.899, s = 0.352, F = 22.029			

* Statistically not significant. † Consistently not significant.

Table (3.23) : Cytotoxic activity contributions of statistically significant substituents of analogues of **15** (Table (3.15)). The parent structure is defined as:

$R_1 = R_2 = H$, (X = N, Y = CH), Q = O, (A = OH, B = H)

R_1	μ
i-Pr = 0.551(±0.093) t-Bu = 0.548(±0.083)	4.163
n = 23, r = 0.968, s = 0.068, F = 146.47	

Table (3.24) : Cytotoxic activity contributions of statistically significant substituents of analogues of **16** (Table 3.16). The parent structure is defined as:

R = H, (X = N, Y = S, Z = CH), Q = O, (A = OH, B = H)

R	X, Y, Z	Q	μ
i-Pr = 0.341(±0.134)	(X = N, Y = O, Z = CH) = −0.283(±0.167)	NCH_3 = −0.135(±0.092)	4.262
t-Bu = 0.618(±0.165)	(X = CH, Y = O, Z = N) = −0.380(±0.190)	CH_2 = −0.244(±0.244)	
	n = 24, r = 0.910, s = 0.083, F = 13.611		

Table (3.25): Cytotoxic activity contributions of statistically significant substituents of analogues of 19 (Table (3.18)). The parent structure is defined as:

R_1 = H, (X = S, Y = CH), Q = O, (A = OH, B = H), (Z = S, U=CH), AA = Val

R_1	X, Y	AA	Q	R_2	μ
c-Bu = 0.419(±0.199)	(X = O, Y = CH) = −0.193(±0.133)	Ala = −0.215(±0.094)	NCH_3 = −0.139(±0.098)	CH_3= 0.291(±0.196)	4.365
i-Pr = 0.187(±0.119)		β-Ala = −0.221(±0.139)	N*i*Bu=0.471(±0.203)	i-Pr = 0.305(±0.184)	
		Gly = −0.241(±0.188)	N*n*Bu= 0.432(±0.259)		
			NPr = 0.266(±0.203)		
			CH_2NEt = 0.240(±0.184)		
			CH_2NPr = 0.305(±0.184)		
			n = 52, r = 0.901, s = 0.103, F = 11.360		

The calculation shows that except **22** , the other three compounds predicted (**20**, **21**, **23**) may have the activity higher than ritonavir (**18**, pEC_{50} = 7.60) (compound **4** in Table (3.18), the compound which has been licensed for use. All the four compounds are predicted to have much higher activity than **13** and **14**, which were found by Kempf et al.[29,30] to possess adequate anti-HIV activity (pEC_{50} = 6.60 for each).

We also analyzed from Fujita-Ban approach the cytotoxic activity ($CCIC_{50}$) of the compounds of all the tables, except those of Table (3.17), where the $CCIC_{50}$ values reported for all compounds except one (**5**, 58μM) were uncertain (>100 μM for all). The activity contributions of substituents of different series of compounds are listed in Tables (3.23) – (3.25). Only statistically significant contributions are reported. From these data, the $pCCIC_{50}$ values for compounds **20**, **21**, and **23**, belonging to the series of Tables (3.15), (3.16), and (3.18), respectively are found to be 4.16, 4.60 and 4.36, respectively. If we compare the toxicity of **23**, which can be a member of ritonavir series (Table (3.18)), with that of ritonavir $pCCIC_{50}$ = 4.24 (**4**, Table (3.18)), both are found to be essentially same, but the anti-HIV activity of **23** (8.22) is predicted to be much higher than that of ritonavir (7.60). Therefore compound **23** can be synthesized are tried. Compound **21** can also be a better choice, as its anti-HIV activity is slightly higher than that of ritonavir and the toxicity slightly lower than that of ritonavir.

CONCLUSIONS

The following conclusions have been drawn from the QSAR studies:

1. The HIV-1-PR has four hydrophobic pockets near its active sites and the favorable hydrophobic interactions with these pockets are desirable for an inhibitor to achieve nanomolar potency.

2. Both antiviral and enzyme inhibition activities of cyclic ureas as well as cyclic cyanoguanidines are the function of hydrophobic property of P2/P2' substituents. These substituents can be of further advantage if they contain OH– or NH_2– like hydrogen bond donor groups. However, a cyclic urea derivative is found to have better antiviral potency or PR-inhibition activity than corresponding cyclic cyanoguanidine derivative.

