

**I. STRONG AND WEAK HYDROGEN BONDS IN THE  
RECEPTOR–LIGAND INTERFACE  
II. MOLECULAR MODELING OF SOME THERAPEUTICALLY  
IMPORTANT TARGETS**

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

**By  
SUNIL KUMAR PANIGRAHI**



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

**February 2007**

*To  
My family and  
Teachers*

## STATEMENT

I hereby declare that the matter embodied in this thesis entitled “**Part I. STRONG AND WEAK HYDROGEN BONDS IN THE RECEPTOR–LIGAND INTERFACE** and **Part II. MOLECULAR MODELING OF SOME THERAPEUTICALLY IMPORTANT TARGETS**” 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  
February 2007

**Sunil Kumar Panigrahi**

## **CERTIFICATE**

Certified that the work “**Part I. STRONG AND WEAK HYDROGEN BONDS IN THE RECEPTOR–LIGAND INTERFACE** and **Part II. MOLECULAR MODELING OF SOME THERAPEUTICALLY IMPORTANT TARGETS**” has been carried out by **Sunil Kumar Panigrahi** 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 gives me immense pleasure to express my deep sense of gratitude and profound thanks to my supervisor, **Prof. Gautam R. Desiraju** for his inspiring guidance and constant encouragement. I have been able to learn a great deal from him through many thought provoking discussions and lectures, and consider my association with him a gratifying experience.

I would like to thank Dr. J. A. R. P. Sarma, GVK Biosciences Pvt. Ltd. and Dr. B. Gopalakrishnan, TATA Consultancy Services, Hyderabad, for their constant help and guidance at both academic and personal levels. I truly enjoyed the rewarding debates and suggestions with both of them during various stages of my Ph. D. work.

I wish to thank Prof. M. Jaskolski, Adam Mickiewicz University, Poznan, Poland for hosting me in his laboratory during April–June, 2004. He showed me the first glimpse of macromolecular crystals in a laboratory. This incident is one of the memorable events of mine during my Ph. D. tenure. My special thanks to all the members at Poznan for their concern regarding my stay there.

I thank the Dean, School of Chemistry and former Deans of the School for their co-operation in providing facilities. I extend my sincere thanks to all the faculty members of the School for their help on various occasions.

I extend my thanks to all my teachers who taught me valuable lessons from my schooling till post graduation.

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

I am thankful to BBC world service (radio), London, for constantly updating me with news and stories over the years, especially in the days of writing manuscripts and this thesis.

I wish to thank all my friendly and cooperative labmates Drs. S. Sarkhel, R. Thaimattam, R. K. R. Jetti, P. K. Thallapally, V. S. S. Kumar, S. George, Vishweshwar, V. R. Vangala, Rahul, Dinu, Basavoju, Srinivasulu, Balu, Sreenivas Reddy, Malla Reddy, Archan, Binoy and Sairam, Aparna, Tejender, Ravi, Jeevan, Saikat, Prashant, Jagadeesh, Sreekanth, Bipul, Naren, Sandip, Sanjeev, Nabakamal, Ranjit, Abhishek, and Kazu 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 thank all my friends and well wishers for making my stay in the campus a memorable one.

I thank the CSIR for fellowship support during my research tenure and DST for providing computational facility in the department and at the CMSD. I thank all the non-teaching staff of the School of Chemistry and CMSD building for helping me on various occasions.

I extend my deep sense of gratitude towards my parents and family members for constant encouragement and patience. I am beholden to my beloved wife, Shibani for her countless support and encouragement. I dedicate this thesis to my family and teachers.

**Sunil Kumar Panigrahi**

## PREFACE

The present thesis is comprised of two parts. In the first section the nature of hydrogen bonds in the receptor–ligand interface is addressed through statistical analysis carried out on X-ray structures of known receptor–ligand complexes. The second part of the thesis deals with the implementation of molecular modeling techniques to identify novel inhibitors for two therapeutically important targets *in silico*. Essentially the aim of the thesis is to understand the basic nature of receptor–ligand recognition through statistical and molecular modeling studies, which in turn will help to design better and safer drug molecules.

The hydrogen bond is a unique phenomenon in structural chemistry and biology. The critical role of hydrogen bonds is unanimously accepted ever since its inception about 90 years ago. Study on hydrogen bonding is an ever-green area of research in various spheres of chemistry and biology. Hydrogen bonds are manifested in a variety of ways. This multifaceted nature of hydrogen bonds fascinates scientists even today. In this context, the study of hydrogen bonding in the receptor–ligand interface is an active area of research. Chapter 1 gives an overview of strong and weak hydrogen bonds in biology. Chapter 2 describes a software for the analysis of hydrogen bonds in X-ray structures of macromolecules. Chapters 3, 4, and 5 respectively examine the nature of hydrogen bonds in a diverse set of protein–ligand complexes, complexes within the kinase family, and drug–DNA complexes, using this software.

The recent advancements in computational power and a better understanding of the molecular world have revolutionized present day's research in chemistry and biology. Molecular modeling and its application in drug design has reached great heights in the past few decades. The second part of the thesis deals with the optimal use of a collection of rational drug design techniques to identify potential drug candidates. Chapter 6 provides an overview of molecular modeling methods with special emphasis on pharmacophore modeling. Chapters 7 and 8 portray application of pharmacophore modeling and virtual screening in two therapeutically important targets, namely cancer and Alzheimer's disease. Chapter 9 is a brief conclusion of the work and gives some future perspectives.

The extended information related to chapters 3, 7, and 9 is presented in appendix I, II, and III respectively.

*Sunil Kumar Panigrahi*

## CONTENTS

Statement	v
Certificate	vi
Acknowledgement	vii
Preface	ix

### CHAPTER 1

<b>STRONG AND WEAK HYDROGEN BONDS IN BIOLOGY</b>	<b>1-12</b>
1.1 Introduction	1
1.2 Definition of a hydrogen bond	1
1.3 Defining strong and weak hydrogen bond	2
1.4 Classification of hydrogen bonds	2
1.4.1 Very strong hydrogen bond	3
1.4.2 Strong hydrogen bond	4
1.4.3 Weak hydrogen bonds	4
1.4.4 Other weak interactions in biology	4
1.5 Methods of studying hydrogen bonds	5
1.5.1 Crystallography	5
1.5.2 Crystallographic databases	6
1.5.3 Statistical analysis	6
1.6 Some technical glitches	7
1.7 Geometrical parameters	7
1.7.1 Distances and angles	7
1.7.2 Geometric criteria for other weak interaction	8
1.7.3 Furcation in hydrogen bond	9
1.8 The weak hydrogen bond in protein–ligand complexes	9
1.8.1 Evidence of weak hydrogen bonds in protein–ligand complexes	10

## Chapter 2

<b>HBAT: A COMPLETE PACKAGE FOR THE ANALYSIS OF STRONG AND WEAK HYDROGEN BONDS IN MACROMOLECULAR CRYSTAL STRUCTURES</b>	<b>13-20</b>
2.1 Introduction	13
2.2 Program description	14
2.2.1 Hardware and software requirement	14
2.2.2 Program availability	14
2.2.3 Input details	15
2.3 Methodology	15
2.3.1 Identification and classification of interactions	15
2.3.2 Hydrogen bonds with single atom acceptors	16
2.3.3 Hydrogen bonds with multi-atom acceptors	16
2.3.4 Halogen bonds	17
2.3.5 Donor and acceptor furcation	17
2.3.6 Cooperativity	18
2.4 Example	18
2.5 Conclusions	20

## CHAPTER 3

<b>STRONG AND WEAK HYDROGEN BONDS IN THE PROTEIN–LIGAND INTERFACE</b>	<b>21-44</b>
3.1 Introduction	21
3.2 Materials and methods	22
3.3 Results and discussion	25
3.3.1 Hydrogen bond geometry. Lengths and angles.	25
3.3.2 Hydrogen bond geometry. Furcation.	30
3.3.3 Hydrogen bond geometry. The resolution problem	33
3.3.4 Residue frequency	37
3.3.5 Interactions involving water	40
3.3.6 Lipinski's rule extended	41
3.3.7 Protein–ligand interactions in kinases	41
3.3.8 Other weak interactions	41
3.4 Conclusions	43

**CHAPTER 4****STRONG AND WEAK HYDROGEN BONDS IN PROTEIN–LIGAND COMPLEXES OF KINASES: A COMPARATIVE STUDY**

45-72

4.1	Introduction	45
4.2	Materials and Methods	46
4.2.1	Dataset	46
4.2.2	Geometry optimization	47
4.2.3	Hydrogen bond analysis	47
4.2.4	Water in the active sites of kinase	47
4.3	Results and discussion	48
4.3.1	Residue frequency	48
4.3.2	Hydrogen bond motif and Synthons	54
4.3.3	Role of conserved residue	66
4.3.4	Active site solvation	69
4.4	Conclusions	71

**CHAPTER 5****STRONG AND WEAK HYDROGEN BONDS IN DRUG–DNA COMPLEXES: A STATISTICAL ANALYSIS**

73-91

5.1	Introduction	73
5.2	Materials and methods	74
5.2.1	Drug–DNA complexes from PDB	74
5.2.2	Hydrogen bond analysis tool (HBAT)	76
5.2.3	Docking of HAT Inhibitors	76
5.3	Results and discussion	79
5.3.1	Overview of strong and weak hydrogen bond in drug–DNA complexes	79
5.3.2	Hydrogen bond analysis	79
5.3.3	Donor furcation	84
5.3.4	Hydrogen bond geometry	85
5.3.5	Human African Trypanosomiasis (Docking)	87
5.3.6	Hydrogen bonds in HAT inhibitors	89
5.4	Conclusions	90

**CHAPTER 6****INTRODUCTION TO MOLECULAR MODELING AND PHARMACOPHORE**

<b>MODELING</b>	93-103
6.1 Introduction	93
6.2 Tools for molecular modeling	93
6.2.1 Quantum mechanics	94
6.2.1.1 <i>Ab initio</i>	94
6.2.1.2 Semiempirical methods	94
6.2.1.3 Density functional theory	95
6.2.2 Molecular mechanics	95
6.2.3 QM/MM	96
6.3 Simulation tools	96
6.4 Rational drug design another face of molecular modeling	97
6.4.1 Quantitative structure activity relationship	98
6.4.2 Docking and scoring functions	98
6.4.3 <i>De Novo</i> ligand design	98
6.4.4 Virtual screening	99
6.4.5 ADMET prediction	99
6.4.6 Chemoinformatics	99
6.5 Pharmacophore modeling	100
6.5.1 Historical perspectives	100
6.5.1.1 Pharmacophore models in the early years	101
6.5.1.2 Pharmacophore modeling after the use of computers	101
6.5.2 Pharmacophore modeling methods	102
6.5.3 Application of pharmacophore modeling in rational drug design	102
6.6 As we move on to following chapters	103

**CHAPTER 7****PHARMACOPHORE MODELING ON EGFR KINASE INHIBITORS: A NOVEL**

<b>STRATEGY FOR LIGAND BASED VIRTUAL SCREENING</b>	105-126
7.1 Introduction	105
7.2 Materials and methods	106
7.2.1 Database profile	106
7.2.2 HipHop	107

7.2.3	HypoGen	108
7.2.4	LigandScout and GOLD	112
7.3	Results and discussion	113
7.3.1	HipHop	113
7.3.2	HypoGen	115
7.3.3	Model validation	117
7.3.4	Mining	120
7.3.5	Active site and structure based pharmacophore	120
7.3.6	Docking	122
7.3.7	Ligand based pharmacophore model versus structure based model	124
7.4	Conclusions	126

## **CHAPTER 8**

### **PHARMACOPHORE MODELING, DOCKING AND VIRTUAL SCREENING OF**

#### **ACETYLCHOLINESTERASE INHIBITORS** 127-149

8.1	Introduction	127
8.2	Materials and methods	128
8.2.1	Structure Building	128
8.2.2	HypoGen	129
8.2.3	Docking (GOLD)	131
8.2.4	Database screening	131
8.3	Results and discussion	133
8.3.1	HypoGen	133
8.3.2	Docking analysis	138
8.3.3	Comparative study of pharmacophore and docking	145
8.3.4	Database screening	146
8.4	Conclusions	149

## **CHAPTER 9**

### **CONCLUSIONS AND FUTURE PROSPECTS** 151-152

9.1	PART I	151
9.2	PART II	151

<b>REFERENCES AND NOTES</b>	153-180
<b>APPENDIX I</b>	181-197
<b>APPENDIX II</b>	199-208
<b>APPENDIX III</b>	209-223
<b>ABOUT THE AUTHOR</b>	225
<b>LIST OF PUBLICATIONS</b>	227

## CHAPTER 1

---

# STRONG AND WEAK HYDROGEN BONDS IN BIOLOGY

---

### 1.1 Introduction

Thousands of different molecules make up the intricate internal structures of a cell [1.1]. Each has its characteristic sequence of subunits, its unique three-dimensional structure, and the highly specific selection of binding partners in the cell [1.2–1.5]. The basic constituents of these subunits are made up of amino acids, sugar, nucleic acid and lipids [1.3, 1.4]. These components are held together through many forces out of which the hydrogen bond is the universal glue. Nevertheless the aqueous environment hosts an array of biological events for the smooth functioning of the cell. Water and hydrogen bonds are inseparable in all respects [1.6]. The typical biomolecules carry many groups that form strong and weak hydrogen bonds. The functional groups at the surfaces may be involved in weak hydrogen bonds that operate in water-biomolecule interactions and also in recognition processes and structural stabilization of the molecular peripheries [1.7].

### 1.2 Definition of a hydrogen bond

To define the hydrogen bonds to its exact term is been a long-standing problem for scientific communities ever since its discovery. Recently the International Union of Pure and Applied Chemistry (IUPAC) have formed a core group of expert to define the hydrogen bond [1.8]. The core group has recommended a modern definition of hydrogen bonds after two meetings, one at Pisa, Italy in 2005 and other at Bangalore, India in 2006. According to the modern definition “The hydrogen bond is an attractive interaction between  $X-H$  and an atom or a group of atoms  $Y$ , in the same or different molecule(s), where there is evidence of bond formation’. The most important criteria for a hydrogen bond are: (i) the H in the  $X-H$  group is more electropositive than  $X$  and (ii) the physical forces involved in hydrogen bonding should include attractive electrostatic forces, i.e. it should not be primarily dispersive forces”. Note that the acceptor will be annotated as A in the present thesis instead of  $Y$ , as recommended by the IUPAC core group. Prior to this, definitions proposed by Pauling (1939) and Pimentel, McClellan (1960) are noteworthy [1.9, 1.10]. The

recommended definition is very similar to the definition proposed by Pimentel, McClellan (1960).

