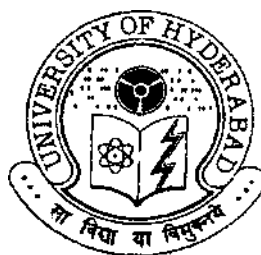


**APPROACHES TO RATIONAL DRUG DESIGN: INSIGHTS INTO
MOLECULES OF THERAPEUTIC INTEREST**

**A Thesis
Submitted for the Degree of
Doctor of Philosophy**

By
VEMA APANA



**School of Chemistry
University of Hyderabad
Hyderabad 500 046
India**

September 2006



Dedicated to

Amma

&

Nanna

STATEMENT

I hereby declare that the matter embodied in this thesis entitled “**APPROACHES TO RATIONAL DRUG DESIGN: INSIGHTS INTO MOLECULES OF THERAPEUTIC INTEREST**” is the result of investigations carried out by me in the School of Chemistry, University of Hyderabad under the supervision of Prof. Gautam R. Desiraju.

In keeping with the general practice of reporting scientific observations due acknowledgments have been made wherever the work described is based on the findings of other investigators.

Hyderabad
Sept 2006

V. Aparna

CERTIFICATE

Certified that the work “**APPROACHES TO RATIONAL DRUG DESIGN: INSIGHTS INTO MOLECULES OF THERAPEUTIC INTEREST**” has been carried out by **Vema Aparna** under my supervision and that the same has not been submitted elsewhere for a degree.

Dean
School of Chemistry

Prof. Gautam R. Desiraju
Thesis Supervisor

ACKNOWLEDGEMENT

It is indeed a great fortune to express my deep sense of gratitude and sincere thanks to my supervisor **Prof. Gautam R. Desiraju** for introducing me to this fascinating multi-faceted field of research. I have been able to learn a great deal from him and consider my association with him a rewarding experience.

It gives me an immense pleasure to thank Dr. B. Gopalakrishnan, TATA Consultancy Services, Hyderabad for his encouragement on various occasions. It has been my privilege to work with him in various projects throughout my Ph. D. tenure.

I would like to thank Dr. J. A. R. P. Sarma, GVK Biosciences Pvt. Ltd., for the fruitful discussions and suggestions during the earlier stages of my Ph. D. work.

I thank the Dean, School of Chemistry and former Deans of the School for their co-operation in providing facilities in the School. I extend my sincere thanks to all the faculty members of the School for help in various occasions. I am thankful to all the teachers and lecturers, who taught me throughout my career.

I thank CMSD for providing the computational facilities without which this work would have been impossible.

I am very thankful to my teachers in NIPER, Chandigarh Prof. A. K. Charaborti, Prof. P. Ramesh and in NCP, Shimoga Dr. K. P. C. Basavaraju for their encouragement to join Ph.D. program.

I am grateful to DST and CSIR, New Delhi for financial support. I thank all the non-teaching staff of the School of Chemistry and CMSD building for their assistance on various occasions.

I wish to thank all my friendly and cooperative labmates Drs. R. Thaimattam, R.K.R. Jetti, P. K. Thallapally, V.S.S. Kumar, V.R. Vangala, Rahul, Dinu, S. George, Vishweshwar, S. Basavoju, Srinivasulu, Balu, Sreenivas Reddy, and Mr. Sairam, Malla Reddy, Sunil, Tejender, Ravi, Jeevan, Archan, Binoy, Saikat, Prashant, Jagadeesh, Sreekanth, Bipul, Naren, Sandeep, Sanjeev, Nabakamal, and Ranjeet for creating a cheerful working atmosphere in the lab. My stay on this campus has been pleasant with

the association of all the research scholars at the School of Chemistry. I am thankful to Subbalakshmi, Sailaja, Satyen, Shankaran, Prema, Pavan, and Sarita.

It would be too formal to thank my bosom friends Aparna, Aruna, Neeraja, Kalpana, Nimmi, Vani, Taufeeq, Ganga, Surya, Kumar, Gullapalli, Raghupathi, Shruti and Srinatha for their unreserved encouragement. I would also like to thank Saisailaja, Jyothi and Shruti Kanta for creating homely atmosphere in the hostel.

I thank my brother for his boundless support and encouragement.

The blessings and best wishes of my parents keep me active throughout my life. They made me what I am and I owe everything to them. Dedicating this thesis to them is a minor recognition for their invaluable support and encouragement.

V. Aparna

PREFACE

Computational approaches for lead discovery and optimization have proven successful in many recent drug research programs. These methods have grown in their effectiveness not only because of improved understanding of the basic science, the biological events and molecular recognition, but also because of advances in the mathematical procedures for studying such processes. Chapter 1 gives an overview of various rational drug design approaches based upon the molecular recognition involved in protein–ligand complexation. It also briefs the reader on various advances in computer technology and molecular simulations.

Quantitative Structure Activity Relationship (QSAR) studies like Molecular Field Analysis (MFA) and Receptor Surface Analysis (RSA) using a large set of Epidermal growth factor receptor (EGFR) kinase inhibitors revealed the structural requirements of the inhibitor for an enhanced biological activity. Comparison of analog based studies with the structure based docking is described in chapter 2.

Docking studies of EGFR kinase inhibitors revealed the role of conventional strong hydrogen bonds like N–H...N and O_w–H...N along with weak hydrogen bonds like C–H...O in molecular recognition. Interaction based analysis has been used to obtain better docking poses. Chapter 3 demonstrates the use of both cross scoring and consensus scoring to develop a virtual screening model against EGFR kinase.

Chapter 4 compares the widely used analog based methods like MFA, pharmacophore and structure based methods like docking to identify the most appropriate method for the virtual screening against Thymidine Mono Phosphate Kinase of *mycobacterium tuberculosis* (TMPK_{mt}). The pharmacophore method, which occupies a middle ground between ligand based MFA and structure based docking, is the most appropriate method for screening large virtual libraries against TMPK_{mt}.

Chapter 5 deals with the virtual screening of TMPK_{mt} inhibitors by applying a protocol of several hierarchical filters. A composite model, that combines both the

accuracy of the docking and the speed of the pharmacophore method, and Lipinski's rules constitute these filters.

Chapter 6 deals with the molecular dynamic simulations of ligand complexes with Nitric Oxide Synthase (NOS) isoforms viz., iNOS and eNOS. Simulations with the substrate, L-Arg, identify water to play a critical role in masking the active site differences between these isoforms. Juxtaposition of the carboxylic and ammonium groups in the substrate and the inhibitor serve as a modulating key in binding to the isoforms. These provide guidelines to the design of selective inhibitors of the isoforms.

Chapter 7 outlines a summary of the significant results presented in this thesis and some implications for further studies.

An extended appendix lists the metrics of hydrogen bonds formed by 128 molecules with the EGFR kinase (chapter 3) and hits obtained from virtual screening (chapter 5) that could be possible leads for further development as novel TMPK_{mt} inhibitors.

CONTENTS

Statement	v
Certificate	vii
Acknowledgement	ix
Preface	xi

CHAPTER 1

RATIONAL DRUG DESIGN: STRATEGIES AND APPLICATIONS 1-15

1.1	Introduction	1
1.2	Rational Drug Design (RDD)	1
1.3	Role of Computer Aided Molecular Design in Drug Discovery	2
1.4	Approaches to RDD	3
1.4.1	Analog Based Studies	3
1.4.2	Structure Based Studies	4
1.5	ADMET Prediction	11
1.6	Principles of Theoretical Chemistry	12
1.6.1	Molecular Mechanics (MM)	13
1.6.2	Quantum Mechanics	13
1.7	Advances in Computational Power	14
1.8	Simulations One Step Ahead–Future Directions	14

CHAPTER 2

EGFR KINASE INHIBITORS: A LIGAND BASED APPROACH AND ITS CONFIRMATION WITH STRUCTURE BASED STUDIES 19-39

2.1	Introduction	19
2.2	Methods	21
2.2.1	Selection of Molecules	21
2.2.2	Initial Setup	22

2.2.3	MFA	28
2.2.4	RSA	29
2.2.5	Docking	30
2.3	Results and Discussion	30
2.3.1	MFA–Results	32
2.3.2	RSA–Results	34
2.3.3	MFA vs. RSA	36
2.3.4	Structural Requirements at the 6- and 7-Positions	36
2.3.5	Correlation of RSA Results with Docking	37
2.4	Conclusions	38

CHAPTER 3

FABRICATION OF VIRTUAL SCREENING MODEL FOR EGFR KINASE INHIBITORS: IMPORTANCE OF HYDROGEN BONDS IN THE EVALUATION OF POSES AND SCORING FUNCTIONS

41-63

3.1	Introduction	41
3.2	Methods	42
3.2.1	Generation of Ligand and Enzyme Structures	42
3.2.2	Docking	42
3.2.3	Docking Methods	43
3.2.4	Scoring Functions	44
3.3	Results and Discussion	48
3.3.1	Direct Docking	49
3.3.2	Molecular Recognition of Erlotinib in Crystal Structure	51
3.3.3	Pose Analysis with RMSD	52
3.3.4	Pose Analysis with Intermolecular Interactions	54
3.3.5	Cross Scoring	60

3.3.6 Consensus Model	61
3.4 Conclusions	62

CHAPTER 4

IDENTIFICATION OF APPROPRIATE METHOD FOR VIRTUAL SCREENING OF THYMIDINE MONOPHOSPHATE KINASE INHIBITORS 67-95

4.1 Introduction	67
4.2 Materials and Methods	68
4.2.1 Selection of Molecules	68
4.2.2 Docking	70
4.2.3 Pharmacophore Modeling	70
4.2.4 3D-QSAR	73
4.3 Results and Discussion	75
4.3.1 Docking	75
4.3.2 Pharmacophore	81
4.3.3 MFA	83
4.3.4 Comparison of the Methods	90
4.4 Conclusions	92

CHAPTER 5

VIRTUAL SCREENING FOR THYMIDINE MONOPHOSPHATE KINASE INHIBITORS BASED ON COMPOSITE PHARMACOPHORE MODEL 93-100

5.1 Introduction	93
5.2 Methods	94
5.2.1 Database Generation	94
5.2.2 Database Screening	94
5.3 Results and Discussions	94
5.3.1 Composite Model	95
5.3.2 Validation of Composite Model	96

5.3.3 Screening	97
5.4 Conclusions	100

CHAPTER 6

INSIGHTS INTO LIGAND SELECTIVITY IN NITRIC OXIDE SYNTHASE ISOFORMS: A RATIONAL APPROACH FOR INHIBITOR DESIGN 103-136

6.1 Introduction	103
6.2 Overview of Molecular Dynamics	106
6.3. Methodology	108
6.3.1 Initial Structure Preparation	108
6.3.2 Molecular Docking	108
6.3.3 Molecular Dynamics	109
6.4 Results and Discussion	110
6.4.1 Overall Structural Changes	110
6.4.2 Mode of Substrate Binding in iNOS and eNOS Isoforms	112
6.4.3 Differences in Substrate Recognition in the Isoforms	116
6.4.4 Heme-H ₄ B Interaction in Isoforms	117
6.4.5 Role of Water in Substrate Binding	120
6.4.6 Mutation Studies of iNOS and eNOS	120
6.4.7 Mode of Inhibitor Binding in iNOS and eNOS Isoforms	122
6.4.8 Rational Design of More Selective Ligands	127
6.4.9 Mode of Binding of Molecule 178 in iNOS and eNOS	128
6.4.10 Mode of Binding of Molecule 179 in Isoforms	130
6.4.11 Comparison of Binding of Ligands in iNOS and eNOS	131
6.4.12 Energetics of Ligand Binding in iNOS and eNOS	132
6.5 Conclusions	135

CHAPTER 7

CONCLUSIONS AND FUTURE PROSPECTS 137-140

REFERENCES AND NOTES	141-154
APPENDIX I	155-163
APPENDIX II	165-175
ABOUT THE AUTHOR	177
LIST OF PUBLICATIONS	179

CHAPTER 1

RATIONAL DRUG DESIGN: STRATEGIES AND APPLICATIONS

1.1 Introduction

Drug discovery is a multidisciplinary approach wherein drugs are designed and/or discovered. The R&D expenditure incurred to bring a new chemical entity (NCE) to the end of phase III clinical trials is estimated to be around \$1.3 billion today in the US [1.1]. The chief contribution to this cost increase is the decrease in efficiency of transforming lead → preclinical candidates from 75% to 50% and the rate of retrogression of compounds from phase 2 → phase 3 clinical trials from 50% to 30% [1.2]. Despite advances in technology and understanding of biological systems, drug discovery is still a long process (~ 15 years) with low rates of new discoveries. This scenario demands alternate procedures/techniques that cut down both the cost and the time period involved and simultaneously increase the success rate.

1.2 Rational Drug Design (RDD)

A glance into the history of drugs shows that many early discoveries in the pharmaceutical industry were serendipitous; they fail to explain why a compound is active or inactive or how it may be improved. The advent of new knowledge of physiological mechanisms has made it possible to take a mechanistic approach and start from a rationally argued hypothesis to design new chemical entities (NCE's). The concept of RDD could be traced to the findings of Paul Ehrlich (*chemoreceptor*) and Emil Fischer (*lock and key model*) in 1872 and 1894 respectively [1.3]. Advances in molecular biology, protein crystallography and computational chemistry since the 1980s have greatly aided the RDD paradigms. The advent and development of combinatorial chemistry and high throughput screening (HTS) in the 1990s led to a paradigm shift in drug research. Combinatorial chemistry uncorked the chemical bottleneck in drug research shifting the question in lead optimization from “what can we make” to “which

should we make” [1.4]. Contemporarily, HTS makes it possible to screen huge libraries of molecules within a short time span [1.5]. Nevertheless, initial euphoria that designated these techniques as universal lead generators subsided as a result of the considerable costs involved and disappointingly low hit rates [1.6]. Lessons learnt from these strategies seek a complete shift of drug research paradigms from an empirical science to structure based analysis of macromolecule–ligand interactions. Figure 1.1 shows a flow chart that describes different approaches that enable RDD to evolve new NCE’s with greater biological activity.

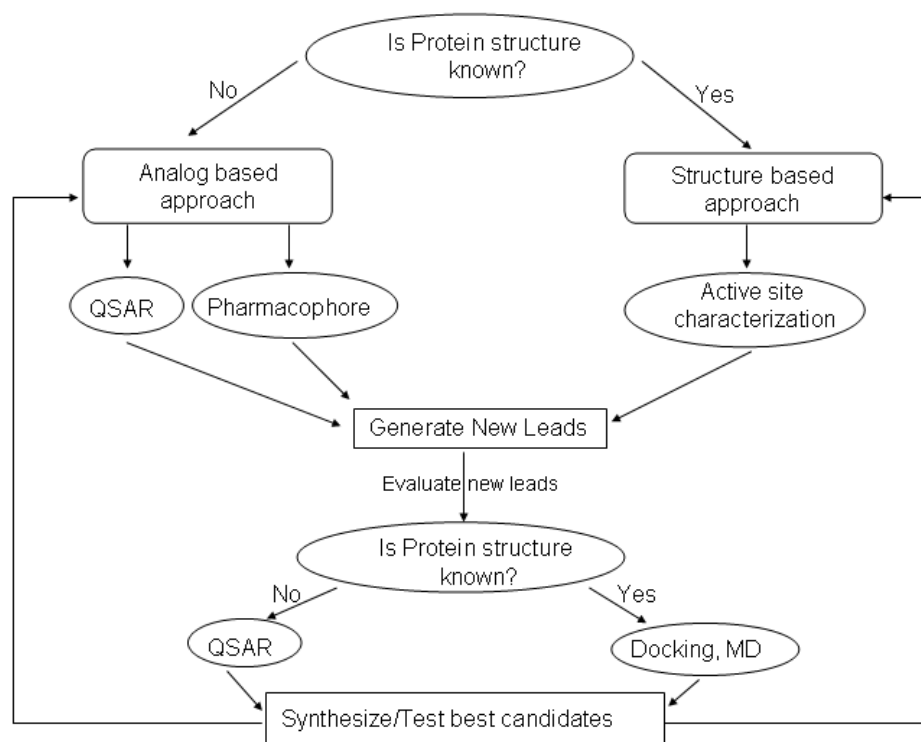


Figure 1.1: Different Approaches of Rational Drug Design (adopted from ref. 1.7).

1.3 Role of Computer Aided Molecular Design in Drug Discovery

It was recognized in the 1960s, that computer-based methods can be of help in the discovery of new leads and can potentially eliminate chemical synthesis and screening of many irrelevant compounds. An ideal computational method for lead

discovery should be able to generate structurally diverse leads rapidly and should give estimates of binding affinities that would correlate with experimental values. Generation of chemical diversity *in silico*, is easily achieved using existing computational resources and algorithms: putative ligands can either be extracted from large databases of compounds, or they can be “grown” computationally by joining molecular fragments. On the other hand, accurate prediction of binding affinities has been a more difficult task [1.8]. Because of the multitude of energetic and entropic factors involved, the thermodynamics of binding cannot be analytically modeled without first simplifying the problem [1.9]. Computational methods that attempt to design leads vary in the nature and in the degree of the simplifying assumptions they use.

1.4 Approaches to RDD

The state of the art in RDD or Computer Aided Molecular Design (CAMD), can be divided into two broad categories: analog based study and structure based study based on the availability of three dimensional structure of the target.

1.4.1 Analog Based Studies

In a broad sense, analog based studies gather information from already existing drugs/ligands that are active against target biological molecule (protein or DNA/RNA) of interest. Based on this information a set of rules are framed to either design a new ligand or modify an existing ligand in order to enhance its biological activity.

Quantitative Structure Activity Relationship (QSAR)

QSAR is one of the most widely used analog based methods. It aims at correlating structural features of a series of known compounds with their biological activities. From these correlations empirical equations are derived and subsequently used to guide the design of new leads. Early QSAR methods related biological activity to the presence (or absence) of functional groups in a series of structurally related compounds (Free-Wilson model), or to the physicochemical properties (lipophilicity, electronic properties) of the compounds in the training set (Hansch analysis) [1.10]. More recently,

three-dimensional QSAR methods have been developed [1.11]. Although QSAR models reproduce binding affinities of ligands more accurately than other methods they have three major shortcomings [1.12]: (i) sufficient number of ligands active against target of interest should be available to develop the structure activity relationships; (ii) the equations that are parameterized for one target do not apply to another; and (iii) are of limited use in understanding the nature of protein–ligand interactions and thermodynamics of binding.

Pharmacophore

Pharmacophore is one another analog based method. The word “pharmacophore” was coined by Paul Ehrlich in the early 1900s referring to a molecular framework that carries the essential features (*phoros*) responsible for a drug’s (*pharmacon*) biological activity [1.13]. Later in 1977 Peter Gund redefined it as “a set of structural features in a molecule that is recognized at a receptor site and is responsible for that molecule’s biological activity”. Pharmacophore models are constructed based on molecules of known biological activity and are refined as more data are acquired in an iterative process. Alternatively, a pharmacophore can also be generated from the receptor structure. One step ahead, the *dynamic pharmacophore model* based on molecular dynamics trajectories takes care of the the binding site dynamics [1.14]. These models can be used for optimizing known ligands or for screening databases to find potential novel leads suitable for further development [1.15].

1.4.2 Structure Based Studies

Structure based approaches, based on the three-dimensional structure of the target overcome many of the limitations of analog based studies. These methods help to develop a general theoretical description of the protein–ligand interactions that would enable an a priori design of new leads for a particular biological target [1.16]. The first success story in structure based design is the antihypertensive drug Captopril, an inhibitor of Angiotensin Converting Enzyme (ACE) [1.17]. Table 1.1 lists other

examples of drugs derived from structure based approaches. Different approaches used for the structure based design are as follows:

Table 1.1: List of drugs resulting from structure based studies

Enzyme	Disease	Drugs	Trade name
Neuraminidase	Influenza	Oseltamivir Zanamivir	Tamiflu® Relenza®
Carbonic Anhydrase II	Glaucoma	Dorzolamide	Cosopt®
5-Hydroxy Tryptamine 1B	Migraine	Zolmitriptan	Zomig®
Angiotensin II	Hypertension	Losartan	Cozaar®
EGFR Kinase	Cancer	Erlotinib	Tarceva®
Bcr-Abl Kinase	CML	Imatinab	Gleevac
HIV-protease	AIDS	Indinavir	Crixivan®
		Nelfinavir	Viracept®
		Saquinavir	Invirase®
		Ritonavir	Norvir®
		Lopinavir	Kaletra®
		Amprenavir	Agenerase®
		Tipranavir	Aptivus®
HIV-Reverse transcriptase		Rilpivirine Etravirine	Phase II* Phase III*

* in clinical trials

Docking

Docking in a true sense is the formation of non-covalent protein–ligand complexes *in silico*. Given the structure of a protein and a ligand, the task is to predict the structure of the complex. Conceptually, docking is an energy optimization process concerned with the search of the lowest free energy binding mode of a ligand within a protein binding site [1.10]. Docking constitutes two components: pose searching and scoring. Inclusion of protein flexibility is computationally expensive; therefore much of the existing docking programs treat the protein either as rigid or allow flexibility only to the side chain functional groups. On the other hand, ligand handling can be broadly

classified as: whole molecule approach as shown in variant 1 and 2 and fragment based approach seen as variant 3 in Figure 1.2. A good docking method estimates the forces involved in the protein–ligand recognition viz. electrostatic, van der Waals and hydrogen bonding and places the ligand appropriately in the active site [1.18]. Table 1.2 lists all the existing docking methodologies and the strategies they use.

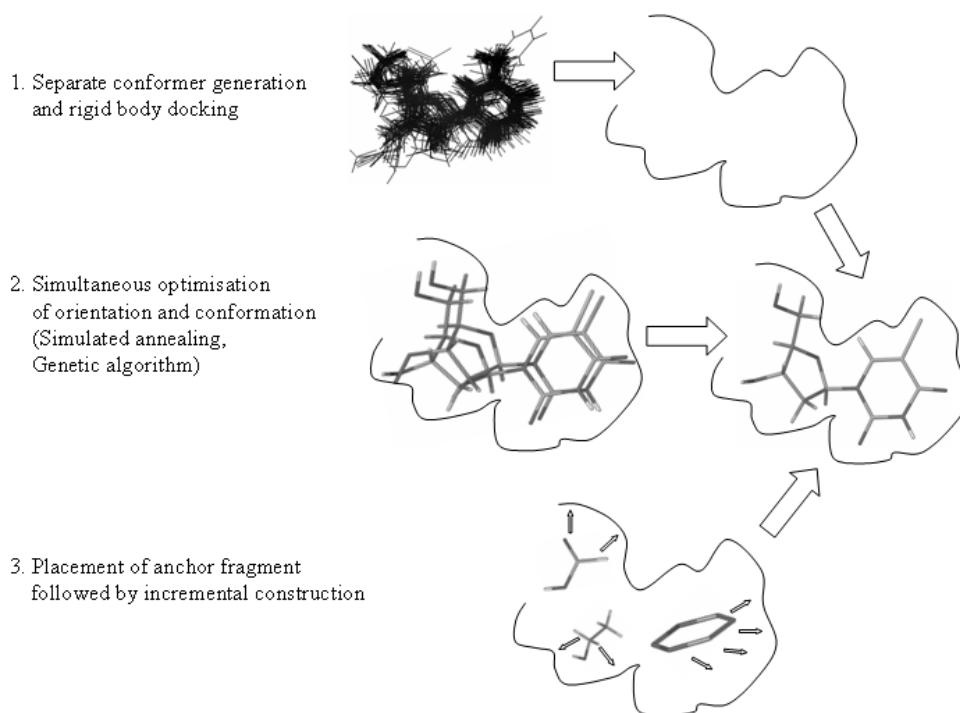


Figure 1.2: Strategies for flexible ligand docking

Table 1.2: Different search strategies for docking

Searching Algorithm	Description	Examples
Monte Carlo (MC)	Stochastic method of generating conformations. Selection based on Metropolis criterion	Ligand Fit
Simulated Annealing (SA)	Random thermal motions are induced, through high temperatures, to explore the local search space. System is driven to a minimum energy conformation by decreasing temperature. SA usually combined with MC	MC-DOCK, ICM-DOCK, AutoDock
Genetic Algorithm (GA)	Based on Darwin principles of evolution. 'chromosome' encoding model parameters (like torsion angles) is varied stochastically. Populations generated through genetic operations (crossover, mutation, migration). The fittest survives in the population.	GOLD, DARWIN, AutoDock
Tabu search	Stochastic method of generating conformation keeping a record of previous conformations (tabu). Generated conformation is retained if it is not tabu or if it scores better than that in tabu	PRO_LEADS
Incremental construction	Systematic method where ligand is broken into rigid fragments at rotatable bonds. Fragments docked in all possible ways and assembled piecewise to regenerate ligand	Flex-X, DOCK, HOOK, LUDI, Hammerhead
Matching methods	Based on clique detection technique from graph theory. Ligand atoms matched to the complimentary atoms in the receptor	FLOG, DOCK
Simulation methods	Molecular dynamics simulations used to generate conformations	DOCK

Scoring

In principle, a scoring function used in docking is a mathematical function whose values are proportional to the binding affinities of the leads. A good scoring function should be able to give reliable estimates of binding affinities of structurally diverse leads for different protein targets while considering the thermodynamic aspects of binding [1.19]. Scoring functions are of three types [1.20]: (1) empirical; (2) force field based and; (3) knowledge based functions. *Empirical scoring functions* are regression based functions derived from a large sample of crystal structures with known affinities for the bound ligands. These functions reflect a best fit with respect to the training set used but rarely achieve generality. *Force field based methods* are first principle methods that use force field parameters to score the vdW and electrostatic interactions between protein and ligand. The score includes receptor–ligand interaction

energy and internal ligand energy (such as steric strain induced by binding). These methods do not require calibration or training with experimental binding data. *Knowledge based methods* evaluate the frequencies of particular type of interaction—the mutual distance between particular type of atoms across the interface, in databases of protein–ligand complexes. The sample distribution describes the probability of occurrence of an interaction and is compared with the reference mean. Any deviation from the mean is translated into statistical preferences using mathematical equations and related to energies in a Boltzmann-like fashion.

De Novo Ligand Design

De novo design uses structural information to “grow” a molecule into the active site by sequentially adding or joining molecular fragments instead of using libraries of existing compounds [1.21]. Structure sampling is carried out by different methods like linking, growing, lattice-based sampling, random structure mutation, transitions driven by molecular dynamics simulations, and graph-based sampling. Figure 1.3 gives a schematic description of few of these strategies. Apart from these, the ligand can also be built from recombination of bioactive conformations of known ligands for a particular target. Recombination is carried out by overlaying the known ligands and swapping the fragments of different ligands. This procedure is carried out recursively, so that the compounds that emerge from recombination are added to the pool of known actives and participate in subsequent cycles of recombination. The largest advantage of *de novo* design is its ability to develop novel scaffolds utilizing the whole chemical space [1.22]. However, this method also suffers limitations like: 1) synthetic feasibility is not considered while constructing structures and; (2) the prediction of binding affinities for the designed structures is not so accurate.

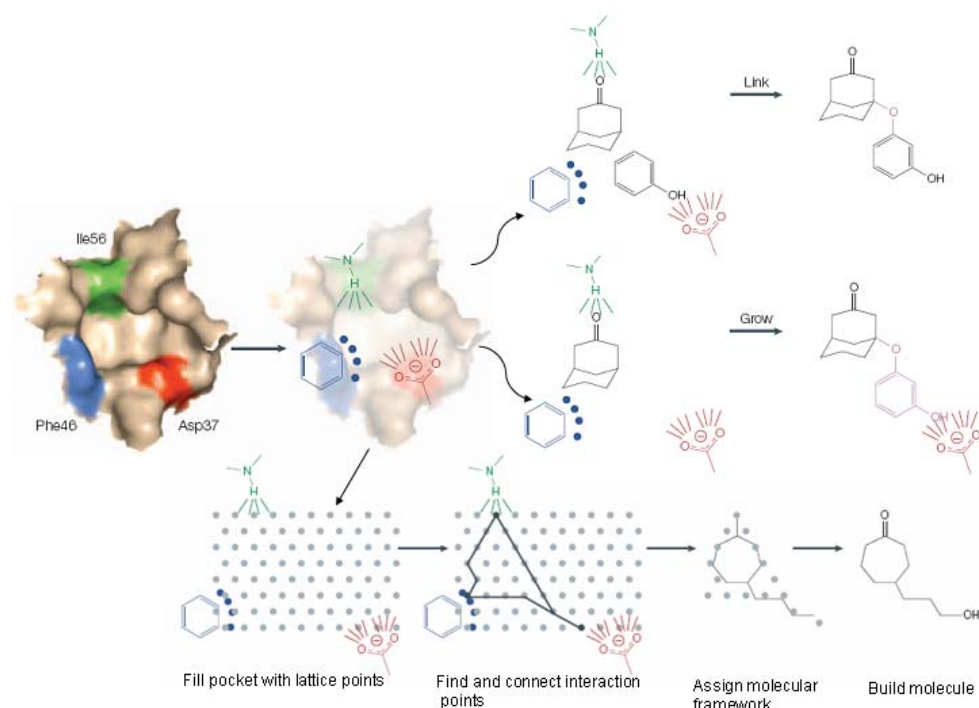


Figure 1.3: Strategies of de novo ligand design

Virtual Screening (VS)

VS is a knowledge driven process that uses computational chemistry techniques to analyze large chemical databases in order to identify possible new leads. VS is used as an initial screen for large databases to prune the number of compounds that are to be screened experimentally [1.23]. This process of finding ‘needles in a haystack’ produces leads that may otherwise not have surfaced and therefore adds immense value to the early drug discovery stages. VS protocols include ligand based screens like 1D filters (e.g. molecular weight), 2D filters (similarity, substructure fingerprints), and 3D filters (3D-pharmacophore [1.24], 3D shape matching) and structure based screens like docking. The potential sources of error contributing to the identification of false positives and false negatives in VS include: 1) approximations in the scoring functions employed; 2) improper solvation terms; 3) neglect of protein flexibility and; 4) poor assessment of the protonation states of active site residues or ligands [1.25]. Significant improvements

in VS have been made by consensus scoring [1.26] of multiple scoring functions and by clustering docking poses, from multiple docking tools before scoring [1.27]. Till date VS was applied to myriad protein targets from G-protein coupled receptors to ion channels, to identify potential leads [1.6, 1.28].

Molecular Dynamics

Molecular Dynamic (MD) simulations are widely used to obtain information on the time evolution of conformations of biological molecules with the associated kinetic and thermodynamic properties. The basic feature of molecular dynamics is the calculation of a trajectory of the molecule, i.e. a series of structures at regular time steps in which the system is moving under the influence of the forces acting on the atoms [1.29]. These are calculated from the first derivative of the potential function with respect to the atom positions. By applying Newton's equations of motion, these forces can then be used to calculate how the atomic positions change with time resulting in a dynamic trajectory. MD can be utilized to quantify the properties of a system and is therefore a valuable tool in understanding the complete profile of a model system. A detailed description of the MD methodology is discussed in chapter 6.

MD methods were originally conceived within the theoretical physics community in the 1950s [1.30]. In 1957, Alder and Wainwright performed the earliest MD simulation using hard sphere model where atoms interacted with perfect collisions [1.31]. Subsequently Rahman applied a smooth, continuous potential to mimic real atomic collisions [1.32]. Breakthroughs in MD lead to its first application on the protein bovine pancreatic trypsin inhibitor for 9.2ps *in-vacuo* in 1977 [1.33]. In 1988, cumulative advances in the MD simulations made it possible to carry out a microsecond MD of a much larger protein in solution [1.34]. Table 1.3 lists different MD methods where the utility of each varies with the aspects of desire.

Table 1.3: List of various MD techniques

MD Methods	Description
Brownian MD	accounts for Brownian motion of molecules particularly evident with solvents of high viscosity
Langevin MD	Uses Langevin equations of motion where additional frictional and random forces are added to Newtonian equation
Activated MD	Simulates activated processes allowing to cross barriers existing between the initial and final stages
Accelerated MD	Time treated as statistical quantity. Conformational sampling through i) modified potential energy surface ii) non-Boltzmann type iii) degrees of freedom at the expense of faster degrees of freedom
Steered MD	Based on Atomic Force Microscopy (AFM) principle. Introduces a time and position dependent force to steer systems along particular degrees of freedom
Targeted MD	Force used to drive simulation towards target conformation
SWARM-MD	Multiple simulations using molecules with each one subjected to force introducing cooperative behaviour and drives the trajectory to the mean trajectory of the entire swarm
Replica Exchange MD (REMD)	Series of simultaneous non-interacting simulations (replica) carried over a range of temperatures and swapped at particular intervals of temperature
Self Guided MD (SGMD)	Introduces an additional guiding force that is a continuously updated time average of the force of the current simulation
Leap Dynamics	A combination of MD and essential dynamics. Leaps applied to the system to force it over energy barriers
Multiple Body Dynamics [MBO(N)D]	Combines rigid body dynamics with multiple time steps where highest frequency harmonic motions are removed retaining the low-frequency anharmonic motions
λ Dynamics	A multiple copy method with reduced interaction potential of the ligand, lowering the barriers of conformational transitions. Fraction of interacting potential, λ_i^2 (additional degree of freedom) allows ligand to explore conformational space due to reduced barriers

1.5 ADMET Prediction

Absorption, Distribution, Metabolism, Elimination, and Toxicity (ADMET) profiles of chemical compounds have become the bottleneck and a major challenge in drug research. Table 1.4 lists few drugs withdrawn from the market due to insufficient

ADMET profile. To overcome the high rate of attrition of active compounds plagued by hidden pharmacokinetic issues/problems, ADMET property evaluation is incorporated into drug design strategies. *In silico* prediction of physicochemical parameters of compound's ionizability (pK_a), and lipophilicity ($\log P$ or $\log D$), provide an indication of its likely absorption in the gut. Various computational techniques using *in silico* models are emerging that attempts to give the ADMET profile of a given compound [1.35, 1.36].

Table 1.4: Drugs withdrawn from market due to ADMET failure

Drug	Therapeutic Utility	Year of withdrawal	Withdrawn due to
Thalidomide	Morning sickness in pregnancy	1960s	Teratogenicity leading to deformities in fetal development
Ticrynafen	Diuretic	1982	Hepatitis
Sumatriptan	Migraine		drug-drug interactions with MAO inhibitors
Terfenadine	Antihistamine	1997	Cardiotoxicity
Mebifradil	Anti hypertension Angina	1998	Interferes with metabolism of other drugs used for hypertension
Troglitazone	Anti diabetic	2000	Hepatotoxicity
Cerivastatin	Anti hyperlipidemic	2001	Rhabdomyolysis
Rofecoxib	Anti inflammatory	2004	Myocardial Infarction
Phenyl propranolamine	In cough as decongestant	2005	Hemorrhagic stroke

1.6 Principles of Theoretical Chemistry

All the above described methods of RDD are highly dependent upon the potential energy function that describes the energy landscape of the molecular systems [1.37]. These landscapes are derived with the application of the following principles.

1.6.1 Molecular Mechanics (MM)

This method uses the laws of classical physics to predict the structures and properties of the molecules. These calculations consider the nuclei and electrons as unified spherical particles while the bonds between these particles are viewed as springs. The mathematics of spring deformation (Hooke's law) is used to describe the ability of bonds to stretch, bend, and twist. Interactions between these particles are treated using potential functions, derived using classical mechanics, to describe different types of interactions: bond stretching, angle bending, and torsional energies, as well as non-bonded interactions. The sum of interactions determines the spatial distribution of atoms, i.e. conformation. The total potential energy surface of a molecule can be described by the summation of interactions through a set of equations called *force-fields*. A typical form of a force field is given by equation 1.1

$$E_{total} = \Sigma E_{stretch} + \Sigma E_{bend} + \Sigma E_{torsion} + \Sigma E_{vdW} + \Sigma E_{Coulomb} + \Sigma E_{cross} \quad (1.1)$$

where E represents the energy of bond distortion of that particular bond. The potential energy functions rely on empirically derived parameters like force constant. Although MM cannot handle chemical properties with predominantly electronic effects, they can usually take care of large systems like proteins.

1.6.2 Quantum Mechanics

These approximations were developed because of the need to describe electronic structures of molecules. There are at least four important areas of RDD where quantum mechanical calculations can be applied: charge calculations, molecular electrostatic potential, parameter development for molecular mechanics and modeling chemical reactions and design of transition-state inhibitors [1.38]. Quantum mechanical methods are broadly classified as semi-empirical and *ab initio* methods. Both use approximations to solve the Schrödinger equation.

Semi-Empirical methods: This uses a simpler Hamiltonian. The parameters used are adjusted to fit experimental data or results from *ab initio* calculations. These can be successfully applied to systems containing heavy atoms up to 60 and total atoms of 120.

Ab initio methods: Are non-parameterized molecular orbital treatments for the description of the chemical behaviour. This method takes into consideration even the electrons of the innermost orbitals. Its application is limited to small molecules containing tens of atoms.

1.7 Advances in Computational Power

All the methods described above are computer intensive and therefore demand immense increase in the availability of computational power. Consequently, Linux farms allowing many hundreds of parallel calculations in acceptable timescales substitute for much expensive supercomputers. GRID computing is one other advancement to high computational needs. An elegant example of applications of such grid computation is the screen saver project SETI@home (Search for Extra Terrestrial Intelligence) over PC networks [1.39]. In this project about 1.2 million household PCs across 215 countries were used to constitute a 65-teraflop equivalent machine with more than 100,000 years of CPU. Further, this network was used to screen 3.2 billion virtual structures in 13 protein active sites in just 24 days searching for novel anticancer agents and anthrax inhibitors. Folding@Home is a similar kind of project used in the prediction of protein folding [1.40].

1.8 Simulations One Step Ahead—Future Directions

Advances in computations in structural biology have made it possible to carry out virtual cell simulations that mimic the cell environment and the cellular events therein [1.41]. A number of programs have been developed in this area. For example *NEURON* and *GENESIS* simulate the electrophysiological behavior of single neurons and neuronal networks. One step ahead of these, *E-CELL* constructs a model of a hypothetical self-sustaining whole-cell with 127 genes sufficient for transcription, translation and energy production [1.42]. Alternately, *MCell* provides a modeling tool for realistic simulation of cellular signaling in the complex 3-D cellular microphysiology—subcellular microenvironment in and around living cells, using Monte Carlo algorithms to track the stochastic behavior of discrete molecules in space and time

[1.43]. *Virtual Cell* is another program that models cell biological processes. Future prospects of such virtual cell simulations would hopefully reduce the limitations of simulations using isolated proteins which will never exist in real situations and interrelated ADMET properties.

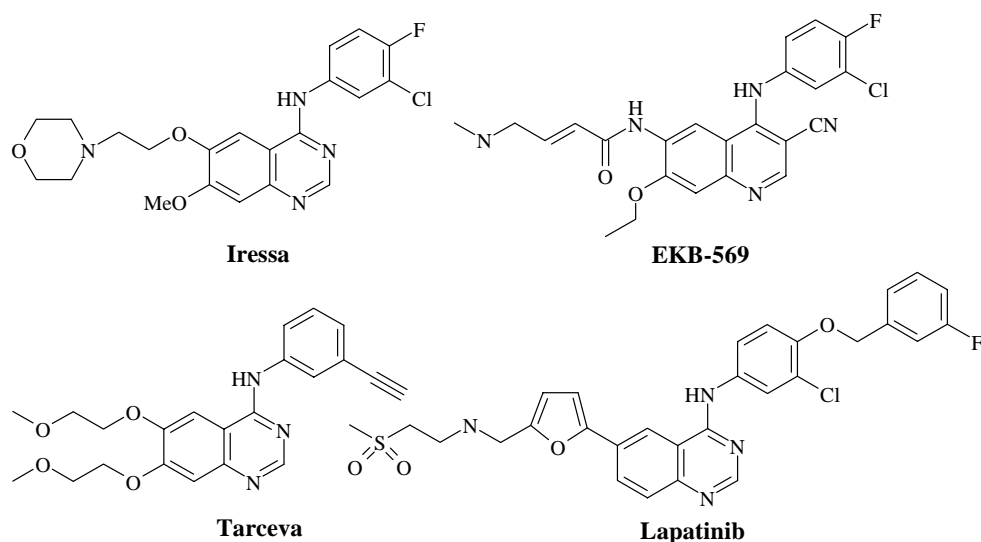
CHAPTER 2

EGFR KINASE INHIBITORS: A LIGAND BASED APPROACH AND ITS CONFIRMATION WITH STRUCTURE BASED STUDIES

2.1 Introduction

In normal physiology Epidermal Growth Factor Receptor (EGFR) is activated by binding to its cognate ligands, epidermal growth factor or transforming growth factor alpha (TGF- α), at the extracellular domain causing receptor dimerization. This in turn leads to activation of the intracellular kinase domain and triggers a cascade of signal transduction events responsible for regulation of cell proliferation [2.1]. However, EGFR is frequently overexpressed, amplified and/or mutated in many human solid tumors such as breast, ovarian, non-small cell lung, and squamous cell cancers [2.2]. It has also been shown to affect proliferation, angiogenesis and cancer metastasis [2.3]. Thus it is being pursued as a molecular target for the treatment of selective cancers [2.4].

Drugs targeting EGFR fall into three main categories depending on the receptor region targeted: extracellular, intracellular and nuclear. Small molecule inhibitors that target the intracellular EGFR appear to be the most promising approach towards treating EGFR mediated cancers. These molecules act by binding either reversibly or irreversibly to the C-terminal tyrosine kinase domain of EGFR, thereby inhibiting autophosphorylation of the receptor and therefore activation [2.5]. Anilinoquinazolines are the most developed class of drugs that inhibit EGFR kinase intracellularly [2.6]. These compounds are being studied actively by many research groups [2.7-2.13]. As a result, drug candidates of this class have already reached various phases of clinical trials (Scheme 2.1).



Scheme 2.1: Anilinoquinazoline derivatives that are currently in clinical trials or marketed.

Structure Activity Relationship (SAR) studies reveal the nature of desirable substituents on the anilinoquinazoline moiety. Electron withdrawing, lipophilic substituents in the 3-position of the aniline ring are favourable with Cl and Br being optimal. Similarly, electron-donating groups in the 6- and 7-positions of the quinazoline skeleton are preferred [2.14]. Bulky substituents appear to be tolerated in the 6- and 7-positions [2.15]. 3D-QSAR studies by Hou *et al.* have described the region around the 7-position as more electronegative than that near the 6-position [2.16]. In the present chapter, 3D-QSAR studies namely, Molecular Field Analysis (MFA) and Receptor Surface Analysis (RSA) [2.17, 2.18] were carried out on a series of EGFR kinase inhibitors in order to provide further insight into the key structural features required to design potential drug candidates of this class.

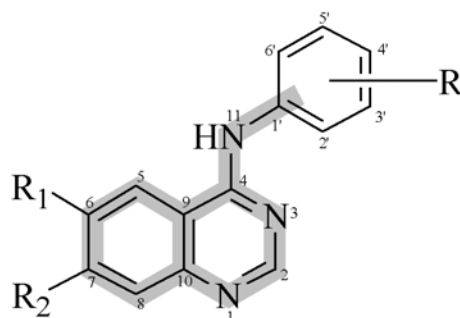


Figure 2.1: Anilinoquinazoline skeleton with atom numbering scheme. The shaded region includes the 12 atoms used for alignment.

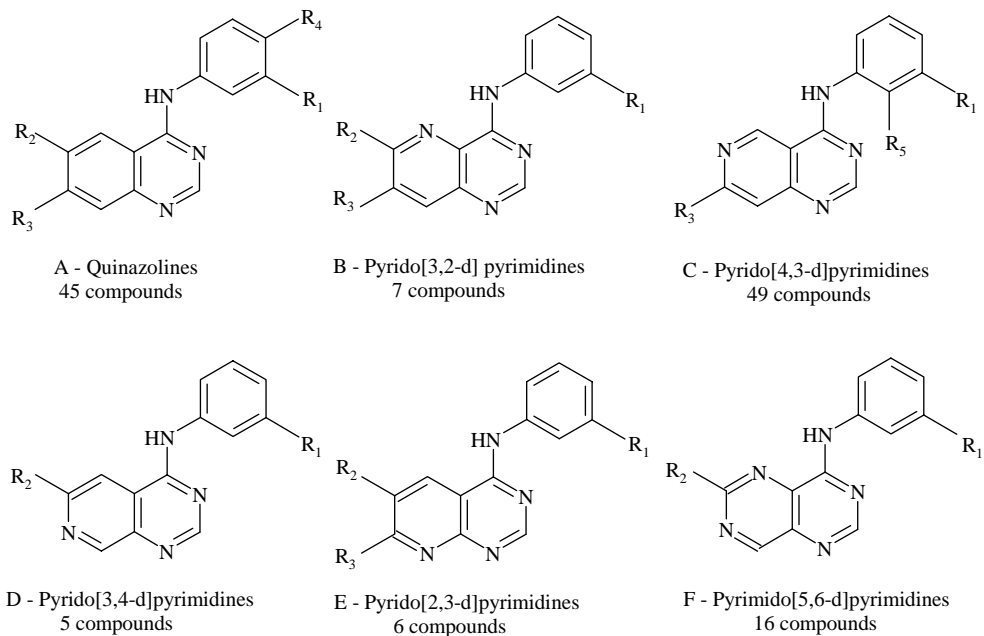
2.2 Methods

2.2.1 Selection of Molecules

A set of 128 compounds reported as EGFR tyrosine kinase inhibitors was compiled (Table 2.1) [2.14, 2.15, 2.19-2.22]. The compounds used in this study may be classified in six structurally different families. These groups are quinazolines (family A: 45 compounds), pyrido[3,2-d]pyrimidines (family B: 7 compounds), pyrido[4,3-d]pyrimidines (family C: 49 compounds), pyrido[3,4-d]pyrimidines (family D: 5 compounds), pyrido[2,3-d]pyrimidines (family E: 6 compounds), pyrimido[5,6-d]pyrimidines (family F: 16 compounds). The biological activities were converted into the corresponding pIC_{50} values ($-\log IC_{50}$), where IC_{50} value represents the drug in molar concentration that causes 50% inhibition of phospholipase $C\gamma 1$ phosphorylation by EGFR. It was verified that all the IC_{50} values had been obtained using the same assay method [2.23]. The IC_{50} values of reference compounds were checked to ensure that no difference occurred between different groups. The pIC_{50} values of the molecules spanned a wide range from 5–11. About 75% of the 128 compounds i.e., 100 compounds were selected as the training set and the remaining 25% i.e., 28 compounds were included in the test set.