3. HIV-1 protease is capable of forming multiple hydrogen bonds with the inhibitors.

4. The confirmational flexibility and the lipophilicity of inhibitor molecules also play important roles in the inhibition phenomenon.

5. The optimal size of cycloalkyl ring in cycloalkylpyranones should be 8-membered. This ring seems to have the maximal hydrophobic interaction with the receptor.

6. In cycloalkylpyranones, enzyme binding affinity of the compounds would be favoured by a cyclooctyl ring, a 3-cyclopropylphenylmethyl substituent at the pyranone ring, and 4-CN-2-pyridine, an N-Me-imidazole-, or a 3- or 4-CN-phenyl-sulfonamide group at the meta position of the phenyl ring of the 3- substituent.

7. In cyclic sulfolanes, most potent compound with a predicted activity of 8.67 can be obtained by changing the 2-substituent of the compound **25** of Table (3.11) with optimum π value 1.32, and its six membered A ring to a five-membered one. Thus a slightly a more potent compound than **26**, the most potent compound listed in Table (3.11), can be expected. The 2-substituent is found to have strong hydrophobic interaction with the receptor.

8. A Fujita-Ban type analysis made on ritonavir analogues leads us to predict the anti-HIV activity of **23** (8.22) to be much higher than that of ritonavir (7.60) (4, Table (3.18)), the compound which has been licenced for use. A few more compounds are also found to be worth trying.

REFERENCES

1. Kohl, N.E., Emini, N.A., Schleif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A.F., Scolnick, E.M. and Sigal, I.S. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 4686 (1988).

2. Seelmeier, S., Schmidt, H., Turk, V. and von der Helm, K. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 6612 (1988).

3. Wlodawer, A. and Erickson, J.W. *Annu. Rev. Biochem.* **62**, 543 (1993).

4. Applet, K. *Perspec. Drug Discovery Des.* **1**, 23 (1993)

5. Boehme, R.E., Borthwick, A.D. and Wyatt, P.G. *Annu. Rep. Med. Chem.*, **30**, 139 (1995).

6. Romines, K.R. and Thaisrivongs, S. *Drugs Future*, **20**, 377 (1995).

7. Olson, G.L., Bolin, D.R., Bonner, M.P., Bos, M., Cook, C.M., Fry, D.C., Graves, B.J., Hatada, M., Hill, D.E., Kahn, M., Madison, V.S., Rusiecki, V.K., Sarabu, R., Sepinwall, J., Vincent, G.P. and Voss, M.E. *J. Med. Chem.*, **36**, 3039 (1993).

8. Maligres, P.E., Upadhyay, V., Rossen, K., Cianciosi, S.J., Purick, R.M., Eng, K.K., Reamer, R.A., Askin, D., Volante, R.P. and Reider, P.J. *Tetrahedron Lett.*, **36**, 2195 (1995).

9. Jacobson, H., Yasargil, K., Winslow, D.L., Craig, J.C., Krohn, A., Duncan, I.B. and Mous, J. *Virology*, **206**, 527 (1995).

10. Lam, P.Y.S., Jadhav, P.K., Eyermann, C.J., Hodge, C.N., Ru, Y., Bacheler, L.T., Meek, J.L., Otto, M.J., Raymer, M.M., Wong, N.Y., Chang, C.-H., Wever, P.C., Jackson, D.A., Sharpe, T.R. and Erickson-Viitanen, S. *Science*, **263**, 380 (1994).

11. Nugiel, D.A., Jacobs, K., Worley, T., Patel, M., Kaltenbach III, R.F., Meyer, D.T., Jadhav, P.K., DeLucca, G.V., Smyser, T.E., Klabe, R.M., Bacheler, L.T., Rayner, M.M. and Seith, S.P. *J. Med. Chem.*, 39, 2156 (1996).

12. Lam, P.Y.S., Ru, Y., Jadhav, P.K., Aldrich, P.E., DeLucca, G.V., Eyermann, C.J., Chang, C.-H., Emmett, G., Holler, E.R., Danekar, W.F., Li, L., Confalone, P.N., McHugh, R.J., Han, Q., Li, R., Markwader, J.A., Seitz, S.P., Sharpe, T.R., Bacheler, L.T., Rayner, M.M., Klabe, R.M., Shum, L., Winslow, D.L., Kornhauser, D.M., Jackson, D.A., Erickson-Viitanen, S. and Hodge, C.N. *J. Med. Chem.*, 39, 3514 (1996).