### 1.3 Defining strong and weak hydrogen bond

The hydrogen bonds are manifested in a variety of strengths and geometries. In hydrogen bonds, hydrogen atoms of O–H, N–H, or S–H groups (known as hydrogen bond *donors*) interact with free electrons of *acceptor* atoms (for example, O, N, or S) [1.11–1.14]. The bonding energies of hydrogen bonds (4–40 kcal/mol) are lower than those of covalent bonds. Sometimes C–H is included in the armory of hydrogen bond donors and  $\pi$  electrons of aromatic ring are also included as the hydrogen bonds acceptors [1.15]. The bonding energies of such hydrogen bonds are  $< 4$  kcal/mol. The hydrogen bond is a group-pair interaction, with an energy limit of 4–40 kcal/mol [1.12]. The hydrogen bond involves all three atoms or groups of atoms, X, H and A. In most cases of hydrogen bonding, one of the two bonds formed by the hydrogen atom, namely X–H, is much stronger than the other, H $\cdots$ A. Accordingly hydrogen bonds like O–H $\cdots$ O, N–H $\cdots$ O, O–H $\cdots$ N and N–H $\cdots$ N may be considered to be strong while interactions like C–H $\cdots$ O, C–H $\cdots$ N, O–H $\cdots\pi$ , N–H $\cdots\pi$  and C–H $\cdots\pi$  are taken as weak. The chemical nature of the donor and acceptor species are considered while defining the strong and weak hydrogen bonds rather than on the basis of the distance between them. This is a subjective definition but it is used consistently in this thesis.

### 1.4 Classification of hydrogen bonds

The wide range of properties of hydrogen bond acceptor and donor species necessitate a classification scheme to emphasize their importance in chemistry and biology. The basis of such classification of hydrogen bonds is geometrical, energetic, thermodynamic and functional in nature. So far three separate attempts have been made independently by Jeffrey and Saenger (1991), Jeffrey (1997), and Desiraju and Steiner (1999) [1.11, 1.12, 1.14]. The properties of hydrogen bonds suggested by Desiraju and Steiner (1999) [1.14] are shown in Table 1.1. They have classified the hydrogen bonds on the basis of the nature of donor and acceptor groups into very strong, strong and weak. This is similar to a proposal made by

Jeffrey (1997) [1.12], who classify hydrogen bonds into strong, moderate and weak types. Such classification should be followed as a guideline rather than on a totalitarian basis.

**Table 1.1:** Properties of very strong, strong and weak hydrogen bonds\*.

	Very strong	Strong	Weak
Bond energy (-kcal/mol)	15–40	4–15	<4
Examples	[F...H...F] <sup>-</sup> [N...H...N] <sup>+</sup> P–OH...O=P	O–H...O=C N–H...O=C O–H...O–H	C–H...O O–H... $\pi$ Os–H...O
IR $\nu_s$ relative shift	> 25%	5–25%	< 5%
Bond lengths	H–A ~ X–H	H...A > X–H	H...A >> X–H
Lengthening of X–H	0.05–0.2 Å	0.01–0.05 Å	≤ 0.01 Å
X...A ( <i>D</i> ) range	2.2–2.5 Å	2.5–3.2 Å	3.0–4.0 Å
H...A ( <i>d</i> ) range	1.2–1.5 Å	1.5–2.2 Å	2.0–3.0 Å
Bonds shorter van der Waals cutoff	100%	almost 100%	30–80%
Angle ( $\theta$ ) range	175–180°	130–180°	90–180°
<i>kT</i> (room temperature)	> 25	7–25	< 7
Effect on crystal packing	dominant	distinctive	variable
Utility in crystal engineering useful	unknown	useful	partly
Covalency	pronounced	weak	vanishing
Electrostatics	significant	dominant	moderate

\*Adopted from Desiraju and Steiner, (1999)

### 1.4.1 Very strong hydrogen bond

Very strong hydrogen bonds are formed by unusually activated donors and acceptors, often in an intramolecular situation. Frequently, they are formed between an acid and its conjugate base,  $X-H\cdots X^-$ , or between a base and its conjugate acid,  $X^+-H\cdots X$ . These types of hydrogen bonds are often described in chemistry literature and recently in biology. The recent resurgence of literature on very strong hydrogen bonds in biology is due to the greater advancement in the understanding of enzymatic reactions [1.16–1.34]. Very strong hydrogen bonds are of great importance in the context of enzymatic reactions and are often referred to as low barrier hydrogen bonds (LBHB). Sometimes these hydrogen bonds are also referred as short, strong hydrogen bonds (SSHB). LBHB and SSHB have been postulated to play a crucial role in enzymatic reactions, particularly those that involve a general acid-general base mechanism, by providing substantial stabilization energy (10–20 kcal/mol) for the intermediate or transition state.

### 1.4.2 Strong hydrogen bond

The strong hydrogen bonds are the normal hydrogen bonds described in chemistry and biology [1.2–1.5, 1.14, 1.35]. The energy range of these types of hydrogen bonds is in between 4–15 kcal/mol. The omnipresence of strong hydrogen bonds are topic of interest ever since the structural biology came into existence in 1960s. The strong hydrogen bond is usually counted in many biological phenomena such as stabilizing three-dimensional structure of biomolecules, membrane permeability, and enzyme substrate recognition. The examples of strong hydrogen bonds are  $\text{O}-\text{H}\cdots\text{O}=\text{C}$ ,  $\text{N}-\text{H}\cdots\text{O}=\text{C}$ .

### 1.4.3 Weak hydrogen bonds

Weak hydrogen bonds in biological structures were observed as early as the 1960s. Sutor (1963) noted the existence of  $\text{C}-\text{H}\cdots\text{O}$  interactions in purine and pyrimidine bases, while it was recognized in nucleosides by Shefter and Trueblood (1965) and later by Sundaralingam (1966) [1.36–1.38]. Ramachandran *et al.* (1965, 1966) observed these hydrogen bonds in collagen and polyglycine II [1.39–1.40]. The existence of weak hydrogen bonds in biology has been strengthened by recent papers in various journals [1.41–1.74]. Weak hydrogen bonds are electrostatic but this characteristic is modified by variable dispersive and charge-transfer components that depend substantially on the nature of the donor and acceptor group. The strongest example in this category are hydrogen bonds like  $\text{C}\equiv\text{C}-\text{H}\cdots\text{O}$  (–2 to –4 kcal/mol). The methyl groups form the weakest hydrogen bond in this series (about –0.5 kcal/mol).

### 1.4.4 Other weak interactions in biology

There are a growing number of reports related to other weak interactions involving  $\pi$ -acceptors [1.75–1.84], halogen atoms (both as electrophiles and nucleophiles) [1.85–1.88], and sulfur-atoms in biology [1.89–1.92]. Short oxygen $\cdots$ halogen interactions have been known since the 1950s. A recent survey of protein and nucleic acid structures reveals similar halogen bonds as potentially stabilizing inter- and intramolecular interactions that can affect ligand binding [1.86]. The acceptor capability of organic halogen, X (X = F, Cl, Br, I), is also important in macromolecules [1.87]. Another weak interaction is the interaction involving sulfur atoms. Sulfur atoms are larger and have a more diffuse electron cloud than oxygen and nitrogen, but are nevertheless capable of participating in hydrogen bonds and are found in macromolecules. While these interactions are weak they seem to

play a definite role in many biological events. In summary, various aspects of the hydrogen bonding phenomenon observed in chemistry are being implicated in biology as well. Hydrogen bonds exist in all respects in the ever complex but still highly organized living world.

### **1.5 Methods of studying hydrogen bonds**

There exist several methods, both experimental and theoretical, to study hydrogen bonds [1.93]. The experimental techniques include X-ray diffraction study, electron diffraction of protein crystals, small-angle X-ray scattering, small-angle neutron scattering, fibre diffraction, electron microscopy, nuclear magnetic resonance and computation. Discussed here briefly are the implications of X-ray diffraction and statistical methods in studying hydrogen bonds in macromolecules in brief. Various experimental and computational techniques to study hydrogen bonds are mentioned in depth elsewhere by Jeffrey and Saenger (1991), Jeffrey (1997), Scheiner (1997), Desiraju and Steiner (1999), Rossman and Arnold (2001) [1.11–1.14, 1.93].

#### **1.5.1 Crystallography**

The rapid growth and development of X-ray and neutron crystallography have enriched biological research greatly [1.93]. However, the complexity of biomolecules has always been a challenge in crystallization and structure solution. The sophistication in crystallization techniques in crystallization methods like cryocrystallography, high-throughput crystallography, automation in the ‘direct methods’ for crystal structure solution and the subsequent refinement, extremely powerful computing facilities, improvements in quality of diffractometers in past few years have overcome many challenges. All this has simplified the overall methods of macromolecular crystallography. The process which was a Herculean task 25 years ago is now a joy among chemists and biologists practising crystallography.

When talking about the detection of the exact location of hydrogen atoms in a molecule, the neutron diffraction method comes to mind [1.94–1.96]. Neutron diffraction technique which was confined to small molecular crystallography earlier is now being used in macromolecular crystallography. Knowing the exact location of hydrogen atoms removes many ambiguities connected with the existence of hydrogen bonds. This is because from the

first visual inspection, chemists/biologists can decipher the important geometric criteria of a putative hydrogen bond.

### **1.5.2 Crystallographic databases**

The results of crystal structure determinations are stored in various file formats. The structural information in MMCIF and PDB file formats are routinely saved [1.97]. This enables the scientists to analyze and visualize the biomolecules in computers using sophisticated software. This helps in the proper understanding of biological intricacies through computer graphics and modeling. The ever increasing number of structures solved through experimental techniques necessitates a proper storage and management system. These requirements are fulfilled by crystallographic databases. Crystallographic databases are very useful resources to study hydrogen bonds in the molecular world. Like the Cambridge Structural Database (CSD) [1.98] for small-molecules, the macromolecular database provides a wealth of information for large biological molecules. The databases related to macromolecules are, (a) The Protein Data Bank at Brookhaven (PDB) [1.99], (b) Nucleic Acid Database (NDB), (c) The Biological Macromolecule Crystallization Database (BMCD) [1.93]. The progress in these databases has later given birth to a second generation of databases e.g. SCOP [1.100], CATH [1.101], PDBSUM [1.102], specialized with respect to the structural and functional aspects of biomolecules.

### **1.5.3 Statistical analysis**

Statistics plays a major role in scientific research, especially when the dataset is large [1.103, 1.104]. Provided with a large sample of crystallographic data it is possible to derive meaningful conclusions about the nature and behavior of crystals. With this backdrop, hydrogen bond research has been constantly enriched, with a parallel in the increase in the number of deposits in these databases. Statistical methods play a major role in the study of weak hydrogen bonds. This is because a large sampling of structures is needed to estimate quantitative information on weak interactions. One such problem is the analysis of hydrogen bond geometries in receptor–ligand interaction. Chapters 3, 4, and 5 of the thesis deals with the application of statistical methods, to study the nature of hydrogen bond geometries in X-ray structures of receptor–ligand complexes.

Most of the statistical studies are carried out with the help of specialized computer software. There exist many programs for hydrogen bond analysis like HBPLUS [1.105],

HBEXPLORE [1.106], BNDLST from Richardson lab [1.107], CONTACT from CCP4 [1.108] and web based servers like LPC [1.109] and NCI [1.110] in the public domain. However “pros and cons” lie in each of these softwares while performing studies of hydrogen bonds in macromolecules. This is because of the huge numbers of atoms and the variety of donor/acceptor functional groups. The next chapter describes a software called hydrogen bond analysis tool (HBAT) to study strong and weak hydrogen bonds in a PDB file [1.111]. Thus HBAT is a new generation software for the convenient study of strong and weak hydrogen bonds in macromolecules.

## **1.6 Some technical glitches**

Despite advances in macromolecular crystallography there still lie some problems in the quality of data it produces [1.14, 1.93]. They are: (a) location of hydrogen atom in macromolecules [1.112], (b) macromolecular crystal is not time stable, (c) the crystallographic resolution problem and (d) unavoidable errors during data collection and refinement.

## **1.7 Geometrical parameters**

In this section, the geometrical characterization of hydrogen bonds is discussed. This is because the first section of the thesis deals with the hydrogen bond geometry in receptor–ligand complexes. The energetic implications of hydrogen bond can be referred elsewhere in the book of Jeffrey and Saenger (1991), Jeffrey (1997), Scheiner (1997), and Desiraju and Steiner (1999) [1.11–1.14].

### **1.7.1 Distances and angles**

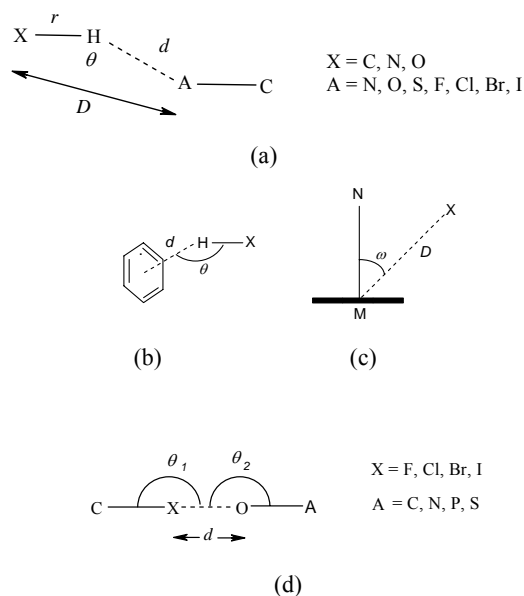
The directional property of the hydrogen bond makes it a unique non-covalent interaction. This directional property is best represented by the hydrogen bond distance and angle [1.11–1.14]. The hydrogen bond is constituted with a donor X–H and an acceptor A, and referred as X–H $\cdots$ A in the present thesis. Scheme 1.1*a* represents a typical hydrogen bond. Various annotations in the scheme are,  $d$  (H $\cdots$ A) the distance between hydrogen atom and the pertinent acceptor,  $D$  heavy atom distance,  $\theta$  (angle X–H $\cdots$ A),  $r$  is X–H covalent distance. These three parameters are important in identifying hydrogen bonds. However,

another parameter angle  $\phi$ , (angle  $\text{H}\cdots\text{A}-\text{C}$ ) is also taken into account to provide a strict geometric criterion for the hydrogen bond. In crystal structures of macromolecules, the H-atoms are usually not defined. In those cases, sometimes the distance between the heavy atoms ( $D$ ) is assumed to be a criterion for putative hydrogen bond, which is a crude way to identify hydrogen bonds. However, there is no reasonable alternative in such cases.

### 1.7.2 Geometric criteria for other weak interaction

The geometrical criterion for a multi atom acceptor like phenyl rings is difficult to derive. The usual practice is the measured distances to the centroid of phenyl ring (M) Scheme 1.1b and c. However  $d \leq 3.5 \text{ \AA}$ ,  $\theta \geq 100^\circ$  and  $\omega \leq 40^\circ$  appear to be satisfactory and this geometric criterion is generally accepted [1.15, 1.75].

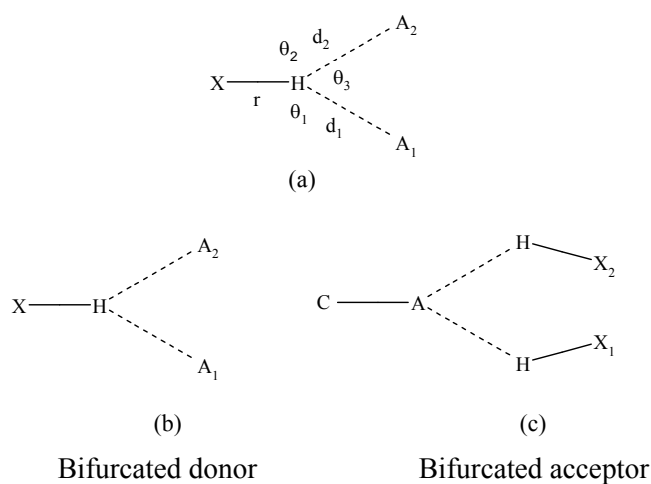
Organic halogens are included in the list of putative donors as well as acceptors (both as electrophiles and nucleophiles). The convention for a halogen bond is retained as  $\text{C}-\text{X}\cdots\text{O}$  (Scheme 1.1d). A detailed description on the geometry of halogen is described by Auffinger *et al.* [1.86]. The geometry described for halogen as acceptors is in the form of  $\text{X}-\text{H}\cdots\text{halogen}$  (here  $\text{X} = \text{O}, \text{N}, \text{C}$ ). The convention for conventional hydrogen bonds can be applied to  $\text{X}-\text{H}\cdots\text{halogen}$  interactions, considering the van der Waals criterion. A similarly criteria can be adopted for sulfur acceptors.