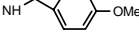
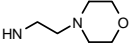
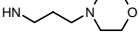
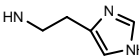
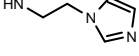
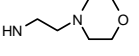
2.2.2 Initial Setup

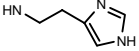
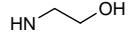
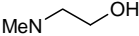
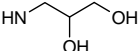
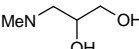
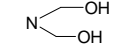


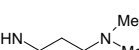
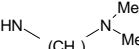
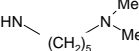
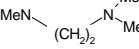
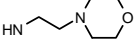
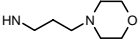
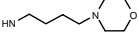
The X-ray crystal structure of the inhibitor (4-[3-hydroxyanilino]-6,7-dimethoxyquinazoline) bound to the active site of CDK2 (PDB entry 1DI8) [2.24] was used as a template to construct the three-dimensional models of all the compounds. All the structures were initially minimized by OFF methods using the steepest descent algorithm with a convergence gradient of $0.001 \text{ kcal mol}^{-1}$. Partial atomic charges were calculated using the Gasteiger method [2.25]. Further geometry optimization was carried out for each compound with the MOPAC 6 package using the semi-empirical AM1 Hamiltonian [2.26]. The most active compound **43** was used as the template for the superimposition of the rest of the molecules. The alignment was carried out using the Align module in *Cerius*² [2.27] with 12 atoms selected for superimposition as shown in Figure 2.1.

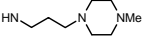
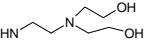
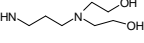
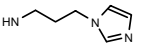
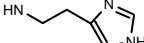
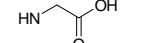
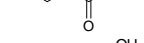

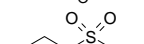
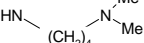
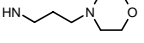
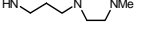
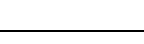

Table 2.1: Molecular structures of 128 EGFR kinase inhibitors used in this study

Molecule No.	Class	Substitution					Actual Activity pIC ₅₀	Predicted Activity	
		R ₁	R ₂	R ₃	R ₄	R ₅		MFA	RSA
1	A	-	-	-	-	-	6.46	6.00	6.09
2	A	Me	-	-	-	-	6.04	7.15	6.57
3	A	Cl	-	-	-	-	7.63	7.32	6.50
4	A	Br	-	-	-	-	7.56	7.96	7.11
5	A	I	-	-	-	-	7.09	6.19	6.93
6	A	CF ₃	-	-	-	-	6.23	6.73	6.93
7	A	Br	NO ₂	-	-	-	6.04	6.35	5.76
8	A	Br	OMe	-	-	-	6.45	7.40	7.40
9	A	Br	-	NO ₂	-	-	6.00	5.78	6.07
10	A	Br	-	OMe	-	-	8.00	8.38	7.67
^a 11	A	Br	OH	OH	-	-	9.76	7.76	8.29
^a 12	A	Br	NH ₂	NH ₂	-	-	9.92	8.08	8.39
13	A	F	-	-	-	-	7.25	8.08	6.83
14	A	-	OMe	-	-	-	7.25	7.45	7.12
15	A	-	NH ₂	-	-	-	6.11	6.28	6.63
16	A	CF ₃	NH ₂	-	-	-	6.24	6.19	6.52
17	A	-	OMe	-	-	-	6.92	6.61	7.53
18	A	-	-	NH ₂	-	-	7.00	6.73	6.74
^a 19	A	CF ₃	-	NH ₂	-	-	8.48	7.06	7.49
20	A	F	-	NO ₂	-	-	5.21	4.58	5.57
21	A	Cl	-	NO ₂	-	-	6.09	6.33	6.11

Molecule No.	Class	Substitution					Actual Activity	Predicted Activity	
		R ₁	R ₂	R ₃	R ₄	R ₅		MFA	RSA
22	A	I	-	NO ₂	-	-	6.26	6.32	6.13
23	A	-	OMe	OMe	-	-	7.53	7.67	8.03
24	A	F	OMe	OMe	-	-	8.42	8.43	8.07
^a 25	A	Cl	OMe	OMe	-	-	9.50	8.47	8.43
^a 26	A	I	OMe	OMe	-	-	9.05	8.50	8.89
27	A	CF ₃	OMe	OMe	-	-	9.61	8.99	8.99
^a 28	A	Br	NHMe	-	-	-	8.39	8.54	7.65
29	A	Br	N(Me) ₂	-	-	-	7.07	7.41	7.22
30	A	Br	NHCOOMe	-	-	-	7.92	8.51	7.23
31	A	Br	-	OH	-	-	8.32	7.81	7.63
32	A	Br	-	NHAc	-	-	7.39	7.67	7.94
^a 33	A	Br	-	NHMe	-	-	8.15	6.22	7.51
^a 34	A	Br	-	NHEt	-	-	7.92	7.76	7.90
^a 35	A	Br	-	N(Me) ₂	-	-	7.95	7.54	7.75
36	A	Br	NH ₂	NHMe	-	-	9.16	8.75	8.74
^a 37	A	Br	NH ₂	N(Me) ₂	-	-	6.79	7.40	8.15
38	A	Br	NH ₂	OMe	-	-	8.42	8.68	8.32
39	A	Br	NH ₂	Cl	-	-	8.18	7.45	7.60
40	A	Br	NO ₂	NHMe	-	-	7.16	7.03	6.96
41	A	Br	NO ₂	OMe	-	-	7.82	7.59	6.96
42	A	Br	NO ₂	Cl	-	-	7.60	7.84	7.79
43	A	Br	OEt	OEt	-	-	11.22	10.16	11.09
44	A	Br	OPr	OPr	-	-	9.76	10.20	9.91
^a 45	A	H	OMe	OMe	Br	-	9.01	7.96	8.04
^a 46	B	Br	-	-	-	-	7.46	7.31	7.18
47	B	Br	NH ₂	-	-	-	8.11	7.98	7.72
48	B	Br	Cl	-	-	-	7.74	7.48	7.34
^a 49	B	Br	F	-	-	-	7.35	7.95	7.42
50	B	Br	NHMe	-	-	-	8.50	8.46	7.85
51	B	Br	N(Me) ₂	-	-	-	8.01	8.51	7.82
52	B	Br	OMe	-	-	-	8.36	8.47	8.03
^a 53	C	Br	-	-	-	-	7.45	7.30	7.28
54	C	Br	-	NHAc	-	-	7.53	7.64	7.68
55	C	Br	-	F	-	-	7.88	7.93	7.40
56	C	Br	-	OMe	-	-	7.40	7.45	8.05
^a 57	C	-	-	NH ₂	-	-	6.60	7.33	7.37
^a 58	C	NO ₂	-	NH ₂	-	-	7.39	7.14	6.71
59	C	-	-	NH ₂	-	Br	6.61	6.35	7.19
^a 60	C	Br	-	NH ₂	-	-	8.00	7.86	7.89
61	C	-	-	NH ₂	Br	-	7.04	6.39	7.21
62	C	-	-	NH ₂	CF ₃	-	5.32	5.33	5.33
63	C	-	-	NH ₂	-	OMe	5.43	6.31	5.17
64	C	OMe	-	NH ₂	-	-	6.88	6.70	6.90
^a 65	C	-	-	NH ₂	OMe	-	6.17	6.55	6.87
^a 66	C	-	-	NH ₂	-	NH ₂	5.27	6.43	6.46
67	C	N(Me) ₂	-	NH ₂	-	-	5.74	5.41	6.12

Molecule No.	Class	Substitution					Actual Activity	Predicted Activity	
		R ₁	R ₂	R ₃	R ₄	R ₅		MFA	RSA
68	C	-	-	NH ₂	N(Me)	-	5.31	5.74	5.05
69	C	F	-	NH ₂	-	-	6.07	6.75	6.83
70	C	Cl	-	NH ₂	-	-	6.92	7.31	7.20
71	C	OH	-	NH ₂	-	-	7.15	7.45	7.46
72	C	Me	-	NH ₂	-	-	7.39	7.29	7.36
73	D	Br	-	-	-	-	7.29	7.33	7.52
74	D	Br	Cl	-	-	-	7.39	7.54	7.64
75	D	Br	F	-	-	-	6.90	7.30	7.58
76	D	Br	OMe	-	-	-	8.58	8.56	8.35
77	D	Br		-	-	-	8.63	8.72	8.58
78	E	Br	-	-	-	-	6.16	6.20	6.02
79	E	Br	-	NH ₂	-	-	6.02	7.35	6.66
80	E	Br	-	F	-	-	6.16	6.26	6.33
81	E	Br	-	NHMe	-	-	7.28	7.41	6.82
82	E	Br	-	N(Me) ₂	-	-	6.48	7.38	7.09
83	E	Br	-	OMe	-	-	6.58	7.34	7.13
84	F	H	NHMe	-	-	-	7.88	8.52	8.38
85	F	Br	Cl	-	-	-	7.08	7.47	8.05
86	F	Br	NH ₂	-	-	-	8.82	7.32	8.31
87	F	Br	NHMe	-	-	-	9.11	8.52	8.38
88	F	Br	N(Me) ₂	-	-	-	9.02	8.51	8.58
89	F	Br	OMe	-	-	-	8.42	8.46	8.62
90	F	Br		-	-	-	9.09	8.45	9.04
91	F	Br		-	-	-	8.53	8.51	9.02
92	F	Br		-	-	-	9.60	9.00	8.95
93	F	Br		-	-	-	8.63	8.65	8.75
94	F	Me	Cl	-	-	-	6.42	7.45	7.53
95	F	Me	NH ₂	-	-	-	7.76	7.25	7.79
96	F	Me	NHMe	-	-	-	8.36	7.41	7.59
97	F	Me	N(Me) ₂	-	-	-	8.39	7.95	8.52
98	F	Me		-	-	-	8.63	8.94	8.85

Molecule No.	Class	Substitution					Actual Activity	Predicted Activity	
		R ₁	R ₂	R ₃	R ₄	R ₅		MFA	RSA
99	F	Me		-	-	-	8.52	8.42	8.51
^a 100	C	Br		-	-	-	9.61	8.10	8.27
101	C	Br		-	-	-	8.58	8.77	8.32
102	C	Br		-	-	-	9.03	8.43	8.54
103	C	Br		-	-	-	8.49	7.99	9.23
104	C	Br		-	-	-	7.85	7.51	7.90
105	C	Br		-	-	-	7.92	8.81	7.87
^a 106	C	Br		-	-	-	7.34	8.36	8.33
107	C	Br		-	-	-	8.05	8.38	8.55
^a 108	C	Br		-	-	-	8.13	9.00	8.74
109	C	Br		-	-	-	8.07	8.37	8.47
110	C	Br		-	-	-	7.39	6.52	7.78
111	C	Br		-	-	-	8.49	8.38	8.48
^a 112	C	Br		-	-	-	8.72	8.93	8.86
113	C	Br		-	-	-	8.26	8.38	8.68

Molecule No.	Class	Substitution					Actual Activity	Predicted Activity	
		R ₁	R ₂	R ₃	R ₄	R ₅		MFA	RSA
114	C	Br		-	-	-	8.30	8.38	8.57
115	C	Br		-	-	-	8.03	8.07	8.38
116	C	Br		-	-	-	8.92	8.37	8.77
117	C	Br	NHNH ₂	-	-	-	8.14	7.90	8.22
^a 118	C	Br		-	-	-	9.29	7.90	8.21
119	C	Br		-	-	-	9.04	8.99	8.99
120	C	Br		-	-	-	8.82	8.98	8.97
121	C	Br		-	-	-	9.21	8.97	8.67
122	C	Br		-	-	-	9.55	8.97	8.67
123	C	Br		-	-	-	7.79	7.35	8.00
124	C	Br		-	-	-	8.85	8.33	8.54
125	C	Me		-	-	-	8.26	8.37	7.97
126	C	Me		-	-	-	8.03	8.45	8.18
127	C	Me		-	-	-	8.25	8.37	8.27
128	C	Me		-	-	-	8.45	8.95	8.76

^aTest set

2.2.3 MFA

At the molecular level, the interactions involved in ligand binding are usually non-covalent in nature. In other words, ligands interact with the receptor through their molecular fields (steric and electrostatic fields) [2.28]. Therefore a suitable sampling of these molecular fields, inherent to 3D structures, surrounding a set of molecules would provide the necessary information for understanding their observed biological properties. This forms the basis of a MFA (Figure 2.2). The most important variable in any 3D-QSAR methodology is the “alignment rule”, i.e. the positioning of a molecule within the fixed lattice. Because the relative interaction energies strongly depend on relative molecular positions, all the study molecules were placed in a 3-dimensional grid with a prior alignment.

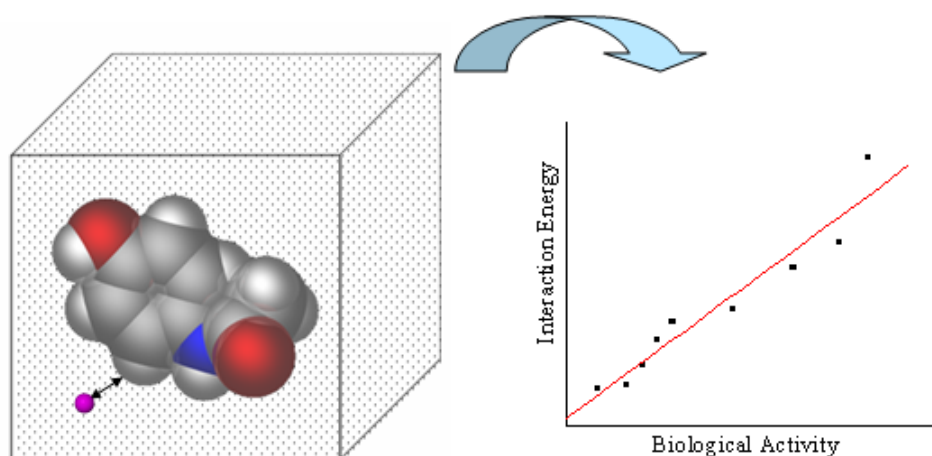


Figure 2.2: A schematic representation of MFA methodology

Model Generation

MFA studies were performed with the QSAR module of *Cerius*². The molecular field was created using proton and methyl groups as probes, which represent electrostatic and steric fields respectively. These fields were sampled at each point of a regularly spaced grid of 1 Å. The coordinates of the MFA grid box were (−3.547, −6.216, −7.450)

and (14.452, 17.783, 6.540) for the lower and upper corners respectively. The total grid points generated were 7125. A number of spatial and structural descriptors such as polarizability, dipole moment, radius of gyration, number of rotatable bonds, molecular volume, principal moment of inertia, AlogP, number of hydrogen bond donors and acceptors and molar refractivity were considered along with the steric and electrostatic fields. Only 10% of the total descriptors whose variance was highest were considered for further analysis. Regression analysis was carried out using Genetic Partial Least Squares (G/PLS) method consisting of over 50,000 generations with a population size of 100. The optimal number of components was set to 5. An energy cutoff of $\pm 30.0 \text{ kcal mol}^{-1}$ was set for both steric and electrostatic contributions. The smoothing parameter, d was set to 1.0 to control the bias in the scoring factors between equations with different number of terms. The length of the final equation was fixed to 9 terms. Cross-validation was performed with the leave-one-out procedure. PLS analysis was scaled with all variables normalized to a variance of 1.0.

2.2.4 RSA

In the absence of three dimensional structure of the target, a hypothetical receptor model can be generated from a series of aligned molecules with associated binding affinities. A surface was generated to represent the molecular shape of the aggregate of aligned molecules. Once the surface was built, putative chemical properties like partial charge, electrostatic potential, hydrogen bonding propensity and hydrophobicity are mapped onto it considering the complementarity between the receptor and the molecules binding to it. The contribution of each molecule in the generation of receptor surface is proportional to its biological activity.

Each molecule was docked into this generated receptor. The interaction energy between the molecule and the receptor was evaluated and added to the study table. Also, the interaction energies of each molecule with different probes such as methyl group (steric) and a proton (electrostatic) positioned along the grid points throughout the receptor surface were added to the study table. Only 10% of the total descriptors whose

variance was highest were considered as independent data to perform further analysis. Regression analysis was carried out using the G/PLS method as mentioned above.

2.2.5 Docking

The crystal structure of EGFR kinase (1M17) [2.29] bound with inhibitor erlotinib (Tarceva[®], Genentech/OSI Pharmaceuticals Inc.) {[6,7-bis(2-methoxy-ethoxy)quinazoline-4-yl]-(3-ethynylphenyl) amine} was taken from the protein data bank. Initially, hydrogen atoms were added to the protein considering all the residues in their neutral form. The added hydrogens and water molecules were minimized while keeping all the heavy atoms fixed. The minimization followed steepest descent and conjugate gradient algorithms for at least 3000 iterations each and a convergence gradient of 0.01 and 0.001 kcal mol⁻¹ respectively, using CHARMM [2.30] force field in *Insight II* [2.31]. The active site was defined to constitute all the amino acid residues within 10 Å radius from the inhibitor.

The most active molecule **43** was docked into the active site of EGFR kinase. Docking was carried out using GOLD [2.32], which uses the genetic algorithm (GA). For each of the 10 independent GA runs, a maximum number of 100000 GA operations were performed on a set of 5 groups with a population size of 100 individuals. Operator weights for crossover, mutation, and migration were set to 95, 95 and 10 respectively. Default cut-off values of 2.5 Å (d_{H-X}) for hydrogen bonds and 4.0 Å for non-bonded interactions were employed. When the top three solutions attained RMSD values within 1.5 Å, docking was terminated. The best ranked solution of the ligand was used for further analysis.

2.3 Results and Discussion

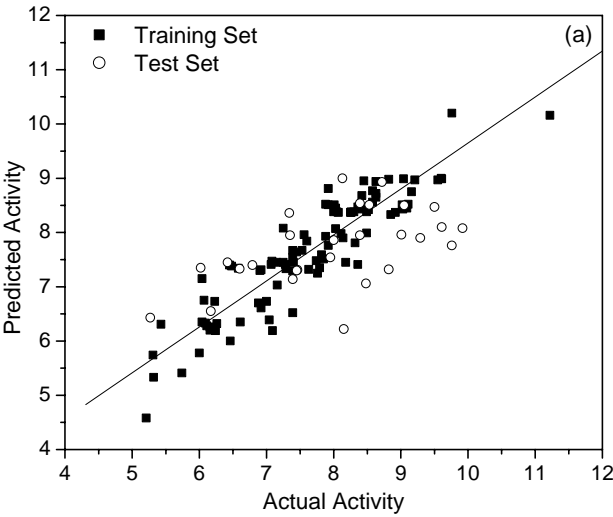
The statistical details of the 3D QSAR models are given in Table 2.2. The actual and predicted activities, obtained from MFA and RSA 3D-QSAR models, of both the training and test sets are listed in Table 2.1. Scatter plots of actual versus predicted

activities for both training and test set molecules obtained from MFA and RSA are shown in Figure 2.3.

Table 2.2: Statistical details for MFA and RSA models

	MFA	RSA
^a r^2_{cv}	0.81	0.79
^b r^2	0.84	0.85
^c PRESS	23.75	26.35
^d N	5.00	5.00
^e r^2_{pred}	0.76	0.84
^f r^2_{bs}	0.84	0.82
^g LSE	0.18	0.19

^across-validated r^2 ; ^bconventional r^2 ; ^cpredicted sum of squared residuals; ^dnumber of components; ^epredictive r^2 ; ^fbootstrap r^2 ; ^gleast square error



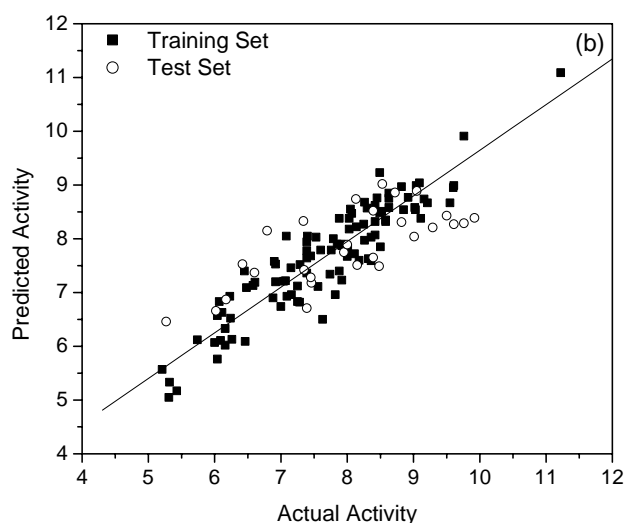


Figure 2.3: Scatter plots of actual vs. predicted activity for both training and test sets using (a) MFA (b) RSA.

2.3.1 MFA–Results

An MFA model with a cross-validated r^2 (r^2_{cv}) value of 0.81 and a conventional r^2 0.84 was developed. The higher r^2_{cv} of MFA than that of RSA makes it possible to interpolate the order of activities of the training set molecules. The steric (CH_3) and electronic (H^+) descriptors in the QSAR equation of MFA (2.1) specify the regions where variations in the structural features (steric or electronic), of different compounds in the training set, lead to increased or decreased activities. The number associated with the descriptors in the equation represents their position in the 3-dimensional MFA grid.

$$\begin{aligned} \text{Activity} = & 6.3273 + 0.018513 (\text{H}^+/1223) + 0.03961 (\text{CH}_3/2871) \\ & - 0.038226 (\text{H}^+/1194) - 0.028351 (\text{H}^+/5124) + 0.03765 (\text{CH}_3/2315) \\ & + 0.035301 (\text{H}^+/1239) + 0.023949 (\text{CH}_3/5111) + 0.038927 (\text{H}^+/5139) \end{aligned} \quad (2.1)$$

The presence of ($\text{H}^+/1239$) and ($\text{H}^+/1223$) descriptors in the vicinity of the *meta* position of the aniline ring emphasizes the electronic environment required at this position (Figure 2.4). Any moderately electron withdrawing group like a halogen at this

position could enhance the EGFR kinase inhibitory activity. Our observation is in agreement with the SAR studies of Rewcastle *et al* [2.14]. The presence of a steric descriptor ($\text{CH}_3/2315$) close to this position indicates the importance of steric interactions. A methyl group at this position would satisfy only the steric requirement but not the electronic requirement while Br, an isostere of methyl, would fulfill both steric and electrostatic stipulations. Hence, the activity of molecules **4**, **60**, **85**, **86**, **87**, **88** and **90** are higher respectively than that of **2**, **72**, **94**, **95**, **96**, **97** and **98**, wherein Br replaces a methyl group. The presence of two electronic descriptors near the 7-position of the quinazoline ring, *viz.*, ($\text{H}^+/5124$) with a negative coefficient and ($\text{H}^+/5139$) with a positive coefficient describes a subtle balance of electronic parameters required at this position. Also, the presence of a steric descriptor ($\text{CH}_3/5111$) at this position implies that it can accommodate bulkier substituents at this position. The need for moderate electron donating groups with appropriate steric parameters is evident with the trend shown by the molecules **79** < **82** < **83**.

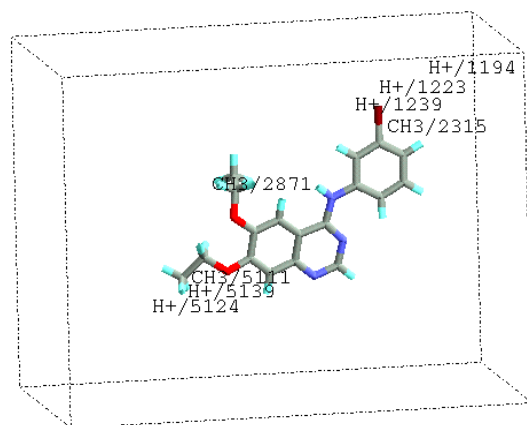


Figure 2.4: Molecule **43** enclosed inside the grid with 3D points of the QSAR equation.

A high value of r_{cv}^2 alone however, is an insufficient criterion for a QSAR model to be robust and highly predictive [2.33]. The predictive power of the model was

therefore validated with the test set molecules. The predictive power of the model generated, is calculated by

$$r_{pred}^2 = \frac{SD - PRESS}{SD} \quad (2.2)$$

where, SD is the sum of squared deviations between the biological activities of each molecule and the mean activity of the training set molecules and PRESS is the sum of squared deviations between the predicted and actual activity values for every molecule in the test set. The prediction of the model was reasonably good with predictive r^2 (r_{pred}^2) value of 0.76.

2.3.2 RSA–Results

A receptor surface model with r_{cv}^2 0.79 and conventional r^2 0.85 was developed. The QSAR equation generated by RSA is given in equation (2.3)

$$\begin{aligned} \text{Activity} = & 6.2786 - 7.34427 (\text{Ele}/1391) + 3.30427 (\text{VDW}/4591) \\ & + 10.2277(\text{VDW}/2861) - 2227.15 (\text{Ele}/3761) + 6.31853 (\text{VDW}/3761) \\ & - 4.42666 (\text{Ele}/3841) + 1.73629 (\text{Ele}/4081) + 0.005889 (\text{MW}) \end{aligned} \quad (2.3)$$

The equation consists of eight molecular field descriptors. The positioning of these descriptors, indicated by a number along with the descriptors, on the model explains the nature of the substituents required. The presence of both steric (VDW/2861) and electronic (Ele/3761) descriptors at the *meta* position of the anilino group explains the need of a bulky and electron withdrawing substituent here. Br, which satisfies both these criteria, is the correct choice. The same was apparent from MFA discussed earlier. The steric descriptor (VDW/4591) present near the 6-position of the quinazoline ring indicates that bulky linear substituents are necessary at this position. Further, the electronic descriptors (Ele/1391) and (Ele/3841) with negative coefficients in the 6- and 7-positions of the quinazoline ring provide further guidelines. Inclusion of MW as one of

the crucial descriptors in the equation is in accordance with the Lipinski's rule of five [2.34].

When charge is mapped on to the receptor surface, it shows a red colour around the *meta* position of the aniline ring (Figure 2.5). This indicates that an electron-withdrawing group in this position would be essential. Further, there are two red regions around the 1- and 3- positions of the quinazoline ring and a blue region at the 2-position. This indicates that electronegative atoms are vital in 1- and 3-positions. The electron withdrawing nature of these atoms makes the adjacent region electron deficient. Thus, the region near the 2-position is blue (electropositive). Similarly, a blue region is in juxtaposition with the red region at the 4-position, indicating that the anilino nitrogen is crucial. The RSA model was further validated using the same test set molecules as in MFA. RSA predicted 93% of molecules correctly with an error less than 1.5.

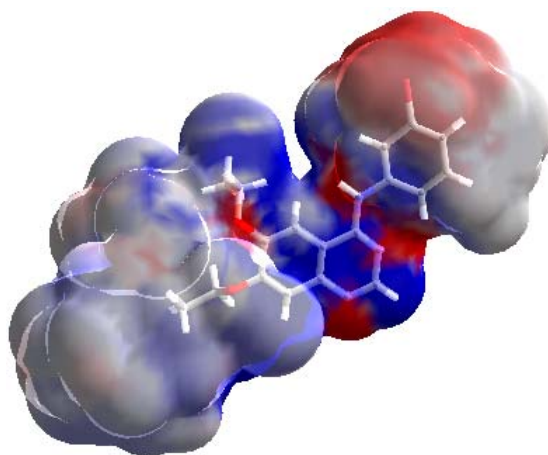


Figure 2.5: Receptor surface model with charge mapped onto it. Positive regions are shown in red and negative regions shown in blue. Molecule **43** is placed within the receptor surface generated.

2.3.3 MFA vs. RSA

The r^2_{pred} values, obtained for both the models are listed in Table 2.2. RSA exhibited higher predictivity with r^2_{pred} values 0.84, while MFA had a lower r^2_{pred} value 0.76. However, MFA had a higher value of r^2_{cv} . This r^2_{cv} of MFA indicates good internal consistency but the lower r^2_{pred} value implies that it cannot predict molecules outside the training set better than RSA. While MFA takes into consideration only steric and electronic parameters, RSA considers hydrophobicity also along with these parameters. Inclusion of this additional parameter to develop the model could be one of the reasons of its higher predictivity over MFA.

2.3.4 Structural Requirements at the 6- and 7-Positions

In the receptor surface model, the presence of steric descriptor (VDW/4591) only near the 6-position but not the 7-position shows that the substitutional requirements for these two positions are different. From this, it is quite evident that bulky substituents are tolerated in the 6-position but not in the 7-position. Also, when the receptor surface model was mapped with various properties like charge, hydrophobic and hydrogen bonding capabilities, certain variations in the structural features of the substituents at 6- and 7-positions were observed.

- 1) Pale red colour (electronegative) near the 6-position as in Figure 2.5
- 2) Light brown colour (less hydrophobic) at the 6-position as in Figure 2.6a
- 3) Pale green colour (hydrogen bond acceptor) at the 6-position as in Figure 2.6b.

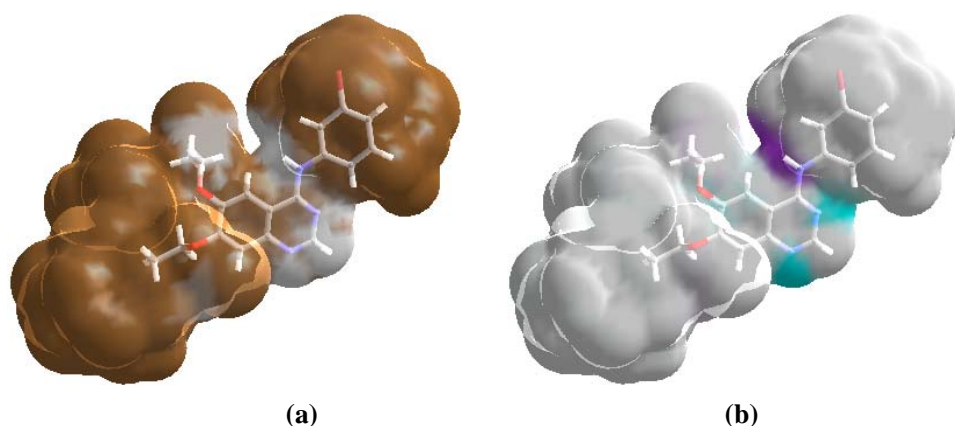


Figure 2.6: Receptor surface model (a) mapping hydrophobic property. Hydrophobic regions are shown in brown and hydrophilic regions in white; (b) mapping hydrogen bonding property. Hydrogen bond donor property is shown in purple and hydrogen bond acceptor property is shown in green. Molecule **43** is placed within the receptor surface generated.

All the above features indicate that substituents in the 6-position should be more electronegative, hydrophilic and possess greater hydrogen bond acceptor propensity than substituents in the 7-position. Though the differences at these positions are seemingly minimal, they appear to be significant in ligand binding. This is contrary to the reports of Rewcastle *et al.* [2.14], which generally state that electron donating groups at both the 6- and 7-positions are acceptable. The above observations clearly differentiate the nature of substitutions at these positions and are different from those reported by Hou *et al.* [2.16]. Till date such differences have not been described although some of the promising drug candidates like Iressa, Lapatinib and EKB-569 satisfy these criteria (see Scheme 2.1).

2.3.5 Correlation of RSA Results with Docking

The results of RSA were compared with docking studies using the crystal structure of EGFR kinase domain that was released during the progress of this work. The most active molecule **43** was docked into the active site using GOLD. The active site of the protein with the docked molecule is shown in Figure 2.7a. The molecule sits in the active site in such a way that its aniline ring protrudes inside one of the hydrophobic

pockets, while the substituent in the 7-position faces another hydrophobic region. Also, the substituent in the 6-position is directed towards the hydrophilic region that is occupied by the phosphate chain of ATP (substrate). The Gauss–Connolly surface (computed in *MOE* [2.35]) of the active site (Figure 2.7b) also shows a slight variation in the regions around the 6- and 7-positions. A pale blue (electropositive) region is seen near the 7-position, while the region around the 6-position is pale red (electronegative). This indicates that the receptor surface around the 6-position is relatively electronegative when compared to that around the 7-position. Thus, all the above observations concur well with that of the crystal structure.

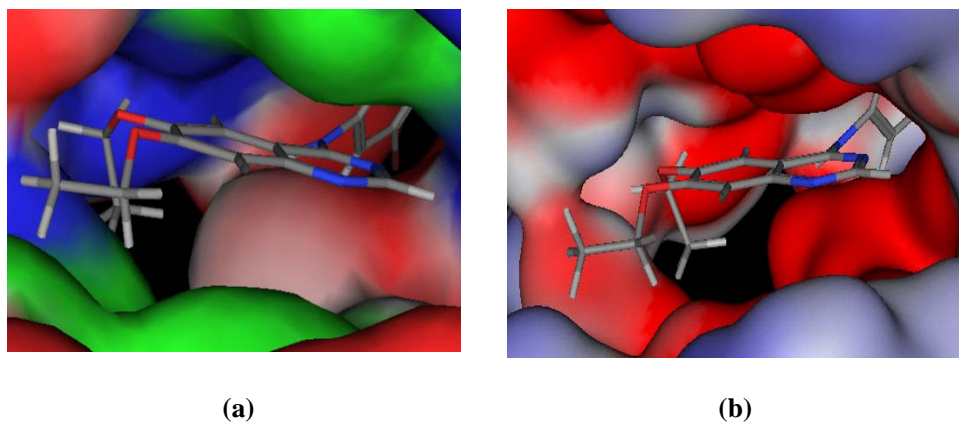


Figure 2.7: Active site of EGFR with molecule **43** docked into it. **(a)** Hydrophilic regions are shown in blue, hydrophobic regions in green and solvent accessible surface in red. **(b)** Gauss–Connolly surface with blue regions indicating electropositive and red regions electronegative character.

2.4 Conclusions

Two different analog based methods, MFA and RSA models were developed and compared for anilinoquinazoline derivatives as EGFR kinase inhibitors. These approaches are different in that while MFA identifies the chemical features of the molecules required for the enhanced biological activity, RSA attempts to postulate and represent the essential features of the receptor site itself. Although, both these methods approach the problem from opposite directions, they provide similar information about

the structural requirements for the biological activity with RSA exhibiting a better predictive capability than MFA. The inclusion of hydrophobic field in RSA in addition to the normal steric and electronic parameters used in MFA could be one of the reasons for its higher predictive capability. The model differentiates the 6- and 7-positions and indicates that substituents at the 6-position should have higher electronegativity, hydrophilicity, hydrogen bonding propensity and bulky in size. These observations were compared and confirmed with structure based studies like docking carried out with the most active molecule **43** of the series. The results provide guidelines for the design of potential inhibitors of EGFR kinase.

CHAPTER 3

VIRTUAL SCREENING MODEL FOR EGFR KINASE INHIBITORS: IMPORTANCE OF HYDROGEN BONDS IN THE EVALUATION OF POSES AND SCORING FUNCTIONS

3.1 Introduction

In order to cope with the dramatic post-genomic scenario in drug design and development, new strategies need to be developed to speed up the discovery and optimization of leads [3.1]. VS represents an efficient strategy that allows selection of possible hits from a very large library of compounds (as high as 10^{12}) by computational methods [3.2]. The reliability and utility of specific VS tools grow with the knowledge available on a particular drug target and the underlying molecular recognition [3.3]. In the cases where there is no structural information of the target, pharmacophore models [3.4] can serve as powerful tools for VS. However, if the 3D-structure is available, structure based tools may be employed directly, or in combination with pharmacophore models and filters such as RO5 [3.5]. The most widely used form of VS employs protein–ligand docking.

VS, especially in the context of docking methodologies, is an evolving challenge and an attractive research proposition. It needs to be studied in depth and transformed into a tool of greater confidence and utility. VS consists of two parts, namely, the accurate prediction of pose and the estimation of tightness of binding (scoring). A number of protein–ligand complexes have been used to validate the performance of the several docking programs [3.6]. Considerable efforts have also been directed to the development and validation of scoring functions, which quantify the protein–ligand interactions [3.7]. While the effects of hydration, polarity, flexibility and entropic contributions have often been neglected in deriving scoring functions, success in deriving these functions depends on the availability of a large amount of good structural data. Studies that seek to compare computational approaches for VS have identified that some approaches are good at predicting poses, while others are good in computing

scores. Arriving at the best docking/scoring combination is still a challenge. The goal of this chapter is to arrive at the best combination of docking and scoring for the EGFR kinase target and to develop a robust VS model.

3.2 Methods

3.2.1 Generation of Ligand and Enzyme Structures

The crystal structure of EGFR kinase domain with its bound inhibitor Erlotinib {[6,7-bis(2-methoxy-ethoxy)quinazoline-4-yl]-(3-ethynylphenyl)amine} was obtained from the Protein Data Bank (1M17) [2.29]. H-atoms were added at pH 7. All five water molecules in the active site were included; the acidic and basic residues in the active site were considered to be in their ionic form. The protein was subjected to minimization using steepest descent (gradient $< 0.1 \text{ kcal mol}^{-1}$) and conjugate gradient algorithms (gradient $< 0.01 \text{ kcal mol}^{-1}$) using the CHARMM force field in *Insight II*. The active site was defined as consisting of all amino acids within a 15 Å radius from the center of the ligand. Histidines (His781 and His811) within the active site were set to the δ -tautomer. However, they are too far from the ligand ($> 12 \text{ Å}$) to have any real impact in the docking studies.

All the docking studies were performed with the same set of 128 anilinoquinazoline derivatives discussed in chapter 2.

3.2.2 Docking

GOLD

Docking using GOLD, that uses genetic algorithm, has been carried out with the standard parameters as discussed in chapter 2. Both GOLDScore and Chemscore were calculated.

LigandFit [3.8]

Docking was performed with Monte Carlo minimization using the CFF95 force field. Two different simulations were performed with 20000 and 99999 iterations. The electrostatic energy was included. A grid resolution was set to 0.5 Å (default), and the

ligand-accessible grid was defined such that the minimum distance between a grid point and the protein is 2.0 Å for hydrogen and 2.5 Å for heavy atoms. The grid extends beyond the defined active site to a distance of 3 Å in all directions. This grid was used to calculate the nonbonded interactions between all the atoms of ligands and protein residues. Nonbonded cut-offs were set to 10 Å while using a distance dependent dielectric constant. After each iteration, 10 iterations of minimization were performed before being considered in the next iteration and the selected configuration was minimized for 100 iterations. Flexible fit, wherein the ligand conformational space is explored with different initial poses, was selected. To avoid identical conformations, an RMSD cut-off of 1.5 and a score cut-off of 20 kcal mol⁻¹ were maintained while saving the final conformations. The top 20 conformations were saved after rigid body minimizations of 1000 steps. Dockscore, Ligscore1 and Ligscore2, LUDI, PMF, PLP1 and PLP2, JAIN, hydrogen bonding and vdW scores were calculated for each of the 20 saved ligand conformations.

3.2.3 Docking Methods

Genetic Algorithm [3.9]

This docking approach adapts the principles of biological competition and population dynamics. Initially a random population is generated in which each member (chromosome) corresponds to a potential solution of the problem. Parameters like translational, rotational and torsion angle degrees of freedom along with hydrogen bond patterns between protein and ligand are encoded as genes in a chromosome. Genetic operators like *crossover* (genes from two distinct chromosomes are interchanged to generate a new population) and *mutation* (genes are stochastically varied) are applied to generate a new population. For each newly generated individual the chromosome is decoded into conformation (genotype → phenotype) and evaluated by fitness function. The new conformation with a better fitness value replaces the least fit members of the population to create an ensemble of ligand conformations.

Monte Carlo Minimization

It is a stochastic search method consisting of two essential components: a random walk of the ligand through the receptor generating a new solution with random translation rotation and/or torsional angle variation and evaluation of new conformation based on Metropolis criterion. This criterion decides whether a new position is accepted or not and hence the conformation from where the search will proceed. If the energy of the new position (E_{new}) scores favorable (lower) than the previous one (E_{old}), the new position is accepted. On the other hand, if it is less favorable then the probability for its acceptance is given by

$$P = \exp\left[-\frac{(E_{new} - E_{old})}{kT}\right] \quad (3.1)$$

where k is the Boltzmann constant and T is the effective temperature. The solutions are collected iteratively until a desired number is obtained. Additionally, a minimization step is included between the random walk and the Metropolis selection in the Monte Carlo minimization method.

3.2.4 Scoring Functions**GOLDscore [3.10]**

GOLDscore, a force field based scoring function, is a sum of protein–ligand complexation term, hydrogen bonding term and an internal energy term.

$$\begin{aligned} E_{total} &= E_{complex} + E_{H-bond} + E_{internal} \\ &= \sum_{protein} \sum_{ligand} \left(\frac{A}{d_{ij}^8} - \frac{B}{d_{ij}^4} \right) + \sum_{protein} \sum_{ligand} [(E_{da} + E_{ww}) - (E_{dw} + E_{aw})] + \\ &\quad \left\{ \sum_{ligand} \left(\frac{C}{d_{ij}^{12}} - \frac{D}{d_{ij}^6} \right) + \sum_{ligand} \frac{1}{2} V \left[1 + \frac{n}{|n|} \cos(n|\omega) \right] \right\} \end{aligned} \quad (3.2)$$

The first term represents the van der Waals energy calculated by the softer 4-8 potential. The parameters A and B are chosen so that the 4-8 potential have the same minimum as the Lennard-Jones 6-12 potential. d_{ij} represents the distance between the interacting atoms i and j . Hydrogen bond energy between the donor and acceptor types are pre-calculated with the model fragments along with accounting for water displacement. Initially the donor (d) and acceptor (a) are in solution, but upon forming hydrogen bonds (da) water is stripped (ww) off. Therefore, the energy term included accounts for each of these factors. The last term indicating the internal energy of the ligand is the sum of steric and torsional energy.

Chemscore

Chemscore is an empirical scoring function (equation 3.3) based on the work of Eldridge *et al.* [3.11]

$$\Delta G = \Delta G_{H-bond} \sum_{H-bond} f(\Delta R, \Delta \alpha) + \Delta G_{metal} \sum_{metal} f(\Delta R, \Delta \alpha) + \Delta G_{lipo} \sum_{lipo} f(\Delta R) + \Delta G_{rotor} \sum_{rotor} f(P_{nl}, P'_{nl}) + \Delta G_0 \quad (3.3)$$

while the first term accounts for hydrogen bonding, the second term represents the coordinate bonding between the ligand and the metal ions residing inside the protein binding site. The hydrophobic effect as given by the third term is calculated by distance-dependent potentials of all the hydrophobic atom pairs formed upon complexation. The fourth term counts rotatable bonds with each rotor scaled to reflect the chemical nature of its environment. The last term is the regression constant. The scoring function is calibrated by reproducing the measured dissociation constants of 82 protein–ligand complexes.

Dockscore [3.12]

Dockscore is a classical force field based energy function, which sums van der Waals (vdW) and electrostatic interactions within the complex.

$$E_{interaction} = \sum_{protein} \sum_{ligand} \left[\left(\frac{A}{d_{ij}^{12}} - \frac{B}{d_{ij}^6} \right) + 332.0 \frac{q_i q_j}{\epsilon (d_{ij}) d_{ij}} \right] \quad (3.4)$$

Ligscore

It is an empirical scoring function used to calculate the pK_i of the ligand by equation (3.5)

$$pK_i = A - B(vdW) + C(C_{+pol}) - D(Totpol^2) \quad (3.5)$$

where vdW is a softened Lennard-Jones 6-9 potential, C_{+pol} is a count of the buried polar surface area between the complex involving attractive protein–ligand interactions. $Totpol^2$ is the square of the buried polar surface area between the complex involving both attractive and repulsive protein–ligand interactions.

LUDI

An empirical scoring function developed by Böhm [3.13] dissects the protein–ligand binding free energy as

$$\begin{aligned} \Delta G_{bind} = & \Delta G_{H-bond} \sum_{H-bond} f(\Delta R, \Delta \alpha) + \Delta G_{ionic} \sum_{ionic} f(\Delta R, \Delta \alpha) + \\ & \Delta G_{hydrophobic} \sum |A_{hydrophobic}| + \Delta G_{rotor} N_{rotor} + \Delta G_0 \end{aligned} \quad (3.6)$$

The first two terms accounts for the hydrogen bonds formed between the complex, where “neutral and “ionic” hydrogen bonds are treated separately. The contribution of each hydrogen bond is scaled by a distance- and angle-dependent function in order to penalize the deviations from an ideal geometry. The third term accounts for the hydrophobic effect, which calculates the buried hydrophobic molecular surface. The fourth term counts all the rotatable bonds (rotors) in the ligand, which is supposed to be related to the torsional entropy loss of the ligand upon protein–ligand complexation. The last term is a regression constant. The scoring function was calibrated by fitting known dissociation constants of 87 protein–ligand complexes.

Potential Mean Force (PMF)

Developed by Muegge *et al.*[3.14] this knowledge based scoring function is derived from distance dependent interaction potentials for various atom pairs obtained from 697 protein–ligand complexes from PDB. It implicitly includes both the enthalpic and entropic effects. In essence, knowledge based scoring functions are designed to reproduce experimental structures rather than binding energies. The protein–ligand interaction energy is defined as a sum of potentials over all heavy atom pairs between the complex given by equation 3.7

$$PMF_score = \sum_{protein} \sum_{ligand} A_{ij}(r)$$

where

$$A_{ij}(r) = -k_B T \ln \left[f_{Vol_corr}^i(r) \frac{\rho_{seg}^{ij}(r)}{\rho_{bulk}^{ij}} \right] \quad (3.7)$$

k_B is the Boltzmann constant, T is the absolute temperature and r is the atom pair distance.

$f_{Vol_corr}^j(r)$ is the ligand volume correction factor, $\rho_{seg}^{ij}(r)$ is the number density of pairs of type ij in a structural database that occur in a certain radius range as indicated by “seg” and ρ_{bulk}^{ij} represents distribution of i and j atoms when no interaction occurs between them.

Pair wise Linear Potential (PLP) [3.15]

This empirical scoring function is a sum of pair wise linear potentials between ligand and protein heavy atoms with parameters dependent on interaction type. Initially all the protein and ligand heavy atoms are classified as hydrogen bond donors, acceptors, donor/acceptor and non polar. Each pair of interacting atoms is then assigned one of the three interaction types: hydrogen bonding between donors and acceptors, repulsive donor-donor and acceptor-acceptor contacts, and generic dispersion of other contacts. The score is given in terms of total Energy as shown in equation 3.8

$$E_{total} = E_{H-bond} + E_{repulsion} + E_{contact} \quad (3.8)$$

Two different versions of scoring function: PLP1 and PLP2 differ in the parameter sets and algorithms used in.