13. Jadhav, P.K., Woerner, F.J., Lam, P.Y.S., Hodge, C.N., Eyermann, C.J., Man, H.-W., Danekar, W.F., Bacheler, L.T., Rayner, M.M., Meek, J.L., Erickson-Viitanen, S., Jackson, D.A., Calabrese, J.C., Schadt, M. and Chang, C.-H. *J. Med. Chem.*, **41**, 1446 (1998).

14. Ala, P.J., Huston, E.E., Klabe, R.M., McCabe, D.D., Duke, J.L., Rizzo, C.J., Korant, B.E., De Loskey, J.R., Lam, P.Y.S., Hodge, C.N. and Chang, C.-H. *Biochemistry*, **36**, 1573 (1997).

15. Thaisrivongs, S., Tomich, P.K., Watenpaugh, K.D., Chong, K.-T., Howe, W.J., Yang, C.-P., Strohbach, J.W., Turner, S.T., McGrath, J.P., Bohanon, M.J., Lynn, J.C., Mulichak, A.M., Spinelli, P.A., Hinshaw, R.R., Pagano, P.J., Moon, J.B., Ruwart, M.J., Wilkinson, K.F., Rusk, B.D., Zipp, G.L., Dalga, R.J., Schwende, Z., Koeplinger, K.A., Kakuk, T.J., Cole, S.L., Zaya, R.M., Piper, R.C. and Jeffry, P. *J. Med. Chem.*, **37**, 3200 (1994).

16. Romines, K.R., Watenpaugh, K.D., Tomich, P.K., Howe, W.J., Morris, J.K., Lovasz, K.D., Mulichak, A.M., Finzel, B.C., Lynn, J.C., Horng, M.-M., Schwende, F.J., Ruwart, M.J., Zipp, G.L., Chong, K.-T., Dolak, L.A., Toth, L.N., Howard, G.M., Rush, B.D., Wilkinson, K.F., Possert, P.L., Dalga, R.J. and Hinshaw, R.R. *J. Med. Chem.*, **38**, 1884 (1995).

17. Romines, K.R., Watenpaugh, K.D., Howe, W.J., Tomich, P.K., Lovasz, K.D., Morris, J.K., Janakiraman, M.N., Lynn, J.C., Horng, M.-M., Chong, K.-T., Hinshaw, R.R. and Dolak, L.A. *J. Med. Chem.*, **38**, 4463 (1995).

18. Romines, K.T., Morris, J.K., Howe, W.J., Tomich, P.K., Horng, M.-M., Chong, K.-T., Hinshaw, R.R., Anderson, D.J., Strohbach, J.W., Turner, S.R. and Mizsak, S.A. *J. Med. Chem.*, **39**, 4125 (1996).

19. Skulnick, H.I., Johnson, P.D., Aristoff, P.A., Morris, J.K., Lovaszk, K.D., Howe, W.J., Watenpaugh, K.D., Janakiraman, M.N., Anderson, D.J., Reischer, R.J., Schwartz, T.M., Banitt, L.S., Tomich, P.K., Lynn, J.C., Horng, M.-M., Chong, K.-T., Hinshaw, R.R., Dolak, L.A., Seest, E.P., Schwende, F.J., Rush, B.D., Howard, G.M., Toth, L.N., Wilkinson, K.R., Kakuk, T.J., Johnson, C.W., Cole, S.L., Zaya, R.M., Zipp, G.L., Possert, P.L., Dalga, R.J., Zhong, W.-Z., Williams, M.G. and Romines, K.R. *J. Med. Chem.*, **40**, 1149 (1997).

20. Hansch, C. and Fujita, T. *J. Am. Chem. Soc.*, **86**, 1616 (1964).

21. Fujita, T. and Ban, T. *J. Med. Chem.*, **14**, 148 (1971).

22. Verloop, A. *The STERIMOL Approach to Drug Design*, Marcel Dekker: New York (1987). Data taken from Hansch, C., Leo, A. and Hoekman, D. *Exploring*

QSAR: Hydrophobic, Electronic, and Steric Constants. American Chemical Society: Washington, D.C. (1995).

23. Thaisrivongs, S., Janakiraman, M.N., Chong, K.-T., Tomich, P.K., Dolak, P.K., Turner, S.R., Strohbach, J.W., Lynn, J.C., Horng, M.-M., Hinshaw, R.R. and Watenpaugh, K.D. *J. Med. Chem.*, **39**, 2400 (1996).