**Scheme 1.1:** (a) Representative hydrogen bond.  $\text{A}-\text{C}$  is a single or double bond, (b), (c) Parameters for  $\text{X}-\text{H}\cdots\pi$  interactions (d) Parameters for halogen $\cdots\text{O}$  interactions.  $\text{O}-\text{A}$  is a double bond.

### 1.7.3 Furcation in hydrogen bond

The hydrogen bond furcation is a situation where a hydrogen atom is shared between more than one acceptor [1.12]. This situation is termed donor furcation. Terms such as bifurcated or ‘3-centered’ and trifurcated or ‘4-centered’ are also used (Scheme 1.2). Analogously, the converse situation where several donors approach a single acceptor is known as acceptor furcation. The conventions for bifurcated hydrogen bonds are presented by the distances  $r$ ,  $d_1$ ,  $d_2$  and the angles  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ .



**Scheme 1.2:** (a) Annotations in a bifurcated hydrogen bond, (b) and (c) represents bifurcated donor and acceptor respectively.

## 1.8 The weak hydrogen bond in protein–ligand complexes

Hydrogen bonds are instrumental in many molecular recognition events in biology [1.113, 1.114]. Very often, the recognition events are specific with conserved orientation. This reminds one of the exact events followed in a relay race, where the baton is exchanged among the fellow players with great precision and in a time-bound manner. Similar is the process of signal transduction and receptor–ligand interactions. The relaying of message becomes easy due to the weak nature of hydrogen bonds. In this context, the hydrogen bond in the molecular recognition event in the receptor–ligand interface is an attractive research proposition. Further, large numbers of receptor–ligand complexes are being deposited in PDB due to various structural initiatives, and this is an essential prerequisite to statistical studies. Apart from the final recognition events, hydrogen bonds are also responsible for physico-chemical properties of a molecule, like solubility, partitioning, distribution, and

permeability, which are crucial to drug development. Thus hydrogen bond research in drug discovery is been an evergreen area of research.

The topic of hydrogen bonds never limits to a particular type of donor and acceptor, rather refers to a variety of donor acceptor association. Unlike other biological events the protein–ligand interactions involves a plethora of strong and weak hydrogen bond. In fact the active sites manifest almost all the hydrogen bond interaction so far discussed. In particular the reversibility in receptor–ligand interactions is often governed by the weak nature of hydrogen bonds. This is because weaker hydrogen bonds are also easier to break. Presented here is an overview of selected studies describing the importance of weak hydrogen bonds in protein–ligand interactions and drug design.

### 1.8.1 Evidence of weak hydrogen bonds in protein–ligand complexes

The functional importance of C–H $\cdots$ O hydrogen bonds at the ligand binding domain of the retinoic acid receptor RAR $\gamma$  was first discussed by Klaholz and Moras (2002) [1.115]. The ligand binding domain of the retinoic acid receptor RAR $\gamma$  complexed with the retinoid SR11254 was revealed with the help of a 1.4 Å resolution crystal structure. They observed that the hydroxyl group of the ligand acts as a hydrogen bond donor and acceptor, leading to a synergy between the strong O–H $\cdots$ O and weak C–H $\cdots$ O hydrogen bonds. This in turn influences the specificity and affinity of the ligand is within the hydrophobic pocket of the receptor.

With the aim of designing possible kinase inhibitors Pierce *et al.* (2002) [1.116] carried out a statistical and theoretical analysis of C–H donors for aromatic ligands in a data set of 184 kinase crystal structures and 358 high-resolution structures from the PDB. The donor capacity of a variety of C–H donor in the protein–ligand complexes was analyzed. The activated C–H groups adjacent to the positively charged nitrogen of nicotinamide exhibit geometric preferences strongly suggested the hydrogen bond. This was also observed in the heterocyclic C–H groups in kinase ligands. Other aromatic C–H groups did not show such characteristics. *Ab initio* calculations by HF/6-31G\*\* level revealed a considerable range of C–H $\cdots$ O hydrogen bonding potential among the different aromatic ring systems, with nicotinamide and heterocycles preferred in kinase inhibitors. These workers showed in particular many favorable hydrogen bond interactions. Further, in a subsequent work they showed that C–H $\cdots$ O hydrogen bonds play an important role in the binding of several analogues of a pyrazol-3-ylquinazolin-4-ylamine inhibitor to the

glycogen synthase kinase 3 (GSK3) [1.117]. Understanding of these C-H $\cdots$ O and C-H $\cdots$ N hydrogen bonds led them to design a novel quinazolin-4-ylthiazol-2-ylamine inhibitor of GSK3.

Denessiouk and Johnson (2003) [1.118] showed that the molecular recognition events for adenine (and some other nucleotides and nucleotide analogs) occur typically through three key hydrogen bonds. These three hydrogen bonds consists of two conventional (N-H $\cdots$ O and N-H $\cdots$ N) and a weak C-H $\cdots$ O hydrogen bond. Notably the weak C-H $\cdots$ O hydrogen bond was found to be an integral part of the adenine protein interaction. Adenosine as a ligand has a unique donor-acceptor-donor (DAD) pattern and the pertinent motif in the proteins is an acceptor-donor-acceptor (ADA). In fact the work carried out by Pierce *et al.* (2002, 2005) [1.116, 1.117] and Denessiouk and Johnson (2003) [1.118] has prompted me to pursue further study in this direction. This has been dealt in Chapter 4, which deals with a qualitative analysis of the interplay between strong and weak hydrogen bonds in the protein–ligand complexes of kinases. Kinases are a family of proteins having ATP as their natural substrate.

A statistical analysis of N-H $\cdots$ O, O-H $\cdots$ O, and C-H $\cdots$ O hydrogen bonds was carried out by Sarkhel and Desiraju (2004) [1.119] in a group of 28 high-resolution crystal structures of protein–ligand complexes from the PDB. The geometries obtained were compared *vis-à-vis* with the interactions found in small-molecule crystal structures from the Cambridge Structural Database (CSD). Some of the important conclusions derived from this study are, (a) both strong and weak hydrogen bonds are involved in ligand binding, (b) the restrictive geometrical criteria set up for hydrogen bonds in small molecule crystal structures may need to be relaxed in macromolecular structures due to extensive multifurcation, (c) the formation of C-H $\cdots$ O hydrogen bonds is enhanced by the activation of the C $_{\alpha}$ -H atoms and by the flexibility of the side chain atoms, (d) in contrast to small-molecule structures, anti-cooperative geometries were found to be common in the 28 macromolecular structures, (e) there is a gradual lengthening of hydrogen bond as the extent of furcation increases, (f) the C-H $\cdots$ O bonds formed by Gly, Phe, and Tyr residues are noteworthy, (g) number of hydrogen bond donors and acceptors agree with Lipinski's rule-of-five that predicts drug-like properties, (h) hydrogen bonds formed by water were also seen to be relevant in ligand binding and ligand C-H $\cdots$ O<sub>w</sub> interactions are abundant when compared to N-H $\cdots$ O<sub>w</sub> and O-H $\cdots$ O<sub>w</sub>. This study has initiated my interest to pursue my study on strong and weak hydrogen bonds in receptor-ligand complexes. Chapter 3 of the

thesis is essentially the extension of the work carried out by Sarkhel and Desiraju (2004) [1.119], which describes a holistic view of strong and weak hydrogen bonds in a diverse set of protein–ligand complexes.

Cashin *et al.* (2005) [1.120] carried out a mutation study to elucidate the key interaction between three agonist acetylcholine, nicotine, epibatidine and nicotinic acetylcholine receptor. They suggested that the cation- $\pi$  interaction is important in acetylcholine and main chain hydrogen bond is instrumental in nicotine binding to the receptor. However in the case of epibatidine both cation- $\pi$  and main chain hydrogen bonds determine ligand binding. Further they augmented on the basis of theoretical calculations that the binding of agonist epibatidine to the receptor is strengthened by aromatic C–H $\cdots$ O hydrogen bonds.

The importance of strong and weak hydrogen bonds in protein–ligand docking is shown through molecular modeling study [1.121]. This was illustrated through a virtual screening (VS) study, targeted against epidermal growth factor receptor (EGFR) kinase domain. Acceptable results were obtained when the outputs from the commercial software packages were analyzed and modified on the basis of a chemical model that incorporates specific hydrogen bonds. We have shown that for 4-anilinoquinazoline type ligands, inclusion of a hydrogen bonded water molecule was indispensable to obtain meaningful VS results. Consideration of protein–ligand hydrogen bonds of the N–H $\cdots$ N, O<sub>w</sub>–H $\cdots$ N and above all the C–H $\cdots$ O type was important to obtain accurate poses and binding affinities in the study.

In summary, the active site of the receptor is a unique environment where macromolecules and small molecules leave their footprints through a variety of strong and weak hydrogen bonds. The better understanding of these interactions in turn will help medicinal chemists and structural biologist to design better and safer drugs in the future.

## CHAPTER 2

---

---

# ***HBAT: A COMPLETE PACKAGE FOR THE ANALYSIS OF STRONG AND WEAK HYDROGEN BONDS IN MACROMOLECULAR CRYSTAL STRUCTURES***

---

---

### 2.1 Introduction

Recent progress in structural biology with advanced techniques like high-throughput crystallography, cryogenic cooling for crystallization and data collection with synchrotron radiation has enabled researchers to determine large numbers of crystal structures of proteins at atomic resolution. The number of structures deposited in the Protein Data Bank (PDB) at or near atomic resolution is around 241 [2.1]. The availability of such high quality structural data has meant that it is possible to accurately characterize the geometrical properties of hydrogen bonds. The hydrogen bond is an attractive electrostatic interaction of the type  $X-H\cdots A$ , where the H atom is attached to a donor atom, X (assumed to be more electronegative than H), and is directed towards an acceptor, A [2.2]. The acceptor A is normally an electronegative atom, usually O or N. Although most hydrogen bonds in proteins and nucleic acids are of the  $N-H\cdots O$  or  $O-H\cdots O$  (less often,  $N-H\cdots N$ ) type, weak hydrogen bonds such as  $C-H\cdots O$ ,  $N-H\cdots S$ ,  $O-H\cdots S$  are also known sometimes to play significant roles in biological processes. Likewise, the  $\pi$ -electron clouds of aromatic rings can also act as acceptors for appropriately oriented  $X-H$  groups [2.3]. In general, there is a one to one correspondence between a hydrogen bond donor and acceptor. There however, also exist several examples where a hydrogen atom is shared between more than one acceptor. This situation is termed donor furcation. Terms such as bifurcated or ‘*3-centered*’ and trifurcated or ‘*4-centered*’ are also used. Analogously, the converse situation where several donors approach a single acceptor is known as acceptor furcation [2.2]. Cooperativity is another important stabilization feature in hydrogen bond arrangements and it occurs whenever the formation of a hydrogen bond facilitates the formation of another, so as to lead to sequences like  $O-H\cdots O-H\cdots O-H$  and so on. Cooperativity is well exemplified in the three-dimensional networks formed by water [2.4]. Halogen atoms are found to act as both electrophiles and nucleophiles interacting with electronegative O, N and  $\pi$  systems [2.4]. In summary, there is a plethora of strong and weak interactions observed in

macromolecules. Analyzing these interactions manually in any macromolecular structure is not practicable and the analysis of interactions in groups of crystal structures rapidly becomes impossible.

Hydrogen Bond Analysis Tool (HBAT) is an attempt to automate the analysis of all non bonded interaction present in a PDB file. There exist many programs like HBPLUS [2.6], HBEXPLORE [2.7], CONTACT [2.8], BNDLST and many web based servers like LPC [2.9], NCI [2.10], which are available in the public domain for analyzing these interactions. The advantage of HBAT over the other available software is its compactness to deliver useful information pertaining to all types of interactions in a more user-friendly way. The program does not require additional information to deal with ligand bound PDB structures. The result can be easily exported to a MSOFFICE Excel file for further analysis. The program finds a wide range of applications from crystallography to structure based drug design and molecular dynamics simulations.

## **2.2 Program description**

HBAT is written using PERL language and the Graphical User Interface (GUI) was built using the Tk library [2.11]. The phenomenon of cooperativity is addressed using a graph theory approach and the visualization of the cooperativity graph is obtained using the open source Graphviz software which is included in HBAT [2.12].

### **2.2.1 Hardware and software requirement**

HBAT can be easily installed and run on any personal computer. Operating systems: Windows 98/2002/Me/NT/XP with minimal memory requirement of 512MB RAM.

### **2.2.2 Program availability**

The program is available free of charge for academic users in non-profit organization after the license agreement has been signed and faxed. More about the program can be found at HBAT home page (<http://202.41.85.161/~grd/HBAT.html>).

### 2.2.3 Input details

HBAT requires structural information in the form of a PDB format. The crystal structure should be free from major errors, i.e. it should not contain any missing residues or atoms. The input structure should contain H-atoms for a full analysis of the interaction. A quick force field based minimization of the hydrogen atoms before using HBAT is recommended in order to get more reliable results. The program is not responsible for accurate generation of H-atom positions.

## 2.3 Methodology

Initially the nearest neighbor list for each atom is generated with a default search radius (5.0 Å) or any user defined cut-off value. The nearest neighbor list is further screened for hydrogen bonds and van der Waals interactions based upon the sum of the van der Waals radii of the interacting atoms. The putative hydrogen bonds are identified using default or user defined cut-off values for distance and angles. A separate cut-off parameter is defined to search for X-H $\cdots$  $\pi$  bonds, halogen bonds, and cooperative interactions in proteins.

HBAT gives a user-friendly, click-and-use facility to calculate matrices for hydrogen bonds from a PDB file. It allows users to easily define geometric parameters for all types of interactions based upon individual user requirements. It also gives a quick summary of various types of interactions in the form of pie charts and histograms so that the user can easily visualize and compare the results.

### 2.3.1 Identification and classification of interactions

Hydrogen bonds are tabulated more informatively according to the type of hydrogen bond, donor atom, acceptor atom, donor residue name/number, acceptor residue name/number, distances  $r$ ,  $d$  and  $D$ , and hydrogen bond angle  $\theta$ .