JAIN [3.16]

This is a regression based scoring function that is both fast and tolerant to inaccurate ligand orientations. This approach has a master equation 3.9 that includes terms tuned with neural network-based functions like Gaussian (g), sigmoidal (s) (equation 3.10), and distance (d) (equation 3.11) functions

$$F = \sum_{i,j} f_0(d(i,j)) + \sum_{i,j} f_1(d(i,j), i, j) + \sum_{i,j} f_2(d(i,j), i, j) + (l_5.phbe) + (l_6.lhbe) + (l_7.n_rot) + (l_8.log(mol.wt)) \quad (3.9)$$

$$g(x, \mu, \pi) = e^{-(x+\mu)^2/\sigma} \text{ and } s(x, \mu) = 1/(1 + e^{10(x+\mu)}) \quad (3.10)$$

$$d(i, j) = [(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2]^{1/2} - r_i - r_j \quad (3.11)$$

where f_0 is the hydrophobic term, f_1 is the polar contribution, f_2 estimates the repulsive term. Solvation terms are given by $(l_5.phbe)$ and $(l_6.lhbe)$ where $phbe$ and $lhbe$ represent the difference between potential hydrogen bonding equivalents and the actual polar interaction term in both protein and ligand respectively. Entropic term includes $(l_7.n_rot)$ and $(l_8.log(mol.wt))$. Tunable linear parameters are denoted by l , whereas nonlinear parameters are denoted n .

3.3 Results and Discussion

EGFR kinase is one of the few well explored and validated targets in anti-cancer therapy [3.17]. In order to develop a good VS model for EGFR, the choice of 128 ligands was based on the following criteria:

1. The compounds must have good structural diversity
2. Binding studies should have been carried out by the same biological assay

3. Structural similarity with erlotinib (in the crystal structure) should be present. This enables to evaluate the RMSD of the poses with that of the experimental structure.

3.3.1 Direct Docking

The main objective of the present work is to perform a docking analysis of selected EGFR kinase inhibitors. Initially, the docking studies were carried out by excluding crystallographic water molecules using GOLD. These studies resulted in poor docking of ligands in the active site and therefore, showed poor correlation between the GOLD scores and the activities. Careful observation of the crystal structure shows a water mediated interaction between the receptor and the ligand [2.29]. Figure 3.1 shows the hydrogen bonds and the hydrophobic environment around the ligand in the receptor.

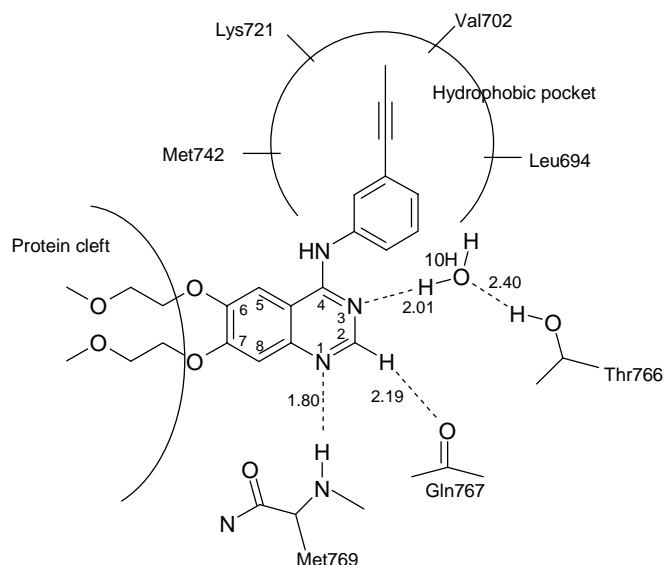


Figure 3.1: Schematic representation of protein–ligand interactions in the crystal structure.

The N₃ of the pyrimidine ring accepts an O_w–H...N hydrogen bond from a water molecule (10H) which in turn form a cooperative [3.19] O–H...O_w hydrogen bond involving Thr766 of the protein [3.20]. Since, the water molecules were not considered,

these crucial hydrogen bonds were not included resulting in poor correlation between the scores and the observed activities. Therefore, all the molecules were re-docked while including this water. Although improvement was observed, the correlation between the inhibitory activities and the scores was still disappointing (Figure 3.2).

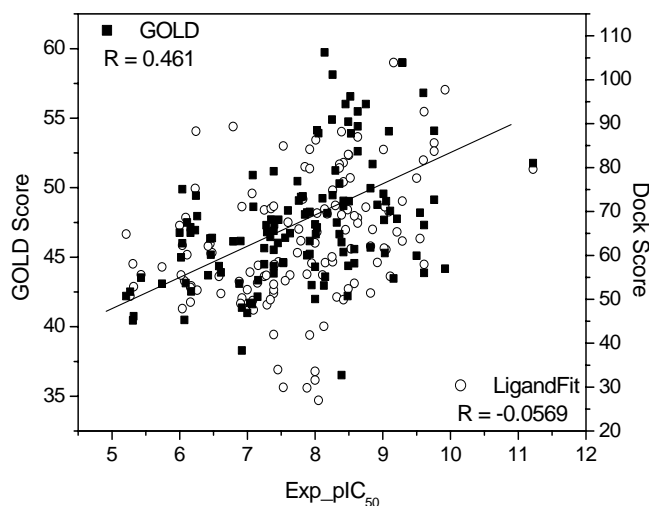


Figure 3.2: Scatter plot of GOLDscore and Dockscore (with 99999 Monte Carlo Iterations) vs activity.

The discouraging results from GOLD lead to the use of alternative docking programs like LigandFit. Scatter plot of scores obtained from each of these methods against the activity (pIC_{50}) of the ligands is shown in Figure 3.2. GOLD seems to be slightly better than LigandFit. Initially, LigandFit was used with only 20000 Monte Carlo iterations resulting in unacceptable results with the positive nonbonding energies for most of the ligands as shown in Figure 3.3. However, when the docking was rerun with a maximum number of iterations (99999), there was a slight improvement but the data was still too scattered with little to no correlation (Figure 3.2). This could be due to inaccuracy in either the prediction of poses and/or the computation of the scores.

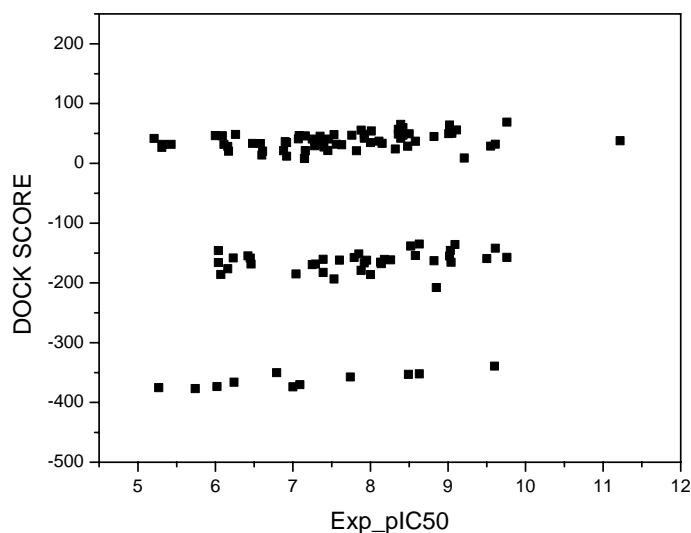


Figure 3.3: Scatter plot of Dockscore (Ligand Fit after keeping 20000 Monte Carlo iterations) vs Experimental pIC_{50}

3.3.2 Molecular Recognition of Erlotinib in Crystal Structure

A thorough investigation of the crystal structure (Figure 3.4) showed that in addition to the hydrogen bonds mediated by water discussed in the previous section, there are two other significant hydrogen bonds. The N_1 of quinazoline ring accepts an $N-H\cdots N$ hydrogen bond from the backbone $N-H$ of Met769 and a $C-H\cdots O$ hydrogen bond formed between the activated C_2-H group and the backbone carbonyl oxygen of Gln767. In effect, these three hydrogen bonds, which are worth around $3-5 \text{ kcal mol}^{-1}$ each, (between $10-15 \text{ kcal mol}^{-1}$ collectively, or roughly a quarter of the total binding enthalpy) anchor the ligand in the active site. Additionally, the 6- and 7-substituents project outside the active site (actually, in the cleft region between the C-terminal and N-terminal lobes) and the *meta*-substituted 4-anilino group fits well in the Thr766–Leu764–Met742–Lys721–Val702 hydrophobic pocket. Effectively, these weak [3.21] hydrogen bonds provide a template effect (chemical recognition) but further activity depends on the insertion of the 4-anilino ring into the hydrophobic pocket (geometrical recognition).

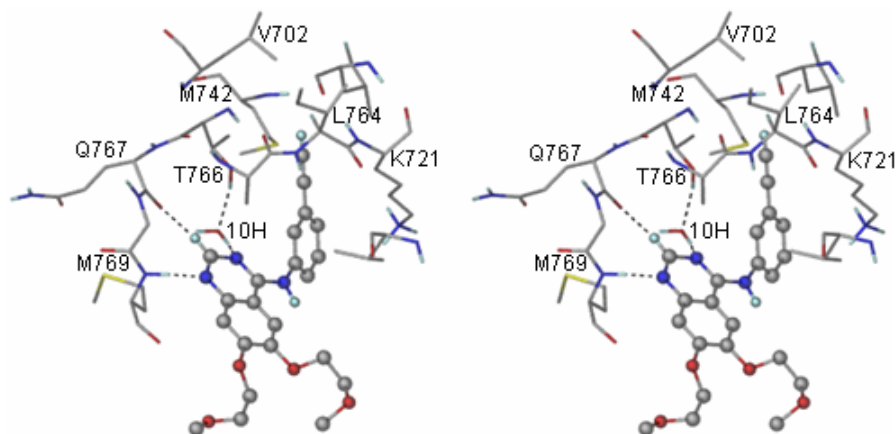


Figure 3.4: Stereoview of active site residues and erlotinib ligand in the crystal structure of 1M17. Three hydrogen bonds which lock the heterocyclic ring are shown as dashed lines. For clarity only some of the active site residues are shown.

3.3.3 Pose Analysis with RMSD

To evaluate the capabilities of the various programs to model these hydrogen bonds, erlotinib was docked using each of the two programs and the relative orientations of the ligand are shown in Figure 3.5. The RMSDs were calculated on the basis of the 4-anilinoquinazoline fragment. Although LigandFit predicted the orientation correctly, it displaced the molecule significantly away from the active site. GOLD performed better with a pose with the minimum RMSD of 1.07. In this respect, these programs were able to reproduce the hydrogen bond arrangement of the experimental crystal structure. Therefore, RMSDs were generated for the 128 molecules in the study for both GOLD and LigandFit poses (appendix I). A scatter plot of these RMSD values vs experimental pIC_{50} values is shown in Figure 3.6. The performance of the two programs in docking the molecules in the active site is analyzed in terms of RMSD and is summarized in Table 3.1. Based on the RMSDs alone, it could be stated that GOLD performs slightly better in that 118 out of 128 molecules have RMSD <1.0 Å while LigandFit has 94 molecules with <1.0 Å RMSD.

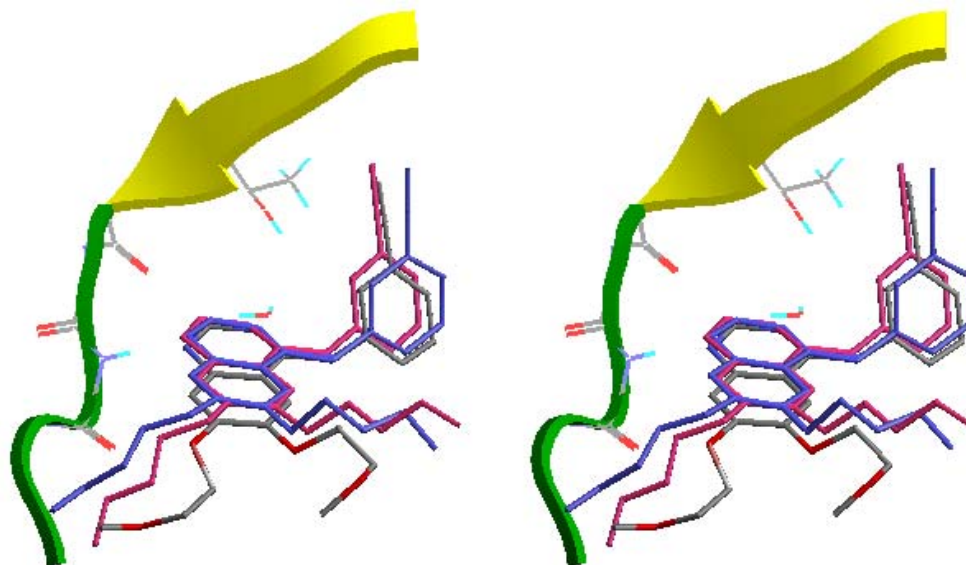


Figure 3.5: Overlay picture of erlotinib (experimental) and the docked conformations obtained from GOLD (pink, RMSD: 1.07), LigandFit (blue; 1.56) in the active site.

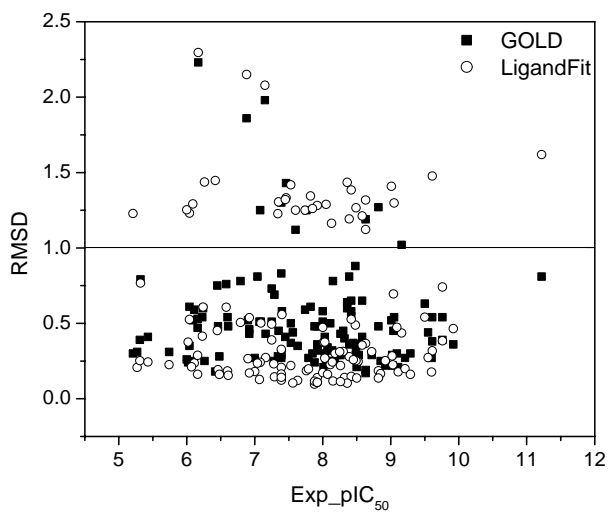


Figure 3.6: Scatter plot of RMSDs of docked conformations obtained from GOLD and LigandFit vs activity.

Table 3.1: Summary of RMS deviations of conformations obtained from GOLD and LigandFit.

RMSD	GOLD	LigandFit
≤0.5	86	83
0.5–1.0	32	13
1.0–1.5	7	28
1.5–2.0	2	1
2.0–2.5	1	3

3.3.4 Pose Analysis with Intermolecular Interactions

A second type of analysis was carried out to further discriminate between GOLD and LigandFit in the prediction of correct poses. Interaction-based evaluation of the generated poses may add more meaning to the measurement of docking accuracy as compared to the normal RMSD based classifications [3.22]. The best poses for the ligands were analyzed in terms of the three key hydrogen bonds mentioned above. The hydrogen bond metrics are listed in Table 2 of appendix I. Based on these interactions, the poses were classified into one of the three categories, namely *close*, *shifted* and *misoriented*. A *close* category corresponds to conformations with low RMSD and has all the three crucial hydrogen bonds intact. A *shifted* category refers to conformations that have medium RMSD values and with the ligand translated with respect to the erlotinib position and naturally, some of the hydrogen bonds are missing or distorted. Finally, the *misoriented* category refers to cases where the orientation itself is completely different from that of erlotinib. Needless to say, all the three hydrogen bonds have disappeared in the misoriented cases. Table 3.2 lists the numbers of close, shifted and misoriented molecules obtained from GOLD and LigandFit. Figure 3.7 illustrates these three situations.

Table 3.2: Classification of GOLD and LigandFit poses of ligands (active, medium, inactive) based on alignment of interactions (close, shifted, misoriented) in the active site.

Category	GOLD			LigandFit		
	Active (pIC ₅₀ > 8)	Medium (pIC ₅₀ 6 – 8)	Inactive (pIC ₅₀ < 6)	Active (pIC ₅₀ > 8)	Medium (pIC ₅₀ 6 – 8)	Inactive (pIC ₅₀ < 6)
Close	42	40	4	1	0	0
Shifted	17	7	3	0	3	0
Misoriented	0	15	0	58	59	7

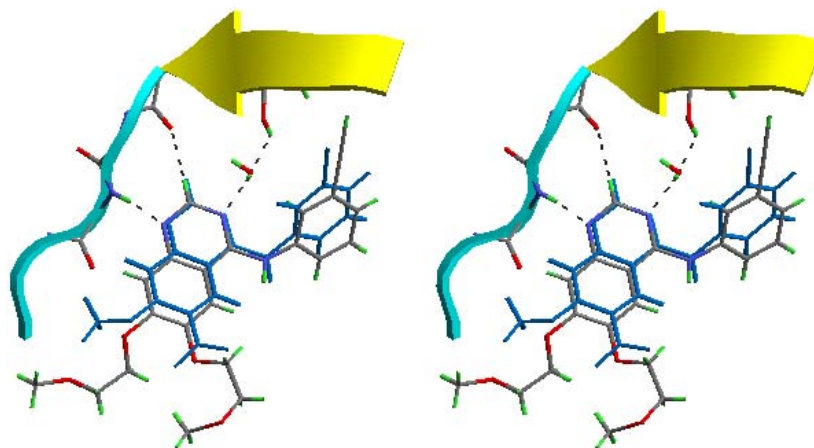


Figure 3.7a: Stereoview of a close category ligand, **25** (blue). Erlotinib is also shown

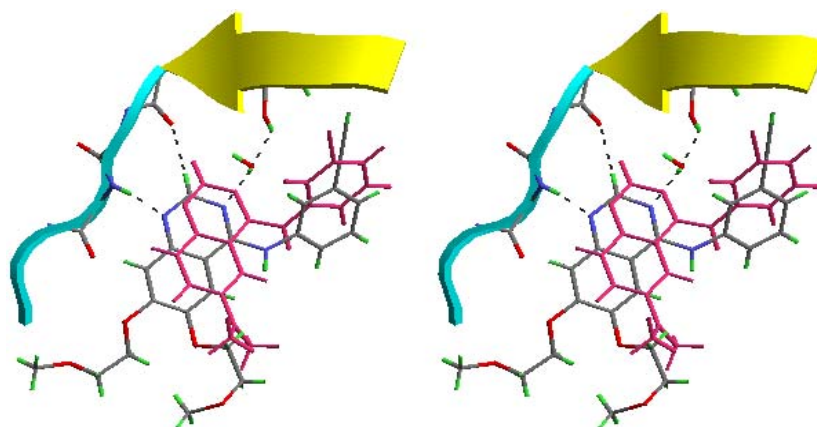


Figure 3.7b: Stereoview of a shifted category ligand, **108** (pink) shown with erlotinib

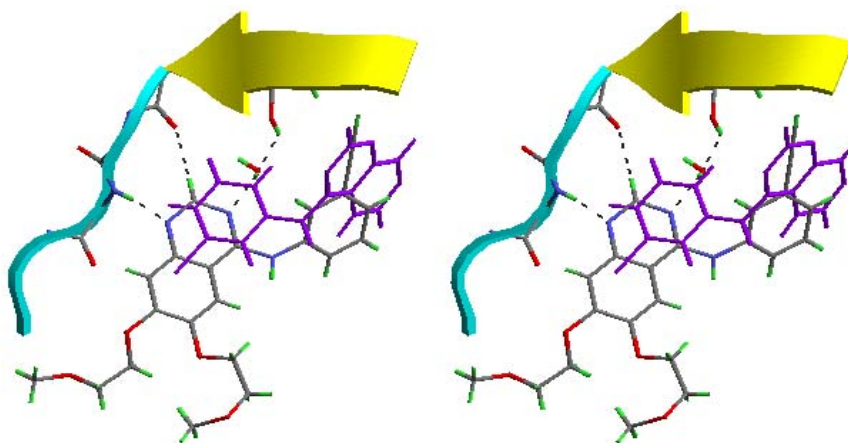


Figure 3.7c: Stereoview of a misoriented category ligand, **75** (purple) with erlotinib

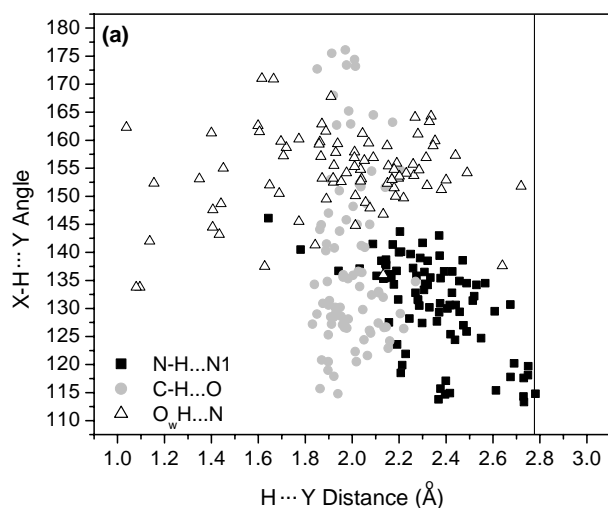
Although the RMSDs of the poses given by LigandFit were comparable to those generated by GOLD (Table 3.1), visual inspection revealed that all the ligands were completely misoriented! This could be due to: (1) the methodology used to find the correct poses is not good enough so that the ‘best’ pose obtained at the end of the stipulated number of iterations is routinely accepted by the program or; (2) improper or excessive weightage has been given for the vdW interactions with respect to the hydrogen bonds. In GOLD, however, nearly 67% of the ligands are in the close category, whilst 21% and 12% are in the shifted and misoriented categories respectively (see

appendix I). Thus, while GOLD certainly outperforms LigandFit in generating good poses, it is still far from perfect. Further, these results show the dangers of relying on RMSDs alone.

A further analysis of the (three) hydrogen bonds is in order. The N–H \cdots N between Met769 and N₁ and the O_w–H \cdots N between water10H and N₃ involve strong donors and acceptors, while the C–H \cdots O between Gln767 and C₂–H is weak. All these hydrogen bonds lie in a continuum between the very weak (~ 0.25 kcal mol⁻¹) to very strong (~ 40 kcal mol⁻¹) [3.23]. In general, the N–H \cdots N is one of the weakest types of hydrogen bonds involving ‘strong’ donors and acceptors (N, O, halide). Also, the hydrogen bonds with water as donor lie at the border between strong and weak. In the present case, the N–H \cdots N distance (all distances are for d_{H-X}) is 1.81 Å while O_w–H \cdots N is 2.01 Å (Figure 3.1). Strikingly, the C–H \cdots O bond involves a donor that is activated by the two flanking heterocyclic N-atoms, and is as short as 2.19 Å. Therefore, it is very significant for an interaction that involves a ‘weak’ donor. Such interactions are not uncommon in protein complexes [3.24]. It is estimated that the C–H \cdots O is of the same, if not of greater significance than the N–H \cdots N and O_w–H \cdots N interactions, in this particular case. All in all, it may be safely said that the three key hydrogen bonds that pin the ligand in the active site are nearly comparable in their overall importance, being nearly equal in distance. This means, in the VS context, it is imperative that all three weak interactions need to be modeled correctly in order to obtain the exact binding affinity. Clearly, this can only happen for ligands in the *close* category.

In GOLD, all the molecules in this *close* category interact correctly with the key residues in the immediate vicinity of the ligand, not only with respect to the hydrogen bonds but also the hydrophobic interactions formed by the 4-anilino group. Figure 3.8a is a representation of the three hydrogen bonds ($d-\theta$ plot) in the active site. Note that the profusion of interactions and the number of (C–H \cdots O), (O_w–H \cdots N) and (N–H \cdots N) points is nearly the same. Notably, the C–H \cdots O distribution is the tightest in terms of length (shorter than N–H \cdots N) and angle. This is a clear demonstration of the fact that the C–H \cdots O interaction is the most significant of the three hydrogen bonds. This is true for ligands in all activity ranges.

In the *shifted* category the overall hydrogen bond situation is less satisfactory (Figure 3.8b). The C–H \cdots O interactions are largely conserved, again indicating that they are the most important among the three hydrogen bonds. The O_w–H \cdots N interactions become much shorter, and therefore repulsive, while the N–H \cdots N interactions become longer. Why did this happen? Almost all the ligands in this shifted category have a lengthy alkyl substituent on C₇ (the converse is not true) and no substituent on C₆. The effect of this is to ‘push’ the ligand closer to water 10H and Thr766 leading to an unfavorable short contact. Some compounds showing this behaviour are **17**, **20**, **27**, **44**, **54**, **91**, **102**, **105**, **108**, **115**, **117**, **118**, **119**, **120** and **127**. Concomitantly, the N–H \cdots N interaction elongates.



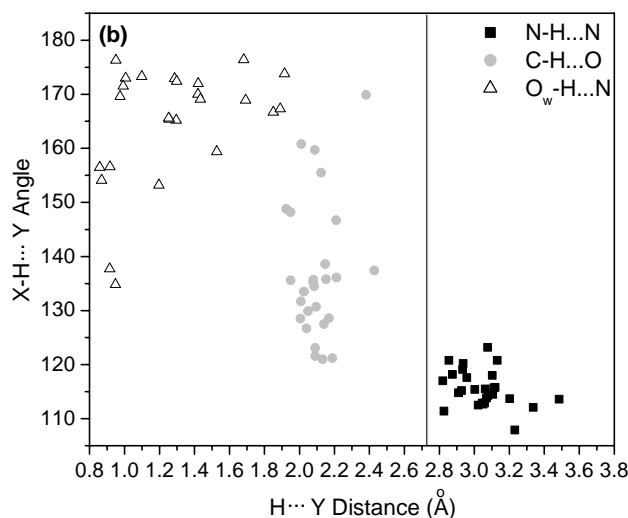


Figure 3.8: d - θ scatter plot of the three key active site hydrogen bonds in the (a) close category and (b) shifted category. Note that the N-H...N (Met769) interactions have elongated in (b) compared to (a), while the O_w-H...N (water 10) interactions have shortened unacceptably. The C-H...O (Gln767) interactions are largely conserved in both the cases.

The *misoriented* category is more drastically in error. For all the molecules in this category, there is hardly any substitution at the 6- or 7-positions. At most, there is a small substituent like fluoro, amino or nitro at one of these two positions; this is exemplified by ligands **1**, **2**, **3**, **4**, **5**, **6**, **7**, **8**, **13**, **55**, **73**, **75**, **78**, **80** and **112**. In effect, the molecule is flipped around to obtain an orientation that is a sheer shape mimic. There are no hydrogen bonds and this *pose* has no chemical or biological significance.

Considering the importance in modeling the hydrogen bonds, all ten solutions offered by GOLD were examined for the *shifted* and *misoriented* cases, in particular to ascertain if there are any poses with correct hydrogen bond geometry even if the affinity was not modeled properly. The aim of this exercise was to identify solutions with the correct pose, for if the pose is incorrect, there is no question of proceeding further. The hope was that if a solution with the correct pose were found (hydrogen bonds modeled correctly), the scores (for that solution) would also be correct. Gratifyingly, for the *shifted* category, 23 out of 27 ligands had other solutions with correct poses. Ligands

which do not improve after this analysis are **17**, **44**, **119** and **127**. For the *misoriented* category, however, there was no solution, as might have been expected. The scores for the 23 (improved) ligands in the *shifted* category and all the ligands in the *close* category were analyzed (Figure 3.9). Although there is an improvement in the correlation it is not significant enough and there are still a number of false positives and false negatives. This leads to an inescapable feeling that these disagreements arise from the lack of C–H···O and other hydrogen bond parameterization. Further absence of electrostatic terms may also affect the Fitness function (GOLDScore). Some of these interactions play a crucial role in locking the molecule in the active site and also contribute significantly to the scores.

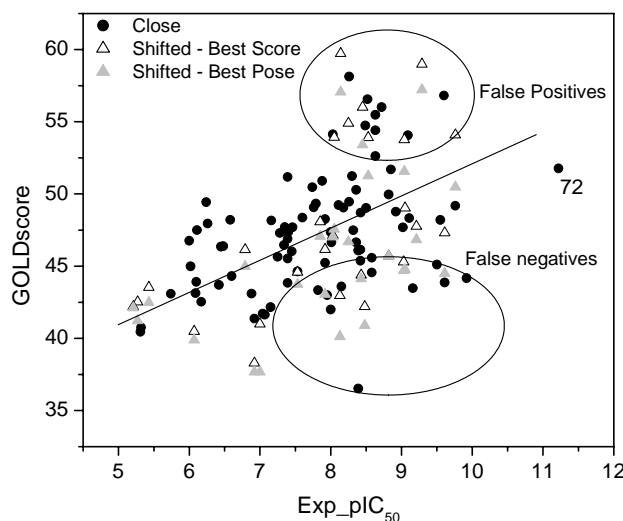


Figure 3.9: Experimental pIC₅₀ values vs GOLD scores of compounds which have all three key hydrogen bonds between heterocyclic and protein residues and water in the active site. Note that the false positives improve more than the false negatives because of this treatment. Compound **72** (far right) is a common outlier in all the docking studies

3.3.5 Cross Scoring

With the observation that the poses from GOLD were much better and that each program has its own methods to calculate scores, these GOLD poses were used as inputs to calculate scores like Dockscore, Ludi, Ligscore, PLP and PMF in LigandFit. This

procedure is termed as *cross scoring*. Thus the GOLD poses were held rigidly in the active site and the different scores were calculated. The Dockscore of LigandFit produced an improved correlation shown in Figure 3.10. The correlation was much improved but the (small number of) false positives and false negatives are still nagging.

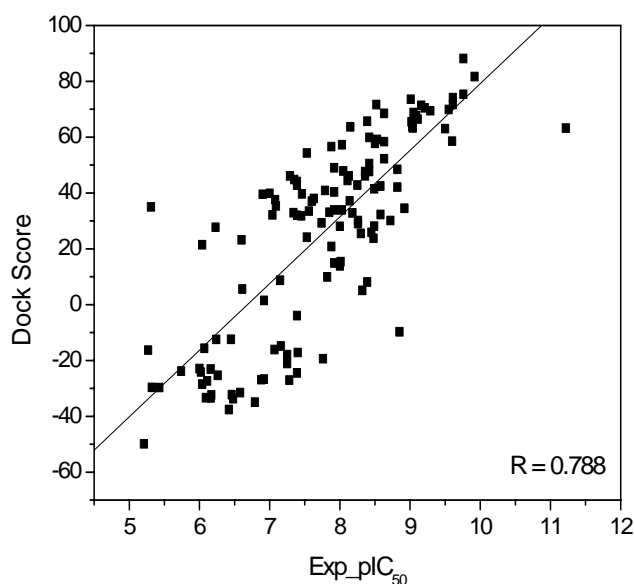


Figure 3.10: Scatter plot of Dockscore (LigandFit) vs activity obtained after docking GOLD conformations rigidly into the active site.

3.3.6 Consensus Model

In order to get a better VS model, a multiple regression analysis was carried out using 12 different scores (consensus scoring) [1.26a]. For each ligand the best solution from GOLD (ligand, protein and water) was directly read into LigandFit and rescored without any further optimization. All such regenerated scores along with GOLD scores were used in a multivariate analysis to produce a consensus model using a training set of 31 molecules. A combination of GOLDScore, Dockscore, Ludiscore, and PLPscore gave a better model than any other (Figure 3.11) combination of scores. The final equation generated by this multiple linear regression is given below:

$$\begin{aligned} \text{Activity} = & -0.3351 + 0.1263 (\text{GOLDScore}) + 0.0087 (\text{Dockscore}) + 0.0074 (\text{Ludiscore}) \\ & - 0.00246 (\text{vdWscore}) - 0.0041 (\text{PLPscore}) \end{aligned} \quad (3.12)$$

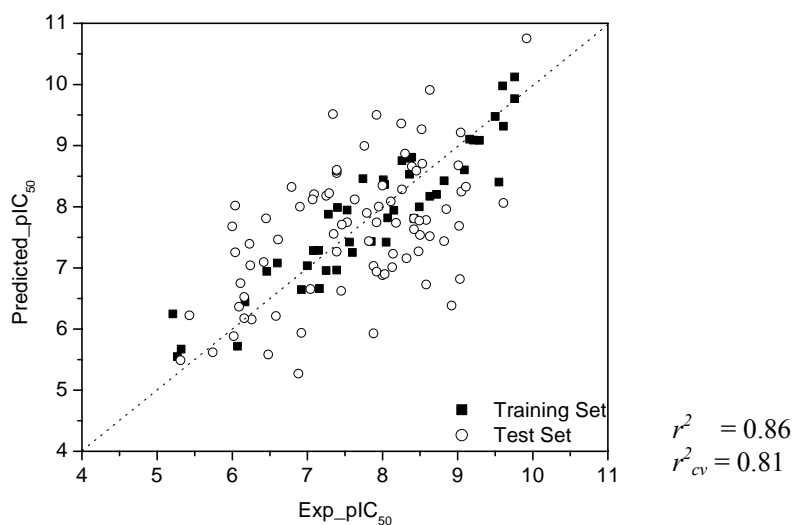


Figure 3.11: Scatter plot of experimental vs predicted activities using consensus scoring.

The coefficient of each term is a measure of the contribution of each scoring function towards the final model. It is well known that Dockscore estimates the hydrophobic interactions correctly. PLP of LigandFit is known to have a high success rate in scoring hydrophobic complexes. The final model ($r^2 = 0.86$, $r^2_{cv} = 0.81$, PRESS = 11.23, $r^2_{bs} = 0.86$) was validated by the test set of 97 molecules. The model was able to correctly predict the activity of 88% of the molecules within a range of 1.5 pIC_{50} value.

3.4 Conclusions

Docking is quite often used as a VS tool in the lead discovery and optimization process. Accurate prediction of binding affinities of the compounds using docking depends on its efficacy in pose prediction and scoring functions. This chapter compares the efficiency of the docking programs GOLD and LigandFit using a library of 128 EGFR kinase inhibitors. It has been shown that acceptable results may be obtained when

the poses are analyzed on the basis of chemical interactions. Docking analysis reveals that for the 4-anilinoquinazoline type of ligands, inclusion of a hydrogen bonded water molecule near N₃ is indispensable to obtain meaningful VS results. GOLD and LigandFit have variable success in identifying correct poses and binding affinities. It has been demonstrated that poses from GOLD are invariably better and that these may be input into LigandFit to get more accurate results. Consideration of ligand–protein hydrogen bonds of the N–H...N, O_w–H...N and above all the C–H...O type seems to be very important to obtain accurate poses and binding affinities. The improvement in the correlation of the scores, of the *close* and *shifted* categories, with the biological activity demonstrates the need of interaction based analysis of the docking poses. However, such correlation is still insufficient in the context of VS. Better poses can be obtained from interaction based analysis that includes adequate modeling of both conventional and weak hydrogen bonds. Further, cross scoring and consensus scoring seem to be more suitable methods when no single scoring function proves to be effective in modeling all types of interactions. The present chapter uses both these methods to develop a VS model against EGFR kinase.

CHAPTER 4

IDENTIFICATION OF APPROPRIATE METHOD FOR VIRTUAL SCREENING OF THYMIDINE MONOPHOSPHATE KINASE INHIBITORS

4.1 Introduction

It is estimated that approximately one third of the world's population is infected with Tuberculosis (TB) bacillus, *Mycobacterium tuberculosis* [4.1]. Each year more than 8 million people contract the disease and nearly 2 million people die of it [4.2]. A peculiar aspect of its pathogenicity comes from the fact that it can remain latent and become active decades later. One of the most significant risk factors for developing tuberculosis is human immunodeficiency virus (HIV) infection [4.3]. The World Health Organisation (WHO) reports that in Africa 80% of the cases with TB are HIV infected. The current treatment of active TB includes a dosage regime of four drugs (isoniazid, rifampicin, pyrazinamide and ethambutol) for at least six months. As a consequence of the prolonged duration and irregular treatment and the highly adaptive nature of the organisms to their surroundings multi-drug resistant (MDR) strains of *M. tuberculosis* have developed. There is therefore, an urgent need to develop new drugs to meet the global emergency.

M. tuberculosis thymidine monophosphate kinase (TMPK_{mt}) belongs to a large super family of nucleoside monophosphate kinases (NMPK). It catalyses the phosphorylation of deoxythymidine monophosphate (dTMP) to deoxythymidine diphosphate utilizing ATP as a phosphoryl donor [4.4]. Therefore, TMPK_{mt} inhibitors can be promising leads to treat tuberculosis by inhibiting DNA synthesis in *M. tuberculosis*. Additionally, they have gained much attention due to their prospects as antimetabolites. Also, the sequence of TMPK_{mt} when compared with its human isozyme shows only 22% sequence identity [4.5]. These characteristics make TMPK_{mt} as one of the potential targets for the design of new antitubercular drugs.

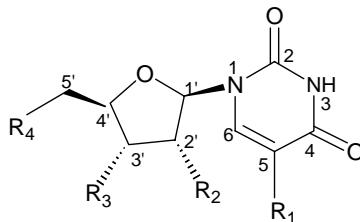
Several dTMP derivatives were synthesized and studied for their effect on the TMPK_{mt} [4.6–4.9]. Also, the availability of a high resolution crystal structure [4.10] is handy for comparing different approaches in the present study that can help to subsequently develop a robust VS model. The scope of the present work lies in the determination of the best VS tool for the system of choice. To accomplish this various widely used methods like docking, pharmacophore modeling, and MFA were carried out against TMPK_{mt} target. Since the reliability and the efficiency of 3D-QSAR methods depend on the orientation of the molecule [4.11], different alignment methods were used to generate MFA models.

4.2 Materials and Methods

4.2.1 Selection of Molecules

A set of 47 compounds reported as thymidine monophosphate kinase inhibitors was compiled (Table 4.1) [4.6–4.9]. The inhibitory activities were converted into the corresponding pK_i ($-\log K_i$) values, where the K_i value represents the drug concentration that causes inhibition of dTMP phosphorylation by TMPK_{mt}. It was verified that all the K_i values had been obtained using the same assay method [4.7]. pK_i values of all the molecules spanned a sufficiently wide range from 2 to 6.

All the molecules were built from the coordinates of dTMP structure (1G3U) [4.10] using the builder module of *Cerius*². All the structures were minimized using the steepest descent algorithm with a convergence gradient value of 0.001 kcal mol⁻¹. Partial atomic charges were calculated using the Gasteiger method. Further geometry optimization was carried out for each molecule with the MOPAC 6 package using the semi-empirical AM1 Hamiltonian.

Table 4.1: Molecules used in the study

Molecule No.	R ₁	R ₂	R ₃	R ₄	Actual activity (pKi)	Predicted activity (pKi)
129	CH ₃	H	OH	NHCOCH ₃	4.04	^c 5.20
130	CH ₃	H	OH	N ₃	5.15	^b 5.09
131	CH ₃	H	OH	NH ₂	4.92	^b 5.04
132	CH ₃	OH	OH	OH	3.14	^c 4.01
133	CH ₃	OH ^a	OH	OH	3.62	^c 4.10
134	CF ₃	H	OH	OH	4.01	^b 3.69
135	C ₂ H ₅	H	OH	OH	2.94	^b 3.30
136	F	OH	OH	OH	3.28	^b 3.31
137	F	OH	OH	H	3.25	^c 3.22
138	CH ₃	H	NH ₂	OH	3.63	^c 3.46
139	CH ₃	H	F	OH	4.55	^c 3.32
140	CH ₃	F	OH	OH	3.91	^c 4.38
141	CH ₃	OH	N ₃	I	3.30	^c 4.82
142	CH ₃	OH	NH ₂	H	2.72	^b 2.92
143	CH ₃	OH	NHC(NH)NH ₂	OH	3.85	^c 3.30
144	CH ₃	H	CH ₂ N ₃	OPO ₃ ²⁻	4.92	^c 3.32
145	CH ₃	H	CH ₂ NH ₂	OPO ₃ ²⁻	4.97	^c 3.63
146	CH ₃	H	CH ₂ F	OPO ₃ ²⁻	4.82	^c 3.52
147	CH ₃	H	CH ₂ OH	OPO ₃ ²⁻	4.53	^c 3.56
148	CH ₃	OH	CH ₂ N ₃	OPO ₃ ²⁻	3.93	^b 3.79
149	CH ₃	OH	CH ₂ NH ₂	OPO ₃ ²⁻	3.50	^c 3.67
150	CH ₃	H	CH ₂ N ₃	OH	4.39	^c 3.82
151	CH ₃	H	CH ₂ NH ₂	OH	4.24	^c 3.31
152	CH ₃	H	CH ₂ F	OH	4.34	^c 3.30
153	CH ₃	H	CH ₂ OH	OH	4.38	^b 4.52
154	CH ₃	OH	CH ₂ N ₃	OH	3.11	^b 3.31
155	CH ₃	OH	CH ₂ NH ₂	OH	3.40	^b 3.32
156	CH ₃	H	CH ₂ CH ₂ OH	OH	3.80	^b 4.00
157	CH ₃	H	NH ₂	OPO ₃ ²⁻	3.62	^c 3.55
158	CH ₃	OH	NH ₂	OPO ₃ ²⁻	4.56	^c 3.45
159	CH ₃	NH ₂	OH	OPO ₃ ²⁻	4.25	^b 4.53
160	CH ₃	Cl	OH	OPO ₃ ²⁻	4.72	^c 4.11
161	CH ₃	F	OH	OPO ₃ ²⁻	4.36	^b 4.49
162	C ₆ H ₅ CH ₂	H	OH	OPO ₃ ²⁻	4.55	^c 3.63
163	CH ₃	H	N ₃	OPO ₃ ²⁻	5.00	^b 4.85
164	CH ₃	H	N ₃	OH	4.55	^c 4.65
165	Br	H	N ₃	OH	4.97	^c 4.88
166	Br	H	OH	OH	5.30	^b 5.22

Molecule No.	R ₁	R ₂	R ₃	R ₄	Actual activity (pKi)	Predicted activity (pKi)
167	CH=CHBr	H	OH	OH	3.20	^c 4.46
168	CH ₂ OH	H	OH	OH	3.08	^c 3.24
169	Cl	H	N ₃	OH	4.79	^b 4.65
170	CH ₃	H	OH	OH	4.56	^b 4.40
171	H	H	OH	OH	2.99	^c 3.07
172	F	H	OH	OH	3.67	^b 3.30
173	I	H	OH	OH	4.48	^b 4.56
174	OH	H	OH	OH	3.56	^b 3.34
175	H	H	N ₃	OH	3.09	^c 3.25

^ahydroxyl in β -position; ^bTraining set molecules; ^cTest set molecules

4.2.2 Docking

The crystal structure of TMPK_{mt} (1G3U) with its bound substrate dTMP was used in the study. Hydrogen atoms were added to the protein at pH 7. Initially, the hydrogen atoms were minimized keeping all the heavy atoms fixed. The whole complex including the bound water was energy minimized by steepest descent followed by conjugate gradient methods to attain a convergence gradient of 0.001 kcal mol⁻¹ using CVFF in *Insight II*. All the amino acids within 13 Å radius from the center of the bound ligand were considered to constitute the active site. Docking was carried out using GOLD as described in chapter 2. The first ranked solutions of the ligands were taken for further analysis.

4.2.3 Pharmacophore Modeling

Initial Setup

Multiple acceptable conformations were generated for all molecules within *Catalyst* [4.12] ConForm module using the “Poling” algorithm [4.13]. A maximum of 250 conformations were generated for each molecule within an energy threshold of 20 kcal mol⁻¹ above the global energy minimum. The training set molecules (20) associated with their conformations were submitted to the *Catalyst* hypothesis generation (HypoGen) (Table 4.1). Based on common features present in the study molecules,

features like hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), hydrophobic feature (HY) and negative ionizable feature (NI) were included for the pharmacophore generation.

Model Generation

This includes three phases: (1) pharmacophore features that were common among the most active compounds are elaborated in the *constructive phase*; (2) those models that fit the inactive training set members were abolished in the *subtractive phase* and; (3) the remaining hypotheses were refined in the *optimization phase*. The optimization was carried out by random transition of features, rotation of vectored features and removal or addition of features from the preferred models using simulated annealing algorithm. The models obtained were selected based on statistical parameters.

Statistical Parameters

The statistical parameters like cost values determine the significance of the model [4.14].

Total Cost

It is the sum of three components: a weight, an error and a configuration component. Each feature in the model is assigned certain weight. The weight component is a value that increases in a Gaussian form as the feature weight deviates from an ideal value (2.0). The error component increases with the increase in the root mean squared (rms) difference between estimated and measured activity between training set molecules. The weight and error components of the total cost are estimated by equation 4.1

$$-\ln P = \frac{1}{2} \ln 2\pi + \ln \sigma - \ln \Delta + \frac{x^2}{2\sigma^2} \quad (4.1)$$

where P is the probability function, and σ and Δ are constant parameters derived from expected values of weights and errors. The deviation x of the error component is the difference between the actual and the predicted activities of the training set and the

deviation x of the weight component is the difference between the average of the feature weights and the ideal weight (2.0).

The configuration cost quantifies the entropy of the pharmacophore space. The value specifies log to the base 2 of the number of trial models that are attempted for optimization. For training sets with a configuration cost higher than 17 (corresponds to $>2^{17}$ starting hypotheses), the molecules may be too flexible with a very large conformational space and therefore likely to generate a pharmacophore with only a chance correlation.

$$\text{Total Cost} = eE + wW + tT + cC \quad (4.2)$$

where e , w , t , c are the coefficients associated with the error (E), weight (W), tolerance (T) and configuration (C) components, respectively.

Fixed Cost

The fixed cost is the lowest possible cost representing a simple model that fits the data perfectly. It is termed fixed as only the fourth term in equation (4.2) varies while all the other three terms remain constant.

$$\text{Fixed Cost} = eE (x = 0) + wW (x = 0) + tT (x = 0) + cC \quad (4.3)$$

Null Cost

It represents the cost of a pharmacophore model with no features but estimates activity of every molecule to be the average activity of the training set molecules.

$$\text{Null Cost} = eE (x_{est} = \bar{x}) \quad (4.4)$$

where x_{est} is the scaled activity value of the training set molecules

Fixed cost and null cost are the theoretical cost values, measured in units of bits, providing useful guidelines for estimating the chances for a successful model. The greater the difference between these values, the higher the probability for finding useful models. Generally, a model with a cost difference of 40–60 bits has 70–90% chance of representing a true correlation in the data. On the other hand, if the cost difference is less

than 40 bits the likelihood of the model representing a true correlation rapidly drops below 50%. The total cost of the pharmacophore model should lie in between the fixed and the null costs. The closer is the value of total cost to the fixed cost and the further it is from the null cost, the more statistically significant the pharmacophore is believed to be.

Ten pharmacophore models with the lowest total cost values were generated. The best model was selected based on a high correlation coefficient (r) and lowest total cost. The final model was further validated by a test set of 27 molecules.

4.2.4 3D-QSAR

Alignment Methods

Three different alignment methods that were used for MFA are detailed as follows:

Model I (Least squares alignment): This alignment was carried out using the shape reference method in the QSAR module within *Cerius²* using the six thymidine ring atoms. The most active compound **166** was used as a template for superposing the rest of the molecules. The aligned molecules are shown in Figure 4.1(a).

Model II (Pharmacophore based alignment): In this method all the molecules were aligned using the pharmacophore developed. Figure 4.1(b) shows the alignment obtained.