24. Ghosh, A.K., Lee, H.Y., Thompson, W.J., Culberson, C., Holloway, K., McKee, S.P., Munson, P.M., Duong, T.T., Smith, A.M., Darke, P.L., Zugay, J.A., Emini, E.A., Schleif, W.A., Huff, J.R. and Anderson, P.S. *J. Med. Chem.*, **37**, 1177.

25. Vara Prasad, J.V.N., Para, K.S., Tummino, P.J., Ferguson, D., McQuade, T.J., Lunney, E.A., Rapundalo, S.T., Batley, L.B., Hingorani, G., Domagala, J.M., Gracheck, S.J., Bhat, T.N., Liu, B., Baldwin, E.T., Erickson, J.W., Sawyer, T.K. *J. Med. Chem.*, **38**, 898 (1995).

26. Bernstein, F.C., Koetzle, T.F., William, G.T.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. *Arch. Biochem. Biophys.*, **185**, 584 (1978).

27. Wang, S., Milne, G.W.A., Yan, X., Posey, I.J., Nicklaus, M.C., Graham, L., Rice, W.G. *J. Med. Chem.*, **39**, 2047 (1996).

28. Kempf, D.J., Norbeck, D.W., Codacovi, L., Wang, X.C., Kohlbrenner, W.E., Wideburg, N.E., Paul, D.A., Knigge, M.F., Vasavanonda, S., Craig-Kennard, A., Saldivar, A., Rosenbrook, W.J., Clement, J.J., Plattner, J.J. and Erickson, J. *J. Med. Chem.*, **33**, 2687 (1990).

29. Kempf, D.J., Codacovi, L., Wang, X.C., Kohlbrenner, W.E., Wideburg, N.E., Saldivar, A., Vasavanonda, S., Marsh, K.C., Bryant, P., Sham, H.L., Green, B.E., Betebenner, D.A., Erickson, J. and Norbeck, D.W. *J. Med. Chem.*, **36**, 320 (1993).

30. Kempf, D.J., Marsh, K.C., Paul, D.A., Knigge, M.F., Norbeck, D.W., Kohlbrenner, W.E., Codacovi, L., Vasavanonda, S., Bryant, P., Wang, X.C., Wideburg, N.E., Clement, J.J., Plattner, J.J. and Erickson, J. *Antimicrob. Agents Chemother.* 35, 2209 (1991).

31. Kempf, D.J., Marsh, K.C., Denissen, J.F., McDonald, E., Vasavanonda, S., Flentge, C.A., Green, B.E., Fino, L., Park, C.H., Kong, X.-P., Wideburg, N.E., Saldivar, A., Ruiz, L., Kati, W.M., Sham, H.L., Robins, T., Stewart, K.D., Hsu, A., Plattner, J.J., Leonard, J.M. and Norbeck, D.W. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 2484 (1995).

32. Kempf, D.J., Sham, H.L., Marsh, K.C., Flentge, C.A., Betebenner, D., Green, B.E., McDonald, E., Vasavanonda, S., Saldivar, A., Wideburg, N.E., Kati, W.M., Ruiz, L., Zhao, C., Fino, L., Patterson, J., Molla, A., Plattner, J.J. and Norbeck, D.W. *J. Med. Chem.*, **41**, 602 (1998).

LIST OF PUBLICATIONS

1. Quantitative Structure-Activity Relationship Studies on Cyclic urea–Based HIV Protease Inhibitors
   J. Enzyme Inhibn., **13**, 399-407 (1998).

2. Quantitative Structure-Activity Relationship Study on Some Sulfolanes and Arylthiomethanes Acting as HIV-1 Protease Inhibitors
   Bioorg. Med. Chem., 6, 2185-2192 (1998).

3. Quantitative Structure-Activity Relationships of Some HIV-Protease Inhibitors
   J. Enzyme Inhibn., **14**, 109-123 (1999).

4. Quantitative Structure-Activity Relationship Studies on Cyclic Cyanoguanidines Acting as HIV-1 Protease Inhibitors
   Bioorg. Med. Chem., 7, ... (1999).

5. Comparative Quantitative Structure-Activity Relationship Studies on Anti HIV-1 Agents
   Chem. Rev., in press.

6. A Fujita-Ban Type Analysis on Structure-Activity Relationships of Some Important HIV Protease Inhibitors
   Bioorg. Med. Chem., Communicated.