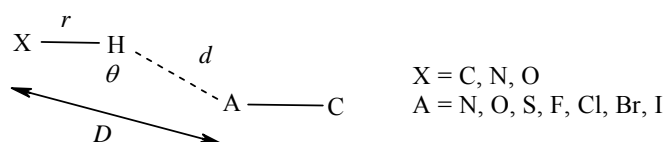
All the hydrogen bonds are further subjected to statistical analysis by classifying them in terms of backbone donor/acceptor, side chain donor/acceptor, ligand donor/acceptor, water donor/acceptors, and nucleic acid donor/acceptors. The respective frequencies of their occurrence are tabulated for all types of hydrogen bonds. Further, a detailed frequency distribution for each type of bond is listed in a 10° degree interval starting at  $90^\circ < \theta < 180^\circ$ . Similarly the distance  $d$ , distribution is specified from 1.2 Å to 3.2 Å with an interval of 0.4 Å. In a macromolecular structure the interacting partners are

protein, ligand, water and sometimes nucleic acid. Along with distance versus angle distributions, are also listed the frequencies with which particular amino acid residues, water, ligands, or nucleic acids residues are involved in various type of hydrogen bond (e. g.  $\text{N-H}\cdots\text{O}$ ,  $\text{C-H}\cdots\text{O}$  etc.). Accordingly the user can easily identify the number of contributions from each donor and/or acceptor. For interactions involving  $\pi$ -acceptors and halogen bonds only a list of interactions is given because the number of interactions is rather small.

The following brief description of the interaction types is meant to define the various geometrical parameters used in the program.

### 2.3.2 Hydrogen bonds with single atom acceptors

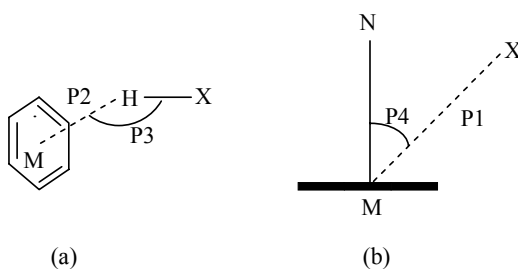
The general hydrogen bond of this type may be described in terms of the  $d$ ,  $D$ ,  $\theta$  and  $r$  values as shown in Scheme 2.1.



**Scheme 2.1:** Representative hydrogen bond.

### 2.3.3 Hydrogen bonds with multi-atom acceptors

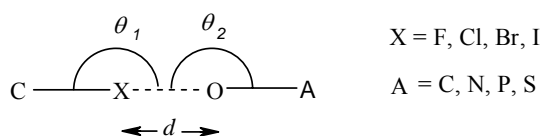
In a  $\pi$ -electron acceptor, the distances are usually measured to the centroid (M) of multiple bonds or phenyl rings. The unit vector is defined by MN. An aromatic ring is represented as a hexagon. In the case of Trp, the five-membered and six-membered ring systems are considered separately. The donor group is represented as  $\text{X-H}$ , where X is similar to the situation described above for single acceptor hydrogen bonds. For a halogen “ $\pi$ ” interaction, the H atom may be replaced by a halogen (F, Cl, Br, I). P1 and P2 are distances from X and H respectively to M. P3 is the angle between vectors  $\text{X-H}$  and  $\text{H-M}$  while P4 is the angle between the XM and MN. The geometry is adapted from earlier work of Babu, 2003 [2.10]. Scheme 2.2.



**Scheme 2.2:** Parameters for X-H $\cdots\pi$  interactions.

### 2.3.4 Halogen bonds

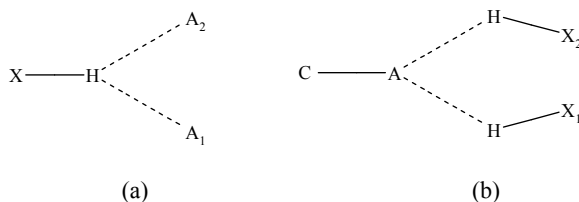
Organic halogens are included in the list of putative donors as well as acceptors (both as electrophiles and nucleophiles). The convention for a halogen bond is retained as C-X $\cdots$ O, Scheme 2.3 [2.5]. The geometry adopted for halogen as acceptors is as in the case of hydrogen bonds.



**Scheme 2.3:** Parameters for halogen $\cdots$ O interactions

### 2.3.5 Donor and acceptor furcation

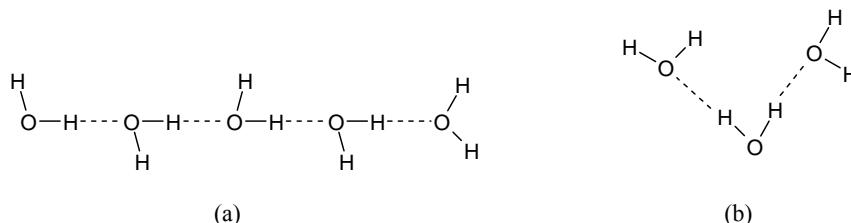
Donor and acceptor '*n-centered*' furcations are listed separately. The list contains '*2-centered*' (not furcated) interaction to '*7-centered*' (six furcated) interactions. A schematic representation of *3-centered* (bifurcated) donor/acceptor furcation is represented in Scheme 2.4.



**Scheme 2.4:** (a) Bifurcated donor (*3-centered*), (b) Bifurcated acceptor (*3-centered*)

### 2.3.6 Cooperativity

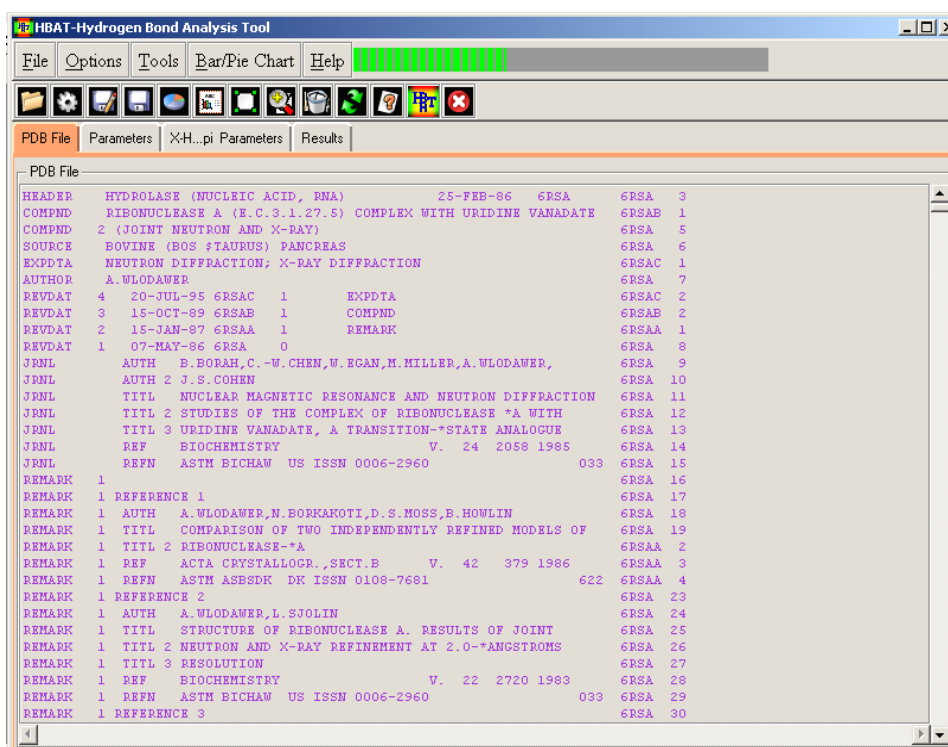
Cooperative and anti-cooperative geometries for water molecules are shown in Scheme 2.5a and b respectively. However similar arrangements can be observed in a system involving protein, ligand and water.



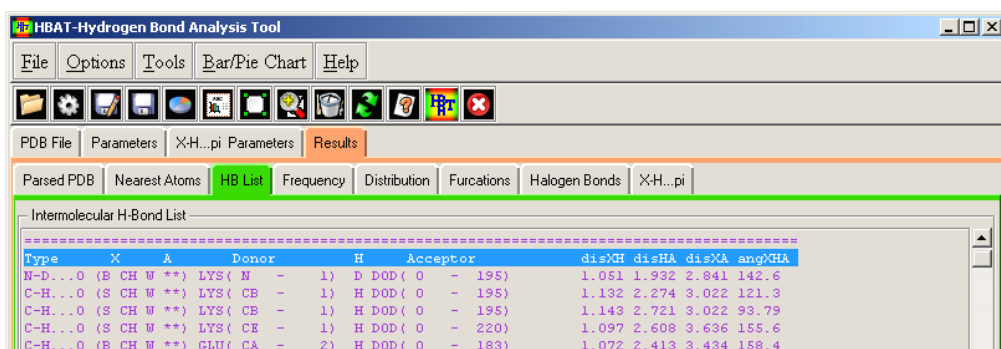
**Scheme 2.5:** Cooperative and anticooperative geometries.

### 2.4 Example

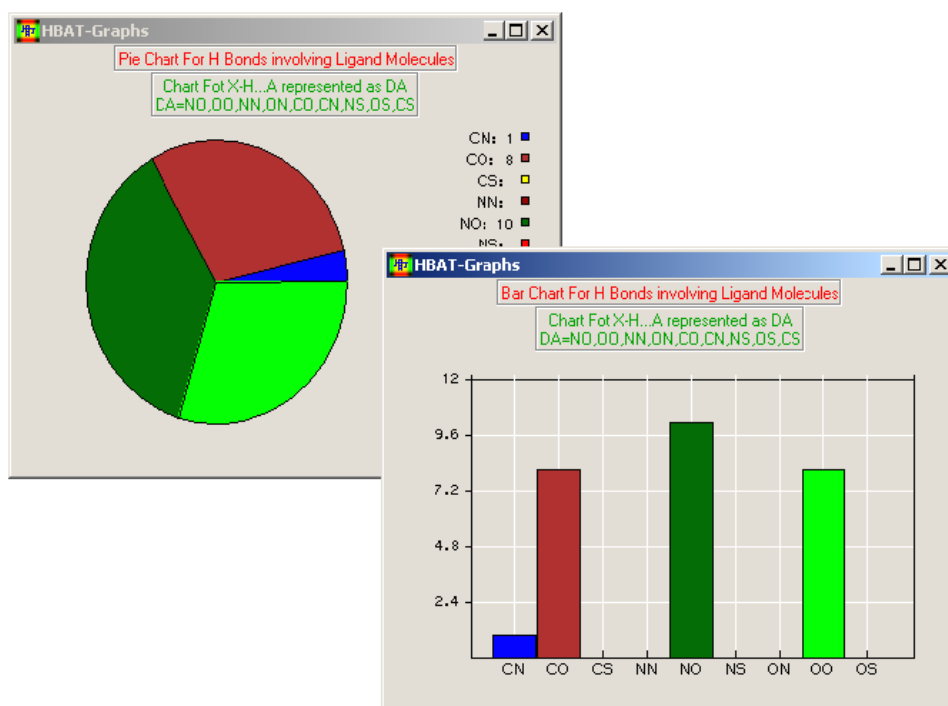
To evaluate the accuracy of the program, HBAT was used to reproduce the published geometries in several recent papers [2.12–2.16]. Some of the utilities of the program are listed in the following section taking the example of PDB ID 6RSA.



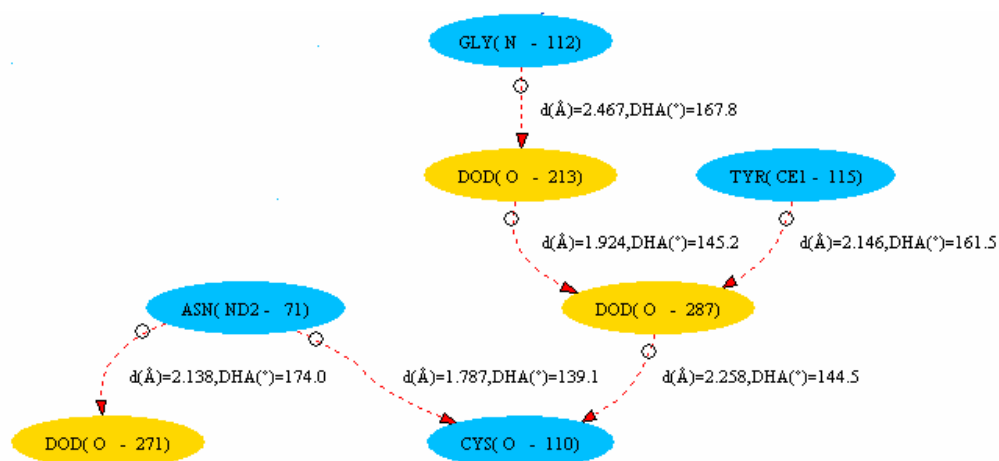
**Figure 2.1:** Input PDB file and the content of the HBAT input window. Also shown is the main GUI in the process of calculating various interactions.



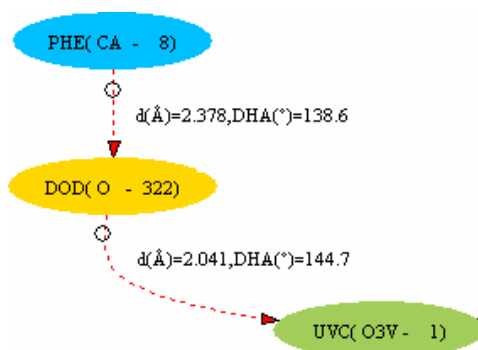
**Figure 2.2:** Section of result window. The active sub page is highlighted in green listing the segment of observed interactions. Other inactive sub pages contain results like frequency, geometry distribution, furcation, cooperativity, halogen bond, interactions involving  $\pi$ -acceptors.



**Figure 2.3:** Various types of hydrogen bond interaction between protein and ligand are presented through pie chart and histogram depictions.



**Figure 2.4:** Cooperative interactions in 6RSA. Amino acid and water are represented in blue and yellow colored boxes respectively. Two cooperative chains are shown: (1) consists of the sequence Gly 112, Water 213, Water 287, Cys 110, (2) Thr 115, Water 287, Cys 110.



**Figure 2.5:** Water mediated protein–ligand cooperative interactions. The ligand is colored green.

## 2.5 Conclusions

Described here is a new software that allows easy analysis of a variety of hydrogen bonds and other interactions in macromolecular crystal structures.

## CHAPTER 3

---

# STRONG AND WEAK HYDROGEN BONDS IN THE PROTEIN–LIGAND INTERFACE

---

### 3.1 Introduction

The three-dimensional architecture of proteins is stabilized to a substantial degree by hydrogen bonds. Because of their strength these interactions are specific, with conserved orientation [3.1, 3.2]. Because of their weakness, however, they are also made and broken rapidly during complexation, conformational change and folding [3.3]. Accordingly, hydrogen bonds in biomolecules may be switched on or off with energies that are within the range of thermal fluctuations. This is one of the prime factors that facilitate ligand binding in the active site, and biological activity. Effectively, the dual strong/weak nature of hydrogen bonds is exploited by Nature to achieve specificity of both structure and function. The importance of weak interactions also varies with the type of biomolecules in the same way that molecules of structural importance might differ from molecules of enzymatic importance [3.4, 3.5].

The literature on hydrogen bonding in biomolecules is voluminous. A seminal review by Hubbard and Baker in 1984 was followed in 1991 by the book of Jeffrey and Saenger which provides much valuable information [3.6, 3.7]. The subject was reviewed in depth by Glusker in 1995 [3.8]. Work by Sundaralingam on nucleic acids and Derewenda on globular proteins in the mid to late 1990's widened the scope of this field [3.9–3.11]. Since 2000, there have been a number of papers that have attempted to analyze the systematics of hydrogen bonds in biological structures [3.12–3.18].