Model III (Receptor based alignment): After docking analysis, the final orientation of each molecule in the active site was selected. Figure 4.1(c) shows the alignment obtained.

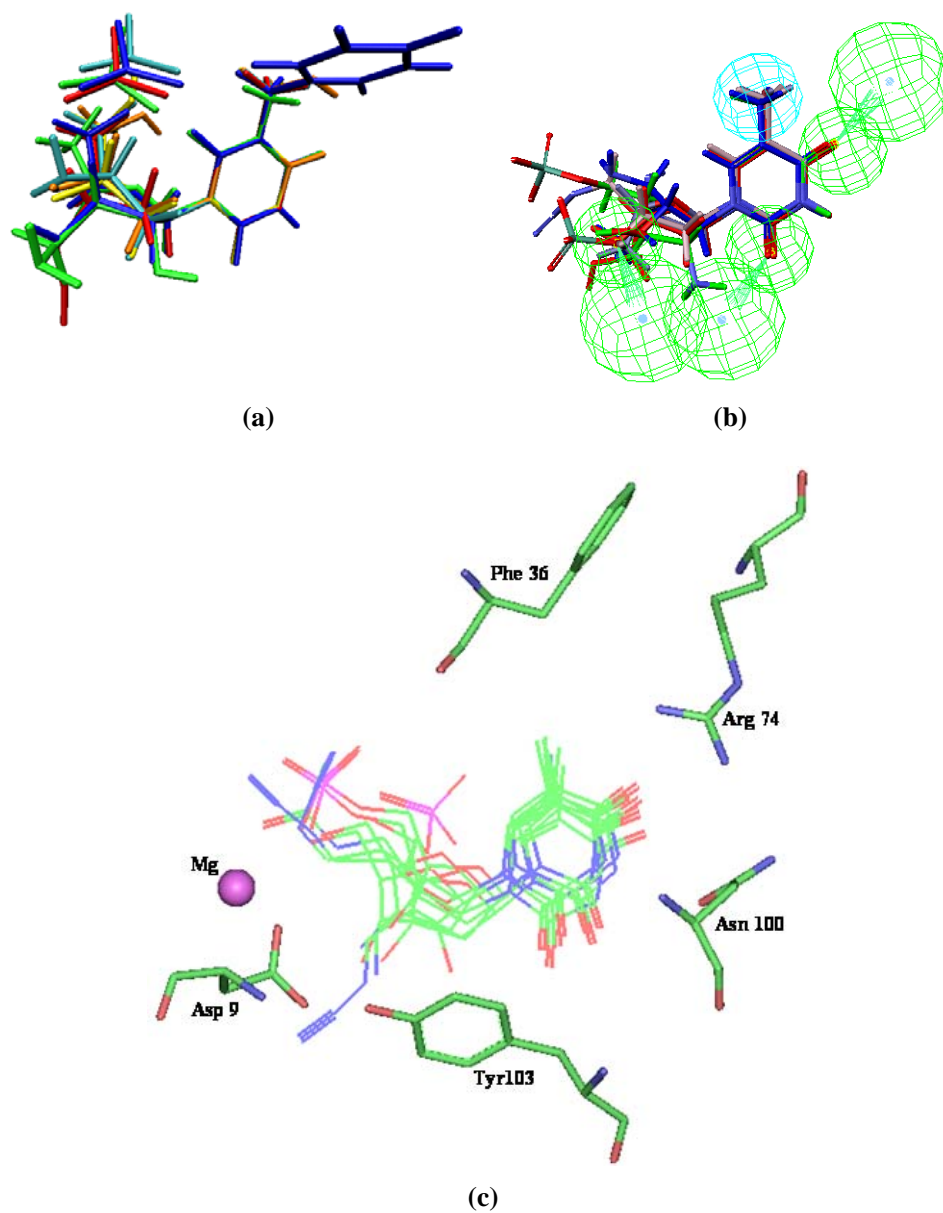


Figure 4.1: Alignment of study molecules using (a) least squares, (b) pharmacophore and (c) docked conformations

Determination of Probe Geometry

Alternate probes like NH_2 and H_2O were used to calculate the electrostatic interaction energy. The geometry of the H_2O probe was optimized at the 6-31G* level using Gaussian 03 [4.15] and the corresponding charges were used for the calculation of interaction energies. In the case of the NH_2 probe, an arginine molecule was taken as template and optimized at 6-31G* level. The charges corresponding to the side chain guanidinium NH_2 of arginine were considered while adjusting the total charge of the probe to + 0.1.

Molecular Interaction Fields

MFA studies were performed with the QSAR module of *Cerius*² as described in chapter 2. From the 47 compounds above, 35 were used in the training set and the remaining 12 compounds were included in the test set. The molecular field was created using CH_3 , NH_2 , and H_2O as probes representing steric and electrostatic fields, respectively.

4.3 Results and Discussion

4.3.1 Docking

Molecular Recognition of Substrate dTMP

The substrate, dTMP was docked into the active site of TMPK_{mt} . The RMSD between the docked pose of dTMP and its bound conformation in the crystal structure (1G3U) is 0.2 Å, indicating that GOLD was able to reproduce the correct pose. The detailed interactions of dTMP– TMPK_{mt} are shown in Figure 4.2.

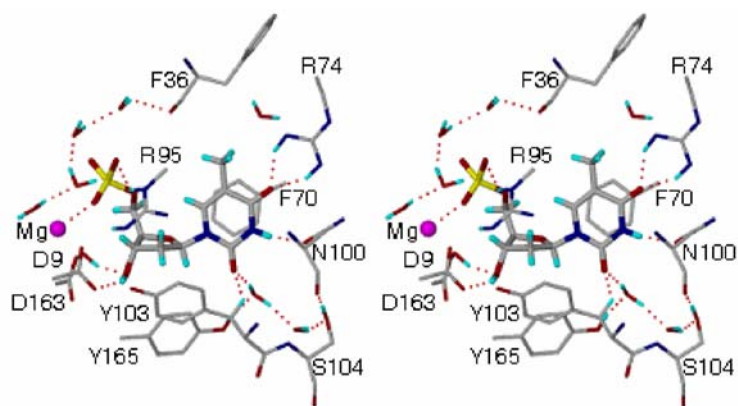


Figure 4.2a: Stereoview of substrate **dTMP**–**TMPK_{mt}** complex. Hydrogen bonds are shown as red dotted lines. Only those protein H-atoms involved in hydrogen bonds are shown. Single letter amino acid codes are used.

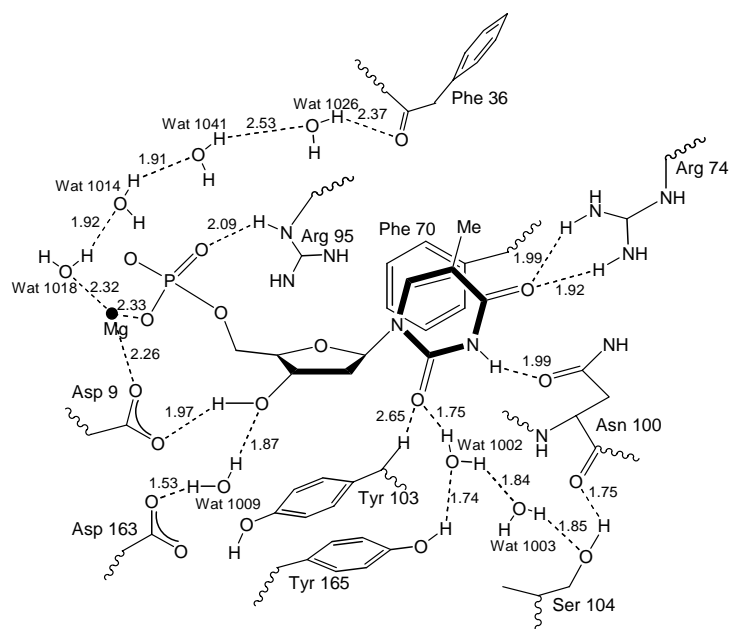


Figure 4.2b: Schematic representation of interactions of substrate **dTMP**–**TMPK_{mt}** complex in the active site. Hydrogen bonds are shown as dashed lines and in Å.

It can be seen that binding forces between dTMP and the enzyme manifest in: (i) a stacking interaction between the thymidine ring and Phe70 (3.70 Å); (ii) a multicenter

(bifurcated) hydrogen bond between carbonyl O-atom at thymidine C₄ and the terminal guanidino nitrogen atoms of Arg74 (N₆H...O=C, 2.06 Å, N₆H...O=C, 2.11 Å); (iii) a hydrogen bond between the C₃-OH of dTMP and the side chain carbonyl of Asp9 (C₈=O...HO, 1.968 Å); (iv) a hydrogen bond between phosphate oxygen OP₂ and guanidino N-atom of Arg95 (N₈H...O=P, 2.0 Å); (v) an ionic interaction between OP₁ and Mg²⁺ (Mg²⁺...O-P, 2.33 Å) where OP₁ occupies the coordination site of Mg²⁺; (v) a hydrogen bond between the side chain carbonyl oxygen of Asn100 and thymidine N₃ (C₈=O...HN, 1.98 Å); and (vi) a weak hydrogen bond between C_β of Tyr103 and carbonyl oxygen at thymidine C₂ (C_βH...O=C₂, 2.65 Å). The electrostatic interaction between OP₁ and Mg²⁺ is probably responsible for positioning the phosphate oxygen of dTMP in the active site for the kinase activity. All the interactions observed above are in complete agreement with the crystal structure [4.10]. In addition to these interactions, the substrate dTMP is stabilized in the active site by a water mediated finite σ-bond cooperative chain, OH (Tyr165)···OH (Wat1002)···OH (Wat1003)···OH (Ser104)···O=C (Asn100).

All the molecules in the study were docked using the same method and their interactions were analyzed. Only the interactions of the most active and the least active molecules are discussed in detail. The orientation and hydrogen bonds of the most active molecule **166** within the active site of TMPK_{mt} are shown in Figure 4.3.

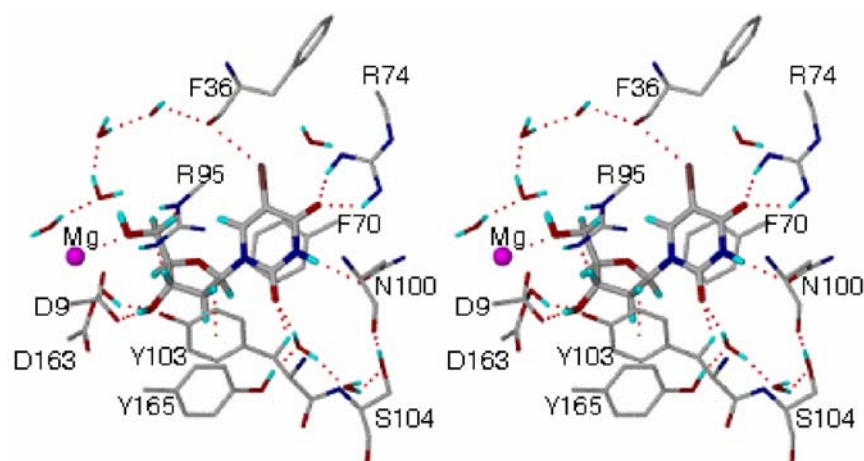


Figure 4.3a: Stereoview of the most active molecule **166**–TMPK_{mt} complex.

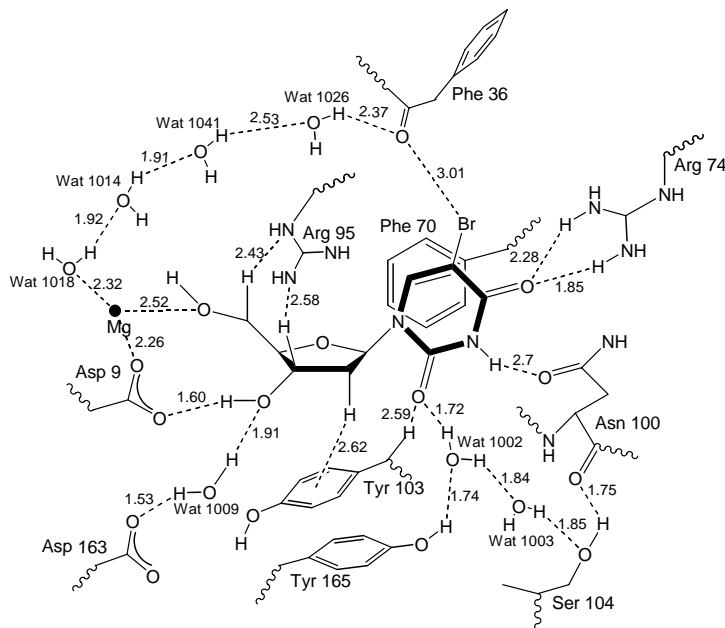


Figure 4.3b: Schematic representation of the most active molecule **166**–TMPK_{mt} complex and the interactions within.

Molecule **166** interacts with the active site residues similar to that of dTMP. Additionally, **166** also forms weak interactions like: (i) a CH...N hydrogen bond with the side chain of Arg95 (N_δ...HC_{5'}, 2.43 Å; N_ω...HC_{3'}, 2.58 Å); (ii) a C–H...O hydrogen bond between carbonyl oxygen at C₂ and the C_βH of Tyr103 (2.35 Å); and iii) a C–H...π interaction between the π system of Tyr103 and the C_{2'} of the ribose ring (2.62 Å). It has been reported that Tyr103 is crucial for the enzyme selectivity towards deoxyribonucleotides [4.10]. The weak interactions with Tyr103 appear to have a significant role in this selectivity.

Electrostatic interactions are observed between the hydroxyl oxygen atom at ribose C_{5'} and Mg²⁺ (2.52 Å). Br at thymine C₅ also forms a polarization type contact with the backbone carbonyl oxygen of Phe36 (Br^{δ+}...δ⁻O=C, 3.01 Å). Replacing Br with a methyl group removes this interaction resulting in a decrease in activity (Table 4.1). The halogen substitution at C₅ was mainly interpreted as space filler [4.16]. However, the

role of halogen at C₅ can be seen as the most stabilizing factor in the molecular recognition. The halogen here is not just a space filler but it is involved in halogen bonding. The properties and the energetics associated with halogen bonds are discussed in great detail by Resnati and co-workers [4.17]. The non linear angle ($\theta_{\text{C-Br}\cdots\text{O}}$, 133.8°) is attributed to the additional polarization of the halogen induced by the contacting oxygen that causes distortion in the cylindrical shape of the electron distribution. This in turn modifies the line of approach of the lone pair electrons of the interacting oxygen atom [4.18]. Also, the nearly perpendicular approach of the halogen ($\phi_{\text{Br}\cdots\text{O-C-N}}$, 78.2°) reflects the involvement of the π electrons of the peptide bond. Further exploration of Br \cdots O interactions with the Cambridge Structural Database (CSD, version 5.25, July 2004) revealed 580 hits within the distance range $2.8 \text{ \AA} < d(\text{Br}\cdots\text{O}) < 4.0 \text{ \AA}$ suggesting that such polarization induced contacts are not uncommon in crystal structures of small molecules [4.19]. Such interactions are stabilized to an extent of around 2 kcal mol^{-1} [4.20] making them, in energetic terms, equivalent to C-H \cdots O or C-H \cdots N hydrogen bonds. Halogen \cdots O interactions have not been reported often in analyses of macromolecular crystal structures [4.18, 4.21]. A notable example, however, is the interaction of iodine atoms of thyroid hormones with donor sites in its cognate protein [4.22]. Also, in the case of Aldose-Reductase protein the role of halogen bond seems to be critical so as to account for >1000 fold selectivity increase in the binding of the inhibitor [4.23]. Although this particular interaction is weak, its stabilization with concomitant water mediated cooperative hydrogen bond network makes it worthwhile to increase the inhibitory activity prominently. In summary, it is noted that in addition to the conventional or strong hydrogen bonds and electrostatic $\text{Mg}^{2+}\cdots\text{O}$ interactions, there are also several weaker interactions which, acting in concert, make a contribution to the binding specificity of the ligand in the active site.

The orientation of the least active molecule **142** is similar to that of the most active molecule **166**. However, fewer hydrogen bonds are present when compared with **166** as shown in Figure 4.4. Also, the absence of $\text{Mg}^{2+}\cdots\text{O}$ electrostatic interactions and Br \cdots O polarization induced interactions are probably the reasons for the low activity.

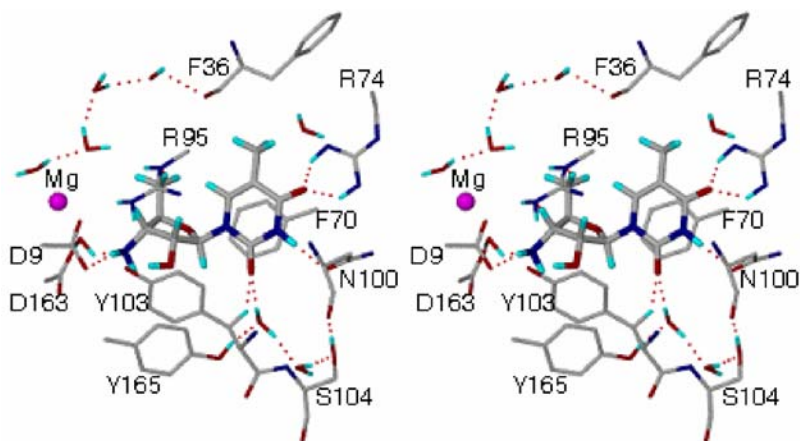


Figure 4.4a: Stereoview of the least active molecule **142**–TMPK_{mt} complex.

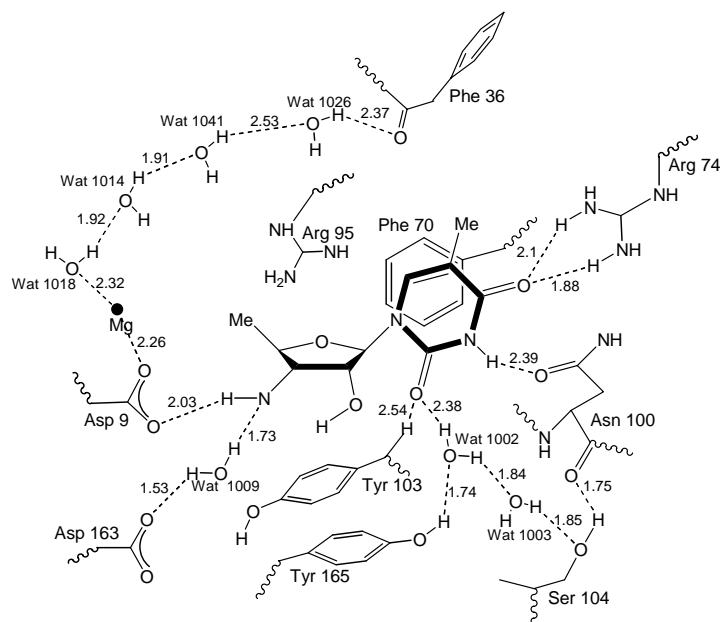


Figure 4.4b: Schematic representation of interactions of the least active molecule **142**–TMPK_{mt} complex in the active site.

It is worthwhile to mention that the ligand–protein complex is well stabilized by two water mediated finite σ -bond cooperative chains [4.24]: (i) OH (Wat1018)···OH

(Wat10142)⋯OH(Wat1041)⋯OH(Wat1026)⋯O=C(Phe36); and (ii) OH (Tyr165)⋯OH (Wat1002)⋯OH (Wat1003)⋯OH (Ser104)⋯O=C (Asn100). The first network is extended to the ligand through the Br-atom in **166** (Figure 4.3). The second network is common to all ligands. Such cooperative networks gain in energy, typically 20% over the isolated interactions. The presence of such water mediated interactions of ligands within the enzyme active sites forming cooperative networks explain the important role of water in the binding affinities of the molecules [3.24c].

4.3.2 Pharmacophore

Ten pharmacophore models with a cost difference of 44.77 were generated. Table 4.2 gives the statistical parameters for all the ten models. Among these, the model with the lowest total cost value (78.68), lowest RMSD (0.66) and highest correlation coefficient (0.963) was chosen for further analysis. All the training set molecules were predicted with an error less than 1 unit. The best model is further validated externally using a test set of 27 molecules, 77% of which were predicted with an error less than 1 unit. The actual and predicted activities of the training and test set molecules are given in Table 4.1.

Table 4.2: Statistical details of the pharmacophore models

Hypothesis ^a	Total cost	RMSD	Correlation Co-efficient	Features ^b
1	78.68	0.66	0.963	HBA, HBA, HBA, HY
2	80.35	0.75	0.951	HBA, HBA, HBA, HY
3	86.74	1.11	0.893	HBA, HBA, HBA, HY
4	87.05	1.12	0.890	HBA, HBA, HBD, HY
5	87.56	1.14	0.885	HBA, HBA, HBA, HY
6	87.87	1.16	0.882	HBA, HBA, HBD, HY
7	87.00	1.16	0.882	HBA, HBA, HBA, HY
8	88.59	1.16	0.881	HBA, HBA, HBA, HY
9	89.15	1.21	0.870	HBA, HBA, HBA, HY
10	89.63	1.23	0.865	HBA, HBA, HBA, HY

^aFixed cost = 74.302; Configuration cost = 15.117; Null cost = 119.075. All cost values are in bits.

^bHydrogen bond acceptor (HBA); Hydrogen bond donor (HBD); Hydrophobic feature (HY).

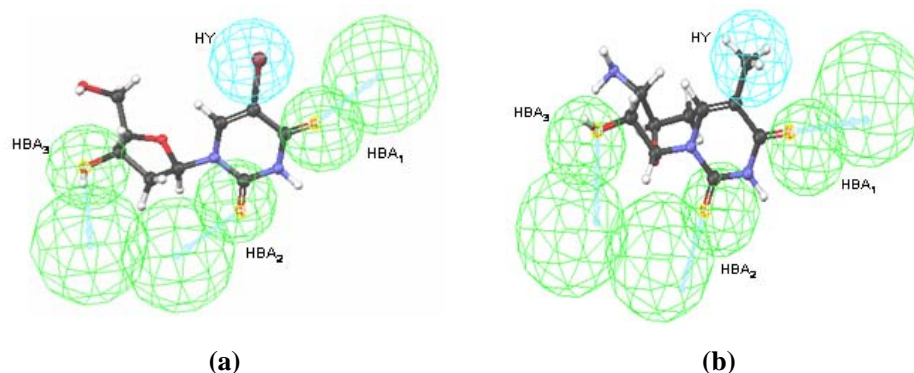


Figure 4.5: Pharmacophore mapping of the compounds (a) most active molecule **166** (b) least active molecule **142**. The green contours represent the positioning of hydrogen bond acceptors (HBA) with their directionality indicated by arrowheads. Cyan contours represent the hydrophobic feature (HY).

The best pharmacophore model contains four features (Figure 4.5). The green coloured contours represent three hydrogen bond acceptors along with the directionality of the hydrogen bond shown as an arrowhead (HBA₁₋₃), while the blue contour represents a hydrophobic feature (HY). The fitness values indicate how well the features in the pharmacophore overlap with the chemical features in the ligand. These are normalized for the number of features and their associated weights in the pharmacophore model. The most active molecule **166** has a fitness score of 7.36 when mapped to the pharmacophore (Figure 4.5a), while the least active molecule **142** maps to a value of 5.05 (Figure 4.5b). In **166** the HY feature corresponds to the Br atom at C₅. Of the three hydrogen bond acceptors, HBA₁ and HBA₂ correspond to the carbonyl O-atoms of C₄ and C₂ respectively while HBA₃ corresponds to the hydroxyl oxygen at C₃. All the four pharmacophore features are mapped in the training set molecules with pKi values greater than 3.7. For molecules with lesser activity (**135**, **136**, **154**, **155**, **172**, and **174**), at least one feature was missing. For example, **136**, **172** and **174** miss the hydrophobic feature. Although all the features are present in the least active molecule **142**, they do not fit the pharmacophore properly. In the test set, molecules **141** and **144** were predicted with an error greater than 1.5 units. Molecule **141** has an azide group at C₃, similar to **163**, **164**

and **165** (which have activity in the 4.5 pKi range). Also, it has a hydroxyl at ribose C₂. Since the pharmacophore does not take this into account the activity is overestimated. In **144**, the spacing of the azide group with a methylene displaces the N-atom beyond the HBA₃ contour. Thus, the activity is underestimated.

4.3.3 MFA

Use of Alternate Probes

MFA was carried out initially with the generic probe, (H⁺) representing the electrostatic interactions. However, the interaction energies between (H⁺) and all the phosphate containing molecules were too high above the cutoff value overemphasizing the electrostatic energy component. The activities of these molecules were therefore predicted incorrectly. To eliminate this artifact NH₂ and H₂O were used as alternate probes, as outlined in the methods section above, to appropriately sample the electrostatic environment. The choice of these probes was appropriate because of the presence of seven positively charged residues (Arg142, Arg74, Arg95, Arg153, Arg156, Arg160 and Lys13) as well as a number of bound water molecules (Wat1002, Wat1003, Wat1009, Wat1012, Wat10142, Wat1018, Wat1022, Wat1024, Wat1026, Wat1050) in the active site of TMPK_{mt}.

Statistical Details of the Models

Three MFA models using different alignments, namely least squares, pharmacophore and receptor based methods were developed. The statistical details of all the three MFA models generated are summarized in Table 4.3. The cross-validated r^2 (r^2_{cv}) of models I, II and III are 0.790, 0.836 and 0.825 respectively. The predictive power of the three models is calculated by equation 2.2.

Table 4.3: Statistical details of the MFA models obtained from three different alignment methods

	Least square (Model I)	Pharmacophore (Model II)	Docked conformations (Model III)
r_{cv}^2 ^a	0.79	0.83	0.82
r^2 ^b	0.95	0.91	0.90
PRESS ^c	3.80	3.14	2.99
r_{bs}^2 ^d	0.84	0.91	0.90
r_{pred}^2 ^e	0.70	0.56	0.72

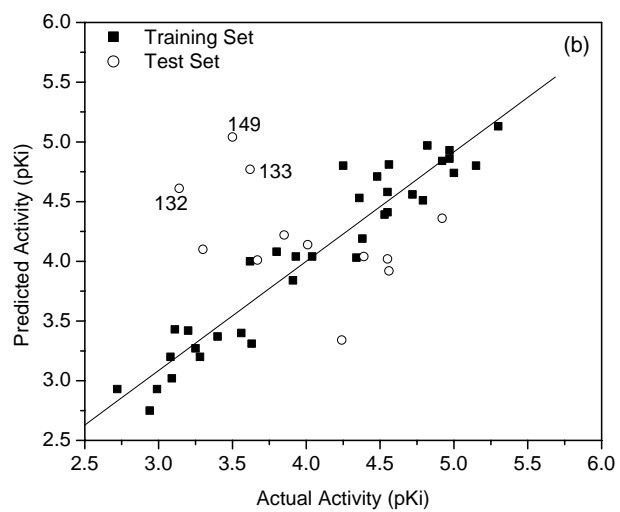
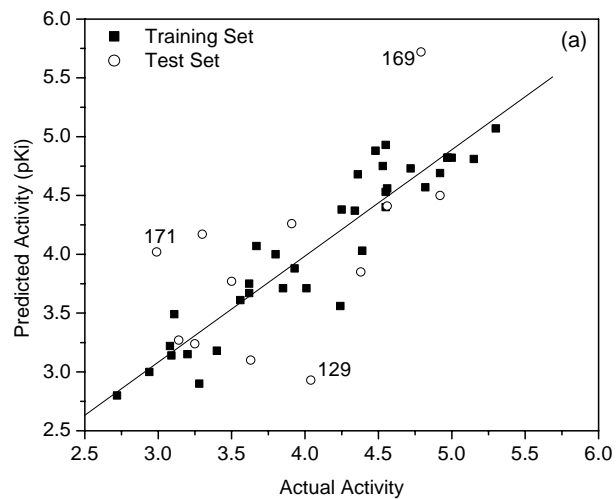
^across-validated r^2 ; ^bconventional r^2 ; ^cpredicted sum of squared residuals; ^dbootstrap r^2 ; ^epredictive r^2

Though the r_{cv}^2 is better for the model II and III than model I, the predictive capability (r_{pred}^2) of model II (0.56) is less when compared to model III (0.72) and model I (0.70). The actual and predicted pKi values of the training and the test set molecules for all the three models are given in Table 4.4. Scatter plots of actual against predicted activities for both the training and the test set molecules for the three different MFA models are given in Figure 4.6. Model III (receptor based) has marginally better statistics than other models. Since receptor based alignment considers all the interactions made by each molecule with the active site residues, it determines the exact orientation of the molecule within the receptor. Thus, these conformations are much closer to the bioactive conformation of the molecules. Accordingly, the r_{pred}^2 values of model III is high.

Table 4.4: Actual and predicted activities of training and test set molecules obtained from the three alignment methods

Molecule No.	Actual activity (pKi)	Least squares (Model I)	Pharmacophore (Model II)	Docked conformations (Model III)
129	4.04	^a 2.93	^b 4.04	^b 4.07
130	5.15	^b 4.81	^b 4.80	^b 5.16
131	4.92	^a 4.50	^b 4.84	^b 4.77
132	3.14	^a 3.27	^a 4.61	^b 3.37
133	3.62	^b 3.75	^a 4.77	^b 3.35
134	4.01	^b 3.71	^a 4.14	^a 3.73
135	2.94	^b 3.00	^b 2.75	^b 3.09
136	3.28	^b 2.90	^b 3.20	^b 3.52
137	3.25	^a 3.24	^b 3.27	^b 2.90
138	3.63	^a 3.10	^b 3.31	^a 3.24
139	4.55	^b 4.53	^a 4.02	^b 4.31
140	3.91	^a 4.26	^b 3.84	^b 3.68
141	3.30	^a 4.17	^a 4.10	^b 3.37
142	2.72	^b 2.80	^b 2.93	^b 3.20
143	3.85	^b 3.71	^a 4.22	^a 5.23
144	4.92	^b 4.69	^a 4.36	^b 5.01
145	4.97	^b 4.82	^b 4.86	^a 4.58
146	4.82	^b 4.57	^b 4.97	^b 4.56
147	4.53	^b 4.75	^b 4.39	^b 4.47
148	3.93	^b 3.88	^b 4.04	^b 3.62
149	3.50	^a 3.77	^a 5.04	^a 3.93
150	4.39	^b 4.03	^a 4.04	^b 4.48
151	4.24	^b 3.56	^a 3.34	^b 4.23
152	4.34	^b 4.37	^b 4.03	^a 3.55
153	4.38	^a 3.85	^b 4.19	^b 4.52
154	3.11	^b 3.49	^b 3.43	^a 3.86
155	3.40	^b 3.18	^b 3.37	^b 3.46
156	3.80	^b 4.00	^b 4.08	^b 3.83
157	3.62	^b 3.67	^b 4.00	^a 3.96
158	4.56	^a 4.41	^a 3.92	^b 4.50
159	4.25	^b 4.38	^b 4.80	^b 4.33
160	4.72	^b 4.73	^b 4.56	^b 4.55
161	4.36	^b 4.68	^b 4.53	^b 4.49
162	4.55	^b 4.93	^b 4.41	^b 4.67
163	5.00	^b 4.82	^b 4.74	^a 5.70
164	4.55	^b 4.40	^b 4.58	^a 4.67
165	4.97	^b 4.82	^b 4.93	^b 5.13
166	5.30	^b 5.07	^b 5.13	^b 5.19
167	3.20	^b 3.15	^b 3.42	^a 3.36
168	3.08	^b 3.22	^b 3.20	^b 3.26
169	4.79	^a 5.72	^b 4.51	^b 4.62
170	4.56	^b 4.56	^b 4.81	^a 4.32
171	2.99	^a 4.02	^b 2.93	^b 3.47
172	3.67	^b 4.07	^a 4.01	^b 3.26
173	4.48	^b 4.88	^b 4.71	^b 4.74
174	3.56	^b 3.61	^b 3.40	^b 3.35
175	3.09	^b 3.14	^b 3.02	^a 2.43

^aTest set; ^bTraining set



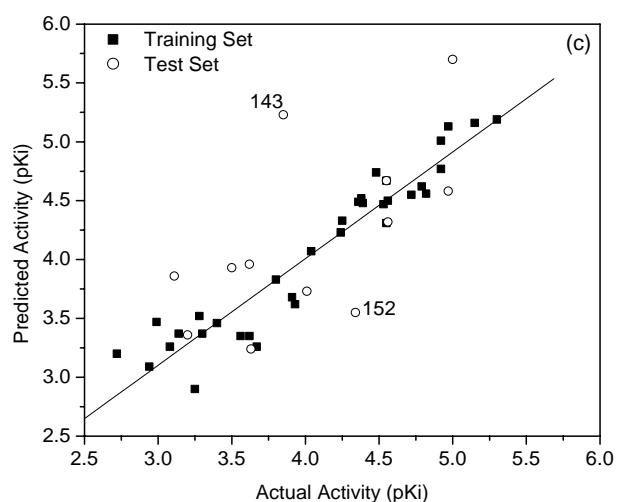


Figure 4.6: Scatter plots of Actual Vs predicted activities of training and test molecules in (a) Model I (b) Model II (c) Model III. Molecules with higher residual values (outliers) are labeled.

The QSAR equations (4.6–4.8) for all the three MFA models are described below. The steric (CH_3) and electrostatic (NH_2 , H_2O) descriptors in the equations specify the regions where structural variations in the ligands result in variation in the biological activity (Figure 4.7). The numbers associated with the descriptor specify its location in the 3D-grid around the molecule.

Model I (*Least Squares alignment*):

$$\begin{aligned} \text{Activity} = & 3.31949 - 0.040242 (\text{CH}_3/187) + 0.036577(\text{H}_2\text{O}/171) - 0.026125 (\text{H}_2\text{O}/99) \\ & - 0.045066 (\text{H}_2\text{O}/246) + 0.027542 (\text{CH}_3/246) - 0.047336 (\text{H}_2\text{O}/117) + \\ & 0.083325 (\text{CH}_3/124) \end{aligned} \quad (4.6)$$

Model II (*Pharmacophore based alignment*):

$$\begin{aligned} \text{Activity} = & 4.19631 + 0.03868 (\text{H}_2\text{O}/218) - 0.026315 (\text{CH}_3/685) - 0.031807 (\text{NH}_2/398) \\ & + 0.023836 (\text{CH}_3/675) - 0.016316 (\text{H}_2\text{O}/476) - 0.027176 (\text{NH}_2/298) - \\ & 0.023874 (\text{H}_2\text{O}/605) \end{aligned} \quad (4.7)$$

Model III (*Receptor based alignment*):

$$\begin{aligned} \text{Activity} = & 2.37154 - 0.070375 (\text{H}_2\text{O}/463) + 0.054086 (\text{CH}_3/534) + 0.036304 (\text{H}_2\text{O}/348) \\ & + 0.043564 (\text{CH}_3/311) + 0.029729 (\text{CH}_3/339) + 0.045245 (\text{CH}_3/125) - \\ & 0.0114228 (\text{NH}_2/331) \end{aligned} \quad (4.8)$$

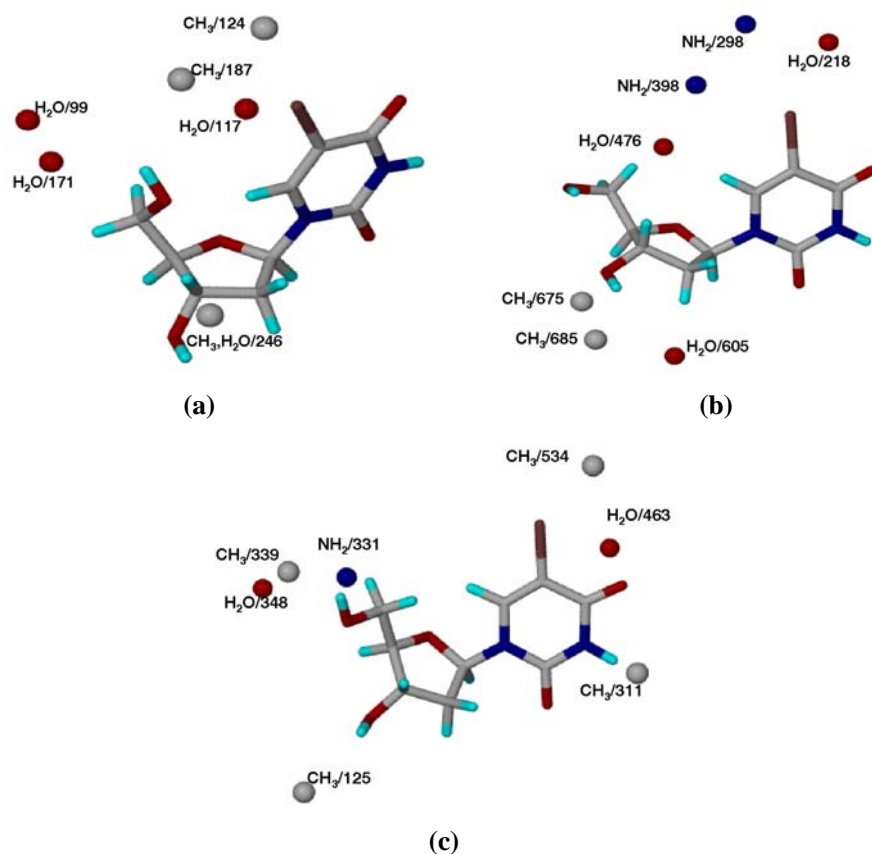


Figure 4.7: Mapping of 3D-descriptor points in the QSAR equations with most active molecule: (a) Model I (Least squares), (b) Model II (Pharmacophore) and (c) Model III (Docked conformations).

The results of all the three MFA models were compared and correlated with the predictive ability of the models. The appearance of steric descriptors ($\text{CH}_3/124$), ($\text{CH}_3/187$) in model I and ($\text{CH}_3/534$) in model III at C_5 indicates that moderately bulky substituents are favored at this position. However, in model II electronic descriptors like

(NH₂/298), (NH₂/398) with a negative coefficient occupy this position indicating that electronegative groups are favored. Br fulfils these criteria and therefore molecule **166** shows the highest activity. The Br...O interaction at Phe36, and the extension of the cooperative network in the active site concurs with the positioning of these descriptors. In model II the presence of (H₂O/605) near ribose C_{2'} with a negative coefficient indicates that a hydrogen bond donor in this position may lead to decrease in activity. Hence, the higher activity of **150** and **151** with respect to **154** and **155** (which have a C_{2'} hydroxyl) is rationalized. This position is not well defined by the model I. The presence of steric descriptors (CH₃/246) in model I, (CH₃/675) in model II and (CH₃/125) in model III with positive coefficient near C_{3'} indicates that bulky substituents are favored. Thus, molecules **144**, **145** and **164** have higher activities than **147**, **157** and **138** respectively (Table 4.4). However, model II shows another steric descriptor (CH₃/685) with a negative coefficient at the same position indicating that bulkier substituents are not preferred. The extent of steric tolerance in this position is not well defined in models III and I. Nevertheless, the coefficient of the descriptor in this position in model I is comparable with that in model II. Thus, molecule **143** with a bulky guanidine group is overestimated only in model III. The presence of (H₂O/246) at C_{3'} in model I also indicates that hydrogen bonding substituents are favored. The appearance of these descriptors is in good agreement with the hydrogen bonds formed by the hydroxyl of dTMP with Asp9 and Wat1009 in the active site (Figure 4.2). The presence of (H₂O/171), (H₂O/99) in model I, (H₂O/476) in model II and (H₂O/348) in model III near ribose C_{5'} indicates that substituents with hydrogen bonding capacities can be accommodated favorably in this position. These represent the water molecules viz., Wat1050, Wat1018 occupying the coordination sites of the Mg²⁺ ion in the active site. Further, the appearance of steric (CH₃/339) and electronic descriptors (NH₂/331) in model III indicates that bulky and highly polar substituents are favored in this position. This is evident by the higher activities of **144**, **145**, **146**, **161** and **163** containing phosphate groups at C_{5'} vis-à-vis molecules **150**, **151**, **152**, **140** and **164** that contain a hydroxyl in this position. Model I does not give any account of the steric requirements in this position. The outliers in each model were rationalized with reference to the location

of the molecular descriptors. Based on the inability of the model I to define the nature of substituents at C₂ and C₅, the least squares method was considered inferior than the other two alignment methods.

In the docking procedure, the conformation of each molecule is dictated by its interactions with the active site residues. Alternately, in case of pharmacophore based alignment, low active molecules are not mapped properly to the pharmacophoric features. Therefore, the conformations of the less active molecules obtained from the pharmacophore based alignment may not be reasonable [4.25]. In the case of least squares method, it is assumed that all the molecules with a similar scaffold binds to the protein in a similar orientation. However, the substituents on the molecule may force a change in the orientation within the active site. Thus, neither of the alignment methods would give the molecular orientation accurately. This is reflected by the low r^2_{pred} value of models I and II. Therefore, receptor based alignment turns out to be superior to the other alignment methods.

4.3.4 Comparison of the Methods

All the three methods used in this study performed well in the analysis of structure-activity relationships. Although MFA, as a 3D-QSAR method, is expected to be sensitive to the alignment procedure employed, it showed only marginal deviations (Table 4.3) between the three alignment criteria that were used for generating the models. Arguably, the MFA model III that used the docking assisted alignment of the ligands is the best predictive model. All the 3D descriptors obtained in the model III concur with the active site residues of TMPK_{mt}. The steric descriptor (CH₃/534) at C₅ corresponds to the cavity formed by the residues Arg74, Phe36, Pro37 and Arg95. This cavity defines the volume of the substituent in this position. Also, the orientation of the backbone carbonyl of Phe36 makes it available for an electrostatic interaction with the substituent in this position (Figure 4.2). Additionally, an electronic descriptor (H₂O/463) in this position corresponds to a water molecule (Wat1012) in the active site. At C₃, the appearance of steric descriptor (CH₃/125) corresponds to the cavity lined by Asp9 and Asp163 indicating a bulkier positively charged substitution favourable in this position.

AZTMP with an azido substitution in this position, therefore, shows higher inhibitory activity [4.26]. The presence of steric and electronic descriptors like (CH₃/339), (NH₂/331) and (H₂O/348) at C₅⁻ signifies that a bulky electronegative substituent is required to interact with the Mg²⁺ ion present in this position and fill the cavity therein. Also, bulkier substituents in this position may displace the water molecules that occupy the co-ordination sites of the metal atom and thereby increase the inhibitory activity. Additionally, the scope of substitutions occupying the water positions and releasing the otherwise interlocked water molecules into the bulk would be entropically favourable [4.27].

The pharmacophoric features obtained by using the *Catalyst* HypoGen method, maps well to the positions of the descriptors in the three dimensional grid around the molecules. These features on the ligands were also found to be complementary to the active site residues and bound water molecules. The pharmacophore feature HBA₁ is complementary to the guanidino group of Arg74, while HBA₂ accepts a hydrogen bond from Wat1002 that in turn forms a hydrogen bond with the side chain hydroxyl of Tyr165. HBA₃ accepts a hydrogen bond from Wat1009 and in turn forms a hydrogen bond with the side chain carbonyl oxygen of Asp163; finally, the HY feature represents the cavity lined by the side chains of Arg95, Pro37 and Ser99.

Although docking is the method of choice to understand the interaction pattern of various molecules in detail, correlation of activities with scores is not at all straightforward [4.28]. Since the scoring function in GOLD has been optimized mainly for the prediction of ligand poses a correlation between the scores and the activities of the TMPK_{mt} ligands could not be obtained. In effect, the MFA method could be confined to predictions for TMP analogs. Also, the additional step of alignment in MFA may dampen its use as a tool for VS. Pharmacophore modeling, on the other hand, can be used as an efficient VS tool as the fitness scores (RMSD) may be directly correlated with the activity and also the pharmacophore model generated may be used for lead-hopping (screening molecules with different scaffolds).

4.4 Conclusions

Virtual screening of large compound libraries for lead discovery can be an attractive and complementary alternative to high-throughput screening. Accurate prediction of binding affinities of the compounds is crucial to such VS strategies, but is computer intensive. Therefore, rational methods to reduce the number of compounds that are assessed for target binding without sacrificing potentially useful leads are of great interest. On these grounds, selection of an optimal method for initial screening of the system of choice is pivotal for a successful VS.

In efforts to find a robust virtual screening tool for identifying putative inhibitors of TMPK_{mt} three well-known methods namely docking, pharmacophore and MFA (3D-QSAR) were compared. All the three methods performed well in the analysis of structure activity relationships. They could provide, within their limitations, an accurate and detailed description of the features required for ligand binding. The pharmacophore features and molecular field descriptors map to important active site residues. Docking confirmed the role of weak interactions in promoting enzyme selectivity towards deoxy ribonucleotides. It also highlighted the role of water mediated cooperative networks and hydrogen bonds to ligand binding affinities. Halogen bonds seem to play a crucial role in increasing the binding affinity of the ligands. Although, all the methods proved efficient, the focus is to be confined to a robust and fast technique. The pharmacophore method, which occupies a middle ground between ligand based MFA and structure based docking, is the most appropriate method for screening large virtual libraries against TMPK_{mt}. The use of pharmacophore portrays remarkable advantages like: 1) providing useful information of the chemical functionality prioritizing the compounds to be synthesized; 2) gaining insights into the features responsible for binding affinity; 3) significantly reducing the number of molecules to be carried forward for further studies; and 4) lead hopping. In this context, initial lead identification can be efficiently carried out by pharmacophore, while docking may be used to prioritize the leads so obtained.

CHAPTER 5

VIRTUAL SCREENING FOR THYMIDINE MONOPHOSPHATE KINASE INHIBITORS USING COMPOSITE PHARMACOPHORE MODEL

5.1 Introduction

Computational procedures have become integral components of most drug discovery programs. Structure based VS methods are emerging as reliable and complementary approaches to high throughput screening (HTS) in the lead discovery process, from hit identification to lead optimization [5.1]. A variety of computational techniques that allow the reduction of enormous virtual libraries ($\sim 10^{12}$) to more manageable proportions constitute the VS protocols [5.2]. In the absence of any knowledge of the target, VS includes ligand-based screens like 1D filters (e.g. molecular weight), 2D filters (similarity [5.3], substructure fingerprints [5.4]) and 3D filters (3D-pharmacophore, [1.24, 5.5, 5.6] 3D shape matching [5.7]). Some of these filters like Lipinski's rule of five can be used as coarse virtual screens for drug likeness [2.34]. Tools based on 2D molecular descriptors cannot adequately gauge the molecular similarity of pairs of compounds that are structurally different but behave alike in binding to a target. In such cases pharmacophore models can be used for 'lead hopping'. However, if conformational flexibility is not considered, the pharmacophore method may miss some of the hits. Structure based VS tools include receptor–ligand docking [3.3, 3.7g, 5.8]. Docking, although computationally intensive, is emerging as a reliable high resolution tool to study recognition between the ligand and its cognate receptor [1.15c, 5.9].