Previously the characteristics of strong ( $\text{N-H}\cdots\text{O}$ ,  $\text{O-H}\cdots\text{O}$ ) and weak ( $\text{C-H}\cdots\text{O}$ ) hydrogen bonds in a group of 28 high resolution crystal structures of protein–ligand complexes [3.19] have been examined from the Protein Data Bank (PDB) [3.20]. These interactions were compared with the interactions found in small molecule crystal structures from the Cambridge Structural Database (CSD) [3.21]. Some of the important conclusions derived in this study are: (a) both strong and weak hydrogen bonds are involved in ligand binding, (b) the restrictive geometrical criteria set up for hydrogen bonds in small molecule crystal structures may need to be relaxed in macromolecular structures due to extensive

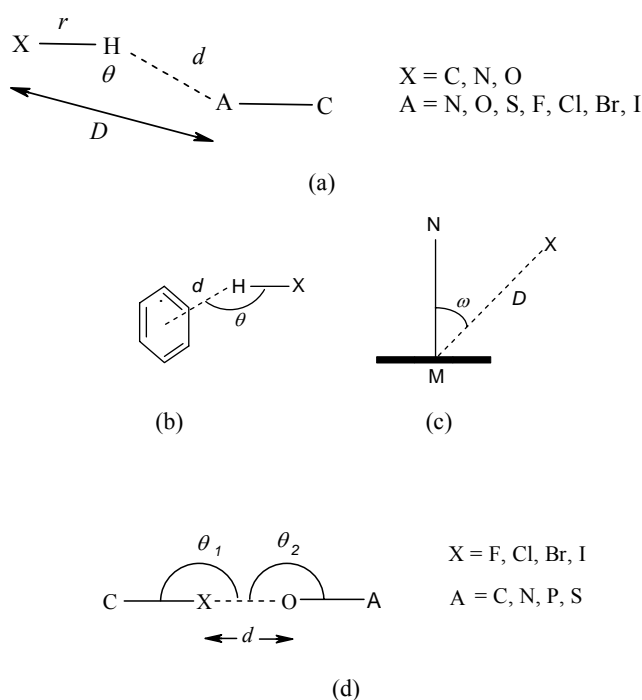
multifurcation, (c) the formation of C–H $\cdots$ O hydrogen bonds is enhanced by the activation of the C $_{\alpha}$ –H atoms and by the flexibility of the side chain atoms, (d) in contrast to small-molecule structures, anti-cooperative geometries were found to be common in the 28 macromolecular structures, (e) there is a gradual lengthening of hydrogen bond as the extent of furcation increases, (f) the C–H $\cdots$ O bonds formed by Gly, Phe, and Tyr residues are noteworthy, (g) number of hydrogen bond donors and acceptors agree with Lipinski's rule-of-five that predicts drug-like properties, (h) hydrogen bonds formed by water were also seen to be relevant in ligand binding and ligand C–H $\cdots$ O<sub>w</sub> interactions are abundant when compared to N–H $\cdots$ O<sub>w</sub> and O–H $\cdots$ O<sub>w</sub>.

Among the limitations of the earlier study [3.19] was the fact that the number of crystal structures examined was relatively small, only 28. Furthermore, several proteins in this group of structures are homologous to each other and so the number of truly independent observations is even smaller. Therefore it was felt that the conclusions of the previous study should be re-evaluated with a larger set of structures. This would necessitate the use of a computer program to evaluate hydrogen bond geometries (the smaller set of 28 crystal structures was analyzed manually). In the present work, a large dataset of 251 protein–ligand complexes, from the PDB, was analyzed with respect to intermolecular interactions with a new in-house computer program, Hydrogen Bond Analysis Tool (HBAT), which was discussed in the previous chapter. The analysis was further extended to an external test set of 233 X-ray crystal structures of protein–ligand complexes in the kinase family. As an extension of the previous study [3.19], other new interactions involving halogen atoms (both as electrophiles and nucleophiles),  $\pi$ -acceptors and sulfur-atom acceptors are examined. Eventually, it is hoped to evaluate the extent to which these weak interactions (hydrogen bonds and others) influence the protein–ligand interface in terms of behavior and function. The initial aim in this chapter is to document these interactions reliably in a representative group of protein–ligand crystal structures.

### 3.2 Materials and methods

A set of 251 X-ray crystal structures of protein–ligand complexes from the PDB was used in this study. The dataset comprises 27 structures from the earlier study by Sarkhel and Desiraju [3.19] and 224 from Nissink *et al.* [3.22]. Because of the wrong assignment of ligand, PDB ID 1G2Y.pdb from the previous study [3.19] is not included in the present

study. The external kinase test set of 233 protein–ligand complexes was taken from the PDB (Appendix I, Table 8). This was done to assess the general applicability of important conclusions derived from the analysis of the 251 X-ray crystal structures. In general, the available macromolecular crystallographic data are prone to two types of errors (1) systematic errors caused by biases during the structure determination and refinement procedure, and (2) random errors which affect the precision of the model. Additionally, the quality of the structure varies in different regions, due to higher local conformational and thermal disorder in certain parts [3.23]. In this respect, the present dataset is free from any major abnormality. The active site was defined by selecting amino acid residues within a 10 Å radius of the ligand molecule. The standard H-bonding criteria were set as  $d(\text{H}\cdots\text{A}) \leq 3.0$  Å and  $\theta(\text{X}-\text{H}\cdots\text{A}) \geq 90^\circ$ . For other weak interactions, the criteria are mentioned in the respective sections. A schematic description of the various interactions is given in Scheme 3.1.



**Scheme 3.1:** (a) Representative hydrogen bond. A–C is a single or double bond, (b), (c) parameters for X–H $\cdots\pi$  interactions (d) Parameters for halogen $\cdots$ O interactions. O–A is a double bond.

Macromolecular crystal structures rarely contain H-atom positional data with the precision required to properly evaluate hydrogen bond geometry. Therefore a method must be found to add or modify all the H-atom positions. H-atoms were added to the protein, water and ligand using the program MOE [3.24]. The H-atom positions were then refined (energy minimization) keeping the position of the non-H atoms fixed using the MMFF94x force field. It is important to note here the basis of selecting this force field for optimization of the H-atom positions. Initially four different types of force field viz. CHARMM22, AMBER96, OPLSAA, and MMFF94x were used to derive standard hydrogen bond geometries ( $d$  and  $\theta$ ) in the earlier studied 28 crystal structures by Sarkhel and Desiraju [3.19, 3.24, 3.25]. The MMFF94x force field outperformed all other force fields with respect to optimization of the protein geometry. Programs like REDUCE from the Richardson group and HGEN from the CCP4 package could have been used for protein H-atoms but these programs are not efficient in generating H-atoms for the ligand and water [3.26, 3.27].

The calculated H-atom positions (MMFF94x optimized) were benchmarked against an experimental neutron crystal structure namely, 6RSA.pdb. In addition to the assessment of protein–ligand geometries, this benchmarking was also expected to be useful in the validation of H-atom positions in the water molecules. Fixing H-atom positions in water molecules has been a long-standing problem in macromolecular crystallography and in this regard the use of MMFF94x achieves reasonable accuracy and is of general utility.

The large number of structures and more personalized requirements led the design of HBAT software to carry out this study. There exist many programs for interaction analysis like HBPLUS, HBEXPLORE, CONTACT from CCP4 and web based servers like LPC and NCI in the public domain [3.27–3.31]. These programs did not fit my requirements for several reasons, notably the fact that weak interactions like C–H $\cdots$ O and C–H $\cdots$  $\pi$  are not considered. The advantage of HBAT over the above-mentioned programs is its compactness in delivering all possible interactions in a single package, thus avoiding server based applications. Among other advantages, an MSOFFICE Excel compatible output file for statistical analysis provides distance-angle distributions across various geometry ranges, while tabulation of frequencies for each residue, ligand, water, and also nucleic acids can be done easily for any kind of interactions. The program is written using PERL and TK languages. It is a user-friendly desktop tool, which offers the freedom to choose several parameters. To evaluate the accuracy of the program, HBAT was used to reproduce the

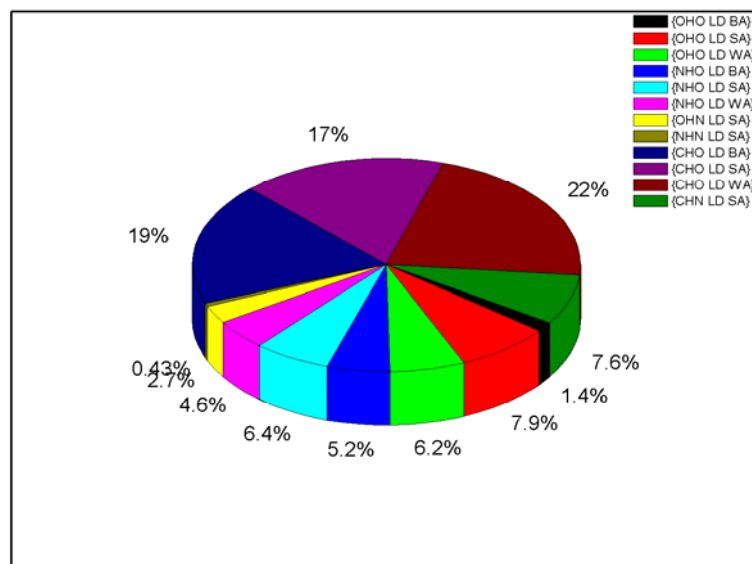
geometries in the earlier manual study of 28 complexes by Sarkhel and Desiraju [3.19] and also in other recent papers [3.32–3.35]. Results obtained are in excellent agreement with the earlier studies [3.19, 3.32–3.35]. This exercise has given me enough confidence to carry out the present study of 484 protein–ligand complexes.

### **3.3 Results and discussion**

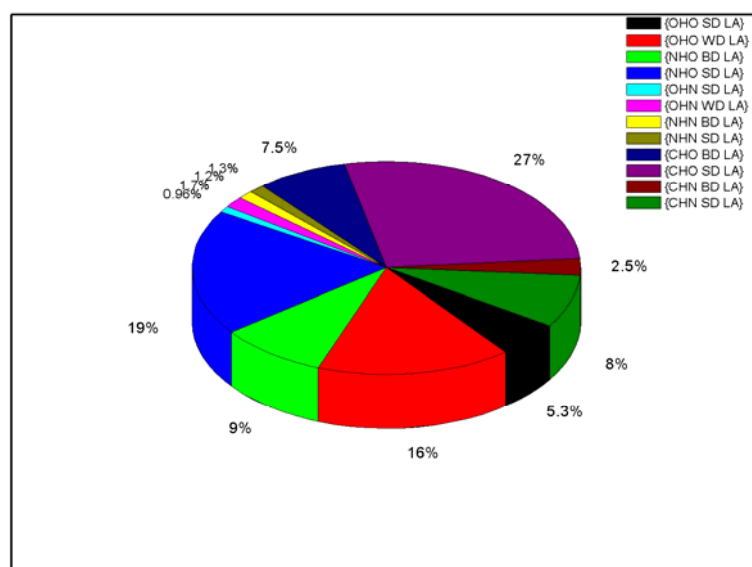
#### **3.3.1 Hydrogen bond geometry. Lengths and angles.**

In any database analysis, evaluation of large data sets provides a more unbiased identification of a chemical signal in the presence of crystallographic noise [3.36, 3.37]. A data set consisting of more than 100 representatives is ideal for the study of interaction geometry in crystal structures [3.38, 3.39]. However analysis of hydrogen bonds in macromolecules is still difficult and requires classification based on backbone, side chain, ligand and water. The geometries observed for these various situations could be different in terms of their lengths, angles and scatter [3.40]. The involvement of many types of hydrogen bond donors and acceptors increases the overall complexity at the protein–ligand interface.

Considering all this, it was felt that a classification of hydrogen bonds based on the participating groups and/or residues would better address the geometrical issues. The interacting partners at the interface are protein, ligand and water. All are able to donate and accept hydrogen bonds. Further, the donors and acceptors fall into different classes based on the strength/weakness of the resulting hydrogen bonds. The percentage contribution of various types of hydrogen bonds in the total protein–ligand interfaces in our 251 crystal structures is shown in Figure 3.1*a* and *b*. The hydrogen bond abbreviation consists of three parts: hydrogen bond type, donor, acceptor. B stands for backbone, S is side chain, L is ligand, W is water, D is donor, and A is acceptor. For example {NHO BD LA} signifies an N–H···O hydrogen bond involving a backbone N–H donor and a ligand O-atom acceptor.



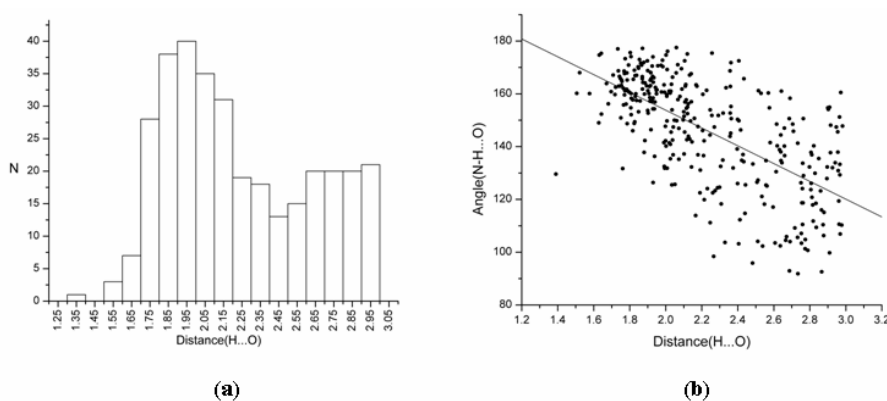
(a)



(b)

**Figure 3.1:** Distribution of all possible of hydrogen bond types in the active sites of protein–ligand complexes (a) ligand as donor (b) ligand as acceptor.

When the ligand is a donor, the percentage of strong hydrogen bonds is 34% while weak hydrogen bonds account for 65%. This is reversed when the ligand is the acceptor, with 54% and 46% strong and weak hydrogen bonds respectively. These numbers are reasonable: not only is the number of strong donors in ligand small but ligands also generally have more acceptors than donors. The data also shows that if there are more acceptors than can form hydrogen bonds with good donors from the ligand, donors from the amino acids and water are used. The population of interactions at the protein–ligand interface from the main chain and the side chain is 32% and 68 % respectively. The types of strong hydrogen bonds observed in the protein–ligand interface are {NHO LD BA}, {NHO SD LA}, {OHO LD BA} and {OHO SD LA}. The median  $H\cdots O$  distances,  $d$ , in all the above cases are less than 2.0 Å and the hydrogen bonds are linear (Figure 3.2). (See Appendix I, Figure 9).

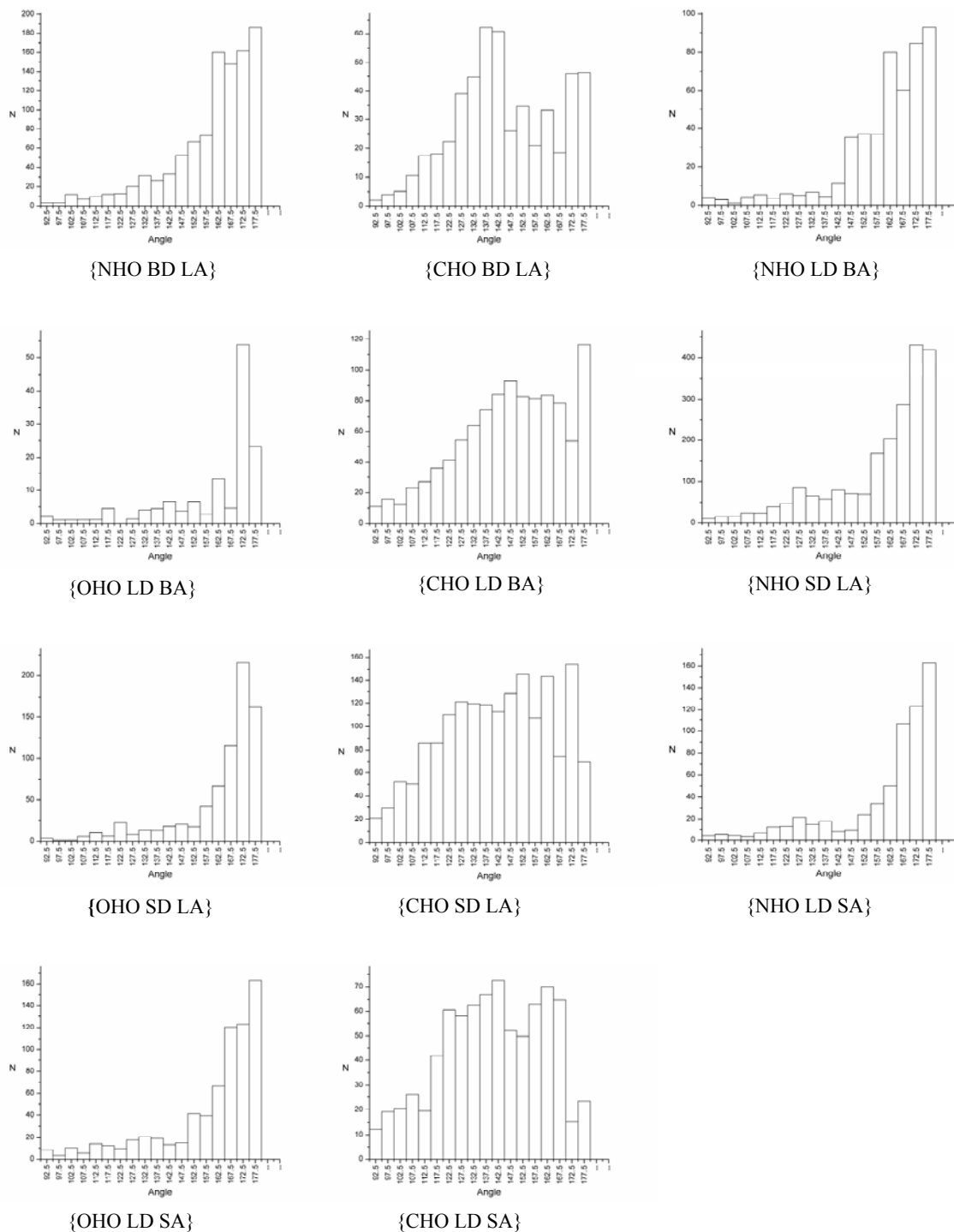


**Figure 3.2:** Distance  $H\cdots O$  histogram and  $d$ - $\theta$  scatterplot ( $R^2 = -0.65$ ) for {NHO BD LA} hydrogen bonds in the active sites of the 251 protein–ligand complexes considered in this study.

It was suggested earlier [3.19] based on PDB cone-corrected angular distributions, that there are small deviations from linearity for both  $N-H\cdots O$  and  $O-H\cdots O$  interactions. This observation does not hold good in the larger dataset of the present study. In both {NHO LD BA} and {NHO BD LA} cases, the cone-corrected angular maxima occur at  $180^\circ$ . The cone-corrected angular distributions for {OHO LD BA} and {OHO SD LA} are similar with maxima in the range  $175$ - $180^\circ$ . The inverse length-angle correlations are also well behaved in all these cases (Figure 3.3). These observations are reassuring and show that the fundamental property of hydrogen bonds, namely linearity, holds by and large for all

categories of strong hydrogen bond in macromolecular structures. Of course, the main chain hydrogen bond might be slightly more linear than the side chain interactions but, all in all, the geometries for strong hydrogen bonds observed in protein–ligand interfaces are comparable to what is observed in small molecules. Baker and Hubbard have discussed hydrogen bond nonlinearity in their 1984 review [3.6]. However, based on the present observations it can be asserted that  $\text{O}-\text{H}\cdots\text{O}$  and  $\text{N}-\text{H}\cdots\text{O}$  hydrogen bonds tend to linearity in all macromolecular crystal structures.

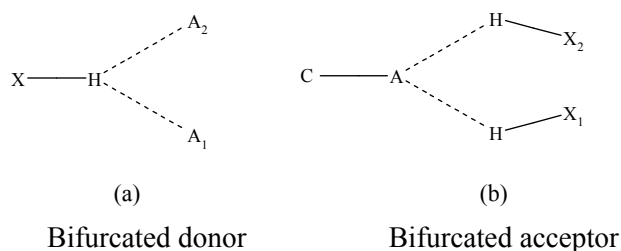
The  $\text{C}-\text{H}\cdots\text{O}$  interactions include {CHO BD LA}, {CHO LD BA}, {CHO SD LA} and {CHO LD SA}. For {CHO BD LA} the angle distribution has a maximum at  $170-180^\circ$ . There is another maximum at  $135-150^\circ$  which corresponds generally to multifurcated geometries. Also similar is {CHO LD BA} with a maximum of  $175-180^\circ$ . For {CHO SD LA} the maximum is still around  $170-175^\circ$ . For {CHO LD SA} the favored angle is around  $140-145^\circ$  with a narrow range of linearity. Unlike strong hydrogen bonds, the cone corrected weak  $\text{C}-\text{H}\cdots\text{O}$  geometries show two distinct maxima at  $130-150^\circ$  and  $170-180^\circ$ . While the weakest {CHO LD SA} have variable geometry, the metrics of the other  $\text{C}-\text{H}\cdots\text{O}$  bonds are surprisingly consistent. This is especially true of bonds donated by main chain  $\text{C}-\text{H}$  groups (Figure 3.3). Also seen are  $\text{C}-\text{H}\cdots\text{O}_w$  bonds formed by main chain and side chain  $\text{C}-\text{H}$  groups to water as acceptor (See Appendix I, Figure 10). These are discussed later.



**Figure 3.3:** Histograms showing cone-corrected angular distribution for strong and weak hydrogen bonds in active sites of 251 protein–ligand complexes.

### 3.3.2 Hydrogen bond geometry. Furcation.

Hydrogen bond furcation is a ubiquitous phenomenon in macromolecular structures. A donor can interact with several acceptors simultaneously or an acceptor can interact simultaneously with many donors. The terms bifurcated and trifurcated are commonly used to describe these arrangements. A bifurcated geometry can also be termed three-centered, and a trifurcated geometry can be termed four-centered (Scheme 3.2). In this study, the H-bonding criteria for furcated geometries were set as  $d \leq 3.0 \text{ \AA}$  and  $\theta \geq 90^\circ$ . These furcated geometries (bifurcated, trifurcated etc.) constitute independent sets in the sense that the trifurcated geometries do not implicitly include the bifurcated ones and so on. Multifurcation was first discussed in the 1960s and 1970s, in small molecule crystal structures. In the modern context, furcation would include all kinds of hydrogen bonds, strong ( $\text{O-H}\cdots\text{O}$ ,  $\text{N-H}\cdots\text{O}$ ) and weak ( $\text{C-H}\cdots\text{O}$ ) and the general idea is that the weak hydrogen bonds fill out or complete the hydrogen bond potential of an acceptor (or donor) which has a small number of strong hydrogen bonds [3.41].



**Scheme 3.2:** Notations for bifurcated donor/acceptor.

In the active sites of protein–ligand complexes, the level of furcation ranges from bifurcated to hexafurcated. Table 3.1 shows that donor and acceptor furcation occur roughly to the same extent (33299 furcated donors and 33038 furcated acceptors in the entire data set). This conveys that furcation is an inherent characteristic of macromolecular crystal structures. It does not arise—as it generally does in small molecule crystal structures—because of a donor–acceptor imbalance. If the analysis is restricted to the ligand, the frequency of furcated acceptors (1104) is more than that of furcated donors (772). This is in accordance with the fact that there are more acceptor atoms in ligands than donor atoms. It could also be due to steric reasons. Furcation levels higher than three are possible in principle, but are rarely found in practice because they require very high spatial densities of

atoms and groups, especially when the donor is furcated. The above are number of furcated donors and acceptors. The total number of interactions is naturally much higher. Also notable are the numbers of nonfurcated geometries. Nonfurcated donors are more numerous than nonfurcated acceptors because: (1) there are more donors overall (C–H is included as a donor) and, (2) acceptors are furcated more easily than donors for steric reasons.

**Table 3.1:** Acceptor and donor furcations in the active sites of 251 protein–ligand complexes.

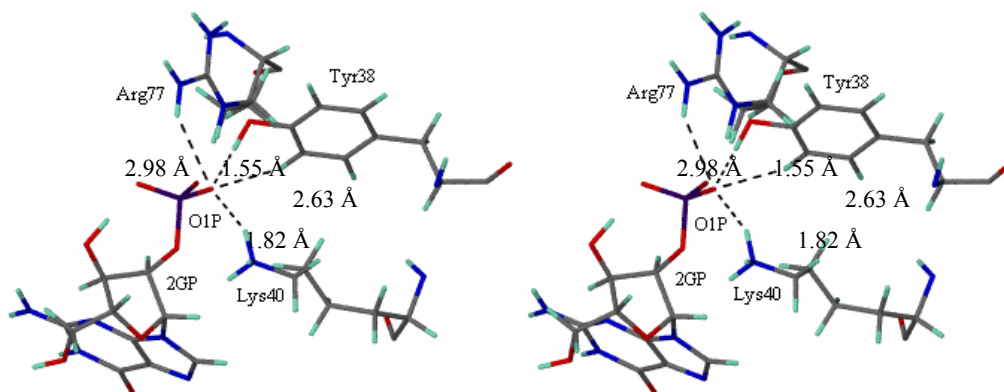
	Furcated acceptors in active site			Furcated donors in active site		
	In protein and water	In ligand	Total	In protein and water	In ligand	Total
Furcation Level						
Bifurcated	12316	438	12754	21452	541	21993
Trifurcated	8672	321	8993	7852	160	8012
Tetrafurcated	6545	211	6756	2383	54	2437
Pentafurcated	3210	94	3304	605	10	615
Hexafurcated	1191	40	1231	235	7	242
<b>Total</b>	<b>31934</b>	<b>1104</b>	<b>33038</b>	<b>32527</b>	<b>772</b>	<b>33299</b>
<i>Nonfurcated</i>	<i>17076</i>	<i>511</i>	<i>17587</i>	<i>41681</i>	<i>1321</i>	<i>43002</i>

**Table 3.2:** Strong and weak hydrogen bonds for ligands at various levels of donor and acceptor furcation.

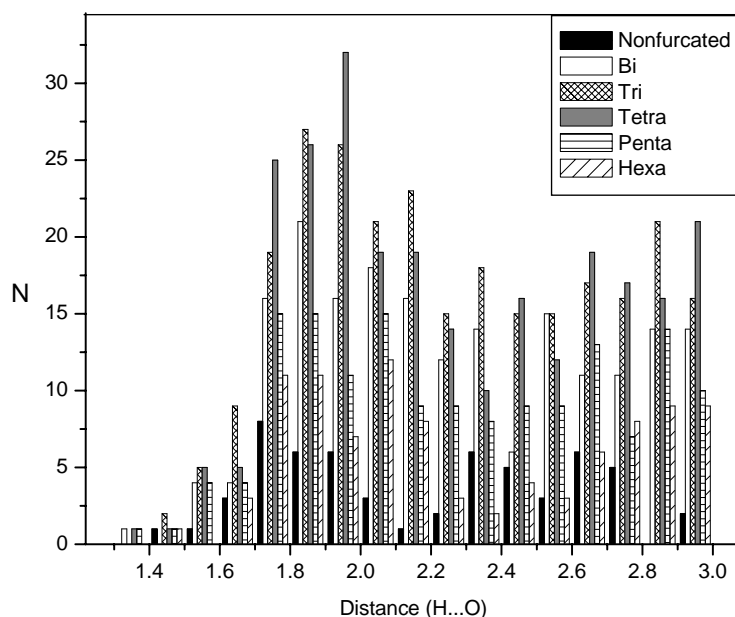
Furcation level	Furcated acceptor			Furcated donor		
	O–H···O	N–H···O	C–H···O	O–H···O	N–H···O	C–H···O
Bifurcated	158	193	283	193	157	562
Trifurcated	175	265	337	68	103	238
Tetrafurcated	176	258	350	7	79	104
Pentafurcated	94	154	192	-	14	22
Hexafurcated	32	97	99	1	21	-
<i>Nonfurcated</i>	<i>127</i>	<i>55</i>	<i>103</i>	<i>178</i>	<i>131</i>	<i>786</i>

It is emphasized here that C–H $\cdots$ O interactions are more common than the strong N–H $\cdots$ O and O–H $\cdots$ O hydrogen bonds in the furcated geometries (Table 3.2). These ideas have been noted by the earlier study of 28 protein–ligand complexes by Sarkhel and Desiraju [3.19]. It is less likely (for electrostatic and statistical reasons) that a strong interaction like O–H $\cdots$ O occurs repeatedly in a furcated interaction. Instead, strong interactions tend towards non-furcated geometries while weak interactions occur in furcated situations. The overall message is that both interaction strength and close packing are important. A furcated geometry typically has one or a small number of strong interactions and many weaker interactions. This optimizes both interaction geometry and efficiency of space-filling. Table 3.2 shows that the total number of hydrogen bonds (O–H $\cdots$ O, N–H $\cdots$ O, C–H $\cdots$ O) to the 1104 furcated acceptors in Table 3.1 is 2863. This corresponds to an average level of furcation of 2.6 interactions to each furcated acceptor in the active site.

In summary, furcation occurs for both donor and acceptor sites on ligands. Acceptor furcation is more common than donor furcation and this could be due to steric reasons. The majority of furcated interactions exhibit longer  $d$  (H $\cdots$ O) distances than the simple non-furcated hydrogen bonds and this is as might have been expected (Figure 3.4a and b).



**Figure 3.4:** (a) Stereo view of tetrafurcated interaction in 2AAD.pdb. The acceptor centre is the O1P atom of the ligand (2GP). O1P interacts with residues Tyr38, Lys40 and Arg77 through C–H $\cdots$ O and O–H $\cdots$ O and N–H $\cdots$ O bonds. The respective H $\cdots$ O distances are also shown.