Successful VS depends on the choice of screening tools and their implementation in each phase of the screen. In the present chapter a robust VS model that can combine both the accuracy of the docking and the speed of the pharmacophore method for the system of interest, TMPK_{mt} was developed. The work describes a strategy to search for possible lead structures applying a protocol of several hierarchical filters.

5.2 Methods

5.2.1 Database Generation

Various publicly available databases, namely, IBS (90,000) [5.10], ChemStar (20,000) [5.11], MDPI (10,655) [5.12], bionet (43,176) [5.13] and Asinex (39,000) [5.14] were compiled. For each molecule in the database 100 conformers were generated using the Fast fit method in *Catalyst*. Other databases in the *Catalyst* module including NCI 2000 (238819) [5.15], Maybridge (55273) [5.16] and MiniMaybridge (2000) [5.16] were also included to build a combined database of about 500,000 molecules.

5.2.2 Database Screening

The pharmacophore model developed using HypoGen, discussed in chapter 4, was used as a 3D structural query in the virtual screening. For this purpose an additional hydrogen bond donor feature was added using the 'Merge hypothesis' option in *Catalyst*. This composite pharmacophore model (screen 1) was used for database searching by the 'best flexible search' method in *Catalyst*. The best flexible search method changes the conformations of the molecules in torsion angle space allowing to screen the complete conformational profile of each compound. In a database screening approach, a molecule must be able to map the features of the pharmacophore model used in the search process to be retained as a hit. The hits obtained were further filtered using Lipinski's rule of five (screen 2) and the fitness score ≥ 3.00 (screen 3). The MACCS fingerprints of the hit molecules were used for cluster analysis using the Tanimoto similarity index. The hits were clustered into any particular group if the similarity between the fingerprints was $\geq 60\%$.

5.3 Results and Discussions

Virtual screening can either be carried out using a pharmacophore model or docking. However, docking is both time consuming and computer intensive. Therefore, the pharmacophore model was selected because it gives a hopefully accurate idea of the necessary and sufficient molecular attributes required in a new lead (chapter 4). Actual application of the model in VS however requires that it be further validated. This

validation and screening were carried out in three steps: (1) Construction of a composite 3D pharmacophore model; (2) Validation of this model by ‘spiking’ with known inhibitors and; (3) Screening a large database to find potential novel leads.

5.3.1 Composite Model

The pharmacophoric features obtained were found to be complementary to the active site residues and bound water molecules within (Figure 4.5). The pharmacophore feature HBA_1 is complementary to the guanidino group of Arg74, while HBA_2 accepts a hydrogen bond from Wat1002 that in turn forms a hydrogen bond with the side chain hydroxyl of Tyr165. HBA_3 accepts a hydrogen bond from Wat1009 and in turn forms a hydrogen bond with the side chain carbonyl oxygen of Asp163; finally, the HY feature represents the cavity lined by the side chains of Arg95, Pro37 and Ser99. However, no feature corresponds to the N1 of thymidine ring involved in a hydrogen bond with the carbonyl oxygen of Asn100. This is mainly due to the limitation of the HypoGen derived pharmacophore model to have a maximum number of 7 points (while all the pharmacophore features have one point, HBD and HBA features have two points). The hydrogen bond to Asn100 is one of the specific features of TMP binding site in *M. tuberculosis* when compared with yeast, human or E.coli enzyme structures. Also, this feature was seen to occur in docking across all the molecules used to develop the model. This feature appeared in models 4 and 6 of the top 10 hypotheses (Table 4.2) corresponding to the N₁ of the thymidine ring in the ligand (see Figure 4.2a). In view of the significance associated with this feature a composite pharmacophore model was built including it as shown in Figure 5.1. The distance constraints between all the features of the composite model are shown in Figure 5.1b.

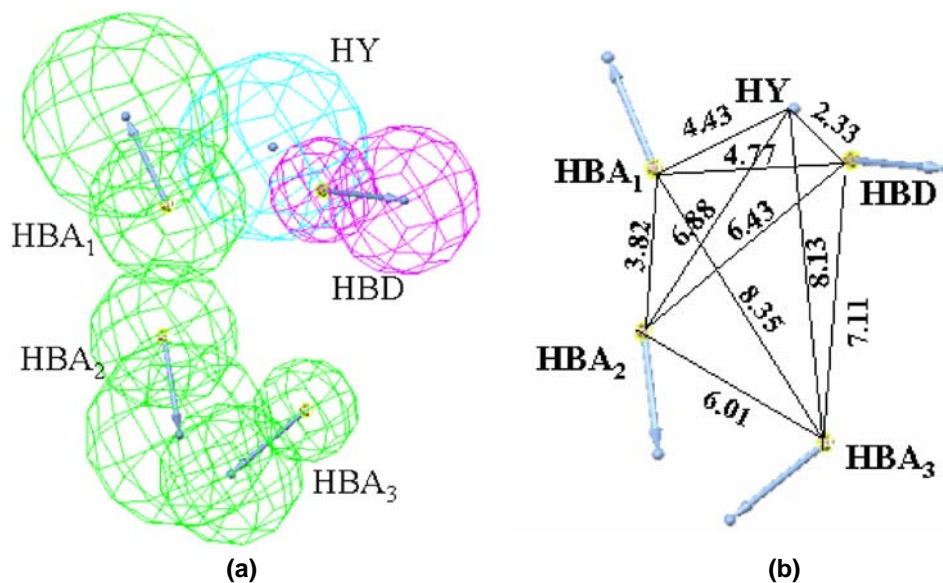


Figure 5.1: Schematic representation of composite pharmacophore model.

5.3.2 Validation of Composite Model

The composite model was then validated by screening a database of compounds with known biological activities and calculating the *enrichment factor* (E) [5.17] using equation (5.1)

$$E = \frac{H_a \cdot D}{H_t \cdot A} \quad (5.1)$$

where, H_t = number of hits retrieved; H_a = number of actives in hit list; A = active molecules present in the database; D = total number of molecules in the database.

For this validation experiment the MiniMaybridge database (containing 2000 molecules) was spiked with the 47 known inhibitors (A) used in this study. When this spiked database (containing 2047 molecules, D) was screened with the pharmacophore model, 110 molecules (H_t) were retrieved from the database as hits. Among these hits 42 (H_a) molecules were from the 47 known actives. Thus, the *enrichment factor* was found to be 17.31, indicating that it is 17 times more probable to pick an active compound from

the hit list than from the database. Further, the hit list obtained is quantitatively scored using “Goodness of Hit list” (GH)²³ using equation 5.2.

$$GH = \left[\frac{H_a (3A + H_t)}{4H_t A} \right] * \left(1 - \frac{H_t - H_a}{D - A} \right) \quad (5.2)$$

GH score gives a good indication of how good the hit list is with respect to a compromise between maximum yield (H_t) and maximum percent of actives retrieved. This is a perfect measure to compare hit lists obtained from alternative methods. The GH score was found to be 0.49 indicating the model to be a good VS screen with 50% efficiency.

5.3.3 Screening

The initial database comprises of 500,000 molecules. VS has been deployed as shown in Figure 5.2. Initially, the validated pharmacophore model was used to screen the large in-house database (*screen 1*). From the initial screen about 5200 molecules were obtained as hits.

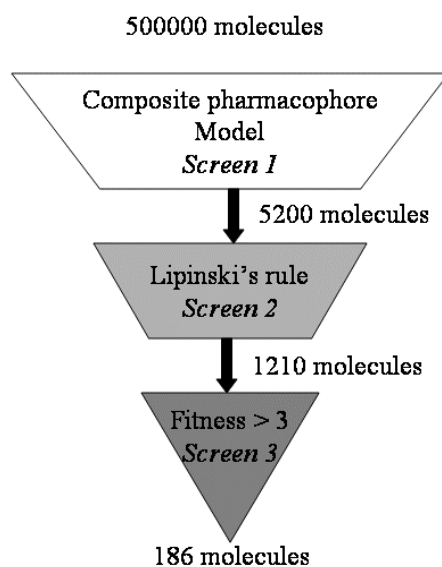


Figure 5.2: Schematic representation of VS strategy

To assess the drug-likeness of these hits a second screen, incorporating Lipinski's rules, was used (*screen 2*). Lipinski's rule of five [2.34], a heuristic guide to assess the oral bioavailability specifies that the compound should possess: i) number of hydrogen bond donors ≤ 5 ; ii) number of hydrogen bond acceptors ≤ 10 ; Mol weight ≤ 500 and; iv) Log P ≤ 5 . A total of 1210 molecules were obtained as hits from this screen. To further increase the probability of the hit to be a lead, a fitness score ≥ 3.00 was used as the third screen. The fitness score indicates how well the features in the pharmacophore overlap with the chemical features in the ligand. A fitness score of 3 indicates that at least three features in the pharmacophore overlap with that of ligand. This ensures that hits obtained include the bioactive conformation within a narrow energy threshold. 186 molecules (listed in appendix II) were obtained as hits from the third screen. These hits comprised mostly of thymidine, benzopteridine phenylsulfonamide and purine derivatives. To understand the chemical space of the final hits, a cluster analysis was carried out using a Tanimoto similarity index of fingerprints. All the hit molecules were grouped into five clusters. Figure 5.3 gives the pharmacophore mapping of representative hit molecules from each cluster. Cluster 1 included purine and pyrimidine derivatives of ribose sugar, while flavanol derivatives constituted cluster 2. Sulfonamide derivatives formed cluster 3 and clusters 4 and 5 included one molecule each that are hybrid purine–sulfonamide structures.

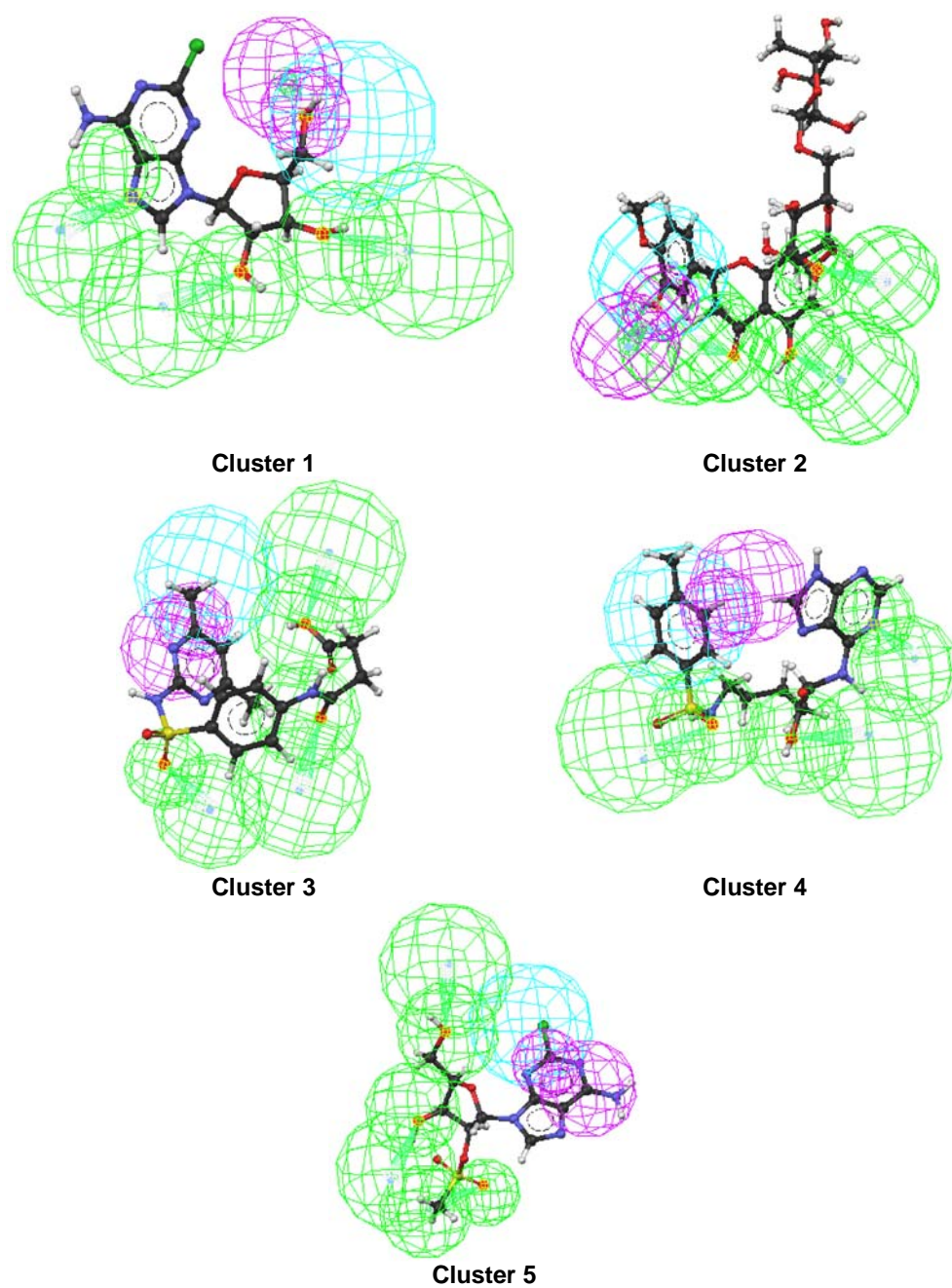


Figure 5.3: Pharmacophore mapping of representative hits from each of the five clusters.

5.4 Conclusions

VS methods are becoming an integral part of the drug discovery process. In the present chapter, a strategy for the screening of large compound libraries to obtain a limited set of prospective hits against TPMK_{mt} has been suggested. A new protocol has been discussed in which successive ‘filters’ were used to reduce the number of compounds from ~5,00,000 to ~200 to be assessed fully for the binding capabilities. With an appreciation of the functionalities involved in molecular recognition acquired from various ligand and structure based methods as described in chapter 4, a composite pharmacophore model was developed with an enrichment factor of 17. The model illustrates important interactions between the TPMK_{mt} and its inhibitors and may serve as a pre-selection tool to check molecules for their inhibitory activity. This composite model was used to screen a database of about 500,000 compounds. Application of subsequent filters ensures that the hit list comprises of drug like molecules. The hits obtained from the VS include molecules with diverse scaffolds that could be possible leads for further development as novel TPMK_{mt} inhibitors.

CHAPTER 6

INSIGHTS INTO LIGAND SELECTIVITY IN NITRIC OXIDE SYNTHASE ISOFORMS: A RATIONAL APPROACH FOR INHIBITOR DESIGN

6.1 Introduction

Alfred Nobel was prescribed nitroglycerin to ease his chest pain. Almost 100 years later it was clarified that the drug acts by releasing nitric oxide (NO). NO is a gaseous, reactive lipophilic molecule that acts as a signal at lower concentrations in many diverse physiological processes including blood flow regulation, neurotransmission, learning and memory [6.1], while at higher concentrations it acts as a defensive cytotoxin against tumor cells and pathogens. In 1998, three researchers Robert Furchgott, Louis Ignarro and Ferid Murad were awarded the Noble prize in medicine for discovering the biological importance of NO. Initially NO was identified as “Endothelium–dependent relaxing factor”. NO is notable among signals for its rapid diffusion, cell permeability and intrinsic instability that eliminates the need for extracellular NO receptors or targeted NO degradation. Pfizer’s blockbuster drug “Viagra” reverses impotence by enhancing an NO-stimulated pathway.

Nitric oxide synthase (NOS) is the key enzyme responsible for the production of NO *in situ* [6.2]. It is expressed as three isoforms: constitutive endothelial NOS (eNOS), neuronal NOS (nNOS), and cytokine inducible NOS (iNOS) isoforms. The isoforms share a common architecture of the N-terminal oxygenase domain and the C-terminal reductase domain connected by a calmodulin (CaM) binding linker. The oxygenase domain binds heme, tetrahydrobiopterin (H₄B) and the substrate, L-arginine while the reductase domain binds to FAD, FMN and NADPH cofactors. While the oxygenase domain provides the site for catalysis, the FAD and FMN of the reductase domain transfer the electrons required for it. Constitutive eNOS and nNOS generate low levels of NO that is imperative for blood circulation and signal transmission in the nervous system respectively [6.3]. The activity of these constitutive enzymes is regulated at the post translational level, through Ca²⁺ dependent CaM binding [6.4]. However, iNOS binds to

CaM irreversibly, independent of Ca^{2+} concentration. It is regulated at the transcription level by pro-inflammatory agents such as endotoxin (bacterial lipopolysaccharide), interleukin-12, and other cytokines in a variety of cell types. Once expressed, iNOS produces high levels of NO over long periods of time as a host defense mechanism.

Generation of NO propagates through a two-step oxidation from L-arginine to L-citrulline [6.5]. The reaction proceeds through a stable intermediate, N^{G} -hydroxy-L-arginine (NHA) (Figure 6.1). The first step involves hydroxylation of guanidinium nitrogen to form NHA, a mixed-function oxidation similar to those catalyzed by the cytochrome P-450. It includes sequential transfer of one electron from $\text{NADPH} \rightarrow \text{FAD} \rightarrow \text{FMN} \rightarrow \text{heme}$ to form a heme-oxygen complex, while H_4B provides the other electron for the formation of heme-peroxy complex [6.6]. In the second step NHA is oxidized to L-citrulline and NO radical.

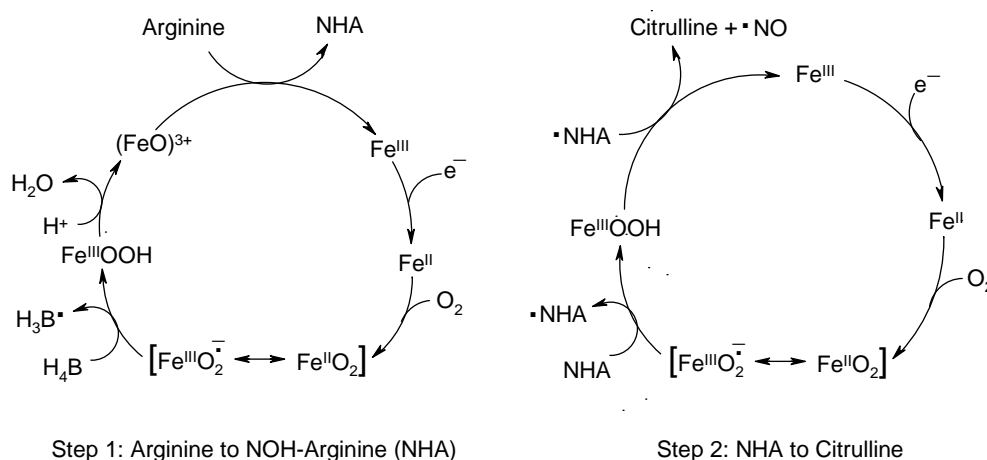


Figure 6.1: Mechanism of NO production in NOS

Excessive production of NO generated either by iNOS, or by the sustained activation of nNOS, elicits cellular toxicity and tissue damage leading to various pathological conditions. Overexpression of iNOS derived NO is implicated in diseases like immune type diabetes, inflammatory bowel disease, rheumatoid arthritis, carcinogenesis, multiple sclerosis, transplant rejection, and septic shock [6.7]. On the

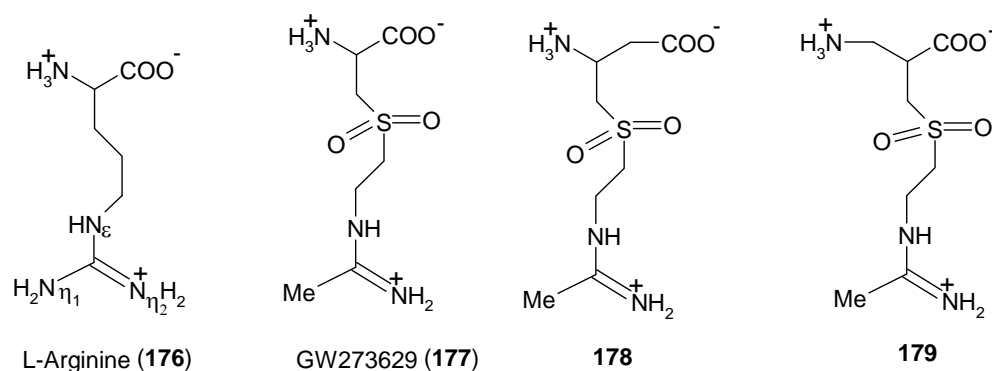
other hand, overproduction of NO by nNOS is linked to the pathogenesis of stroke, Alzheimer's, Parkinson's and Huntington's diseases. However, NO generated from eNOS is shown to be critical for angiogenesis and for maintaining the vascular tone. Also, the role of eNOS derived NO in maintaining the blood circulation is well implicated with the use of NO donating compounds as antihypertensives [6.8]. Therefore, isoform specific inhibitors are desirable to specifically inhibit NO production by iNOS and nNOS under pathological conditions without inhibiting the functions of eNOS.

A number of substrate based analogs like N^ω-methyl-L-arginine methyl ester (L-NMMA), N^ω-nitro-L-arginine (L-NNA), N^δ-iminoethyl-L-ornithine (L-NIO), L-N⁶-(1-iminoethyl)lysine [6.9] and non-amino acid analogues like iminopiperidines [6.10], quinazolinamines [6.11], aminoguanidines, isothiureas, carbamidines has been developed and show little isoform selectivity. Additionally, inhibitors that target the H₄B site of the enzyme have also been developed [6.12]. Nitroindazoles on the other hand bind to both the substrate and H₄B binding sites. However, these ligands show minimal selectivity towards the isoforms. Very few molecules are known to show selective inhibition of NOS isoforms, and the basis of their selectivity remains unknown.

Several crystal structures of human iNOS and eNOS isoforms with various classes of inhibitors have been determined [6.13]. This has helped to identify the structural variations and their implication in the design of selective inhibitors. NOS isoforms share a common topology, with a striking degree of conservation in the active site. The oxygenase domains of iNOS and eNOS structures have 57% sequence identity over 435 residues. The RMSD of the backbone atoms between iNOS and eNOS oxygenase domains is 0.714 Å. The active site of the isoforms is highly conserved with only one residue differing (Asp382 in iNOS mutated to Asn366 in eNOS) among those forming direct contacts with the substrate. Fischmann *et al.* have addressed the issue of utilizing this key difference in designing isoform specific inhibitors [6.14].

The high structural similarity between the active sites of the isoforms presents a challenging problem for the design of selective inhibitors. To circumvent these problems it is essential to critically examine the structural differences in the active site among

different isoforms and to understand how the known isoform selective inhibitors interact with their active sites. In this context the substrate L-arginine (**176**) and GW273629 (**177**), a selective iNOS inhibitor with IC_{50} values of $8\mu\text{M}$ and $>1000\mu\text{M}$ for iNOS and eNOS respectively [6.15], were selected for the present study. Understanding the structural and dynamical features of binding of molecule **177** with these isoforms is expected to further increase our insights towards its selectivity and aid in the development of selective and potent inhibitors of the two isoforms. With this perspective, the molecular recognition pattern of the substrate and inhibitor, **177** was thoroughly explored in iNOS and eNOS isoforms. Based on these investigations, active site differences were exploited to bring in selectivity to the isoforms. To accomplish this, molecular dynamics (MD) simulations were carried out with isoforms iNOS and eNOS complexed with substrate (**176**), GW273629 (**177**) and two designed ligands **178** and **179** as shown in Scheme 6.1.



Scheme 6.1: Ligands used in the MD Simulations

6.2 Overview of Molecular Dynamics

MD simulations give time evolution of conformations of molecules as follows:

For an atom, i , with mass m_i , and position indicated by the 3-dimensional vector \mathbf{r}_i , the relationship between the atom's velocity and momentum, p_i , is

$$\frac{dr_i}{dt} = \frac{p_i}{m_i} \quad (6.1)$$

The net force, F_i , exerted on the atom i by the remainder of the system is given by the negative gradient of the potential energy function with respect to the position of atom i

$$F = -\frac{dV}{dr_i} \quad (6.2)$$

The Newtonian equation of motion for atom i is

$$\frac{dp_i}{dt_i} = F \quad (6.3)$$

Given the position with respect to a single component of vector \mathbf{r}_i , (that is the position along a single dimension, x) at a specific time, t , then the position after a short and finite interval, denoted by Δt , is given by a standard Taylor series expansion

$$x(t + \Delta t) = x(t) + \frac{dx(t)}{dt} \Delta t + \frac{d^2x(t)}{dt^2} \frac{\Delta t^2}{2} + \dots \quad (6.4)$$

The position $x(t)$, the velocity $dx(t)/dt$, and the acceleration $d^2x(t)/dt^2$ are sufficient for numerical solution to the equations of motion. Therefore, approximations are made such that the higher terms sum to zero, effectively truncating the Taylor expansion up to the second derivative, the acceleration.

Newton's second law describes the acceleration as:

$$\frac{d^2x(t)}{dt^2} = \frac{F_x}{m} \quad (6.5)$$

where F_x is the component of the net force acting on the atom parallel to the direction of x .

These equations of motion are integrated with numerous algorithms. The simple Verlet algorithm uses the atomic positions and accelerations at time t and the positions from the prior step, $x(t - \Delta t)$, to determine the new positions at $t + \Delta t$

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + \frac{d^2x(t)}{dt^2} \Delta t^2 \quad (6.6)$$

This gives a trajectory of conformations at definite time intervals with associated kinetic and thermodynamic properties.

6.3 Methodology

6.3.1 Initial Structure Preparation

The crystal structures of iNOS and eNOS isoforms complexed with L-NHA and L-arginine respectively were taken from the protein data bank [1NSI (iNOS); 3NOS (eNOS)]. The isoform exists as a dimer with structural Zn^{2+} ion coordinating with the four cysteines, two cysteines from each monomer: Cys110 and Cys115 of each monomer in iNOS (Cys95 and Cys99 in eNOS). The Zn^{2+} ion was excluded from the simulation because (i) it is situated ~ 22 Å away from the substrate bound site and has no effect on the substrate binding. (ii) Li *et al.* have shown experimentally that in the absence of Zn^{2+} ion, the symmetry related cysteines form a disulphide bridge [6.16]. Only monomer of each isoform (with 425 and 404 residues in iNOS and eNOS respectively) was used in the study. H-atoms were added to the complex at pH 7 using the biopolymer module in the *Insight II* program. The His residues 102, 162, 230, 333, 436, 477, 499 in iNOS and residues 214, 420, 421 and 461 in eNOS were designated as the ϵ -tautomers, while the rest were designated as δ -tautomers based on their hydrogen bonding preferences. Force field charges were applied while assigning a charge of +3 on the Fe atom of heme. The complex was minimized using steepest descent and conjugate gradient algorithms with a convergence criteria of 0.1 and 0.01 kcal mol⁻¹ Å⁻¹ respectively. The energy minimization and MD simulations were performed with ‘Extensible Systematic Force Field’ (ESFF) [6.17] in the *Discover* module of *Insight II*. All the other ligands used in the simulation (Scheme 6.1) were constructed and energy minimized using the OFF methods module in *Cerius*². Geometry optimization was carried out with the MOPAC 6 package using the semi-empirical AM1 Hamiltonian.

6.3.2 Molecular Docking

All the four ligands in the study were docked into the structures of iNOS and eNOS enzymes. Docking was carried out using GOLD as discussed in the methods

section of chapter 2. The best ranked solution of the ligands was used for the MD simulations.

6.3.3 Molecular Dynamics

To study the structural and dynamical properties of the protein, restricted MD simulations of each isoform were performed with the substrate and the inhibitor. The oxygenase domain of NOS is seen to be rigid with an exposed active site. A series of inhibitors co-crystallized with NOS show that the conformational changes in the isoform are confined to side chains of the active site residues [6.13d]. The β strand architecture of the active site has been implicated to impart rigidity to the NOS structure [6.18]. Considering these facts, the simulations were confined to all the protein residues that contain at least one atom within 10 Å sphere from any atom of the ligand to reduce the computation time. Incidentally, this includes all the residues lining the substrate access channel. All the protein–ligand complexes were built in the *Biopolymer* module using the best docking solution obtained from GOLD. The complexes were explicitly solvated using the TIP3P [6.19] water model. The thickness of the hydration layer was selected to be 4 Å such that the total number of atoms in the complex did not exceed 10000 atoms (limitation of ESFF). This resulted in about 1250 water molecules in each system. The crystallographic waters and the water molecules added in the simulation were differentiated by adding ‘H’ and ‘w’ in their nomenclature respectively. Nonbonded interactions were calculated with a cut-off radius of 12 Å. Initially, water molecules were equilibrated with the system for about 50ps while keeping all the protein heavy atoms fixed. All the simulations were carried out at room temperature, 300 K, with the temperature controlled by a Nosé-Hoover chain thermostat [6.20]. The simulations were carried out for 1.2 ns with NVT ensemble and a time step of 1fs. Newton’s equations of motion were calculated using the velocity verlet algorithm. All the complexes were equilibrated for 200ps and the trajectory was collected at each ps for 1ns. The nonbonded interaction energies between the ligands and the active site residues were calculated using *Discover* module in *Insight II*.

6.4 Results and Discussion

6.4.1 Overall Structural Changes

The analysis of the MD simulations for the fully solvated complexes of iNOS and eNOS with substrate (**176**), known inhibitor GW263729 (**177**), and the two hypothetical ligands **178** and **179** is now discussed. The potential energy of the system is a simple measure of its stability. Plots of the potential energy as a function of time are shown in Figure 6.2, which indicate that all the systems in the study were well equilibrated and remain stable throughout the simulation. The stability of each protein complex during the simulation has also been monitored by plotting the RMSD of the complexes with respect to their initial structures as shown in Figure 6.3. The RMSD of the iNOS and eNOS complexes converged with mean values of 0.75 Å and 1.5 Å respectively, which are within acceptable limits. The higher RMSD values of eNOS complexes depicts that eNOS has structural instability on binding with all the ligands when compared to iNOS. The average hydrogen bond distances of all the interactions between the molecules **176-179** with the active site residues of iNOS and eNOS are given in Tables 6.1 and 6.2 respectively.

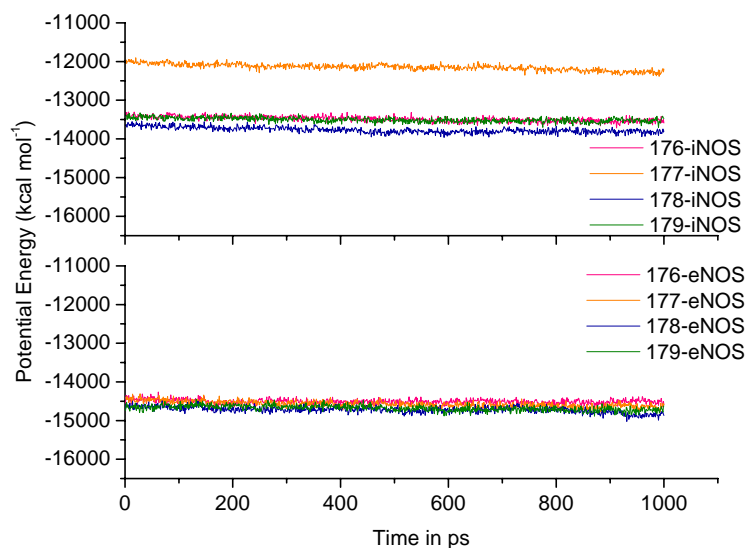


Figure 6.2: Potential Energy plots of ligand complexes with iNOS and eNOS isoforms

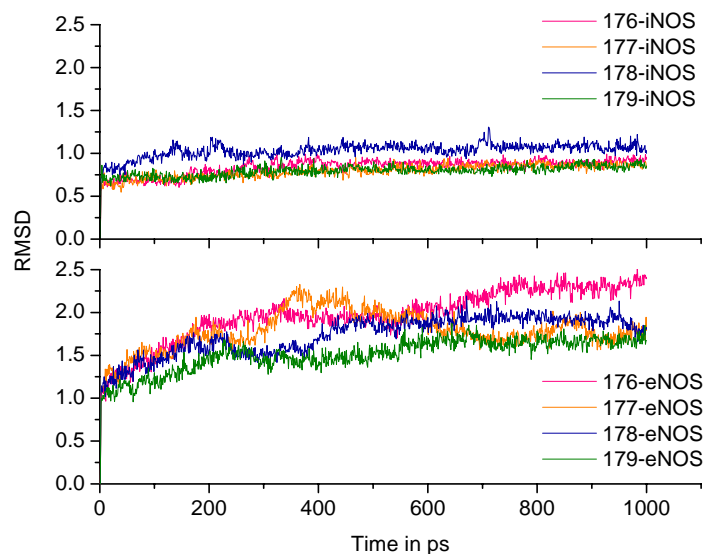


Figure 6.3: Plots of RMSD from initial structures of ligand complexes in iNOS and eNOS isoforms.

Table 6.1: Average hydrogen bond distances between active site residues and ligands **176-179** in iNOS

Active site residues	176	177	178	179
Glu377	1.63±0.08	1.56±0.06	1.57±0.07	1.58±0.06
Glu377	1.59±0.07	1.67±0.14	1.58±0.07	1.60±0.08
Trp372	1.67±0.09	1.70±0.11	1.68±0.11	1.75±0.15
Tyr373	1.78±0.13	—	—	—
Asp382	—	1.96±0.63	2.08±0.38	—
Arg266	1.63±0.08	2.30±0.34	2.24±0.92	2.76±0.75
Gln263	3.45±0.56	1.80±0.12	3.00±0.37	—
Heme	1.89±0.26	—	—	2.78±0.20
Arg388	—	2.06±0.39	2.41±0.76	2.68±0.83
H ₄ B	—	—	—	2.19±0.69

Table 6.2: Average hydrogen bond distances between active site residues and the ligands **176-179** in eNOS

Active site residues	176	177	178	179
Glu361	1.86±0.23	1.61±0.08	1.59±0.09	1.60±0.07
	1.87±0.16	1.63±0.09	1.60±0.08	1.64±0.09
Glu361	1.63±0.23	—	—	—
Trp356	1.93±0.19	1.65±0.10	1.64±0.08	1.66±0.10
Tyr357	1.88±0.19	1.71±0.10	3.09±1.57	1.76±0.20
Asn366	1.89±0.17	2.04±0.27	2.60±0.83	2.57±0.65
Gln247	1.99±0.28	1.95±0.20	2.00±0.36	3.17±0.69
Arg250	3.15±0.16	2.18±0.68	2.19±0.63	1.81±0.29
Heme	1.74±0.24	2.42±0.69	1.60±0.09	—
Arg372	—	1.75±0.18	—	—
Tyr331	—	—	—	2.37±0.69

6.4.2 Mode of Substrate Binding in iNOS and eNOS Isoforms

The substrate L-arginine (**176**), binds to the active site of iNOS and eNOS in a similar manner. All the residues corresponding to eNOS are given in parentheses in this section for ease of comparison. The substrate consists of three functional groups: the ammonium and the carboxylate groups in the head region and the guanidinium group in the tail region. The tail region of the substrate is anchored in the active site through a bidentate interaction between the guanidinium nitrogen atoms (N_{ϵ} , N_{η_1}) and the side chain carboxylate O-atoms of the Glu377 (Glu361) as seen in Figure 6.4.

The importance of this residue has been revealed by site-directed mutagenesis studies where substitution of Glu by Gln, Leu or Ala resulted in loss of activity [6.21]. Additionally, the N_{η_1} atom of the substrate hydrogen bonds to the backbone carbonyl of Trp372 (Trp356). These interactions bring the guanidinium N_{η_2} atom of the substrate closer to the Fe of heme (3.6 Å in iNOS and 4 Å in eNOS). The close proximity of the heme iron to the N_{η_2} atom is consistent with the catalysis models of NOS [6.22]. The guanidinyll group stacks over the heme *meso* carbon and the edge of pyrrole ring A with an angle of ~25° to the heme plane. The hydrogen bonds between the guanidinium group and the carboxylate of Glu377 (Glu361) are relatively short (Table 6.1) due to charge assisted nature of hydrogen bonds. This has also been observed by Li *et al.* who reasoned

that such short interactions were necessary to accommodate NO binding to Fe of heme during catalysis [6.23].

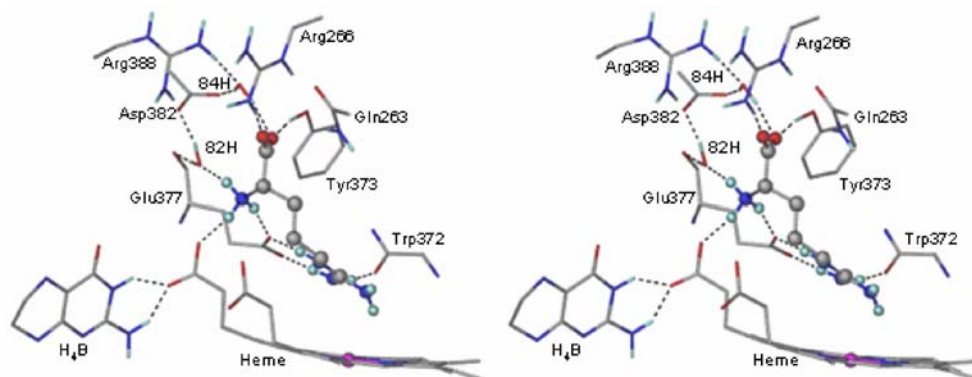


Figure 6.4: Stereoview of active site of the **176-iNOS** complex. Molecule **176** (L-Arg) is shown in ball-and-stick representation. Hydrogen bonds are shown as dotted lines. Only selected H-atoms are shown for clarity.

The binding conformation was also stabilized by hydrogen bonds formed by the ammonium and the carboxylate groups of the substrate in the head region: these groups seem to modulate the differences in molecular recognition of the substrate in the two isoforms. In both **176-iNOS** and **176-eNOS** complexes, the ammonium group of the substrate form hydrogen bonds with the water 82H (168H) [H corresponds to the crystal water], the heme carboxylate of pyrrole ring A and the side chain carboxyl O ϵ_1 -atom of Glu377 (Glu361). Although the interacting residues are common in both the complexes, they differ in the entropy associated with these interactions. In iNOS, the interactions of the ammonium H-atoms interchange with the three acceptors viz., water 82H, heme and Glu377. Plots of the distances between the acceptor and the H-atoms of the ammonium group with time showed that these hydrogen bonds were retained throughout the simulation (Figure 6.5a). The flipping of the three H-atoms of the ammonium nitrogen to form uninterrupted hydrogen bonds can be seen clearly. However, no such flipping was observed in the **176-eNOS** complex (Figure 6.5b).

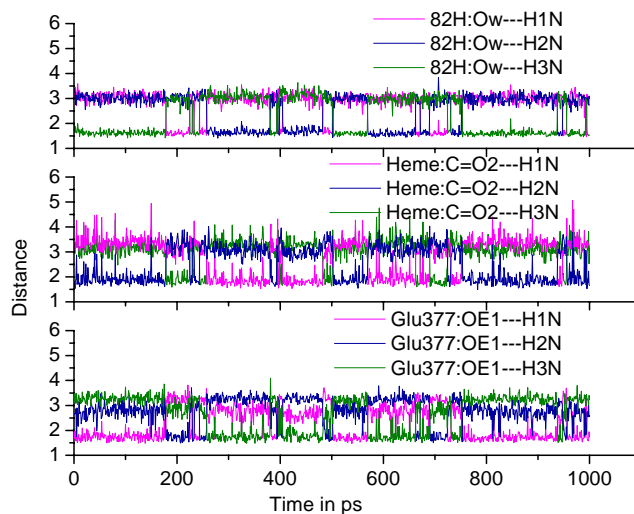


Figure 6.5a: Interactions between the ammonium group of the substrate and the active site residues in the **176-iNOS** complex. The interchanging of the three ammonium hydrogen atoms among the acceptors can be seen.

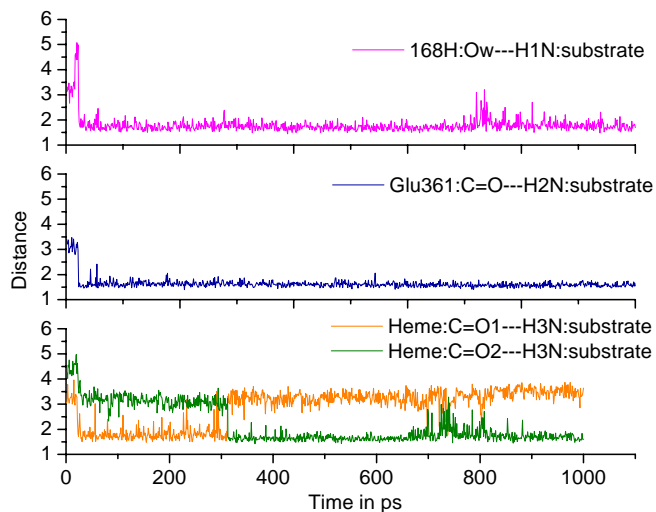


Figure 6.5b: Interactions between the ammonium group of substrate **176** and the active site residues in the **176-eNOS** complex.

Grant *et al.* observed experimentally that the hydrogen bond between ammonium group of the substrate/inhibitor and the enzyme is essential for the catalysis and the absence of ammonium group in the ligand leads to loss of inhibitor activity [6.24]. However, their study limited the observation to iNOS and did not show details of the hydrogen bonds in the isoform. Further in the **176-iNOS**, water 82H was seen to mediate hydrogen bond interaction with Asp382, whereas in the **176-eNOS** complex no such interaction was seen between the ammonium group and the corresponding residue Asn366.

In **176-iNOS**, the carboxylate group of the substrate forms hydrogen bonds with water 84H, Tyr373 and Arg266 (Figure 6.4). Although an interaction with Gln263 was also seen in the crystal structure, it was not retained throughout the MD simulation. In the **176-eNOS** complex, the carboxylate of the substrate interacts with water 168H, Tyr357, Asn366 and Gln247 (Figure 6.6). The electrostatic interaction with Arg250 (corresponding to Arg266 in iNOS) is absent in the **176-eNOS** complex. Interestingly, water 168H binds to both the carboxylate and the ammonium groups of the substrate. This could be one of the reasons for the entropy loss in the **176-eNOS** complex. In addition to the enthalpic contribution, an entropic advantage due to the ammonium group in the **176-iNOS** complex provides extra stabilization when compared to **176-eNOS** complex. Apart from these interactions, van der Waals contacts were observed between the aliphatic chain of the substrate and the residues Pro350 (334), Val352 (336), and the alkyl chain of the heme propionate.

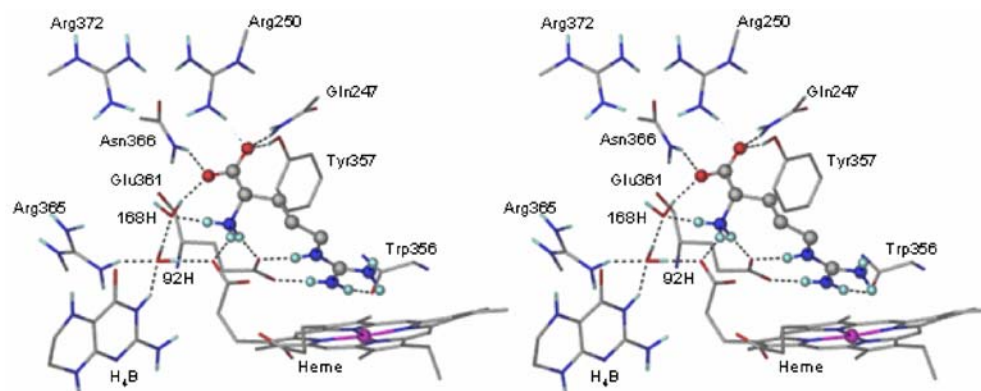


Figure 6.6: Stereoview of the active site in the **176-eNOS** complex. Molecule **176** is shown in ball-and-stick representation.

6.4.3 Differences in Substrate Recognition in the Isoforms

The major difference in substrate recognition among the isoforms is the interaction with Asp382 (Asn366). While the ammonium group of the substrate participates in this recognition event through water mediation in iNOS, the carboxylate group interacts directly with the corresponding residue, Asn366 in eNOS (Figure 6.7). In comparison, substrate in eNOS rotates $\sim 80^\circ$ about C_α -COO⁻ bond to interact with the amide group of Asn366. Because of electrostatic repulsion between the carboxylate groups of Asp382 and the substrate in iNOS, the carboxylate group of the substrate moves away allowing the ammonium group to make water mediated interaction with Asp382. The juxtaposition of both the ammonium and the carboxylate groups in the substrate makes this variation feasible in both isoforms.

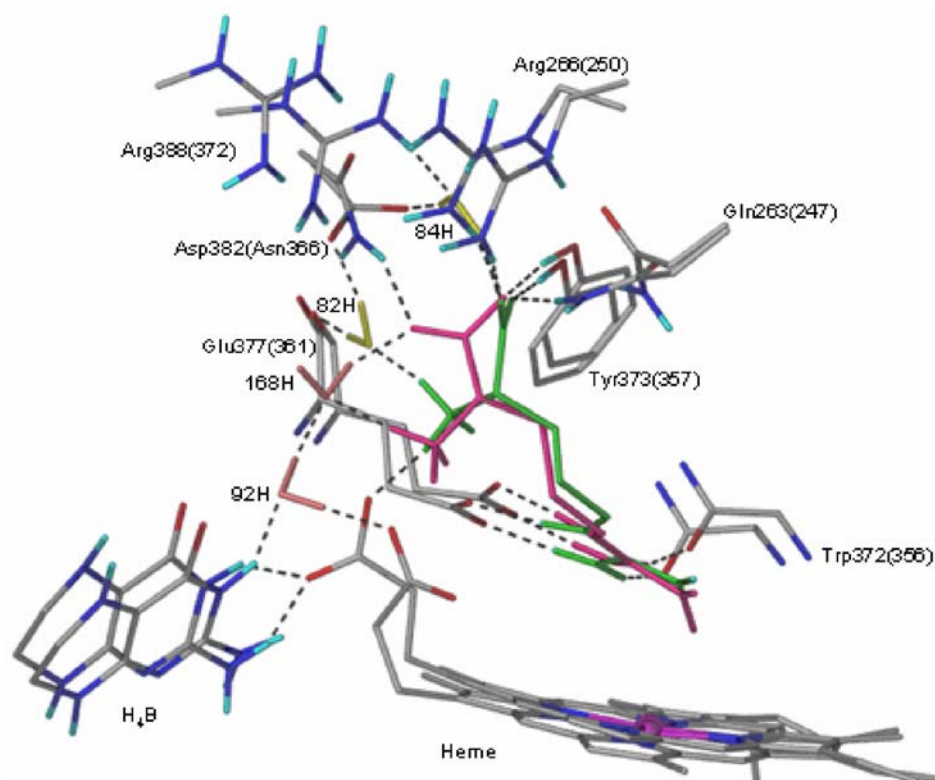


Figure 6.7: Overlay of the active sites of average structures of iNOS and eNOS complexed with substrate **176**. Substrate and water molecules pertaining to iNOS are coloured green and yellow respectively while that of eNOS shown in pink and orange.

6.4.4 Heme–H₄B Interaction in Isoforms

The heme–H₄B interaction is crucial for the H₄B mediated electron transport in the catalysis [6.6b, 6.25]. However, this interaction has been seen to be distinct in the isoforms during the simulation. While the heme carboxylate O-atom makes a direct hydrogen bond with H₄B in iNOS, the connection is interrupted by a water molecule (92H) in eNOS (Figure 6.8). In other words, the direct interaction in eNOS becomes a water bridged interaction at 300ps. This could be attributed to the drift in the hydrogen bond interaction of the ammonium group of molecule **176** between the two carboxylate O-atoms of the heme propionate group at 300ps as seen in Figure 6.5b. Consequently, the RMSD of the heme is relatively high in eNOS compared to that of iNOS (Figure

6.9). A similar observation was found with crystal structures of iNOS and eNOS complexes of nitroindazoles, wherein movements in heme propionate occur in eNOS complexes with the H₄B being displaced from its site. However, the heme in iNOS complexes is relatively stable with the H₄B tethered to it [6.26].

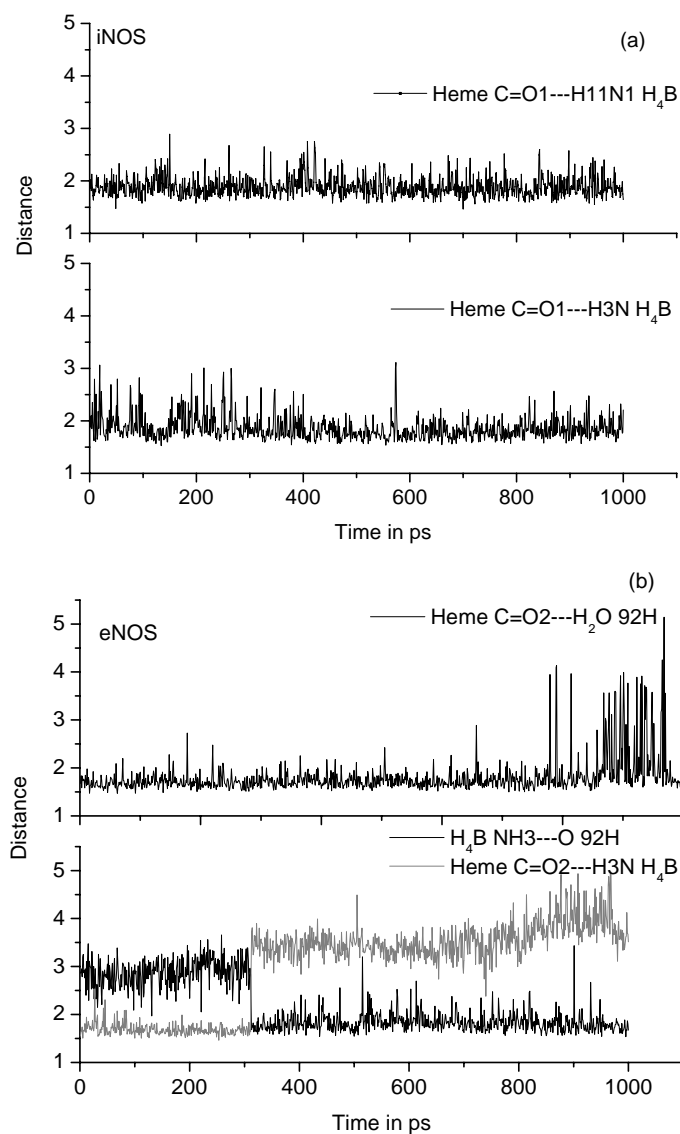


Figure 6.8: Hydrogen bond distance plots of heme and H₄B interactions in (a) 176-iNOS (b) 176-eNOS complexes

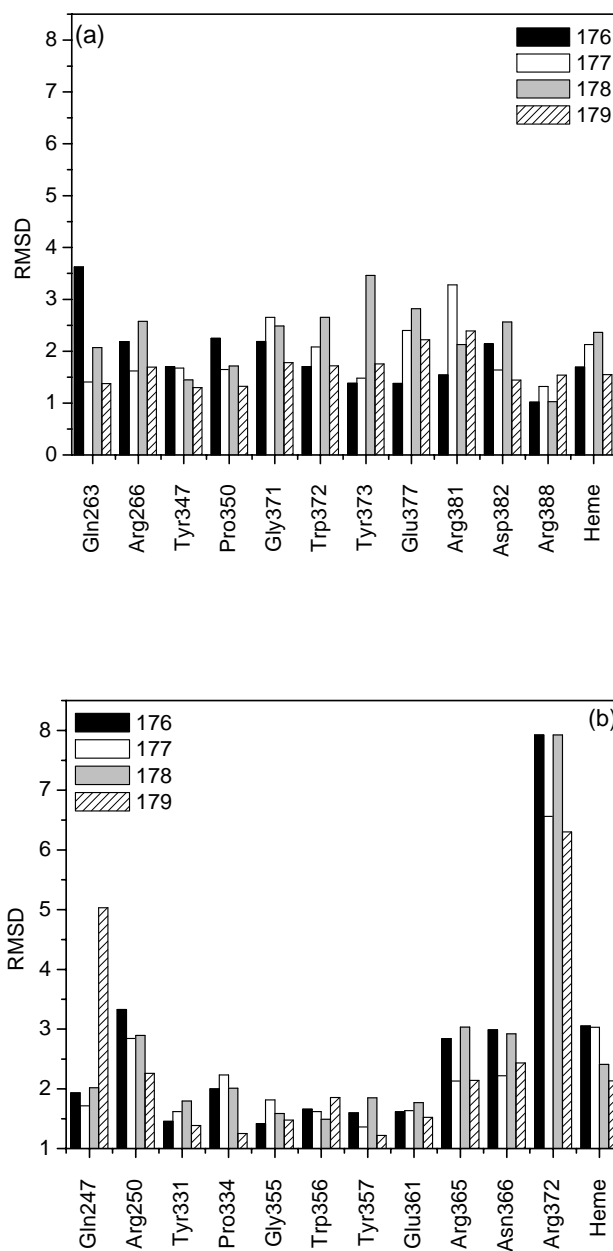


Figure 6.9: RMSD of active site residues from its initial structure in (a) iNOS and (b) eNOS complexes

6.4.5 Role of Water in Substrate Binding

Water molecules trapped at the binding interface ostensibly constitute an unfavourable event owing to the entropic penalty. Conversely, water can contribute to the free energy of binding if the entropic loss is compensated by the enthalpic gain by forming (maximum number of) hydrogen bonds [4.28]. Two such water molecules are trapped at the interface of the **176-iNOS** complex, namely 82H and 84H. A detailed analysis of X-ray structures of NOS isoforms complexed with different ligands revealed that these water molecules are highly conserved. Each of these water molecules forms three hydrogen bonds with the active site residues. Water 82H mediates an interaction between the ammonium group of the substrate and the side chain carboxylate O-atom of Asp382. It also interacts with the backbone carbonyl of Glu377 (Figure 6.4). Water 84H mediates an interaction between the carboxylate group of the substrate and the active site residues Asp382, and partially with Arg388, Arg266, Tyr347 (~ for 500ps). The dielectric property of the water attenuates the electrostatic repulsion between the carboxylate groups of the substrate and Asp382. While water mediated interactions are prominent in the **176-iNOS** complex, the substrate in **176-eNOS** complex directly interacts with the active site residue Asn366. Additionally the **176-eNOS** complex is stabilized by a σ -bond cooperative chain, Heme C=O...HOH (92H)...OH(168H)...O=C(**176**). Participation of water molecules in the molecular recognition of ligand within the active site is commonly found in several proteins [6.27]. However, in this context, water molecules seem to be critical in masking the active site differences, thereby facilitating the substrate binding in the NOS isoforms.

6.4.6 Mutation Studies of iNOS and eNOS

Being the major difference in the substrate recognition across the isoforms, Asp382 and Asn366 residues were interchanged among the isoforms to check the role of these residues and water. The anchor interactions between the tail region of the substrate and the side chain carboxylate of Glu377 (Glu361) and backbone carbonyl of Trp372 (Trp356) were retained whilst the interactions with the head region ammonium and carboxylate groups of the ligand showed no major changes in the recognition event. Only

subtle changes were observed with the water mediated interactions due to the replacement of O-atom \rightarrow N-atom (Asp \rightarrow Asn) and vice versa. In iNOS, to accommodate the donor \rightarrow acceptor change, water 84H flips around to interact with Tyr347 as seen in Figure 6.10. While in eNOS, hydrogen bond between the carboxylate of the ligand and the amide nitrogen of Asn366 becomes water mediated upon mutation in order to minimize the electrostatic repulsion between the carboxylate groups of ligand and Asp as observed in Figure 6.11 (cf. section 6.3.2). This confirms the decisive role of water.

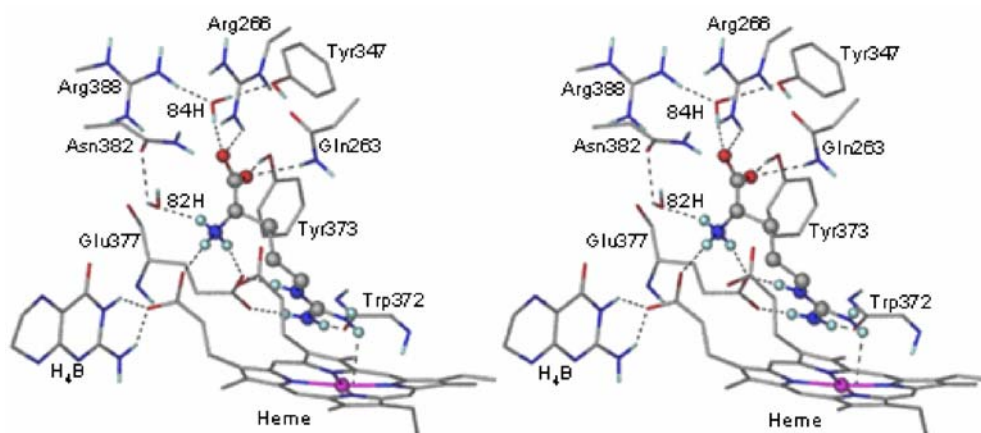


Figure 6.10: Stereo view of active site of **176**-iNOS_{mut} complex. Molecule **176** (L-Arg) is shown in ball-and-stick representation.

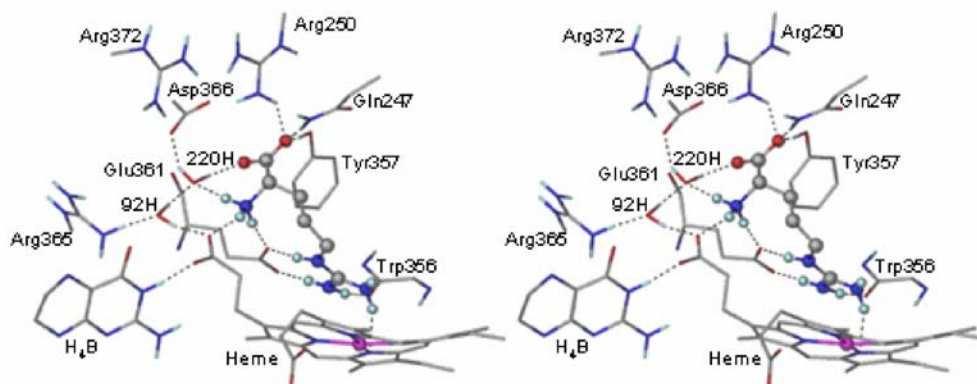


Figure 6.11: Stereoview of active site of **176**–eNOS_{mui} complex. Molecule **176** (L-Arg) is shown in ball-and-stick representation.

6.4.7 Mode of Inhibitor Binding in iNOS and eNOS Isoforms

Replacement of the guanidinium nitrogen $N\eta_2$, required for the production of NO during the catalysis, with an amidino methyl carbon precludes molecule **177** from serving as substrate (see Scheme 6.1). The methyl group sits snugly in the hydrophobic cavity formed by Gly371 (Gly355), Phe369 (Phe353), Val352 (Val337) and Pro350 (Pro335). Figure 6.12 and 6.13 shows that the anchoring interactions between the amidino tail group of molecule **177** and that the carboxylate of Glu377 (Glu361) and Trp372 (Trp356) are retained in both isoforms.

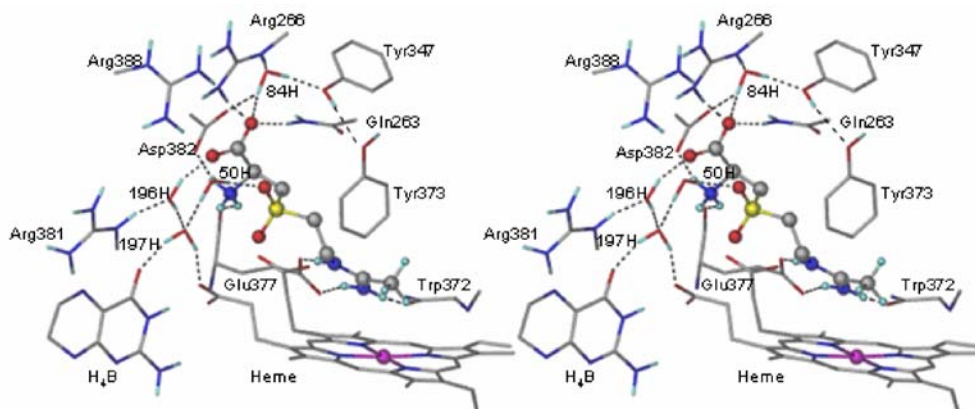


Figure 6.12: Stereoview of the active site of the **177**–iNOS complex. Molecule **177** is shown in ball-and-stick representation.

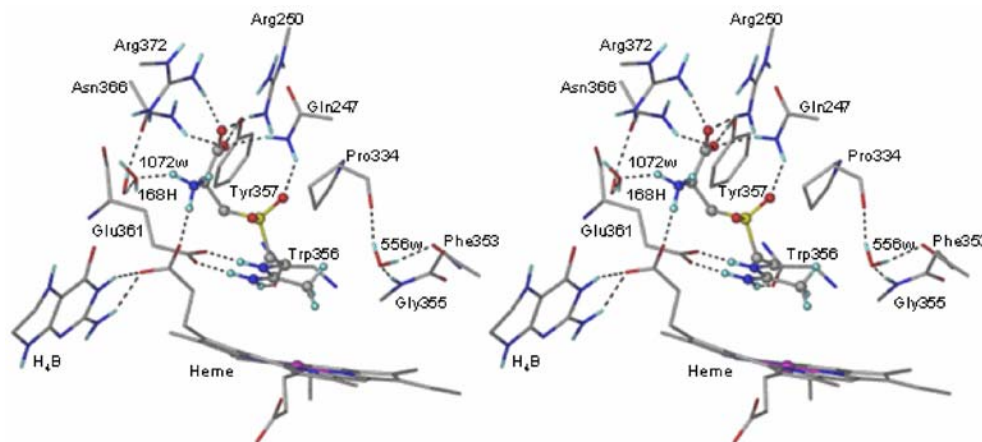


Figure 6.13: Stereoview of the active site of the **177**–eNOS complex. Molecule **177** is shown in ball-and-stick representation.

Molecule **177** contains an additional sulfone moiety that is placed in the hydrophobic cavity lined by Pro350 and Val352. Consequently, the ammonium group of molecule **177** gets displaced by 2.5 Å with respect to substrate (**176**) (Figure 6.14). The displaced ammonium group in the **177**–iNOS complex therefore interacts directly with Asp382 and liberates the otherwise trapped water (82H) seen in the **176**–iNOS complex. Such an interaction is favourable in terms of entropy [6.28]. Also, the inclusion of sulfone shifts the contacts of the ammonium group from the side chain carboxylate to the backbone carbonyl of Glu377. Although the ligand does not interact with the heme carboxylate directly, the σ -bond cooperative chain mediates its interaction with heme and H₄B. Water 197H seems to be crucial in the **177**–iNOS complex as three σ -bond cooperative chains are mediated through it, namely i) (**177**)HN \cdots HO(196H) \cdots HO(197H) \cdots O=C(Heme carboxylate); ii) (**177**)HN \cdots HO(196H) \cdots HO(197H) \cdots O12(H₄B)]; and iii) (**177**) S=O \cdots HOH(50H) \cdots OH(197H) \cdots O=C(H₄B). These interactions impart additional stability to the **177**–iNOS complex suggesting that because molecule **177** has a higher binding affinity to iNOS than the substrate (**176**) it acts as an inhibitor.

On the other hand, the carboxylate group is displaced by 2.2 Å anterior to the pocket thereby interacting with the residue Arg266. When compared with the substrate, molecule **177** forms an additional interaction with Gln263 but loses contacts with Tyr373. In the **177**-iNOS complex, the hydroxyl of Tyr373 rotates about 180° (with respect to that in the **176**-iNOS complex) and interacts with Tyr347. Therefore it does not interact with molecule **177**. Figure 6.9 shows the corresponding changes in the RMSD of the residues. Although Tyr373 loses its interaction with molecule **177**, its RMSD remains low due to its interaction with Tyr347. Apart from these interactions, water 84H mediates interactions between carboxylate of **177** and the active site residues Tyr347, Asp382, Arg266 and Arg388.

In the **177**-eNOS complex, the ligand adopts an S-like conformation such that both the sulfone and carboxylate groups interact with Gln247 (Figure 6.13). Due to this conformational change, the ammonium group of the ligand gets displaced by about 2.3 Å, while the carboxylate group retains its position compared to the substrate in the **176**-eNOS complex (Figure 6.14). The ammonium group therefore loses its interaction with Glu361 but interacts with heme and two water molecules, 1072w and 743w. Although the ammonium group interacts with water 743w throughout the simulation, no stable interactions are mediated to the protein residues. The carboxylate group interacts with Tyr357, Gln247, Arg250, Arg372 and Asn366. The **177**-eNOS complex is further stabilized by a σ -bond cooperative chain (177)NH \cdots OH(1072w) \cdots HO(168H) \cdots HN(Arg365).

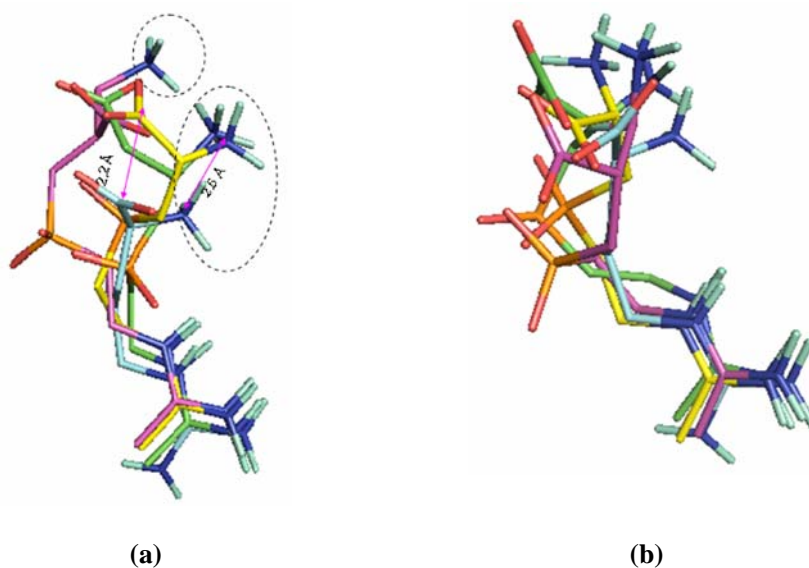
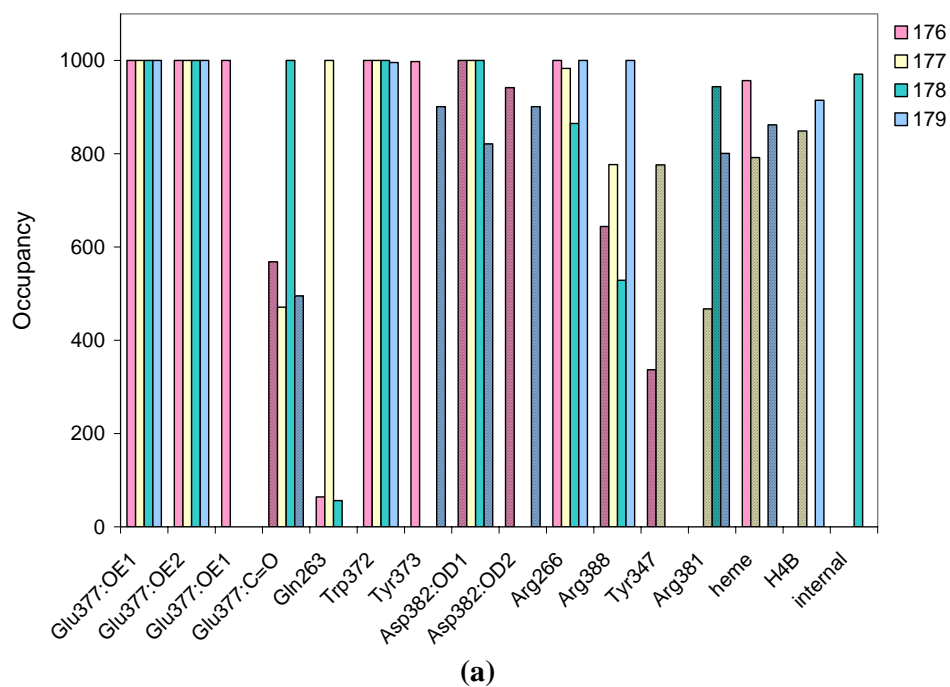


Figure 6.14: Overlay of ligand conformations in (a) iNOS (b) eNOS complexes. The C-atoms of the ligands are coloured as cyan (**176**); yellow (**177**); green (**178**); and pink (**179**). Distances between the ammonium and the carboxylic groups of substrate and molecule **177** are shown. Deviation of ammonium group of molecule **178** from that of molecules **176–178** can be seen (encircled).

In comparison to iNOS, molecule **177** lacks one of the crucial interactions with the carbonyl O-atom of Glu361, a number of water mediated interactions (see occupancy Figure 6.15) as well as the extra stabilization of the two σ -bond cooperative chains in eNOS. It has been suggested that σ -bond cooperative chains can contribute to the energy gain of around 20% relative to the isolated interactions [6.29]. Also, in the **177–iNOS** complex the methyl group of **177**, substituting for the guanidinium $N\eta_2$ of the substrate, is well placed in a hydrophobic cavity as mentioned earlier. In general, hydrophobic interactions have been postulated to contribute about $0.68 \text{ kcal mol}^{-1}$ (or a 3.2 fold increase in binding constant) per methyl group to the binding energy [6.30]. However, such contributions have only been seen in the **177–iNOS** complex. In the **177–eNOS** complex, a water molecule, 556w is trapped in this hydrophobic cavity by interacting with the backbone carbonyl of Pro334, and backbone NH atoms of Phe353 and Gly355. The entropic penalty of trapping water 556w in the hydrophobic cavity could affect the

stabilization energy of the **177**–eNOS complex. All the above interactions and differences contribute to higher selectivity of **177** towards iNOS when compared to eNOS. It is interesting to note that a water molecule bridging the carbonyl O-atom of Pro334 and the guanidine nitrogen of the substrate has been observed in the crystal structure of the eNOS–substrate complex [6.21]. Therefore, inclusion of 556w at this position is not either by chance or due to an artifact of the simulation, but that it has relevance to the structure of eNOS. However, the absence of such water molecules in other isoforms complexes cannot be explained.



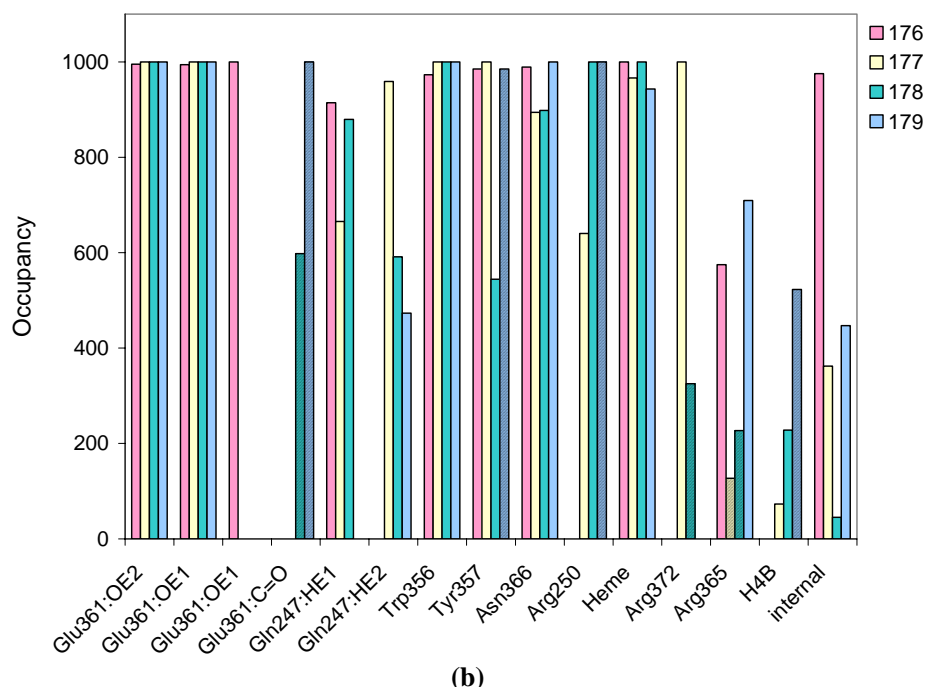


Figure 6.15: Occupancy of the hydrogen bond interactions of ligands **176–179** in (a) iNOS (b) eNOS complexes. Water mediated interactions are shaded.

6.4.8 Rational Design of More Selective Ligands

As observed with the interactions of ligands **176** and **177** in both the NOS isoforms, charge-charge interactions play a prominent role in the molecular recognition event. The juxtaposition of the positively and negatively charged groups in the ligand (α -amino acid head group) facilitates this molecular recognition, and (probably) selectivity among the isoforms. GRID analysis of the active sites of NOS isoforms also concurs that the hydrophobic and charge-charge interactions are the most important determinants of ligand selectivity [6.31]. I presume that selectivity can be improved by increasing the distance between the ammonium and the carboxylate groups. Accordingly, molecule **177** was modified to include a methylene spacer between the oppositely charged groups. This resulted in two hypothetical ligands: **178**, wherein the methylene group was introduced between C_α atom and the carboxylate group and **179**, wherein the methylene was placed

between the C_α atom and the ammonium group. Variations in the interactions of these modified ligands with both the isoforms are studied in detail and described in the following sections.

6.4.9 Mode of Binding of Molecule **178** in iNOS and eNOS

The interactions that anchor the ligand in the active site, namely the bidentate hydrogen bond between the amidino group of ligand and the carboxylate of Glu377 (Glu361) were retained in the complexes of **178** with both the isoforms. The RMSD between the bound conformations of molecule **3** in the two isoforms is 3.32 Å. In the **178-iNOS** complex, the spacer between the charged groups placed the functional groups such that they form an intramolecular hydrogen bond. The ammonium group of **178** interacts with the backbone carbonyl oxygen of Glu377 and the side chain carboxylate of Asp382 as shown in Figure 6.16.

On the other hand introduction of an additional C-atom displaces the carboxylate group about 1.4 Å away from its initial position in molecule **177**. Consequently, the interaction with Gln263 is lost. A corresponding change in the RMSD of the residue Gln263 can be observed in Figure 6.9. The carboxylate of the ligand interacts with Arg266. The occupancy plots of hydrogen bond interactions throughout the simulation shows that the interaction with Arg388 is retained for only ~500ps (Figure 6.15). Apart from these, water (82H) mediated interactions were also observed between the ligand carboxylate and Arg381.

The interactions of the **178-eNOS** complex are shown in Figure 6.17. The ammonium group of **178** interacts with the heme carboxylate and two water molecules 1070w and 1052w. Water mediated (1052w) interactions are formed between the ammonium group and the backbone carbonyl of Glu361 (~500ps). However, no interactions are mediated through 1070w. Molecule **178** interacts with Gln247 through both its sulfone and the carboxylate groups, with a similar conformation as in the **177-eNOS** complex. Hydrogen bond interactions were also observed between the carboxylate group of **178** and the residues Asn366, Arg250 and partly with Tyr357

(~500ps). Although a few water molecules 107H, 757w and 1022w cluster around the sulfone moiety transiently, they do not form any interactions with the active site residues.

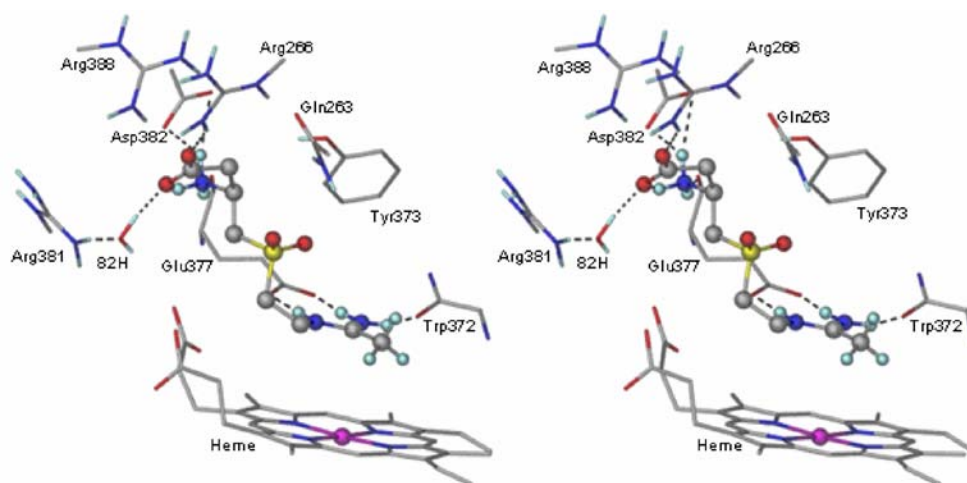


Figure 6.16: Stereoview of the active site of the **178**-iNOS complex. Molecule **178** is shown in ball-and-stick.

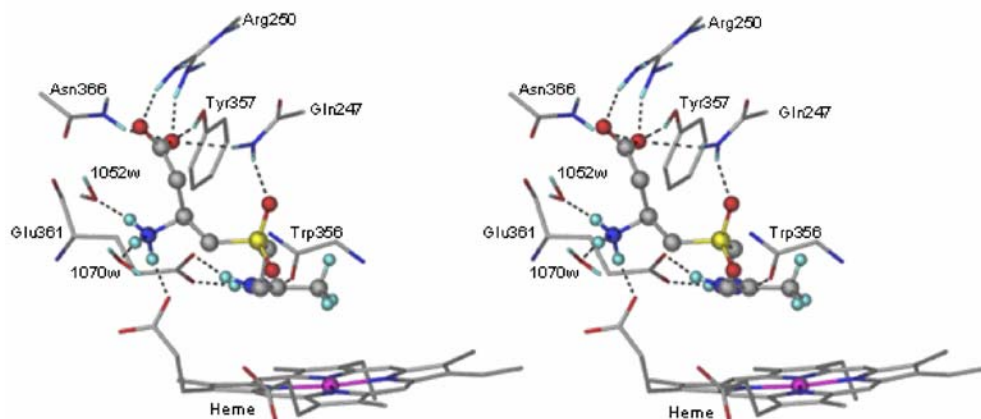


Figure 6.17: Stereoview of the active site of the **178**-eNOS complex.

6.4.10 Mode of Binding of Molecule **179** in Isoforms

The RMSD between the bound conformations of **179** in both the isoforms is 2.79 Å. Molecule **179** adopts a totally different conformation in iNOS when compared with other ligands in iNOS-complexes (Figure 6.14a). The anchoring interactions in the active site remained unaltered in both the isoforms. In the **179**-iNOS complex, the ammonium group of **179** makes a direct interaction only with H₄B, while all the other interactions are mediated through water viz., 82H and 50H as shown in Figure 6.18. Hydrogen bond interactions with H₄B and heme were mediated through water 82H, while interactions with residues Asp382, Arg388 and Arg381 are through water 196H. Electrostatic interactions are seen between the carboxylate group of the ligand and the residues Arg388 and Arg266. The carboxylate group also forms water mediated interactions with Asp382 and Tyr373 residues through 84H and with Glu377 through 196H. Additionally, the **179**-iNOS complex was stabilized by the formation of σ -bond cooperative chain viz., (**179**) NH \cdots OH(82H) \cdots OH (196H) \cdots O=C (Glu377).

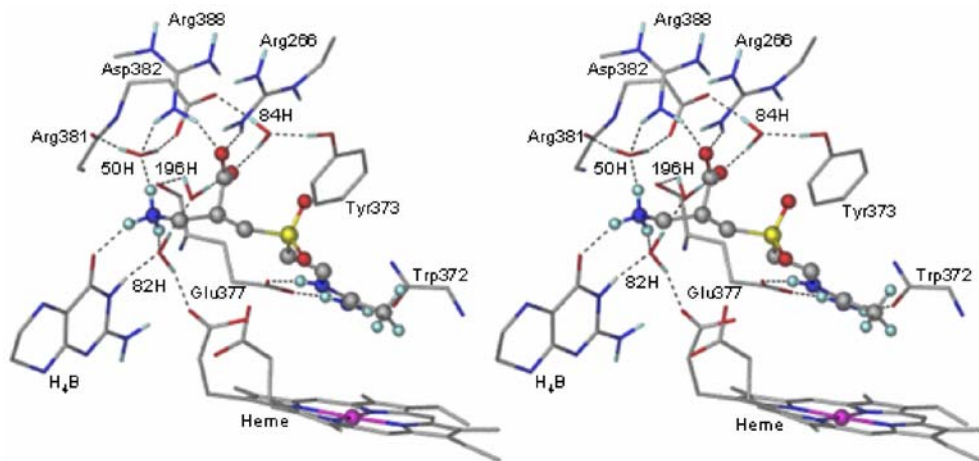


Figure 6.18: Stereoview of the active site of the **179**-iNOS complex.

In the **179-eNOS** complex, no direct interactions are formed with the ammonium group. Water mediated interactions with Asn366 and Glu361 through water 201H and with heme through water 402H are observed as shown in Figure 6.19. The ligand carboxylate interacts with Arg250 and Tyr357. The additional stability due to the σ -bond cooperative chain found in the **179-iNOS** complex was not found in the **179-eNOS** complex. Thus, a thorough inspection of these interactions predict molecule **179** to be a tighter binder to iNOS than to eNOS.

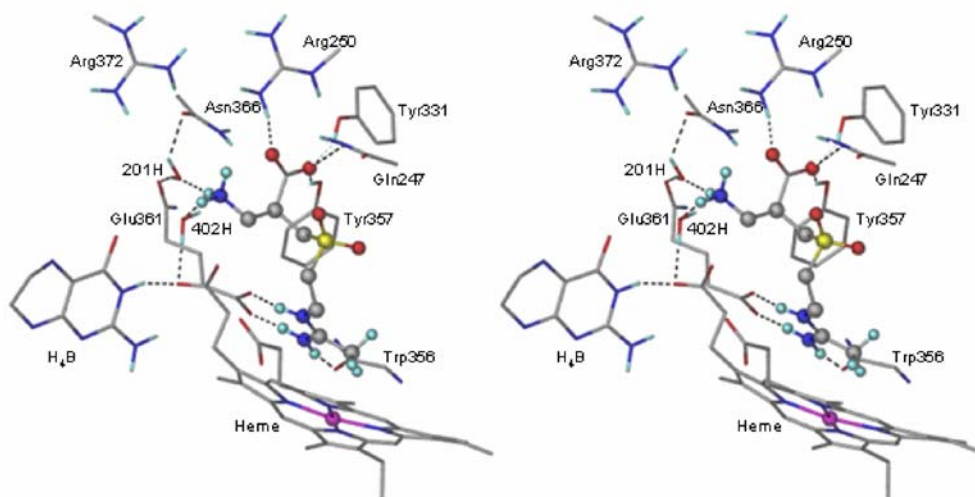


Figure 6.19: Stereoview of the active site of the **179-eNOS** complex.

6.4.11 Comparison of Binding of Ligands in iNOS and eNOS

The occupancy plots of all the hydrogen bond interactions formed by the ligands **176–179** throughout the simulation in both iNOS and eNOS isoforms are shown in Figure 6.15. All the interactions that persisted for a minimum of 500ps duration were considered as being stable. Molecule **177** forms more number of interactions than molecule **176** with both isoforms. Also, molecule **177** forms more water mediated interactions in iNOS than in eNOS. This concurs with the experimental data that molecule **177** is more selective to iNOS than to eNOS [6.13c]. As described above, such

selectivity may be attributed to the presence of water mediated interactions and the extra stabilization of the σ -bond cooperative chains in the **177**-iNOS complex.

Molecule **178**, on the other hand, forms a smaller number of direct as well as water mediated interactions with both the isoforms. Therefore, incorporation of the methylene spacer between the C_α and the carboxylate as in molecule **178** does not improve the ligand binding to both isoforms. Upon a qualitative comparison of all the ligands binding to iNOS, molecule **179** is envisaged to have higher binding affinity with a maximum number of hydrogen bonds (6 direct interactions and 6 water mediated interactions), as each hydrogen bond contributes roughly to 3.5–5 kcal mol⁻¹. Also, when compared across the isoforms, molecule **179** forms a maximum number of stable interactions with iNOS than with eNOS indicating that it is deemed to bind more selectively to iNOS.

6.4.12 Energetics of Ligand Binding in iNOS and eNOS

The electrostatic and van der Waals contributions to the nonbonded interaction energies of ligands **176**-**179** from direct interactions with active site residues in both iNOS and eNOS are summarized in Tables 6.3 and 6.4. From the interaction energies, it is evident that the active site residues Glu377, Trp372, Asp382 and heme have major contributions to the electrostatic energy in the ligand binding to iNOS. Molecule **177** which makes a direct interaction with Asp382 has higher electrostatic interaction energy when compared to molecules **176** and **179**. Although molecule **178** showed a higher electrostatic interaction energy with Asp382 (> 6 kcal mol⁻¹), its interaction energy with heme was approx. 20 kcal mol⁻¹ less than the other molecules. Among all the ligand complexes, the energy contribution due to Asn366 (1–4 kcal mol⁻¹) in eNOS was far less than that due to Asp382 (10–30 kcal mol⁻¹) in iNOS.

Table 6.3: Nonbonded interaction energies of ligands **176-179** with iNOS (direct contacts)

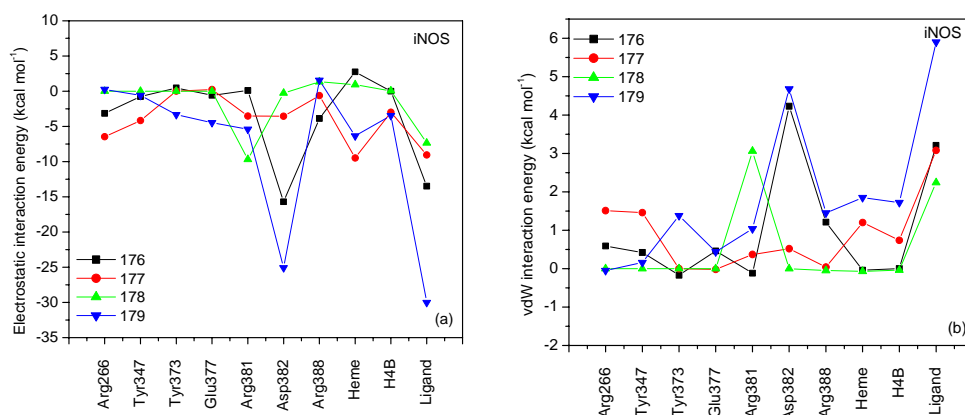
Monomer	176		177		178		179	
	coulomb	vdw	coulomb	vdw	coulomb	vdw	coulomb	vdw
Gln263	0.36	-0.33	-7.28	0.23	2.28	-0.77	3.56	-0.75
Arg266	-1.25	1.85	-4.26	1.37	-3.27	1.42	-3.28	1.36
Pro350	-1.35	-0.96	-0.54	-1.00	-0.78	-1.02	-1.32	-1.02
Gly371	2.51	-0.17	3.57	-0.25	3.56	-0.31	3.76	-0.30
Trp372	-8.64	1.31	-11.61	1.39	-11.11	1.14	-11.43	0.77
Tyr373	0.54	-0.03	2.01	-0.82	5.04	-1.30	2.68	-1.02
Met374	0.86	-0.29	1.03	-0.18	2.54	-0.26	2.81	-0.13
Glu377	-69.17	4.77	-71.57	5.48	-78.69	6.87	-58.30	3.43
Arg381	1.87	0.76	17.35	-0.47	17.91	-0.49	4.95	-0.17
Asp382	-11.25	-0.26	-25.02	2.04	-31.35	1.52	-11.11	-0.38
Arg388	4.69	-0.07	9.30	0.24	11.76	0.16	6.30	0.86
Heme	-35.96	-2.82	-46.47	-4.24	-24.97	-4.37	-41.36	-3.61
H ₄ B	-	-	-	-	-	-	-4.27	0.50

The electrostatic and van der Waals interaction energies of water mediated interactions with various amino acid residues in the active sites of iNOS and eNOS complexes are plotted in Figures 6.20 and 6.21. The plots show that molecule **179** utilizes the active site variation of Asp382 in iNOS to a maximum extent, contributing about -30kcal mol^{-1} to the electrostatic interaction energy. Although the van der Waals interactions have a destabilizing effect, they are compensated by the higher electrostatic contributions. However, no such major contributions were observed with the eNOS complexes indicating a minor role for water with eNOS compared to iNOS.

Also, in iNOS the electrostatic energy contribution due to interactions with water was about -30 kcal mol^{-1} with molecule **179** which concurs with the observation that molecule **179** forms a maximum number of water mediated interactions with iNOS (Figure 6.15). Therefore, molecule **179** may show higher selectivity to iNOS when compared to eNOS.

Table 6.4: Nonbonded interaction energies of ligands **176-179** with eNOS (direct contacts)

Monomer	176		177		178		179	
	coulomb	vdw	coulomb	vdw	coulomb	vdw	coulomb	vdw
Gln247	-0.25	0.02	-0.07	0.16	-1.49	0.35	-1.48	-0.52
Arg250	1.34	1.02	1.97	1.42	-4.61	2.65	-2.08	1.87
Pro334	0.09	-0.92	-1.46	-1.30	-0.50	-1.14	-1.17	-1.38
Gly355	4.21	-0.19	4.49	-0.30	-0.58	-0.24	2.67	-0.32
Trp356	-7.08	0.14	-8.83	1.42	-7.30	1.73	-7.28	1.28
Tyr357	0.76	-0.66	-1.22	1.08	2.46	-0.04	0.99	0.08
Met358	3.26	-0.31	1.49	-0.28	-20.85	-0.25	-5.65	-0.26
Glu361	-61.00	2.64	-64.41	4.79	-45.04	5.21	-57.87	3.45
Arg365	1.57	-0.02	2.59	-0.01	—	—	-0.35	-0.01
Asn366	-1.82	0.79	-3.84	-0.17	-0.87	0.76	-3.39	-0.24
Arg372	—	—	7.45	1.10	-0.16	0.59	0.73	—
Heme	-59.54	-1.21	-61.11	-2.52	-71.53	-1.56	-54.75	-4.17
H ₄ B	—	—	—	—	0.40	-0.06	0.21	-0.02

**Figure 6.20:** (a) Electrostatic and (b) steric contributions to the nonbonded interaction energies of water mediated interactions in **176-179** iNOS complexes.

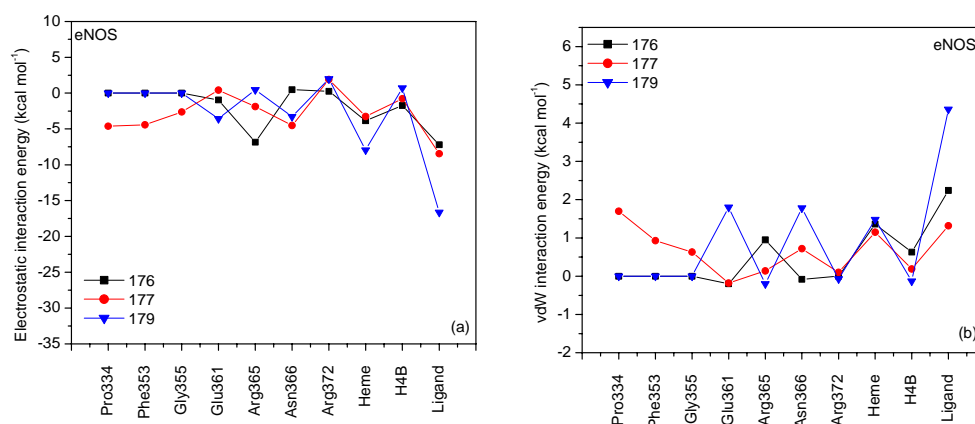


Figure 6.21: (a) Electrostatic and (b) steric contributions to the nonbonded interaction energies of water mediated interactions in 176-179 eNOS complexes. Molecule 178 is not shown in figure as it has no water mediated interactions.

6.5 Conclusions

Molecular dynamics simulations were performed for four ligand complexes with both iNOS and eNOS isoforms. MD simulations made it possible to identify and understand stable interactions in the enzyme–substrate/inhibitor complexes. The major difference in the substrate recognition among the isoforms is the interaction of the ammonium group of the substrate with Asp382 in iNOS that corresponds to Asn366 in eNOS. The substrate in the iNOS seems to have an entropic advantage over eNOS which could be speculated as a plausible reason for comparatively high turnover of the former. While the ammonium group forms a water mediated hydrogen bond in iNOS, the carboxylate group (of the substrate) interacts directly with the corresponding residue, Asn366 in eNOS (Figure 6.7). Analysis of the molecular recognition of the substrate in both the isoforms showed that water molecules seem to be critical in masking these active site differences, thereby, facilitating substrate binding. Especially, in iNOS, water molecules attenuate the electrostatic repulsion between the carboxylate groups of the substrate and Asp382. It is important to note that such repulsion is absent in the case of eNOS where the corresponding residue is Asn366. In eNOS, however, water has a different role to play of just providing hydrogen bonding groups in the active site.