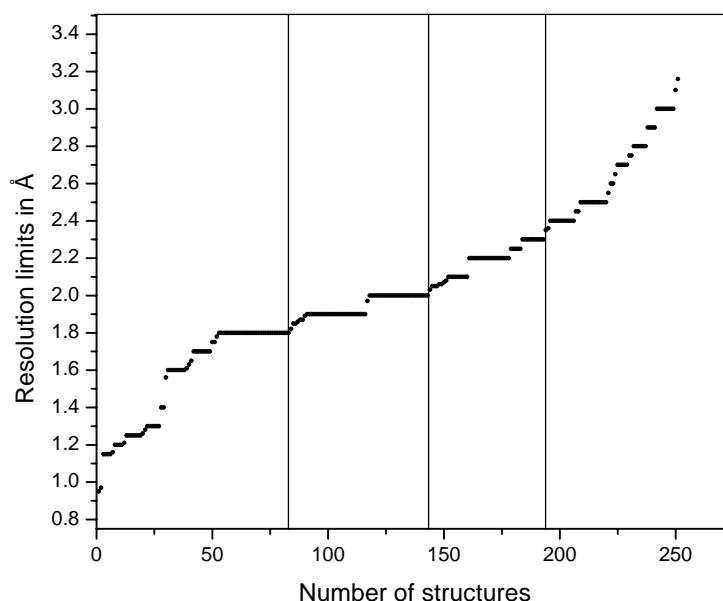
**Figure 3.4:** (b) Histogram of the distance  $d$  of N–H...O interactions to furcated ligand acceptors.

### 3.3.3 Hydrogen bond geometry. The resolution problem.

The resolution in a macromolecular crystal structure determination has a direct effect on the geometry of both strong and weak interactions. The issue of reliability of hydrogen bond metrics as a function of crystallographic resolution is interesting, and could not be addressed by Sarkhel and Desiraju [3.19], which considered only 28 crystal structures, and all of very good resolution. One of the aims of the present study is to identify the resolutions limits for macromolecular crystal structures where the hydrogen bond geometries are as reliable as that obtained in small molecule structures. Therefore dataset of 251 structures classified on the basis of resolution. The categories chosen were as follows: resolution  $< 1.8$  Å (83 structures); resolution 1.8–2.0 Å (60 structures); resolution 2.0–2.3 Å (50 structures); resolution  $> 2.3$  Å (58 structures). Table 3.3 contains a list of structures and Figure 3.5 gives this information pictorially.

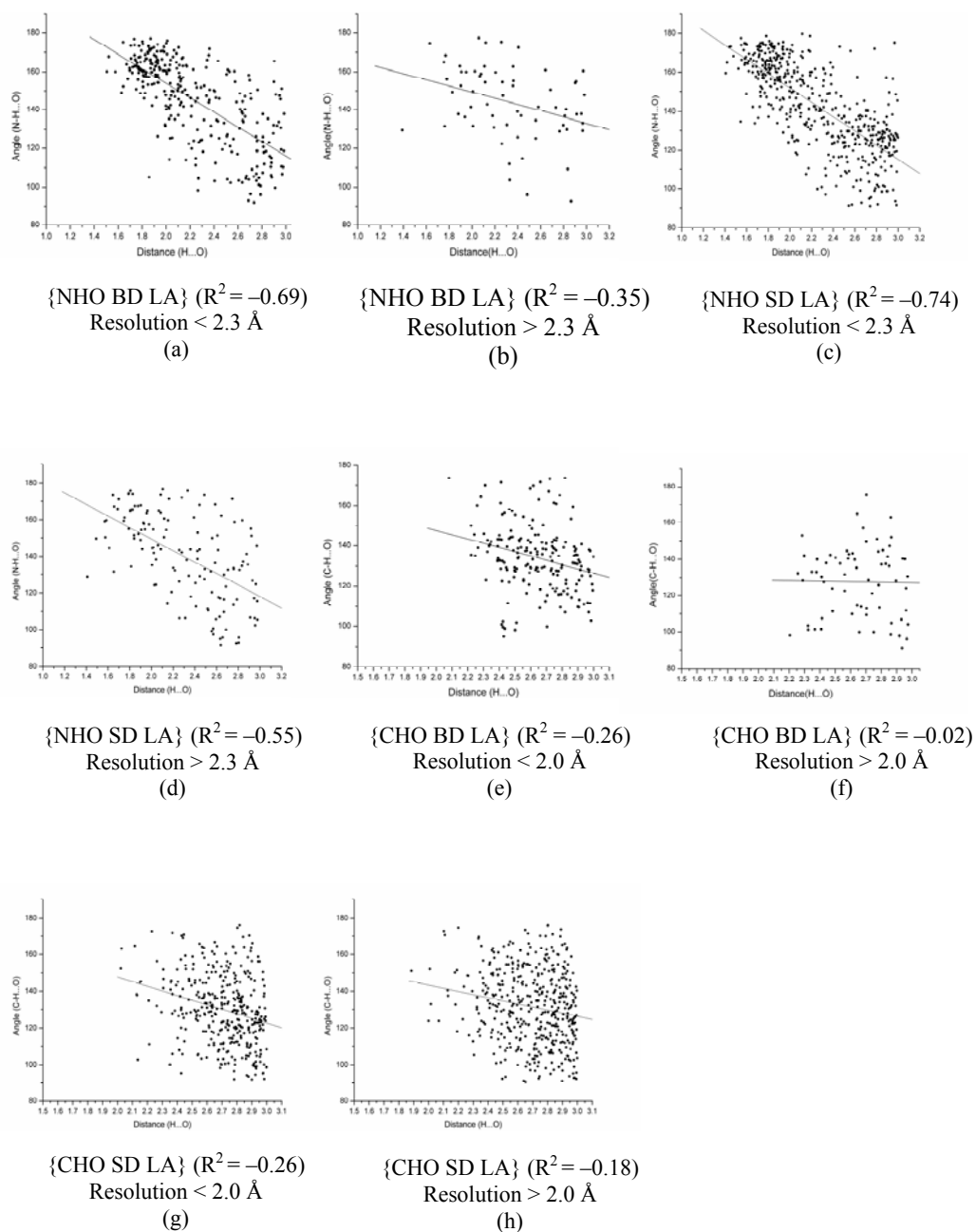
**Table 3.3:** List of entries of PDB structures based on four resolution limits.

Resolution < 1.8 Å (83 Entries)									
1BXO	8A3H	1A6G	1HDO	1HET	1RGE	1MRO	1C0P	1I3H	1I76
2NLR	1D4O	1C1D	1EQO	1OAA	1Q0N	2TPS	2WEA	3CHB	1CTQ
1QKS	1DJR	1FCY	1FK5	1HYO	1I12	1RUV	1TPP	2CTC	1HFC
1AOE	1CIL	1LIC	1MRK	1PHD	1PHG	1ROB	2TMN	1C5C	2CPP
1SNC	1ABE	1B17	1ETA	1HYT	1IDA	1XID	1XIE	4DFR	1C5X
3CLA	4EST	1A28	1A6W	1APT	1APU	1AQW	1ATL	1B58	1B59
1BMA	1C83	1CBS	1COY	1D3H	1EJN	1GLQ	1HVR	1LST	1MRG
1NCO	1PPC	1QBR	1QBU	1RDS	1SRJ	1TNG	1TNH	1WAP	2FOX
2QWK	5ABP	6RNT							
Resolution 1.8-2.0 Å (60 Entries)									
1JAP	1FLR	2AK3	1AEC	1F3D	2CMD	1HSL	1A4Q	1ABF	1BYB
1FEN	1GLP	1HPV	1HSB	1KEL	1LNA	1MLD	1MMQ	1MTS	1PPH
1RNT	1SLT	1TMN	1TNI	1TNL	1TYL	1UKZ	1VGC	2GBP	2H4N
3ERT	3TPI	7TIM	2TSC	1AZM	1BBP	1CBX	1CDG	1CLE	1DO1
1DG5	1DMP	1EED	1EIL	1EPO	1FKG	1FRP	1HIV	1MBI	1POC
1PSO	1QCF	1QPE	1TRK	25C8	2AAD	2IFB	3CPA	5ER1	6RSA
Resolution 2.0-2.3 Å (50 Entries)									
3ERD	1ACO	1CKP	1NIS	1LAH	1LCP	1D4P	5CPP	1DY9	1EBG
1F0R	1F0S	1LDM	1MDR	1OKL	2YHX	4PHV	1A4G	1CL2	1COM
1DR1	1EPB	1ETR	1FKI	1HDC	1HTF	1IMB	1LPM	1OKM	1PDZ
1PPI	1YDR	1YEE	2CHT	2PCP	1A42	1CPS	1DD7	1EOC	2PK4
1BGO	1BLH	1DHF	1ETS	1HOS	1PBD	1PTV	1TLP	1YDT	3GPB
Resolution > 2.3 Å (58 Entries)									
1B9V	1AI5	1BYG	1CVU	1DOG	1IVB	1MUP	1NGP	1RNE	2ACK
2ADA	4AAH	4LBD	1FL3	1QPQ	1A9U	1AAQ	1ASE	1BMQ	1DID
1EAP	1ETT	1FGI	1LYB	1UVT	2YPI	4FBP	1RT2	1CQP	1IVQ
1TDB	1DBB	1DBJ	1IBG	1MCQ	2PHH	1BKO	1ULB	1ACJ	1ACL
1ACM	1LYL	1UVS	2LGS	1BAF	2DBL	3HVT	4CTS	1CX2	1DWC
1DWD	1FAX	1HAK	1HRI	2RO7	4COX	2MCP	1DWB		



**Figure 3.5:** Distribution of crystal structures as a function of resolution.

As prototypes of strong and weak hydrogen bonds, {NHO BD LA}, {NHO SD LA} and {CHO BD LA}, {CHO SD LA} are described here for comparisons of hydrogen bond geometry as a function of resolution. The  $d$ - $\theta$  scatterplots (cone-corrected) were analyzed for these prototype interactions (Appendix I, Figure 11). From a visual inspection of these plots, it was concluded that strong hydrogen bonds {NHO BD LA} and {NHO SD LA} retain acceptable geometries till a resolution of 2.3 Å, whereas for {CHO BD LA} and {CHO SD LA} the threshold is 2.0 Å (Figure 3.6), with bond geometries beyond these limits being poor. Accordingly, crystal structures within a resolution of 2.3 Å may be safely considered for strong hydrogen bonds like N–H $\cdots$ O. For C–H $\cdots$ O the corresponding limit is 2.0 Å and signs of non-linearity are observed above this. (Appendix I, Figure 11).

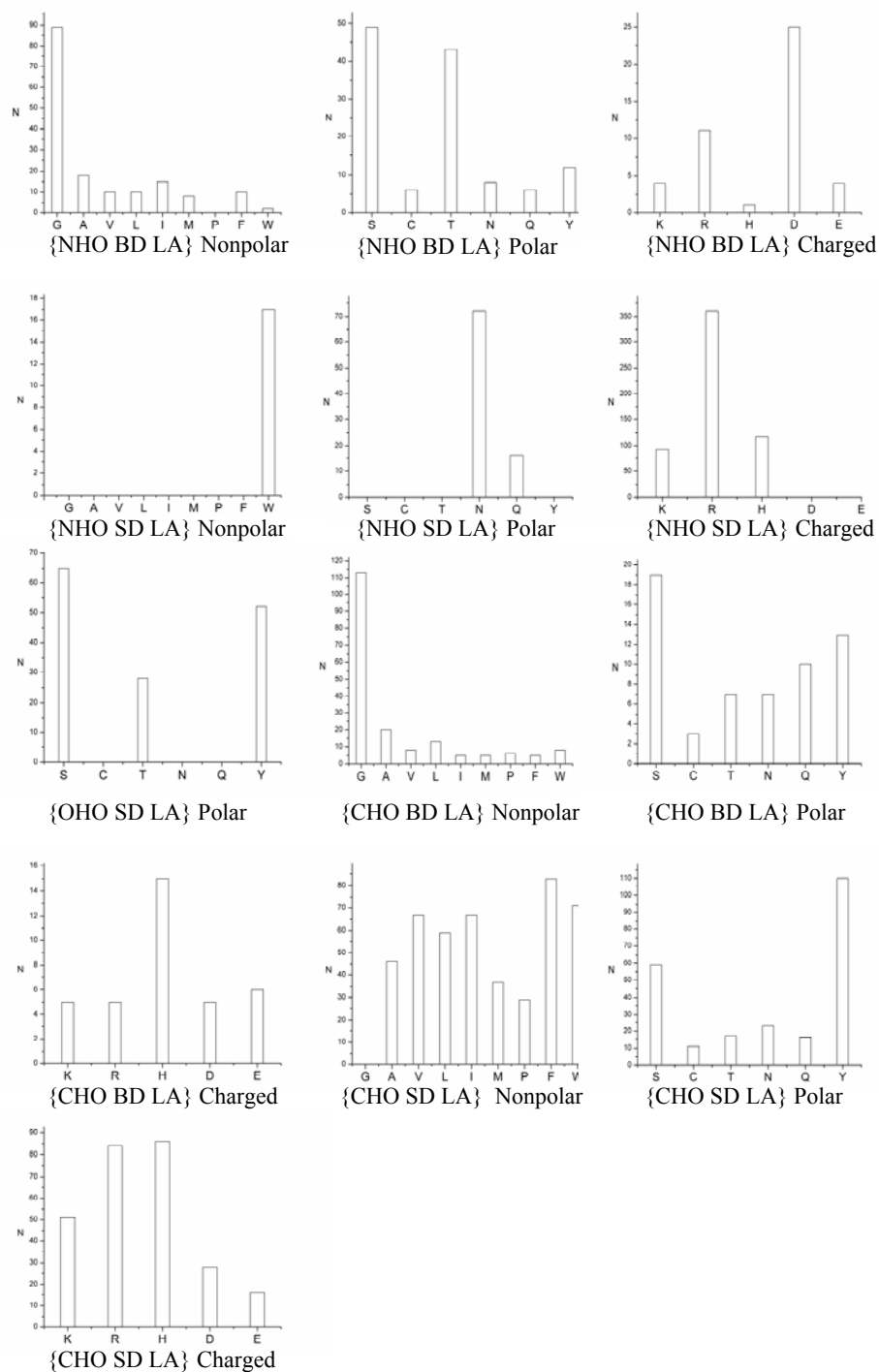


**Figure 3.6:**  $d$ - $\theta$  Scatterplots for  $\{\text{NHO BD LA}\}$ ,  $\{\text{NHO SD LA}\}$  (a)-(d), and  $\{\text{CHO BD LA}\}$ ,  $\{\text{CHO SD LA}\}$  (e)-(h). Notice the different appearance of these two sets of plots below and above the threshold resolution limits.

### **3.3.4 Residue frequency**

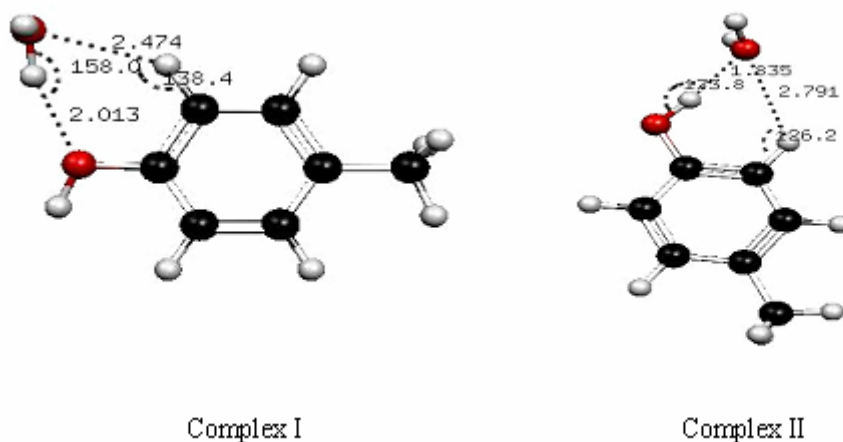
The percentage distributions of various residues in the total protein and in the active site are comparable for nonpolar and polar amino acids (Appendix I, Table 6). Residues like Gly, Ile, Phe and Tyr are among the most common residues. For {NHO BD LA} Gly is the major nonpolar donor, while Ser and Thr are the major polar donors. The percentage of charged residues in the active site (20%) is slightly less than in the overall protein (24%) with Asp and Arg being the major charged donors (Figure 3.7) and this is in keeping with the hydrophobic nature of the protein interior. This result is similar to the earlier observed trend in the smaller set of 28 structures [3.19] except that among the charged residues, Asp is now the major donor instead of Lys. For {CHO BD LA}, Gly, Ser and Tyr, and His are dominant as donors in the three respective classes of amino acids. These observations provide a hint that residues which are smaller in size and have greater flexibility participate well in both strong and weak hydrogen bonds. Sometimes a so-called weak donor like C<sub>α</sub>–H can behave like a strong donor in the presence of a charged side chain; this is observed for His.