MD simulations of an iNOS specific inhibitor **177** bound to both the isoforms indicates that in the **177**-iNOS complex the ammonium group interacts directly with Asp382 and liberates the otherwise trapped water seen in the **176**-iNOS complex. The basis of selectivity of molecule **177** towards iNOS was identified to be due to water mediated interactions and σ -bond cooperative networks. The juxtaposition of the positively and negatively charged groups in the ligand (α -amino acid head group) facilitates molecular recognition, and contributes to the selectivity differences in the various isoforms. The detailed analysis of both the substrate and the inhibitor recognition in the two isoforms provides useful guidelines for the design of potent isoform specific inhibitors. With a motto to explore the applications of MD to make structural variations and survey all possibilities, two hypothetical molecules **178** and **179** were designed by increasing the distance between the ammonium and the carboxylate groups of molecule **177**. The molecular recognition pattern of these ligands within the two isoforms was analyzed. Of the two, molecule **179** was envisaged to be more selective towards iNOS, based both on the number of interactions and the energetics involved therein. Water mediated interactions that remain stable throughout the simulations make a major contribution to the selectivity of molecule **179** to iNOS.

CHAPTER 7

CONCLUSIONS AND FUTURE PROSPECTS

Rational drug design approaches are employed with a hope that the odds of finding a useful drug would be increased in the drug discovery process. The basic idea underlying these methods is to improve the binding affinity between a drug and its target. The improvement is expected to be achieved by first examining the binding forces between the drug and the target and then suggesting ways of modifying the strength of the interaction energy—the binding affinity, between the two binding partners. Different computational approaches were used throughout this thesis to understand the molecular recognition patterns of ligands binding to three different macromolecular systems (EGFR, TMPK_{mt}, NOS).

The 3D-QSAR models developed for EGFR kinase differentiates the 6- and 7-positions and indicate that substituents at the 6-position should have higher electronegativity, hydrophilicity, hydrogen bonding propensity and bulkiness for enhanced biological activity. Additionally, Br with appropriate steric and electronic properties makes it one of the essential requirements at 3'-position of the 4-anilinoquinazoline moiety.

Docking studies with EGFR kinase reveal that N-H \cdots N, O_w-H \cdots N and above all the C-H \cdots O type hydrogen bonds are very important for obtaining accurate poses and binding affinities. The role of weak hydrogen bonds (C-H \cdots O) in predicting the correct poses in the binding site was also highlighted. Although weak, occurrence of such interactions is not uncommon in the molecular recognition events [3.24a]. Therefore, the efficiency of the docking programs in predicting good poses can be enhanced with force fields that can adequately model C-H \cdots O interactions. Hydrophobic interactions are also important for a good fit of the 4-anilino fragment into a hydrophobic pocket of the EGFR binding site. Therefore, adequate modeling of hydrophobic interactions and both weak and strong hydrogen bonds is necessary. Although there are many scoring functions till date none of them can completely score all the interactions. While some are accurate in

scoring the hydrogen bond interactions, others lay emphasis on hydrophobic interactions. Therefore, a consensus VS model was developed that accounts for all the interactions with a combination of scoring functions. During the course of generating a VS model stress was laid upon the need of assessing the strength of weak interactions contributing to binding affinity and their incorporation into the force fields to make VS more accurate and reliable.

All the widely used analog based (pharmacophore and 3D-QSAR) and structure based methods (docking) were compared. The pharmacophore model was identified as an appropriate method for the development of a robust VS tool against TMPK_{mt}. Docking analysis revealed that in addition to the strong hydrogen bonds, non-conventional interactions like halogen bond, water mediated cooperative networks of hydrogen bonds along with weak hydrogen bonds play critical role in the molecular recognition. Halogen bond formed by Br, due to its polarizability, and its subsequent stabilization with concomitant water mediated cooperative hydrogen bond networks makes it worthwhile to increase the inhibitory activity prominently as observed with molecule **166** (chapter 4). Exploiting halogen bonding proved beneficial in several other target proteins [4.21, 4.23].

Lead hopping is the most important advantage of using pharmacophore as a VS tool. However, the use of pharmacophore as the sole tool for VS may miss some potential candidates at times. Therefore, a composite model that includes all the essential features from different methods would be advantageous over a single tool. In chapter 5 a method was described for the development of such a composite model that includes all the functionalities involved in molecular recognition acquired from various ligand and structure based methods. The composite model was also used to screen databases to arrive at few lead molecules. A similar kind of pharmacophore based VS tool has been used to screen ligands against cytochrome P-450 19 (Aromatase) [5.5b]. However the VS tool described in this chapter is distinct with a hydrogen bond donor feature added from the docking analysis.

In contrast to docking, MD simulations give a dynamic picture of the molecular recognition pattern and help to identify stable interactions in the

enzyme–substrate/inhibitor complexes. Therefore, in chapter 6, MD simulations were used to analyze recognition of both the substrate and an inhibitor in two NOS isoforms–iNOS and eNOS. It was noticed that water molecules, owing to their dipole moment, seem to be critical in masking the active site differences among NOS isoforms thus facilitating the substrate binding. The basis of selectivity of molecule **177** towards iNOS was identified to be due to water mediated interactions and σ -bond cooperative networks. The juxtaposition of the positively and negatively charged groups in the ligand (α -amino acid head group) facilitates molecular recognition, and contributes to the selectivity differences in the various isoforms. Based on these observations two hypothetical molecules (**178** and **179**) were designed. Of these, molecule **179** was predicted to be more selective to iNOS based on both the number of interactions and the energetics involved therein.

Another important aspect of the study is the role of water. Water is observed to have major impact in the binding of ligands to their receptor proteins as seen in chapters 3, 4 and 6. There are a number of examples where tightly bound water molecules, seen in the binding sites of proteins, have been targeted or displaced. The bound water in EGFR is targeted in the design approaches, while it is displaced in the case of NOS studies. The targeting and displacement of water molecules in the active site is probably determined by the free energy changes arising from the gain in entropy and enthalpy. The favourable entropy gain results from displacement of the tightly bound water and subsequent transfer to the bulk. And, the enthalpy contributions are from the hydrogen bonding interactions with the protein and/or ligand and other water molecules that the water molecule had. Therefore, tightly bound water molecules in the binding site can be successfully targeted or displaced by appropriately designed ligands.

In summary, various rational drug design (RDD) approaches have been used to explore the role of strong and weak interactions in the molecular recognition in three macromolecular systems. These approaches helped to understand the principles of interactions between receptors, ligands, and conserved water molecules. Basically, all these approaches depend on the assessment of the molecular recognition events, i.e., assessing the strengths of van der Waals and electrostatic interactions. To a little extent

consensus scoring performs better in relating the scores to the binding affinity of the ligands. However, this becomes more complex when interactions like C–H \cdots π , and cooperative networks of hydrogen bonds manifest in the molecular recognition as seen with TMPK_{mt} and NOS isoforms in chapters 4 and 6 respectively. This thesis provides a qualitative analysis of the interactions involved in the ligand binding and highlights their role in the molecular recognition. Research in this area could be further extended to quantify both these strong and weak interactions that would give a direct measure of the binding affinity more precisely with moderate computational costs.

REFERENCES AND NOTES

CHAPTER 1

- [1.1] Hileman, B. C & EN **2006**, 84, 50.
- [1.2] Hodgson, J. *Nat. Biotech.* **2001**, 19,722-726.
- [1.3] (a) Maehle, A. H.; Prüll, C. R.; Halliwell, R. F. *Nat. Rev. Drug Discov.* **2002**, 1, 637-641. (b) Berg, J. M.; Tymoczko, J. L.; Stryer, L., *Biochemistry*, 5th ed.; W. H. Freeman and company: New York, 2000, p 200.
- [1.4] Ganesan, A. *Drug Discov. Today* **2002**, 7, 47-55.
- [1.5] Hertzberg, R. P.; Pope, A. J. *Curr. Opin. Chem. Biol.* **2000**, 4, 445-451.
- [1.6] Klebe, G. *Drug Discov. Today* **2006**, 11, 580-594.
- [1.7] Parrill, A. L.; Reddy, M. R., Overview of rational drug design. In *Rational Drug Design: novel methodology and practical applications*; Rami Reddy, M., Parrill, A. L., Ed.; American Chemical Society: Washington, DC, 1999; Vol. 719, pp 1-11.
- [1.8] Davis, A. M.; Teague, S. J.; Kleywegt, G. J. *Angew. Chem. Int. Ed.* **2003**, 42, 2718-2736.
- [1.9] Murray, C. W.; Verdonk, M. L. *J. Comput. -Aided. Mol. Des.* **2002**, 16, 741-753.
- [1.10] Abraham, D. J., Ed. *Burger's medicinal chemistry and drug discovery*, 6th ed.; vol. 1, John Wiley & Sons Inc.: New Jercey, 2003.
- [1.11] (a) Martin, Y. C., 3D QSAR: current state, scope and limitations. In *Perspectives in drug discovery and design-3D QSAR in drug design: recent advances*, Kubinyi, H.; Folkers, G.; Martin, Y. C., Eds.; Kluwer Academic Publishers: London 1998; Vol. 12/13/14, pp 3-23. (b) Cramer III, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.*, **1988**, 110, 5959-5967.
- [1.12] Grzybowski, B. A.; Ishchenko, A. V.; Shimada, J.; Shakhnovich, E. I. *Acc. Chem. Res.* **2002**, 35, 261-269.
- [1.13] Gund, P., Evolution of pharmacophore concept in pharmaceutical research. In *Pharmacophore perception, development, and use in drug design*, Guner, O. F. Ed.; International University Line: CA, 2000, pp 3-11.

- [1.14] Carlson, H. A.; McCammon, J. A. *Mol. Pharmacol.* **2000**, *57*, 213-218.
- [1.15] (a) Renner, S.; Schneider, G. *J. Med. Chem.* **2004**, *47*, 4653-4664. (b) Singh, J.; van Vlijmen, H.; Liao, Y.; Lee, W.; Cornebise, M.; Harris, M.; Shu, I.; Gill, A.; Cuervo, J. H.; Abraham, W. M.; Adams S. P. *J. Med. Chem.* **2002**, *45*, 2988-2993. (c) Steindl, T.; Langer, T. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1849-1856.
- [1.16] Marrone, T. J.; Briggs, J. M.; McCammon, J. A. *Annu. Rev. Pharmacol. Toxicol.* **1997**, *37*, 71-90.
- [1.17] Cushman, D. W.; Cheung, H. S.; Sabo, E. F.; Ondetti, M. A. *Biochemistry* **1977**, *16*, 5484-5491.
- [1.18] Weber, I. T.; Harrison, R. W., Molecular mechanics calculations on protein-ligand complexes. In *Perspectives in drug discovery and design-3D QSAR in drug design: ligand-protein interactions and molecular similarity*, Kubinyi, H.; Folkers, G.; Martin, Y. C., Eds.; Kluwer Academic Publishers: London 1998; Vol. 2, pp 115-127.
- [1.19] Ajay, Murko, M. A. *J. Med. Chem.* **1995**, *38*, 4953-4967.
- [1.20] Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discov.* **2004**, *3*, 935-949.
- [1.21] Honma, T. *Med. Res. Rev.*, **2003**, *23*, 606-632.
- [1.22] Schneider, G.; Fechner, U. *Nat. Rev. Drug Discov.* **2005**, *4*, 649-663.
- [1.23] Lyne, P.D. *Drug Discov. Today* **2002**, *7*, 1047-1055.
- [1.24] Sirois, S.; Wei, D. Q.; Du, Q.; Chou, K. C. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1111-1122.
- [1.25] Lyne, P. D.; Kenny, P. W.; Cosgrove, D. A.; Deng, C.; Zabudoff, S.; Wendoloski, J. J.; Ashwell, S. *J. Med. Chem.* **2004**, *47*, 1962-1968.
- [1.26] (a) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. *J. Med. Chem.* **1999**, *42*, 5100-5109. (b) Bissantz, C.; Folkers, G.; Rognan, D. *J. Med. Chem.* **2000**, *43*, 4759-4767.
- [1.27] Paul, N.; Rognan, D. *Proteins* **2002**, *47*, 521-533.

- [1.28] (a) Bauera, A.; Kovária, Z.; Keseru, G. M. *J. Mol. Struct. (Theochem)* **2004**, 676, 1-5. (b) Jenkins, J. L.; Kao, R. Y. T.; Shapiro, R. *Proteins* **2003**, 50, 81-93. (c) Flohr, S.; Kurz, M.; Kostenis, E.; Brkovich, A.; Fournier, A.; Klabunde, T. *J. Med. Chem.* **2002**, 45, 1799-1805. (d) Vangrevelinghe, E.; Zimmermann, K.; Schoepfer, J.; Portmann, R.; Fabbro, D.; Furet, P. *J. Med. Chem.* **2003**, 46, 2656-2662.
- [1.29] McCammon, J. A., Harvey, S. C., *Dynamics of proteins and nucleic acids*. Cambridge University Press: Cambridge 1987.
- [1.30] Adcock, S. A.; McCammon, J. A. *Chem. Rev.* **2006**, 106, 1589-1615.
- [1.31] Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957**, 27, 1208-1209.
- [1.32] Rahman, A. *Phys. Rev. A* **1964**, 136, 405-411.
- [1.33] McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, 267, 585-590.
- [1.34] Duan, Y.; Kollman, P. A. *Science* **1998**, 282, 740-744.
- [1.35] Song, M.; Clark, M. *J. Chem. Inf. Model.* **2006**, 46, 392-400.
- [1.36] O'Brien, S. E.; De Groot, M. J. *J. Med. Chem.* **2005**, 48, 1287-1291.
- [1.37] Leach, A. R., *Molecular Modeling: principles and applications*. Addison Wesley Longman Limited: Essex, England, 1996.
- [1.38] Jensen, F., *Introduction to Computational Chemistry*, John Wiley & Sons, New York, 1999.
- [1.39] Richards, W. G. *Nat. Rev. Drug Discov.* **2002**, 1, 551-555.
- [1.40] Shirts, M.; Pande, V. S. *Science* **2000**, 290, 1903-1904.
- [1.41] Slepchenko, B. M.; Schaff, J. C.; Carson, J. H.; Loew, L. M. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, 31, 423-441.
- [1.42] Tomita, M.; Hashimoto, K.; Takahashi, K.; Shimizu, T. S.; Matsuzaki, Y.; Miyoshi, F.; Saito, K.; Tanida, S.; Yugi, K.; Venter, J. C.; Hutchison III, C. A. *Bioinformatics* **1999**, 15, 72-84.
- [1.43] Coggan, J. S.; Bartol, T. M.; Esquenazi, E.; Stiles, J. R.; Lamont, S.; Martone, M. E.; Berg, D. K.; Ellisman, M. H.; Sejnowski, T. J. *Science* **2005**, 295, 446-451.

CHAPTER 2

- [2.1] Yarden, Y. *Ann. Rev. Biochem.* **1998**, *57*, 443-478.
- [2.2] Aaronson, S. A. *Science* **1991**, *254*, 1146-1152.
- [2.3] Barnes, C. J.; Kumar, R. *Cancer and Metastasis Rev.* **2003**, *22*, 301-307.
- [2.4] Cohen, P. *Nature Rev. Drug Discov.* **2002**, *1*, 309-315.
- [2.5] Pedersen, M. W.; Poulsen, H. S. *Science & Medicine* **2002**, *8*, 206-217.
- [2.6] Bridges, A. J. *Chem. Rev.* **2001**, *101*, 2541-2571.
- [2.7] Traxler, P.; Bold, G.; Buchdunger, E.; Caravatti, G.; Furet, P.; Manley, P.; Reilly, T. O.; Wood, J.; Zimmermann, J. *Med. Res. Rev.* **2001**, *21*, 499-512.
- [2.8] Smaill, J. B.; Showalter, H. D. H.; Zhou, H.; Bridges, A. J.; McNamara, D.; Fry, D. W.; Nelson, J. M.; Sherwood, V.; Vincent, P. W.; Roberts, B. J.; Elliott, W. L.; Denny, W. A. *J. Med. Chem.* **2001**, *44*, 429-440.
- [2.9] Smaill, J. B.; Rewcastle, G. W.; Loo, J. A.; Greis, K. D.; Chan, O. H.; Reyner, E. L.; Lipka, E.; Showalter, H. D. H.; Vincent, P. W.; Elliott, W. L.; Denny, W. A. *J. Med. Chem.* **2000**, *43*, 1380-1397.
- [2.10] Smaill, J. B.; Palmer, B. D.; Rewcastle, G. W.; Denny, W. A.; McNamara, D. J.; Dobrusin, E. M.; Bridges, A. J.; Zhou, H.; Showalter, H. D. H.; Winters, R. T.; Leopold, W. R.; Fry, D. W.; Nelson, J. M.; Slintak, V.; Elliot, W.; Roberts, B. J.; Vincent, P. W.; Patmore, S. J. *J. Med. Chem.* **1999**, *42*, 1803-1815.
- [2.11] Kurup, A.; Garg, R.; Hansch, C. *Chem. Rev.* **2001**, *101*, 2573-2600.
- [2.12] Gibson, K. H.; Grundy, W.; Godfrey, A. A.; Woodburn, J. R.; Ashton, S. E.; Curry, B. J.; Scarlett, L.; Barker, A. J.; Brown, D. S. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2723-2728.
- [2.13] Wissner, A.; Floyd, M. B.; Rabindran, S. K.; Nilakantan, R.; Greenberger, L. M.; Shen, R.; Wang, Y. F.; Tsou, H. R. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2893-2897.
- [2.14] Rewcastle, G. W.; Denny, W. A.; Bridges, A. J.; Zhou, H.; Cody, D. R. *J. Med. Chem.* **1995**, *38*, 3482-3487.

- [2.15] Bridges, A. J.; Zhou, H.; Cody, D. R.; Rewcastle, G. W.; McMichael, A.; Showalter, H. D. H.; Fry, D. W.; Kraker, A. J.; Denny, W. A. *J. Med. Chem.* **1996**, *39*, 267-276.
- [2.16] Hou, T.; Zhu, L.; Chen, L.; Xu, X. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 273-283.
- [2.17] Hahn, M. *J. Med. Chem.* **1995**, *38*, 2080-2090.
- [2.18] Hahn, M.; Rogers, D. *J. Med. Chem.* **1995**, *38*, 2091-2102.
- [2.19] Thompson, A. M.; Bridges, A. J.; Fry, D. W.; Kraker, A. J.; Denny, W. A. *J. Med. Chem.* **1995**, *38*, 3780-3788.
- [2.20] Rewcastle, G. W.; Palmer, B. D.; Thompson, A. M.; Bridges, A. J.; Cody, D. R.; Zhou, H.; Fry, D. W.; McMichael, A.; Denny, W. A. *J. Med. Chem.* **1996**, *39*, 1823-1835.
- [2.21] Rewcastle, G. W.; Bridges, A. J.; Fry, D. W.; Rubin, J. R.; Denny, W. A. *J. Med. Chem.* **1997**, *40*, 1820-1826.
- [2.22] Thompson, A. M.; Murray, D. K.; Elliott, W. L.; Fry, D. W.; Nelson, J. A.; Showalter, H. D. H.; Roberts, B. J.; Vincent, P. W.; Denny, W. A. *J. Med. Chem.* **1997**, *40*, 3915-3925.
- [2.23] Fry, D. W.; Kraker, A. J.; McMichael, A.; Ambroso, L. A.; Nelson, J. M.; Leopold, W. R.; Connors, R. W.; Bridges, A. J. *Science* **1994**, *265*, 1093-1095.
- [2.24] Shewchuk, L.; Hassell, A.; Wisley, B.; Rocque, W.; Holmes, W.; Veal, J.; Kuyper, L. F. *J. Med. Chem.* **2000**, *43*, 133-138.
- [2.25] Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219-3228.
- [2.26] Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
- [2.27] Cerius² version 4.8 Molecular Modeling Program, Accelrys: San Diego, CA 92121-3752, USA
- [2.28] Langer T., Molecular similarity characterization using COMFA. In *Perspectives in drug discovery and design-3D-QSAR in drug Design: Recent advances*, Kubinyi, H.; Folkers, G.; Martin Y. C. Eds., Kluwer Academic Publishers, London, **1998**, 215-231.

- [2.29] Stamos, J.; Sliwowski, M. X.; Eigenbrot, C. *J. Biol. Chem.* **2002**, 277, 46265-46272.
- [2.30] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, 4, 187-217.
- [2.31] *INSIGHT II 97* Molecular Modeling Program, Accelrys: San Diego, CA, 1997.
- [2.32] *GOLD* version 3.0, Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK.
- [2.33] Golbraikh, A.; Tropsha, A. *J. Mol. Graphs. Mod.* **2002**, 20, 269-276.
- [2.34] Lipinski, C. A.; Lombardo, F.; Dominy, B.W.; Feeney P. J. *Adv. Drug. Del. Rev.* **1997**, 23, 3-25.
- [2.35] *MOE 2002.03*, Chemical Computing Group Inc., Montreal, H3A 2R7 Canada.

CHAPTER 3

- [3.1] Gershell, L. J.; Atkins, J. H. *Nature Rev. Drug Discov.* **2003**, 2, 321-327.
- [3.2] Oprea, T. I. *Molecules* **2002**, 7, 51-62.
- [3.3] Schneider, G.; Böhm, H-J. *Drug Discov. Today* **2002**, 7, 64-70.
- [3.4] Langer, T.; Krovat, E. M. *Curr. Opin. Drug Disc. Develop.* **2003**, 6, 370-376.
- [3.5] (a) Grüneberg, S; Stubbs, M. T.; Klebe, G. *J. Med. Chem.* **2002**, 45, 3588-3602.
(b) Singh, J.; Vlijmen, H. V.; Liao, Y.; Lee, W.; Cornebise, M.; Harris, M.; Shu, I.; Gill, A.; Cuervo, J. H.; Abraham, W. M.; Adams S. P. *J. Med. Chem.* **2002**, 45, 2988-2993.
- [3.6] (a) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. *Proteins* **2004**, 57, 225-242. (b) Nissink, J.; Willem, M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. *Proteins* **2002**, 49, 457-471. (c) Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks III, C. L. *J. Comput.-Aided Mol. Des.* **2003**, 17, 755-763. (d) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. *J. Med. Chem.* **2004**, 47, 558-565.
- [3.7] (a) Gohlke, H.; Klebe, G. *Cur. Opin. Str. Biol.* **2001**, 11, 231-235. (b) Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. *J. Mol. Graph. Mod.* **2002**, 20, 281-295. (c) Wang, R.; Lu, Y.; Wang, S. *J. Med.Chem.* **2003**,

- 46, 2287-2303. (d) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks III, C. L. *J. Med. Chem.* **2004**, *47*, 3032-3047. (e) Stahl, M.; Rarey, M. *J. Med. Chem.* **2001**, *44*, 1035-1042. (f) Krovat, E. M.; Langer, T. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1123-1129. (g) Wang, R.; Lai, L.; Wang, S. *J. Comput.-Aided Mol. Des.*, **2002**, *16*, 11-26. (h) Terp, G. E.; Johansen, B. N.; Christensen, I. T.; Jørgensen, F. S. *J. Med. Chem.* **2001**, *44*, 2333-2343.
- [3.8] Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. *J. Mol. Graphs. Model.* **2003**, *21*, 289-307.
- [3.9] Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discov.* **2004**, *3*, 935-949.
- [3.10] (a) Jones, G.; Willet, P.; Glen, R. C. *J. Mol. Biol.* **1995**, *245*, 43-53. (b) Jones, G.; Willet, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727-748.
- [3.11] (a) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425-445. (b) Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. *Proteins* **1998**, *33*, 367-382.
- [3.12] Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269-288.
- [3.13] Böhm, H. J. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243-256.
- [3.14] Muegge, I.; Martin, Y. C. *J. Med. Chem.* **1999**, *42*, 791-804.
- [3.15] Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel L. J.; Freer, S. T. *Chemistry & Biology* **1995**, *2*, 317-324.
- [3.16] Jain, A. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 427-440.
- [3.17] Levitzki, A. *Acc. Chem. Res.* **2003**, *36*, 462-469.
- [3.18] Keserü, G. M. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 649-657.
- [3.19] Jeffrey, G. A., *An introduction to hydrogen bonding*, Oxford University Press: Oxford, 1997, p 98.
- [3.20] Assefa H.; Kamath S.; Buolamwini J. K. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 475-493.

- [3.21] (a) Desiraju, G. R.; Steiner, T., *The Weak Hydrogen Bond in Structural Chemistry and Biology*. Oxford University Press: Oxford, 1999. (b) Desiraju, G. R. *Angew Chem. Int. Ed. Engl.* **1995**, *34*, 2311-2327.
- [3.22] Kroemer, R. T.; Vulpetti, A.; McDonald, J. J.; Rohrer, D. C.; Trosset, J. Y.; Giordanetto, F.; Cotesta, S.; McMartin, C.; Kihlén, M.; Stouten, P. F. W. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 871-881.
- [3.23] Desiraju, G. R. *Acc. Chem. Res.* **2002**, *35*, 565-573.
- [3.24] (a) Pierce, A. C.; ter Haar, E.; Binch, H. M.; Kay, D. P.; Patel, S. R.; Pan Li, P. *J. Med. Chem.* **2005**, *48*, 1278-1281. (b) Pierce, A. C.; Sandretto, K. L.; Bemis, G. W. *Proteins* **2002**, *49*, 567-576. (c) Sarkhel, S.; Desiraju, G. R. *Proteins* **2004**, *54*, 247-259.

CHAPTER 4

- [4.1] Stokstad, E. *Science* **2000**, 287, 2391.
- [4.2] WHO web pages on TB www.who.int/entity/tb/en
- [4.3] Dye, C.; Williams, B. G.; Espinal, M. A.; Raviglione, M. C. *Science* **2002**, 295, 2042-2046.
- [4.4] Lavie, A.; Ostermann, N.; Brundiers, R.; Goody, R. S.; Reinstein, J.; Konrad, M.; Schlichting, I. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 142045-142050.
- [4.5] Anderson, E. In *Enzymes*, 3rd ed.; Boyer, P. D., Ed. Academic Press: New York, 1973; Vol.8, p49.
- [4.6] Pochet, S.; Dugué, L.; Labesse, G.; Delepierre, M.; Munier-Lehmann, H. *ChemBioChem* **2003**, *4*, 742-747.
- [4.7] Vanheusden, V.; Munier-Lehmann, H.; Froeyen, M.; Dugué, L.; Heyerick, A.; Keukeleire, D. D.; Pochet, S.; Busson, R.; Herdewijn, P.; Calenbergh, S. V. *J. Med. Chem.* **2003**, *46*, 3811-3821.
- [4.8] Vanheusden, V.; Munier-Lehmann, H.; Pochet, S.; Herdewijn, P.; Calenbergh, S. V. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2695-2698.
- [4.9] Vanheusden, V.; Rompaey, P. V.; Munier-Lehmann, H.; Pochet, S.; Herdewijn, P.; Calenbergh, S. V. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3045-3048.

- [4.10] Sierra, L. D. L.; Munier Lehmann, H.; Gilles, A. M.; Bârză, O.; Delarue, M. *J. Mol. Biol.* **2001**, *311*, 87-100.
- [4.11] Cho, S. J.; Tropsha, A. *J. Med. Chem.* **1995**, *38*, 1060-1066.
- [4.12] Catalyst version 4. 6, Accelrys: Burlington, MA.
- [4.13] Smellie, A.; Teig, S. L.; Towbin, P. *J. Comput. Chem.* **1995**, *16*, 171-187.
- [4.14] Sutter, J.; Güner, O. F.; Hoffmann, R.; Li, H.; Waldman, M. Effect of variable weights and tolerances on predictive model generation. In *Pharmacophore: perception, development and use in drug design*, Güner, O. F., Ed.; International University Line: CA, 2000, pp 501-511.
- [4.15] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, 2004.
- [4.16] Haouz, A.; Vanheusden, V.; Munier-Lehmann, H.; Froeyen, M.; Herdewijn, P.; Calenbergh, S. V.; Delarue, M. *J. Biol. Chem.* **2003**, *278*, 4963-4971.
- [4.17] Metrangolo, P.; Neukrich, H.; Pilati, T.; Resnati, G. *Acc. Chem. Res.* **2005**, *38*, 386-395.

- [4.18] Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. *Proc. Nat. Acad. Sci. USA* **2004**, *101*, 16789-16794.
- [4.19] CSD search was restricted to entries with crystallographic $R < 0.10$ and no disorder. Also, metal complexes were excluded from the search.
- [4.20] Lommerse, J. P. M.; Stone, A. J.; Taylor, R.; Allen, F. H. *J. Am. Chem. Soc.* **1996**, *118*, 3108-3116.
- [4.21] (a) Jiang, Y.; Alcaraz, A. A.; Chen, J.; Kobayashi, H.; Lu, Y. J.; Snyder, J. P. *J. Med. Chem.* **2006**, *49*, 1891-1899. (b) Alder, M.; Kochanny, M. J.; Ye, B.; Rumennik, G.; Light, D. R.; Biancalana, S.; Whitlow, M. *Biochemistry* **2002**, *41*, 155142-15523.
- [4.22] Steinrauf, L. K.; Hamilton, J. A.; Braden, B. C.; Murrell, J. R.; Benson, M. D. *J. Biol. Chem.* **1993**, *268*, 2425-2430.
- [4.23] Howard, E. I.; Sanishvili, R.; Cachau, R. E.; Mitschler, A.; Chevrier, B.; Barth, P.; Lamour, V.; Zandt, M. V.; Sibley, E.; Bon, C.; Moras, D.; Schneider, T. R.; Joachimiak, A.; Podjarny, A. *Proteins* **2004**, *55*, 792-804.
- [4.24] Jeffrey, G. A.; Saenger, W., *Hydrogen bonding in biological structures*. Springer-Verlag: Berlin, 1991.
- [4.25] Zhu, L. L.; Hou, T. J.; Chen, L. R.; Xu, X. J. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1032-1040.
- [4.26] Munier-Lehmann, H.; Chaffotte, A.; Pochet, S.; Labesse, G. *Protein Sci.* **2001**, *10*, 1195-1205.
- [4.27] Ladbury, J. E. *Chem. & Biol.* **1996**, *3*, 973-980.
- [4.28] Aparna, V.; Rambabu, G.; Panigrahi, S. K.; Sarma, J. A. R. P.; Desiraju, G. R. *J. Chem. Inf. Model.* **2005**, *45*, 725-738.

CHAPTER 5

- [5.1] Bajorath, J. *Nature Rev. Drug Discov.* **2002**, *1*, 882-894.
- [5.2] Lyne, P. D. *Drug Discovery Today* **2002**, *7*, 1047-1055.
- [5.3] Shoda, M.; Harada, T.; Kogami, Y.; Tsujita, R.; Akashi, H.; Kouji, H.; Stahura, F. L.; Xue, L.; Bajorath, J. *J. Med. Chem.* **2004**, *47*, 4286-4290.

- [5.4] Xue, L.; Stahura, F. L.; Bajorath, J. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2032-2039.
- [5.5] Schuster, D.; Laggner, C.; Steindl, T. M.; Paluszczak, A.; Hartmann, R. W.; Langer, T. *J. Chem. Inf. Model.* **2006**, *46*, 1301-1311.
- [5.6] Debnath, A. K. *J. Med. Chem.* **2002**, *45*, 41-53.
- [5.7] Srinivasan, J.; Castellino, A.; Bradley, E. K.; Eksterowicz, J. E.; Grootenhuis, P. D. J.; Putta, S.; Stanton, R. V. *J. Med. Chem.* **2002**, *45*, 2494-2500.
- [5.8] Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T. M.; Murray, C. W.; Taylor, R. D.; Watson, P. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793-806.
- [5.9] Perola, E.; Xu, K.; Kollmeyer, T. M.; Kaufmann, S. H.; Prendergast, F. G.; Pang, Y. P. *J. Med. Chem.* **2000**, *43*, 401-408.
- [5.10] <http://www.ibscreen.com>.
- [5.11] <http://www.chemstar.ru>.
- [5.12] <http://www.mdpi.org>
- [5.13] <http://www.keyorganics.ltd.uk>.
- [5.14] <http://www.asinex.com/welcome.htm>
- [5.15] http://www.apps1.niaid.nih.gov/struct_search/an/an_search.htm
- [5.16] <http://www.maybridge.com/html/home.htm>
- [5.17] Sutter, J.; Güner, O. F.; Hoffmann, R.; Li, H.; Waldman, M. Effect of variable weights and tolerances on predictive model generation. In *Pharmacophore: perception, development and use in drug design*, Güner, O. F., Ed.; International University Line: CA, 2000, pp 501-511.

CHAPTER 6

- [6.1] Bredt, D. S. *J. Cell Sci.* **2003**, *116*, 9-15.
- [6.2] Marletta, M. A. *Cell* **1994**, *78*, 927-930.
- [6.3] Schmidt, H. H. H. W.; Walter, U. *Cell* **1994**, *78*, 919-925.
- [6.4] Abu-Soud, H. M.; Stuehr, D. J. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10769-10772.

- [6.5] Alderton, W. K.; Cooper, C. E.; Knowles, R. G. *Biochem. J.* **2001**, *357*, 593-615.
- [6.6] (a) Berka, V.; Yeh, H. C.; Gao, D.; Kiran, F.; Tsai, A. L. *Biochemistry* **2004**, *43*, 13137-13148. (b) Sørli, M.; Gorren, A. F. C.; Marchal, S.; Shimizu, T.; Lange, R.; Andersson, K. K.; Mayer, B. *J. Biol. Chem.* **2003**, *278*, 48602-48610. (c) Hurshman, A. R.; Krebs, C.; Edmondson, D. E.; Huynh, B. H.; Marletta, M. A. *Biochemistry* **1999**, *38*, 15689-15696.
- [6.7] Messerschmidt, A.; Huber, R.; Poulos, T.; Wieghardt, K. Inducible nitric oxide synthase. In *Handbook of metalloproteins*, John Wiley & Sons: Westsussex, 2001, pp 285-299.
- [6.8] Wang, P. G.; Xian, M.; Tang, X.; Wu, X.; Wen, Z.; Cai, T.; Janczuk, A. J. *Chem. Rev.* **2002**, *102*, 1091-1134.
- [6.9] Moore, W. M.; Webber, R. K.; Jerome, G. M.; Tjoeng, F. S.; Misko, T. P.; Currie, M. G. *J. Med. Chem.* **1994**, *37*, 3886-3888.
- [6.10] Webber, R. K.; Metz, S.; Moore, W. M.; Connor, J. R.; Currie, M. G.; Fok, K. F.; Hagen, T. J.; Hansen, Jr., D. W.; Jerome, G. M.; Manning, P. T.; Pitzele, B. S.; Toth, M. V.; Trivedi, M.; Zupiec, M. E.; Tjoeng, F. S. *J. Med. Chem.* **1998**, *41*, 96-101.
- [6.11] Tinker, A. C.; Beaton, H. G.; Boughton-Smith, N.; Cook, T. R.; Cooper, S. L.; Fraser-Rae, L.; Hallam, K.; Hamley, P.; McNally, T.; Nicholls, D. J.; Pimm, A. D.; Wallace, A. V. *J. Med. Chem.* **2003**, *46*, 913-916.
- [6.12] Matter, H.; Kotsonis, P. *Med. Res. Rev.* **2004**, *24*, 662-684.
- [6.13] (a) Crane, B. R.; Arvai, A. S.; Gachhui, A. R.; Wu, C.; Ghosh, D. K.; Getzoff, E. D.; Steuhr, D. J.; Tainer, J. A. *Science* **1997**, *278*, 425-438. (b) Crane, B. R.; Arvai, A. S.; Ghosh, D. K.; Wu, C.; Getzoff, E. D.; Stuehr, D. J.; Tainer, J. A. *Science* **1998**, *279*, 2121-2126. (c) Federov, R.; Vasan, R.; Ghosh, D. K.; Schlichthing, I. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5892-5897. (d) Raman, C. S.; Li, H.; Martasek, P.; Babu, B. R.; Griffith, O. W.; Masters, B. S. S.; Poulos, T. L. *J. Biol. Chem.* **2001**, *276*, 26486-26491.

- [6.14] Fischmann, T. O.; Hruza, A.; Niu, X. D.; Fossetta, J. D.; Lunn, C. A.; Dolphin, E.; Prongay, A. J.; Reichert, P.; Lundell, D. J.; Narula, S. K.; Weber, P. C. *Nat. Struct. Biol.* **1999**, *6*, 233-242.
- [6.15] Young R. J.; Beams, R. M.; Keith C.; Clark, H. A. R.; Coe, D. M.; Chambers, C. L.; Davies, P.I.; Dawson, J.; Drysdale, M. J.; Karl W.; Franzman, K. W.; French, C.; Hodgson, S. T.; Hodson, H. F.; Kleanthous, S.; Rider, P.; Sanders, D.; Sawyer, D. A.; Scott, K. J.; Shearer, B. G; Stocker, R.; Smith, S.; Tackley, M. C.; Knowles, R. G. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 597-600.
- [6.16] Li, H.; Raman, C. S.; Glaser, C. B.; Blasko, E.; Tish, A.; Young, T. A.; John F.; Parkinsoni J. F.; Whitlow, M.; Poulos, T. L. *J. Biol. Chem.* **1999**, *274*, 21276-21284.
- [6.17] Shi, S.; Yan, L.; Yang, Y.; Fisher-shaulsky J.; Thacher, T. *J. Comput. Chem.* **2003**, *24*, 1059-1076.
- [6.18] Poulos, T. L. *Drug Metab. Dispos.* **2005**, *33*, 10-18.
- [6.19] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.
- [6.20] Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635-2643.
- [6.21] Chen, P.; Tsai, A.; Berka, V.; Wu, K. K. *J. Biol. Chem.* **1997**, *272*, 6114-6118.
- [6.22] Griffith, O.W.; Stuehr, D. J. *Annu. Rev. Physiol.* **1995**, *57*, 707-736.
- [6.23] Li, H.; Raman, C. S.; Martásek, P.; Masters, B. S. S.; Poulos, T. L. *Biochemistry* **2001**, *40*, 5399-5406.
- [6.24] Grant, S. K.; Green, B. G.; Stiffey-Wilusz, J.; Durette, P. L.; Shah, S. K.; Kozarich, J. W. *Biochemistry* **1998**, *37*, 4174-4180.
- [6.25] Davydov, R.; Ledbetter-Rogers, A.; Martásek, P.; Larukhin, M.; Sono, M.; Dawson, J. H.; Masters, B. S. S.; Hoffman, B. M. *Biochemistry* **2002**, *41*, 10375-10381.
- [6.26] Rosenfeld, R. J.; Garcin, E. D.; Panda, K.; Andersson, G.; Åberg, A.; Wallace, A. V.; Morris, G. M.; Olson, A. J.; Stuehr, D. J.; Tainer, J. A.; Getzoff, E. D. *Biochemistry*, **2002**, *41*, 13915-13925.

- [6.27] (a) Sanjeev, B. S.; Vishveshwara, S. *J. Biomol. Struct. Dyn.* **2005**, 22, 657-672.
(b) Gopalakrishnan, B.; Aparna, V.; Jeevan, J.; Ravi, M.; Desiraju, G. R. *J. Chem. Inf. Model.* **2005**, 45, 1101-1108.
- [6.28] Dunitz, J. D. *Science* **1994**, 264, 670.
- [6.29] Scheiner, S. *Hydrogen bonding: a theoretical perspective*. Oxford University Press: Oxford, 1997.
- [6.30] Davis, A. M.; Teague, S. J. *Angew. Chem. Int. Ed.* **1999**, 38, 736-749.
- [6.31] Ji, H.; Li, H.; Flinspach, M.; Poulos, T. L.; Silverman, R. B. *J. Med. Chem.* **2003**, 46, 5700-5711.

APPENDIX I

Table 1: Scores and RMSD values of best poses obtained from GOLD and LigandFit

Molecule No.	Activity pIC ₅₀	GOLD		Ligand Fit		Class ^a
		Score	RMSD	Score	RMSD	
1	6.46	45.18	0.48	60.42	0.19	Misoriented
2	6.04	46.00	0.35	47.88	0.52	Misoriented
3	7.63	46.72	0.35	55.50	0.12	Misoriented
4	7.56	46.40	0.44	54.22	0.10	Misoriented
5	7.09	48.62	0.51	47.58	0.50	Misoriented
6	6.23	46.94	0.54	75.24	0.41	Misoriented
7	6.04	49.89	0.61	62.28	1.23	Misoriented
8	6.45	46.35	0.75	62.64	0.45	Misoriented
9	6.00	46.76	0.26	66.79	1.25	Close
10	8.00	47.36	0.58	62.84	0.47	Close
11	9.76	49.13	0.39	85.59	0.38	Close
12	9.92	44.17	0.36	97.71	0.46	Close
13	7.25	44.49	0.51	55.23	0.49	Misoriented
14	7.25	45.65	0.73	70.34	0.38	Close
15	6.11	47.50	0.59	60.15	0.23	Close
16	6.24	49.43	0.60	88.26	0.60	Close
17	6.92	38.29	0.46	71.10	0.53	Shifted
18	7.00	40.99	0.27	52.17	0.17	Shifted
19	8.48	42.20	0.88	70.41	0.48	Shifted
20	5.21	42.20	0.30	64.86	1.22	Shifted
21	6.09	43.13	0.52	68.62	1.29	Close
22	6.26	47.95	0.25	52.07	1.43	Close
23	7.53	44.63	0.50	84.90	1.41	Close
24	8.42	45.37	0.55	76.73	1.38	Close
25	9.50	45.11	0.63	77.58	0.54	Close
26	9.05	49.03	0.54	61.31	1.29	Shifted
27	9.61	47.31	0.54	92.71	1.47	Shifted
28	8.39	36.52	0.81	69.21	0.27	Close
29	7.07	41.63	0.51	74.11	0.12	Close
30	7.92	48.26	0.32	84.15	1.28	Close
31	8.32	47.49	0.40	50.50	0.14	Close
32	7.39	46.89	0.27	41.95	0.12	Close
33	8.15	43.59	0.78	58.34	0.11	Close
34	7.92	45.23	0.33	41.85	0.10	Close
35	7.95	42.99	0.36	58.21	0.13	Close
36	9.16	43.47	1.02	103.9	0.43	Close
37	6.79	46.14	0.78	89.32	0.50	Shifted
38	8.42	49.05	0.65	81.11	0.52	Close
39	8.18	48.16	0.25	69.46	0.30	Close
40	7.16	43.34	0.43	57.72	0.27	Close
41	7.82	49.37	0.61	72.82	1.34	Close
42	7.60	48.36	1.12	67.54	1.25	Close
43	11.22	51.77	0.81	79.64	1.61	Close
44	9.76	54.09	0.54	83.72	0.74	Shifted
45	10.14	49.56	0.52	84.10	1.40	Close

Molecule No.	Activity pIC_{50}	GOLD		Ligand Fit		Class ^a
		Score	RMSD	Score	RMSD	
46	7.46	47.69	1.43	58.52	1.33	Close
47	8.11	49.23	0.50	55.53	0.22	Close
48	7.74	50.47	0.59	60.59	1.24	Close
49	7.35	47.70	0.45	70.66	1.30	Close
50	8.50	49.03	0.21	82.81	0.13	Close
51	8.01	46.65	0.21	86.29	0.37	Close
52	8.36	50.29	0.60	80.73	1.43	Close
53	7.45	46.02	0.41	33.96	1.32	Close
54	7.53	44.58	0.37	29.92	0.21	Shifted
55	7.88	45.14	0.48	29.82	0.09	Misoriented
56	7.40	47.36	0.58	57.87	0.55	Close
57	6.60	44.31	0.54	31.59	0.18	Close
58	7.39	44.35	0.30	51.35	0.14	Close
59	6.61	43.91	0.48	51.31	0.15	Close
60	8.00	41.99	0.51	33.58	0.17	Close
61	7.04	41.71	0.81	60.42	0.23	Close
62	5.32	40.75	0.79	47.88	0.76	Close
63	5.43	43.53	0.41	55.50	0.24	Shifted
64	6.88	43.10	1.86	54.22	2.15	Close
65	6.17	42.53	2.23	47.58	2.29	Close
66	5.27	42.52	0.31	75.23	0.20	Shifted
67	5.74	43.09	0.31	62.28	0.22	Close
68	5.31	40.45	0.39	62.64	0.25	Close
69	6.07	40.49	0.25	66.79	0.21	Shifted
70	6.92	41.37	0.43	62.84	0.17	Close
71	7.15	42.16	1.98	85.59	2.07	Close
72	7.39	43.84	0.29	97.71	0.20	Close
73	7.29	46.80	0.69	55.23	0.14	Misoriented
74	7.39	51.18	1.30	70.34	0.32	Close
75	6.90	46.16	0.52	60.15	0.26	Misoriented
76	8.58	45.58	0.65	88.26	0.35	Close
77	8.63	54.41	0.65	71.10	0.54	Close
78	6.16	46.76	0.53	52.17	0.28	Misoriented
79	6.02	44.99	0.24	70.41	0.37	Close
80	6.16	47.15	0.47	64.86	0.16	Misoriented
81	7.28	47.30	0.25	68.62	0.23	Close
82	6.48	46.38	0.28	52.07	0.16	Close
83	6.58	44.36	0.76	84.90	0.60	Close
84	7.88	48.21	0.24	76.73	0.11	Close
85	7.08	50.90	1.25	77.58	0.24	Close
86	8.82	49.96	1.27	61.31	0.13	Close
87	9.11	48.33	0.21	92.71	0.17	Close
88	9.02	47.68	0.29	69.21	0.28	Close
89	8.42	48.70	0.58	74.11	0.14	Close
90	9.09	54.06	0.30	84.15	0.47	Close
91	8.53	53.91	0.29	50.50	0.75	Shifted
92	9.60	56.82	0.27	41.95	0.17	Close
93	8.63	55.48	0.17	58.34	1.12	Close
94	6.42	43.70	0.18	41.85	1.44	Close
95	7.76	49.07	1.25	58.21	0.18	Close

Molecule No.	Activity pIC ₅₀	GOLD		Ligand Fit		Class ^a
		Score	RMSD	Score	RMSD	
96	8.36	46.65	0.64	103.9	0.10	Close
97	8.39	46.08	0.36	89.32	1.19	Close
98	8.63	52.62	0.19	81.11	0.36	Close
99	8.52	56.56	0.36	69.46	0.24	Close
100	9.61	43.87	0.38	57.72	0.31	Close
101	8.58	44.56	0.41	72.82	1.21	Close
102	9.03	45.30	0.22	67.54	0.22	Shifted
103	8.49	44.38	0.33	79.64	1.26	Shifted
104	7.85	48.07	0.30	83.72	1.26	Shifted
105	7.92	46.16	0.48	84.10	0.10	Shifted
106	7.34	46.46	0.28	58.52	1.22	Close
107	8.05	53.92	0.25	55.53	1.28	Shifted
108	8.13	42.96	0.25	60.59	1.16	Shifted
109	8.07	56.30	0.34	75.94	0.16	Close
110	7.39	45.53	0.83	71.18	0.16	Close
111	8.49	54.74	0.29	82.36	0.25	Close
112	8.72	56.01	0.29	70.96	0.31	Misoriented
113	8.26	58.13	0.29	74.82	0.11	Close
114	8.30	51.24	0.45	71.38	0.31	Close
115	8.03	47.14	0.41	69.63	0.37	Shifted
116	8.92	48.77	0.22	72.56	0.25	Close
117	8.14	59.73	0.32	70.52	0.24	Shifted
118	9.29	58.99	0.30	61.53	0.16	Shifted
119	9.04	53.76	0.45	65.03	0.69	Shifted
120	8.82	45.70	0.48	62.09	0.18	Shifted
121	9.21	47.76	0.27	65.34	0.19	Shifted
122	9.55	48.20	0.44	63.82	0.27	Close
123	7.79	49.33	0.27	63.30	0.19	Close
124	8.85	51.70	0.25	65.91	0.16	Close
125	8.26	49.46	0.43	58.52	0.21	Close
126	8.03	54.13	0.30	60.25	0.26	Close
127	8.25	54.90	0.29	68.32	0.31	Shifted
128	8.45	56.01	0.37	65.83	0.25	Shifted

^aClassification of GOLD poses based on interactions

Table 2: Metrics of hydrogen bonds

Molecule No.	Met769, N–H···N		Gln767, C–H···O		Water10, O _w –H···N	
	Distance	Angle	Distance	Angle	Distance	Angle
1	–	–	–	–	–	–
2	–	–	–	–	–	–
3	–	–	–	–	–	–
4	–	–	–	–	–	–
5	–	–	–	–	–	–
6	–	–	–	–	–	–
7	–	–	–	–	–	–
8	1.94	136.7	2.03	152.2	2.13	136.0
9	2.41	130.6	1.92	130.2	1.93	159.4
10	2.41	114.9	2.02	136.6	1.40	147.6
11	2.10	135.8	1.94	130.2	2.26	153.7
12	2.30	133.3	1.86	115.7	2.05	148.9
13	–	–	–	–	–	–
14	2.13	135.3	1.95	128.6	2.26	155.7
15	2.51	131.4	1.93	114.8	1.92	152.5
16	2.24	139.7	1.89	147	2.07	147.9
17	3.00	115.4	2.14	127.5	1.25	165.6
18	2.43	137.9	2.00	171.6	2.29	162.7
19	2.53	137.3	2.10	170.4	2.14	159.8
20	2.37	137.0	1.93	177.0	2.22	160.9
21	2.42	132.8	2.20	154.6	1.84	141.3
22	2.40	130.0	1.93	129.3	1.93	157.8
23	2.73	113.3	2.08	132.4	1.13	142.0
24	2.78	114.8	2.03	133.9	1.15	152.3
25	2.73	117.6	2.05	140.9	1.61	171.0
26	2.18	143.7	1.94	150.6	2.22	145.1
27	2.55	114.4	1.88	138.6	1.49	168.0
28	2.36	113.8	2.20	129.1	1.45	155.0
29	2.74	118.1	2.06	148.7	1.09	133.8
30	2.47	138.6	1.98	165.2	2.33	164.3
31	2.08	141.5	1.97	176.1	2.38	151.2
32	2.45	134.9	1.96	136.0	2.01	157.9
33	2.17	141.4	2.14	126.7	2.31	151.9
34	2.31	134.4	1.89	120.5	2.01	150.1
35	2.32	138.5	1.91	127.2	2.17	151.5
36	2.30	141.7	2.05	124.4	2.21	149.7
37	3.07	123.2	2.08	159.7	1.89	167.3
38	2.28	130.5	1.89	119.0	2.03	154.8
39	2.47	127.0	1.95	127.1	1.71	158.7
40	2.48	134.5	1.89	125.3	1.95	152.6
41	2.37	115.7	2.11	133.0	1.43	143.2
42	2.14	138.7	1.90	128.5	2.13	146.8
43	2.41	125.4	1.92	117.9	1.70	157.2
44	2.55	114.2	2.00	132.7	1.31	147.0
45	2.65	116.2	2.01	138.6	1.27	148.8
46	2.18	136.7	2.03	122.3	2.19	153.2

Molecule No.	Met769, N-H...N		Gln767, C-H...O		Water10, O _w -H...N	
	Distance	Angle	Distance	Angle	Distance	Angle
47	2.20	140.0	2.13	132.0	2.23	154.1
48	2.32	134.2	1.97	126.5	2.05	156.4
49	2.15	136.1	1.92	128.5	2.16	153.0
50	2.54	124.7	1.96	127.7	1.59	162.6
51	2.37	129.3	1.87	130.9	1.86	157.1
52	2.27	132.8	1.89	130.6	2.09	156.9
53	2.20	143.7	2.01	173.2	2.40	152.9
54	2.69	117.6	2.02	136.9	1.63	169.5
55	—	—	—	—	—	—
56	2.27	13.8	1.98	128.1	2.15	165.2
57	2.29	139	1.96	135.6	2.15	145.3
58	2.46	137.2	1.92	164.5	2.06	158.0
59	2.53	134.2	1.98	135.3	2.07	159.5
60	2.37	139.4	2.08	154.5	2.01	144.8
61	2.03	137.1	2.03	151.7	2.49	154.2
62	2.19	131.6	2.02	145.0	2.35	159.9
63	2.58	137.5	2.02	154.0	1.82	146.9
64	2.19	123.6	1.94	151.0	2.01	155.3
65	1.78	140.5	1.99	162.9	2.72	151.8
66	2.75	114.1	1.97	138.9	1.30	170.5
67	2.37	143.0	2.17	163.2	2.44	157.3
68	1.64	146.1	2.09	164.5	2.64	137.6
69	2.37	141.9	2.06	167.7	2.41	159.6
70	2.56	134.5	2.01	120.5	1.91	153.2
71	2.39	136.6	1.97	173.4	2.14	159.0
72	2.60	129.5	1.89	134.8	1.87	162.9
73	—	—	—	—	—	—
74	2.12	138.5	2.16	124.2	2.28	154.7
75	—	—	—	—	—	—
76	2.44	130.6	1.83	134.3	2.04	161.2
77	2.29	127.4	1.85	172.7	1.62	137.5
78	—	—	—	—	—	—
79	2.43	124.4	1.83	127.2	1.77	160.2
80	—	—	—	—	—	—
81	2.45	129.4	1.90	131.4	1.86	159.7
82	2.39	114.7	1.92	140.8	1.44	148.7
83	2.22	121.9	1.94	143.8	1.68	150.5
84	2.33	135.5	2.13	126.6	1.97	154.2
85	2.17	134.3	2.11	127.4	2.04	153.1
86	2.13	135.4	2.04	129.8	2.19	155.9
87	2.28	131.5	1.96	126.6	1.92	155.5
88	2.37	135.4	2.01	174.4	2.32	163.3
89	2.14	137.7	2.07	126.1	2.20	153.6
90	2.16	135.5	1.93	162.7	2.34	159.0
91	2.65	120.5	1.85	156.2	1.85	174.0
92	2.32	130.2	2.06	121.6	1.87	153.2
93	2.36	127.7	2.15	122.9	1.88	161.6
94	2.14	138.0	2.22	126.4	2.31	156.9
95	2.28	131.6	1.92	128.5	2.16	153.0

Molecule No.	Met769, N-H...N		Gln767, C-H...O		Water10, O _w -H...N	
	Distance	Angle	Distance	Angle	Distance	Angle
96	2.30	136.5	1.92	168.0	2.28	161.1
97	2.21	119.9	1.86	144.1	1.85	159.3
98	2.37	130.9	2.27	134.8	2.26	164.1
99	2.15	127.5	2.00	148.2	2.00	156.9
100	2.75	119.7	2.03	141.2	1.66	170.9
101	2.73	114.3	2.07	145.0	1.08	133.8
102	2.42	129.5	2.01	158.8	1.63	134.8
103	2.55	120.1	1.93	153.4	1.34	142.1
104	2.09	135.0	2.01	168.7	2.33	154.4
105	2.30	130.4	1.94	157.1	1.70	133.5
106	2.69	120.2	1.87	120.8	1.40	161.3
107	1.92	135.5	2.01	161.5	2.16	137.8
108	2.26	126.9	1.94	153.9	1.63	135.7
109	2.39	117.1	1.90	153.0	1.40	144.5
110	2.20	140.1	1.88	140.3	2.18	150.0
111	2.61	115.4	2.04	136.4	1.34	153.1
112	2.98	113.0	2.07	125.6	1.03	162.3
113	2.48	125.9	1.85	129.0	1.69	159.8
114	2.42	136.6	2.00	135.9	1.89	149.5
115	2.26	126.9	1.94	153.9	1.63	135.7
116	2.24	128.2	1.97	145.8	2.03	152.7
117	2.17	140.1	2.01	176.3	2.37	154.2
118	2.08	135.4	2.07	164.7	2.23	148.2
119	2.52	116.6	2.08	137.1	1.54	161.1
120	2.32	129.2	1.91	165.5	2.10	162.2
121	2.55	122.4	1.80	153.5	1.91	171.8
122	2.52	132.2	2.14	151.6	1.77	145.5
123	2.26	137.2	1.86	144.9	2.17	154.8
124	2.67	117.8	2.11	132.8	1.60	161.5
125	2.20	118.5	1.89	141.5	1.64	152.0
126	2.67	130.7	1.91	175.5	1.91	167.8
127	2.69	132.5	2.16	164.4	2.07	165.0
128	2.46	130.2	1.85	137.1	1.66	151.7

Table 3: Experimental and predicted activities of study molecules in consensus model

Molecule No.	Experimental Activity pIC ₅₀	Predicted Activity pIC ₅₀
1	6.46	6.94
2	6.04	7.25
3	7.63	8.12
4	7.56	7.42
5	7.09	8.20
6	6.23	7.39
7	6.04	8.01
8	6.45	7.81
9	6.00	7.67
10	8.00	8.34
11	9.76	9.76
12	9.92	10.75
13	7.25	6.95
14	7.25	8.17
15	6.11	6.75
16	6.24	7.04
17	6.92	6.64
18	7.00	7.03
19	8.48	7.27
20	5.21	6.24
21	6.09	6.36
22	6.26	6.15
23	7.53	7.74
24	8.42	7.80
25	9.50	9.47
26	9.05	8.24
27	9.61	8.06
28	8.39	8.66
29	7.07	8.12
30	7.92	9.50
31	8.32	7.15
32	7.39	8.55
33	8.15	7.94
34	7.92	7.74
35	7.95	8.00
36	9.16	9.10
37	6.79	8.32
38	8.42	7.81
39	8.18	7.73
40	7.16	6.66
41	7.82	7.43
42	7.60	7.25

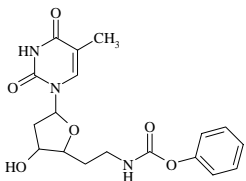
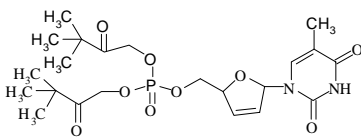
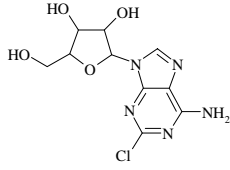
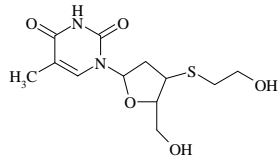
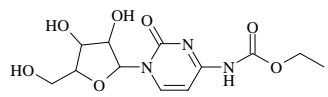
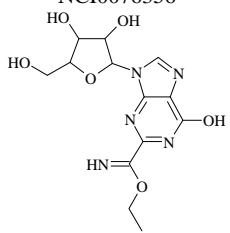
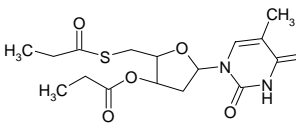
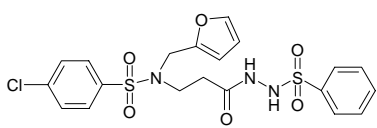
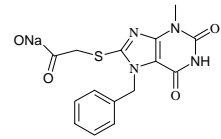
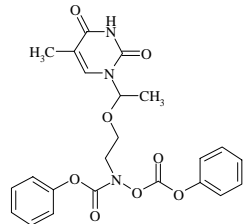
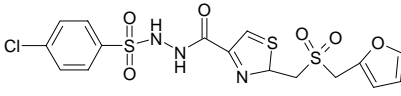
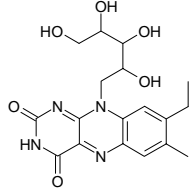
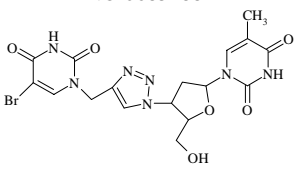
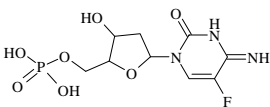
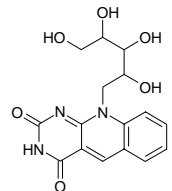
Molecule No.	Experimental Activity pIC ₅₀	Predicted Activity pIC ₅₀
43	11.22	10.57
44	9.76	10.12
45	9.01	8.67
46	7.46	7.70
47	8.11	8.08
48	7.74	8.45
49	7.35	7.55
50	8.50	7.53
51	8.01	6.62
52	8.36	8.44
53	7.45	8.53
54	7.53	7.94
55	7.88	7.03
56	7.40	7.98
57	6.60	6.88
58	7.39	7.26
59	6.61	7.46
60	8.00	6.65
61	7.04	5.67
62	5.32	6.22
63	5.43	5.27
64	6.88	6.44
65	6.17	5.54
66	5.27	5.61
67	5.74	5.48
68	5.31	5.71
69	6.07	5.93
70	6.92	7.28
71	7.15	6.96
72	7.39	8.22
73	7.29	8.58
74	7.39	7.99
75	6.90	7.78
76	8.58	8.17
77	8.63	6.52
78	6.16	5.88
79	6.02	6.17
80	6.16	7.87
81	7.28	5.58
82	6.48	6.21
83	6.58	5.93
84	7.88	7.28
85	7.08	8.42
86	8.82	8.32

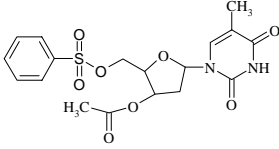
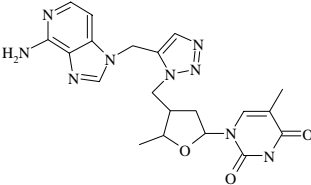
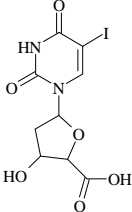
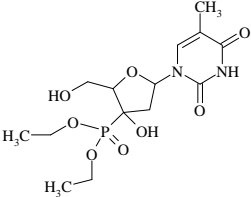
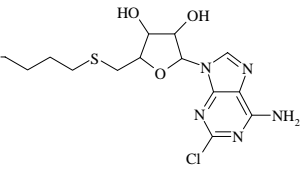
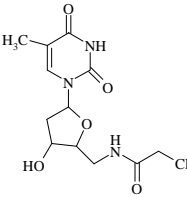
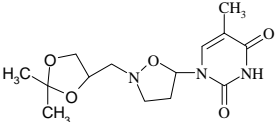
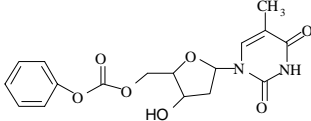
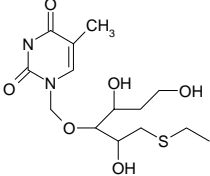
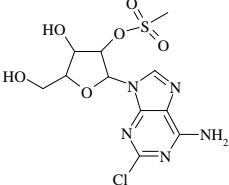
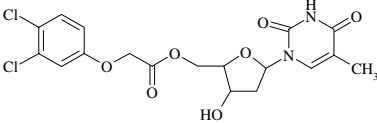
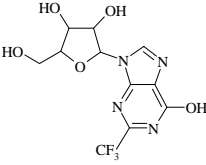
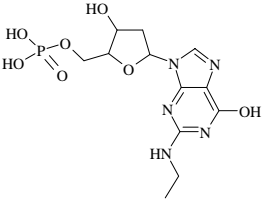
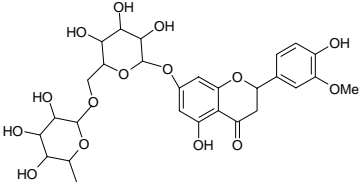
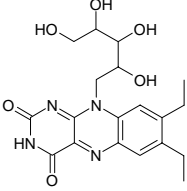
Molecule No.	Experimental Activity pIC ₅₀	Predicted Activity pIC ₅₀
87	9.11	7.68
88	9.02	7.63
89	8.42	8.60
90	9.09	8.70
91	8.53	7.09
92	9.60	9.97
93	8.63	9.91
94	6.42	8.99
95	7.76	8.79
96	8.36	8.80
97	8.39	7.51
98	8.63	9.26
99	8.52	9.31
100	9.61	6.72
101	8.58	6.93
102	9.03	6.81
103	8.49	7.99
104	7.85	7.43
105	7.92	9.51
106	7.34	7.42
107	8.05	7.01
108	8.13	7.81
109	8.07	8.60
110	7.39	7.77
111	8.49	8.20
112	8.72	8.28
113	8.26	8.86
114	8.30	8.36
115	8.03	6.38
116	8.92	7.23
117	8.14	9.08
118	9.29	9.21
119	9.04	7.43
120	8.82	9.08
121	9.21	8.40
122	9.55	7.89
123	7.79	7.96
124	8.85	8.75
125	8.26	6.89
126	8.03	9.36
127	8.25	8.58
128	8.45	7.07

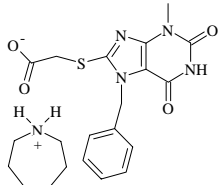
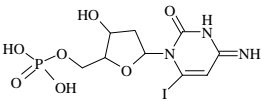
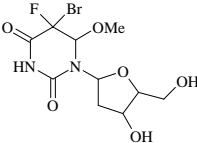
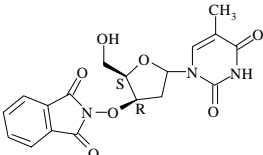
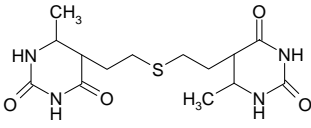
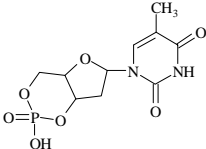
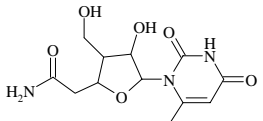
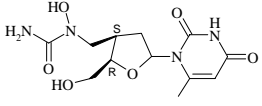
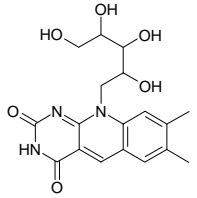
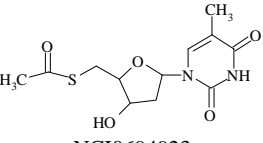
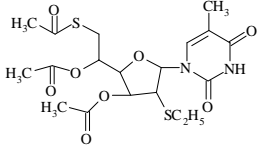
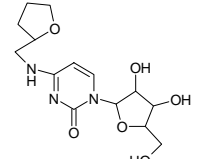
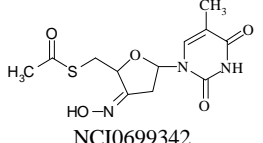
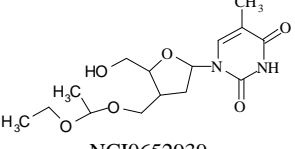
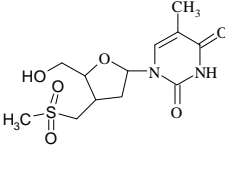
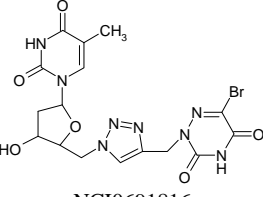
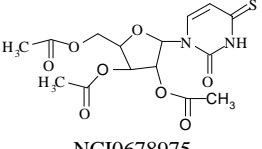
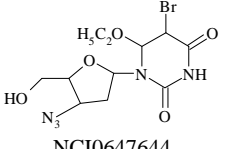
*Training set in **bold**

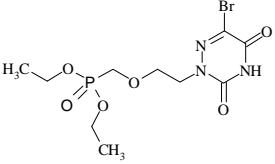
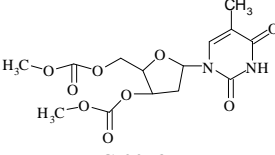
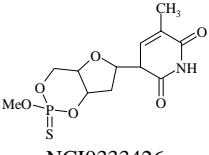
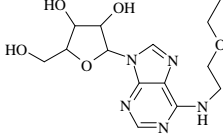
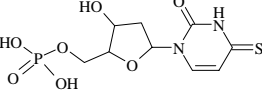
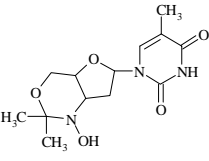
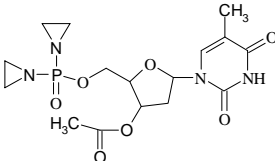
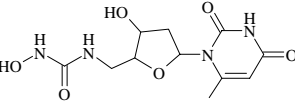
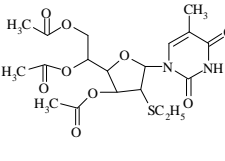
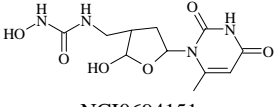
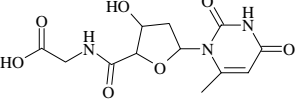
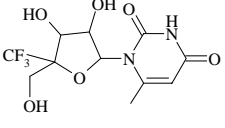
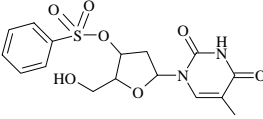
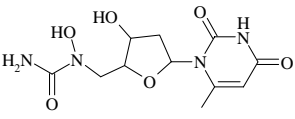
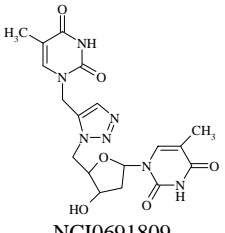
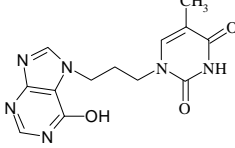
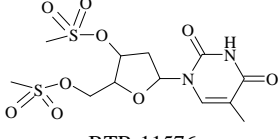
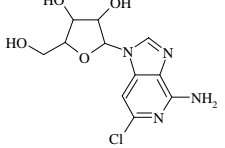
APPENDIX-II

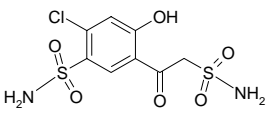
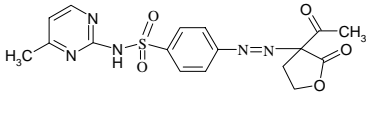
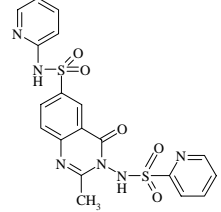
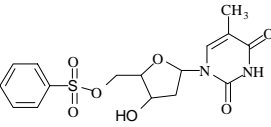
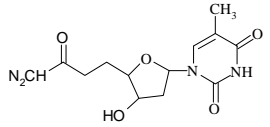
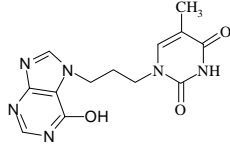
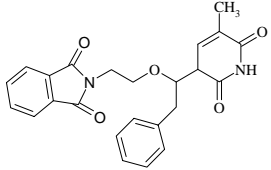
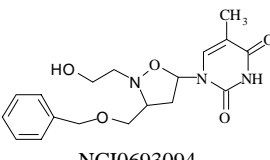
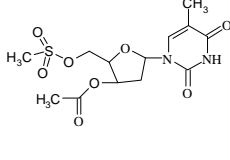
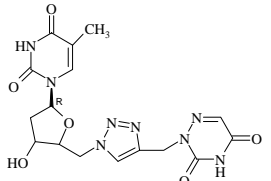
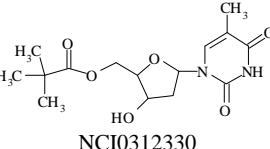
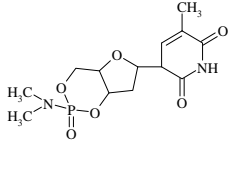
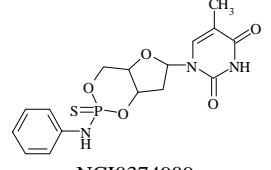
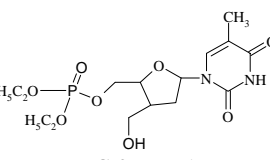
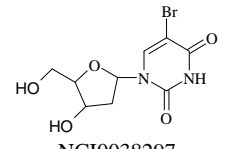
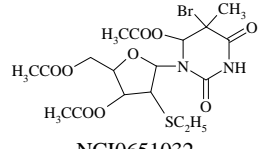
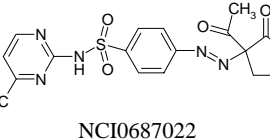
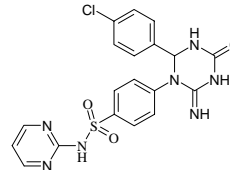
Table1: List of hits obtained from Virtual Screening of TMPK_{mt}

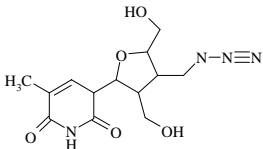
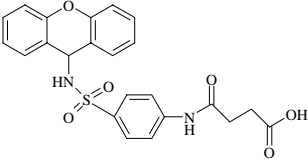
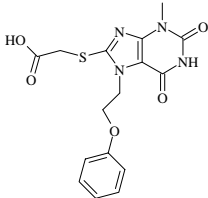
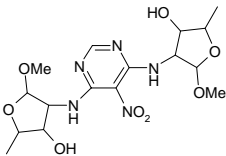
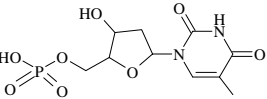
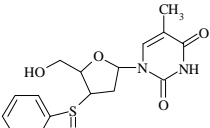
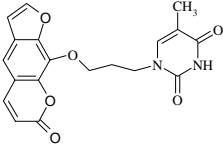
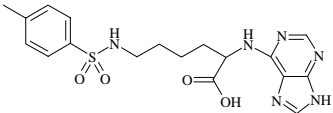
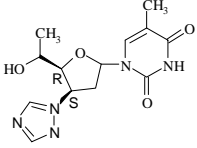
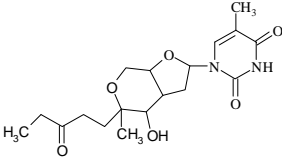
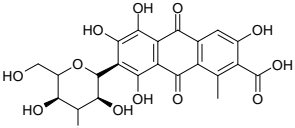
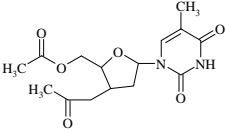
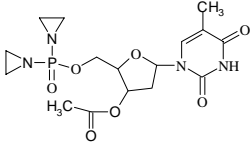
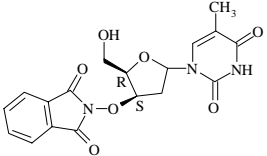
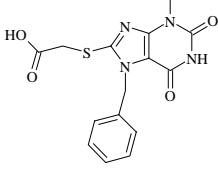
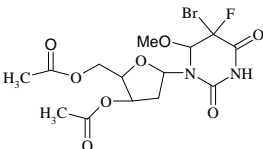
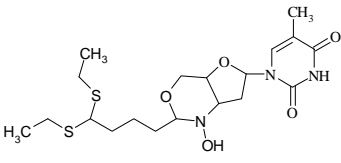
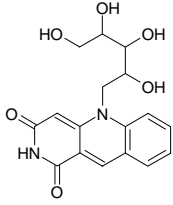
Hits	Hits	Hits
 <p>NCI0327193</p>	 <p>NCI0701202</p>	 <p>NCI0076356</p>
 <p>NCI0700462</p>	 <p>NCI0129781</p>	 <p>NCI0128665</p>
 <p>NCI0077489</p>	 <p>KM-09420</p>	 <p>sdmol-16384</p>
 <p>NCI0685188</p>	 <p>KM-03270</p>	 <p>NCI0068758</p>
 <p>NCI0691824</p>	 <p>NCI0521751</p>	 <p>RH-01439</p>

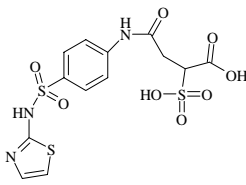
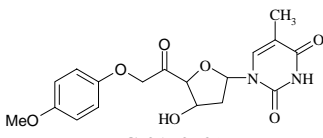
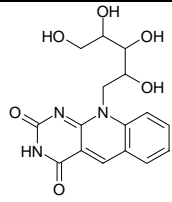
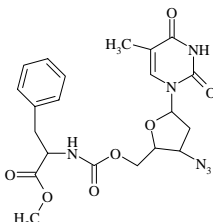
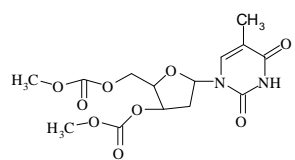
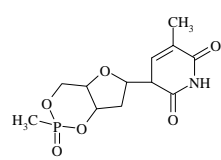
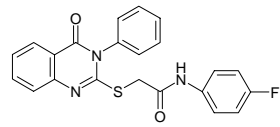
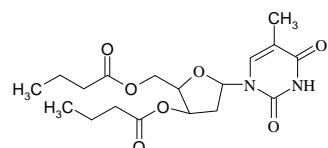
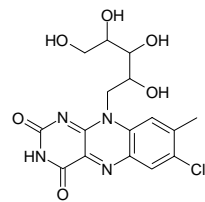
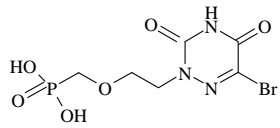
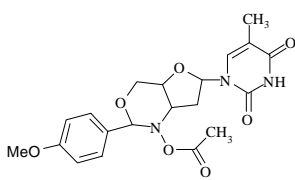
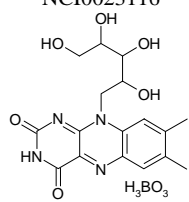
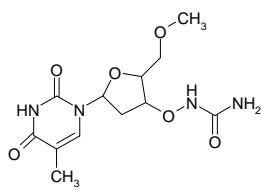
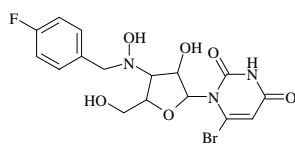
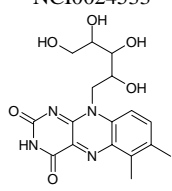
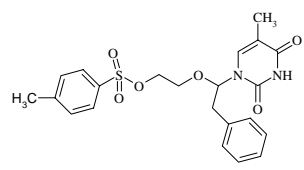
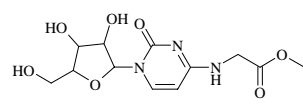
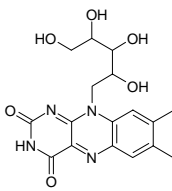
Hits	Hits	Hits
 <p>BTB-14739</p>	 <p>NCI0691812</p>	 <p>NCI0113570</p>
 <p>NCI0667004</p>	 <p>NCI0166151</p>	 <p>NCI0338955</p>
 <p>NCI0666121</p>	 <p>NCI0211721</p>	 <p>NCI0152759</p>
 <p>NCI0080100</p>	 <p>NCI0160911</p>	 <p>NCI0327195</p>
 <p>NCI0685837</p>	 <p>Compound-403</p>	 <p>NCI0036314</p>

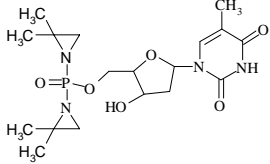
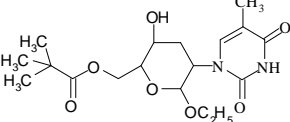
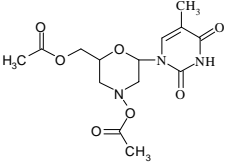
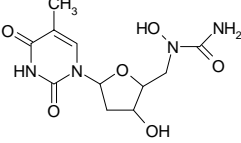
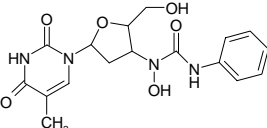
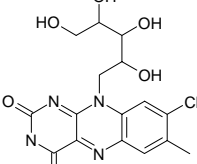
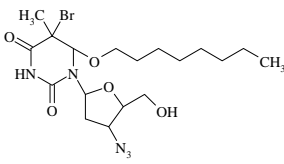
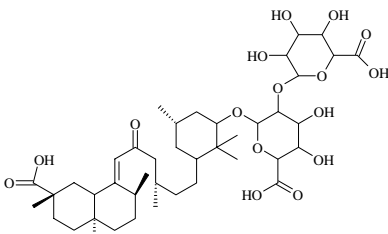
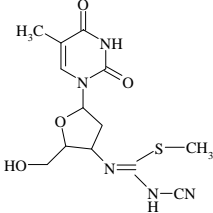
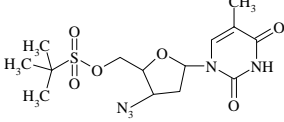
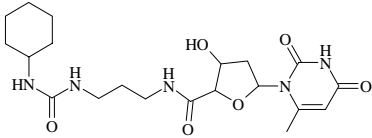
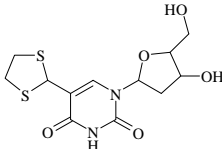
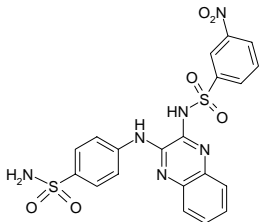
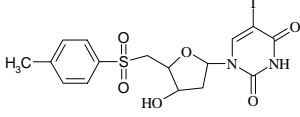
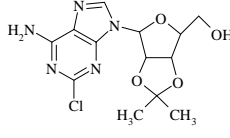
Hits	Hits	Hits
 <p>sdmol-15983</p>	 <p>NCI0124854</p>	 <p>NCI0080870</p>
 <p>NCI0658885</p>	 <p>NCI0111271</p>	 <p>NCI0155796</p>
 <p>NCI0693096</p>	 <p>NCI0693091</p>	 <p>NCI0112202</p>
 <p>NCI0694923</p>	 <p>NCI0157065</p>	 <p>NCI0352273</p>
 <p>NCI0699342</p>	 <p>NCI0652939</p>	 <p>NCI0142846</p>
 <p>NCI0691816</p>	 <p>NCI0678975</p>	 <p>NCI0647644</p>

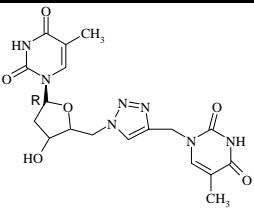
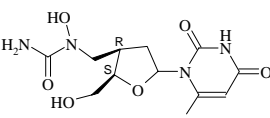
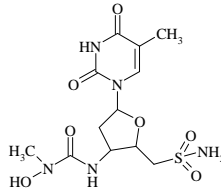
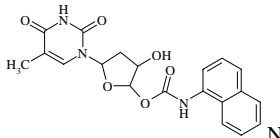
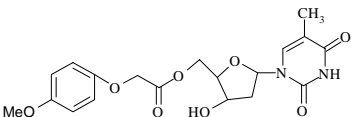
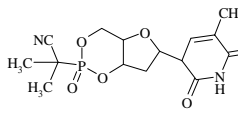
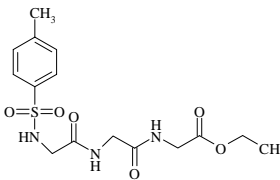
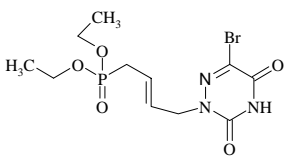
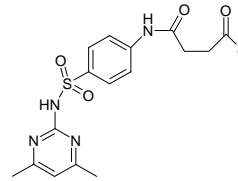
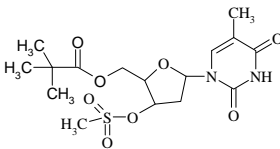
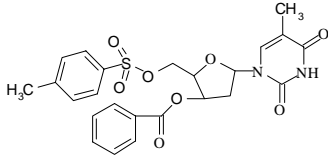
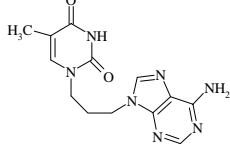
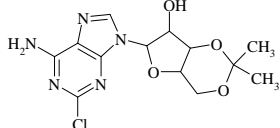
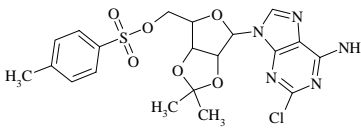
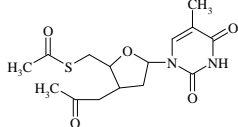
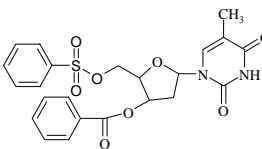
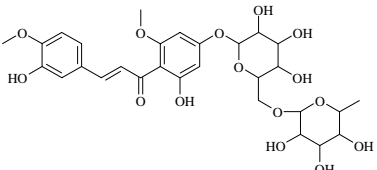
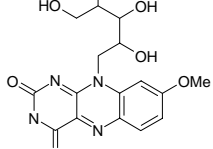
Hits	Hits	Hits
 NCI0678975	 NCI0098676	 NCI0333426
 NCI0120955	 NCI064055	 NCI0664941
 NCI0200726	 NCI0701707	 NCI0157061
 NCI0694151	 NCI0660234	 NCI0690193
 RJC-03236	 NCI0674466	 NCI0691809
 NCI0200728	 BTB-11576	 NCI0158900

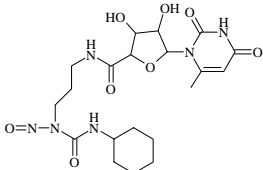
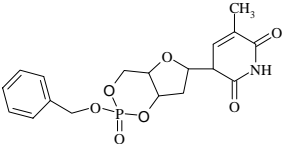
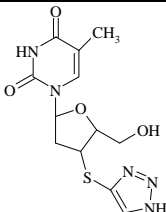
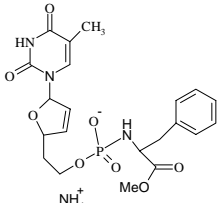
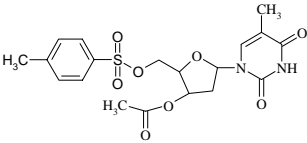
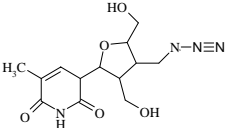
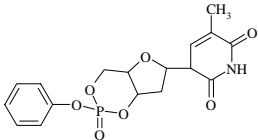
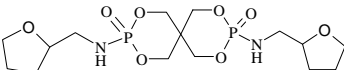
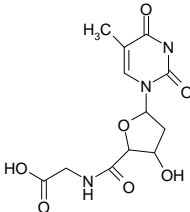
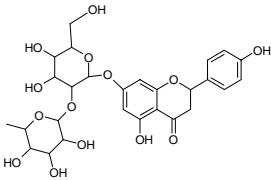
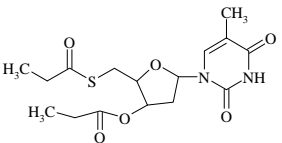
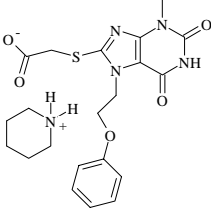
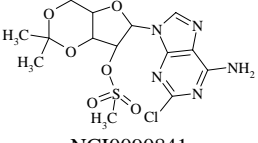
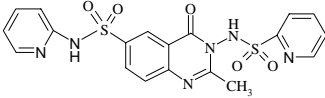
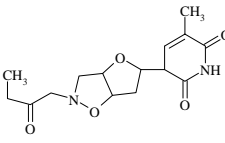
Hits	Hits	Hits
 NCI0156244	 NCI0134124	 NCI0691813
 NCI0069443	 NCI0329228	 NCI0372332
 NCI0706103	 NCI0693094	 NCI0212204
 NCI0691810	 NCI0312330	 NCI0308308
 NCI0374909	 NCI0677991	 NCI0038297
 NCI0651032	 NCI0687022	 NCI0117269

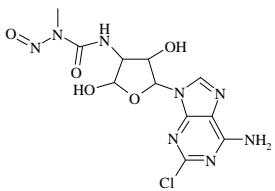
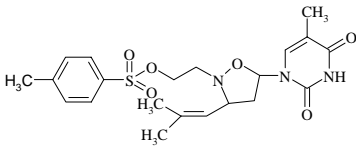
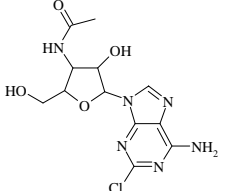
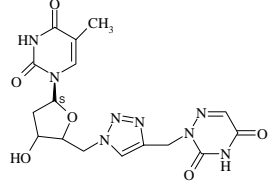
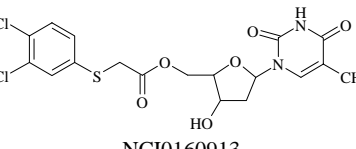
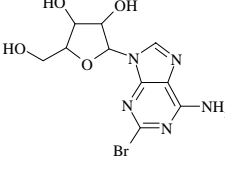
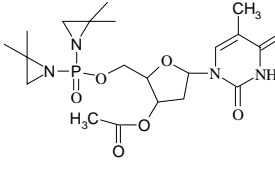
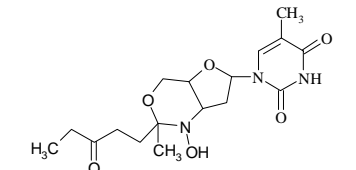
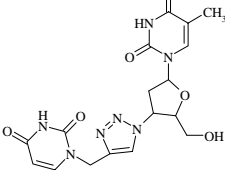
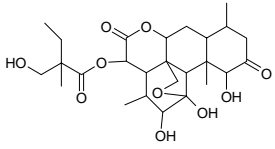
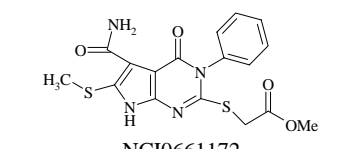
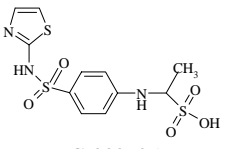
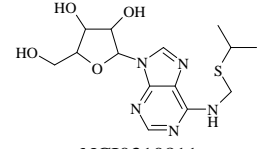
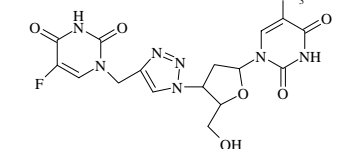
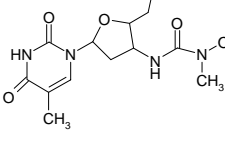
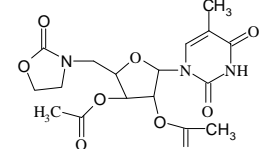
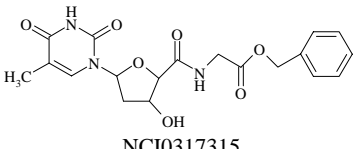
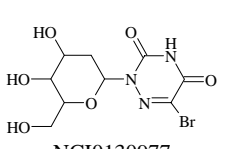
Hits	Hits	Hits
 NCI0670034	 NCI0031229	 sdmol-15263
 NCI0178027	 NCI0046713	 NCI0662430
 NCI0369883	 NCI0060891	 NCI0651031
 NCI0160910	 Compound-922	 NCI0526766
 NCI0087085	 NCI0664944	 Compound-293
 NCI0527965	 NCI0697599	 NCI0275266

Hits	Hits	Hits
 <p>NCI0002458</p>	 <p>NCI0160909</p>	 <p>NCI0112397</p>
 <p>NCI0682697</p>	 <p>NCI0668357</p>	 <p>NCI0333427</p>
 <p>BTB-14717</p>	 <p>NCI0078788</p>	 <p>NCI0023116</p>
 <p>NCI0678974</p>	 <p>NCI0685193</p>	 <p>NCI0024533</p>
 <p>NCI0694921</p>	 <p>NCI0654224</p>	 <p>NCI0003100</p>
 <p>NCI0706098</p>	 <p>NCI0343659</p>	 <p>NCI0033298</p>

Hits	Hits	Hits
 <p>NCI0200729</p>	 <p>NCI0699455</p>	 <p>NCI0656906</p>
 <p>NCI0694913</p>	 <p>NCI0699330</p>	 <p>NCI0025020</p>
 <p>NCI0659017</p>	 <p>Compound-469</p>	 <p>NCI0666990</p>
 <p>NCI0682805</p>	 <p>NCI0317317</p>	 <p>NCI0259912</p>
 <p>sdmol-15260</p>	 <p>BTBS-00069</p>	 <p>NCI0164687</p>

Hits	Hits	Hits
 <p>NCI0691815</p>	 <p>NCI0688958</p>	 <p>NCI069941</p>
 <p>CI0254066</p>	 <p>NCI0160912</p>	 <p>NCI0355706</p>
 <p>NCI0148834</p>	 <p>NCI0678987</p>	 <p>BTB-12149</p>
 <p>BTB-11575</p>	 <p>NCI0069883</p>	 <p>NCI0372333</p>
 <p>NCI0069882</p>	 <p>NCI0164686</p>	 <p>NCI0069441</p>
 <p>JFD-01835</p>	 <p>Compound-418</p>	 <p>NCI0330187</p>

Hits	Hits	Hits
 <p>NCI0319513</p>	 <p>NCI0355705</p>	 <p>NCI0649646</p>
 <p>NCI0682699</p>	 <p>BTB-14738</p>	 <p>NCI0691811</p>
 <p>NCI0157742</p>	 <p>NCI0088659</p>	 <p>NCI0315828</p>
 <p>Compound-861</p>	 <p>NCI0077489</p>	 <p>sdmol-16203</p>
 <p>NCI0090841</p>	 <p>NCI0687784</p>	 <p>NCI0667855</p>

Hits	Hits	Hits
 <p>NCI0238126</p>	 <p>NCI0697602</p>	 <p>NCI0241520</p>
 <p>NCI0691814</p>	 <p>NCI0160913</p>	 <p>NCI0086063</p>
 <p>NCI0200727</p>	 <p>NCI0688964</p>	 <p>NCI0691821</p>
 <p>NCI0305724</p>	 <p>NCI0661172</p>	 <p>NCI0027257</p>
 <p>NCI0310811</p>	 <p>NCI0691823</p>	 <p>NCI0699347</p>
 <p>NCI0661175</p>	 <p>NCI0317315</p>	 <p>NCI0130977</p>