Residues like Trp, Asn and Arg are the major donors in the {NHO SD LA} category. However, Lys and His also interact equally well with the ligand through their side chains. The major participation of charged residues here suggests that ligand binding is dominated by electrostatic interactions. At this point, it is interesting to discuss the donor capability of Gly. For Gly the total number of {NHO BD LA} and {CHO BD LA} interactions are 89 and 113, and this is the highest among all amino acids. This fact is ascribed to Gly being the smallest residue in terms of volume (63.8 Å<sup>3</sup>) and also occurring most frequently in the active site (9.63%). This result reaffirms earlier observations [3.19] and suggests that the highly flexible nature of Gly is very well exploited in biological recognition [3.42, 3.43].



**Figure 3.7:** Frequency distribution of nonpolar, polar, and charged amino acids for strong and weak hydrogen bonds in the protein-ligand interface.

For N–H···O bonds with N–H side chain donors, the major donor residues are Trp (nonpolar) Asn (polar) and Arg (charged). The number of interactions for {OHO SD LA} polar residues follows the order Ser > Tyr > Thr. The alkyl hydroxyl present in Ser has less steric hindrance than in Thr and is therefore more commonly used. For nonpolar residues, the numbers of {CHO SD LA} interactions are greater for Phe and Trp. Similarly, Tyr, Arg and His are the major amino acids in the other categories participating in {CHO SD LA} interactions. An interesting case in this category is the side chain phenyl ring donor capacity of Tyr, wherein a large number of C–H donors are present along with a strong O–H donor. The total number of C–H···O and O–H···O interactions exhibited are 87 and 52 respectively. To explain this phenomenon an *ab initio* calculation between Tyr side chain and water was carried out at the 6-31G\*\* basis set level. Two minima were obtained corresponding to Complex I and Complex II (Figure 3.8). The energies for the O–H···O and C–H···O interactions in Complex I are –6.68 and –1.83 kcal/mol. For Complex II, the values are –9.43 and –0.56 kcal/mol respectively.



**Figure 3.8:** Two possible conformations of the interaction of the Tyr side chain with water.

### 3.3.5 Interactions involving water

The hydrogen bonding capacity of water makes it easy to interact with protein, ligand or neighboring water molecules. The total number of hydrogen bonds (N–H $\cdots$ O, O–H $\cdots$ O, C–H $\cdots$ O) formed by water (as donor or acceptor) in the active sites of the 251 complexes under consideration is 29718, in other words these are nearly 118 such hydrogen bonds for each structure, on average. The frequency of occurrence of these bonds follows the order, side chain > main chain > water > ligand. Table 3.4 shows that there are many more C–H $\cdots$ O<sub>w</sub> interactions than for N–H $\cdots$ O<sub>w</sub> and O–H $\cdots$ O<sub>w</sub> interactions. A small number (15.29%) of hydrogen bonds are of the type O<sub>w</sub>–H $\cdots$ O<sub>w</sub>.

**Table 3.4:** Percent distribution of various bond types among the 29718 hydrogen bonds formed by water in the active site of 251 protein–ligand complexes.

	Type	%
1	{NHO LD WA}	0.42
2	{OHO LD WA}	0.58
3	{CHO LD WA}	2.05
4	{OHO WD LA}	1.99
5	{NHO BD WA}	8.43
6	{CHO BD WA}	8.84
7	{OHO SD WA}	3.99
8	{NHO SD WA}	10.49
9	{CHO SD WA}	25.48
10	{OHO WD WA}	15.29
11	{OHO WD BA}	12.80
12	{OHO WD SA}	9.59

Among the 1503 ligand-water interactions, {CHO LD WA} interactions constitute as many as 66%, while {OHO LD WA} and {NHO LD WA} account for only 19% and 13%, respectively. A similar trend was observed in the previous study [3.19]. For {OHO WD LA} the number of interactions observed is 592 of which 53% are below a H $\cdots$ O distance of 2.2 Å (Appendix I, Table 7). This is in contrast to the earlier study [3.19], where the maximum number (about 66%) of O<sub>w</sub>–H $\cdots$ O hydrogen bonds were observed in the range of 2.2–2.7 Å. A possible reason for this discrepancy could be the inaccuracy in the H-atom addition method for water in our previous study. A total of 8624 water molecules are present in the 251 active sites. For these, the average coordination number is 2.1, if only strong hydrogen bonds are considered. If weak interactions are added, then this number rises to 3.4

per water molecule, which is in good agreement with the earlier study of Steiner [3.44]; it is also chemically reasonable.

### 3.3.6 Lipinski's rule extended

The numbers of hydrogen bond donors and acceptors are known to affect the physico-chemical properties (solubility, adsorption, distribution) of a molecule and hence the efficacy of a drug. Lipinski's rule-of-five states that for better permeation and absorption, the number of donors and acceptor in a ligand should be less than 5 and 10 respectively [3.45]. The present dataset contains 302 ligands. The total number of strong hydrogen bonds made by the various ligands with protein and water molecules stand at 973 (ligand as donor) and 2001 (ligand as acceptor). Accordingly, each ligand has 3.2 donors and 6.6 acceptors on average. This figure satisfies Lipinski's rule-of-five for hydrogen bond donors and acceptors in that the number of acceptors present per ligand is around twice the number of donors.

### 3.3.7 Protein–ligand interactions in kinases

An external test set of 233 protein–ligand complexes of various kinases was compiled from the PDB (Appendix I, Table 8) to assess the general applicability of important conclusions derived so far. The nature of strong and weak hydrogen bonds (linearity of hydrogen bonds), prevalence of multifurcated interactions, distinctiveness of interaction patterns as a function of resolution and other attributes were analyzed in this dataset (Appendix I, Figure 12, Table 8–10). These results suggest a similar trend for hydrogen bond geometries and constitute a useful validation of the principles enunciated in this analysis.

### 3.3.8 Other weak interactions

#### (a) $X-H\cdots\pi$ hydrogen bonds to amino acid residues

The most common  $\pi$ -acceptors in proteins are the side chains of Phe, Tyr, Trp and occasionally His [3.13, 3.32]. In this study, the ligand  $\pi$ -acceptors have not been taken into account and so the results are restricted to  $\pi$ -acceptors in the side chains of Phe, Tyr, Trp and His. But this is not a serious limitation. For Trp, the five membered and six membered rings were treated separately. In the present study, the convention adopted is shown in Scheme 3.1*b* and *c* to locate ligand  $\pi$  interactions. It is difficult to derive an ideal geometry

for these interactions to multi-atom acceptors. However  $d \leq 3.5$  Å,  $\theta \geq 100^\circ$  and  $\omega \leq 40^\circ$  appear to be satisfactory and this geometric criterion is generally accepted [3.13]. In the present study 4 N–H $\cdots\pi$  (two each to Trp and Tyr), 3 O–H $\cdots\pi$  (two to His and one to Trp) and 159 C–H $\cdots\pi$  interactions are observed (Appendix I, Figure 13, Table 11). For the 159 C–H $\cdots\pi$  interactions, the acceptor frequency is Trp (41%), Tyr (28%), Phe (14%) and His (17%). The percentage occurrences of these residues in the active sites are Trp (2.39%), Tyr (5.18%), Phe (5.11%) and His (2.83%). The high frequency of C–H $\cdots\pi$  bonds to Trp is accounted for by the larger accessible areas afforded by the two fused rings. The doubling of the number of interactions to Tyr when compared with Phe, two residues that occur nearly equally in the active sites, is nicely accounted for by the increased acceptor capability of the Tyr aromatic ring.

#### (b) Halogen bonds

Short oxygen $\cdots$ halogen interactions have been known since the 1950s [3.46]. A recent survey of protein and nucleic acid structures reveals similar halogen bonds as potentially stabilizing inter- and intramolecular interactions that can affect ligand binding [3.34]. A typical halogen bond is represented by Scheme 3.1*d*. Protein–ligand complexes were analyzed for possible halogen bonds following the van der Waals radii criterion. Seven halogen bonds are found in the following structures: 1CKP.pdb, 1CLE.pdb, 1BMA.pdb, 4EST.pdb, 2NLR.pdb, 1ETA.pdb listed in Table 3.5. These interactions were observed between C–F, C–Cl and C–I and carbonyl O-atoms in the main chain.

**Table 3.5:** Halogen bonds observed in the protein–ligand interfaces is tabulated along with respective PDB ID.

	Type	Ligand	Residue ID	Acceptor	Residue ID	$d$	$\theta_1$	$\theta_2$	PDB ID
1	C–Cl $\cdots$ O=C	PVB	1	Asp	86	3.0	134.0	78.4	1CKP
2	C–Cl $\cdots$ O=C	ENH	703	Thr	96	2.9	147.7	74.2	1CLE
3	C–Cl $\cdots$ O=C	ENH	703	Arg	100	2.8	152.9	81.0	1CLE
4	C–F $\cdots$ O=C	TFA	256	Cys	199	2.9	133.9	108.2	1BMA
5	C–F $\cdots$ O=C	FPA	5	Pro	3	2.8	102.1	90.0	4EST
6	C–F $\cdots$ O=C	G2F	603	Glu	120	1.9	107.5	121.9	2NLR
7	C–I $\cdots$ O=C	T44	128	Ala	109	3.1	149.3	106.0	1ETA

*(c) Halogen as nucleophile*

The acceptor capability of organic halogen, X (X = F, Cl, Br, I), has not been studied in detail in macromolecules [3.47]. While these interactions are weak they seem to play a definite role in protein–ligand stabilization when halogenated ligands are present. The number of X–H···halogen (here X = O, N, C) interactions for O–H, N–H and C–H donors is 5, 12 and 35 respectively (Appendix I, Table 12). Almost all O–H···Cl interactions are observed between water and ligand. For C–H···Cl and C–H···F interactions side chain C–H groups are frequently used.

*(d) Hydrogen bonds involving sulfur atoms*

Sulfur atoms are larger and have a more diffuse electron cloud than oxygen and nitrogen, but are nevertheless capable of participating in hydrogen bonds [3.48]. The acceptor functionality of sulfur atoms have been studied here. Sulfur is present in amino acid residues like Met and Cys or it may occur in the ligand. In all these situations, a hydrogen bond is presumed to exist if the distance  $d$  (H···S) is  $\leq 2.9$  Å. The numbers of such cases are 12, 15 and 24 for O–H, N–H and C–H respectively with these donors belonging to either ligand, protein or water (Appendix I, Table 13). For hydrogen bonds of the type O–H···S and N–H···S, the acceptor is found more often in the ligand than in the protein.

**3.4 Conclusions**

The nature of strong (O–H···O, N–H···O) and weak (C–H···O) hydrogen bonds in the protein–ligand interface has been studied in a dataset of 251 protein–ligand complexes using a new in-house computer program (HBAT). Reasonable accuracy in locating hydrogen atoms positions in these complexes were achieved using the MMFF94x force field in the MOE software. The fundamental property of hydrogen bonds, namely linearity, holds by and large for all strong hydrogen bonds in these structures. Strong hydrogen bonds have more consistent distance and angle attributes, while the weak C–H···O interactions have variable geometry. Main chain hydrogen bonds are, in general, shorter and more linear than those formed by side chain donors and acceptors. Furcated ligand–receptor interactions are manifested by both donors and acceptors. Acceptor furcation is more common than donor furcation. The majority of furcated interactions exhibit longer  $d$  (H···O) distances when

compared to simple non-furcated hydrogen bonds. Resolution limits are important with respect to the hydrogen bond geometry. Strong hydrogen bonds retain good geometries up to a resolution of 2.3 Å, whereas for weak bonds the limit is 2.0 Å. Residues like Gly and Ala, which are smaller in size and have greater flexibility, participate well in both strong and weak hydrogen bonds. In this respect, Gly frequently interacts with the ligand. The side chain donor capacity of Tyr, with respect to both O–H···O and C–H···O interactions, is noteworthy. Other weak interactions involving halogen atom (both as electrophiles and nucleophiles),  $\pi$ -acceptors and sulfur atom acceptors are also important in the protein–ligand interface. Strong and weak hydrogen bonds involving water are ubiquitous in the active sites. Water is found to interact with amino acid residues and ligands forming O–H···O, N–H···O and C–H···O bonds. The hydrogen bond donor-acceptor ratio for the ligands is in accordance with Lipinski’s rule-of-five. I conclude that the results of the previous study of 28 structures are largely applicable to a set of structures that is nearly twenty times as large. An encouraging aspect of this study is that macromolecular crystal structures with resolutions up to 2.0 Å may be used to analyze hydrogen bond geometry provided a reliable way is found to fix H-atom positions.

## CHAPTER 4

# STRONG AND WEAK HYDROGEN BONDS IN PROTEIN–LIGAND COMPLEXES OF KINASES: A COMPARATIVE STUDY

### 4.1 Introduction

The hydrogen bond is one of the most important interactions between biologically important molecules [4.1–4.3]. The three-dimensional architecture of proteins and nucleic acids is stabilized by hydrogen bonds, biological recognition operates through hydrogen bonding, and the molecular mobility required for biological processes is directly connected with rapid formation and breaking of hydrogen bonds. Importance of hydrogen bonds in substrate/ligand recognition in macromolecules is an active area of research [4.4–4.8]. Hydrogen bonds are observed with a variety of strengths and geometries in the active sites of protein–ligand complexes [4.4, 4.5, 4.9].

The existence of strong and weak hydrogen bonds in the protein–ligand complexes has been demonstrated in the previous chapter in a dataset of 251 protein–ligand complexes [4.5]. Among the important conclusions derived in previous chapter are: (a) ubiquitous presence of strong and weak hydrogen bonds in the protein–ligand interface; (b) linearity of  $\text{N–H}\cdots\text{O}$  and  $\text{O–H}\cdots\text{O}$  bonds; (c) occurrence of multifurcated hydrogen bonds; (d) resolution limits are crucial in studying hydrogen bond geometries; (e) hydrogen bond geometry of water and amino acid residues like Gly and Tyr are significant in the active sites. These important conclusions derived from this dataset were also validated against 233 protein–ligand complexes of kinase family. In this context, there is a growing amount of literature which emphasizes the importance of strong and weak hydrogen bonds in the protein–ligand interface in protein kinases (PKs) [4.10, 4.11]. The present chapter aims at analyzing strong and weak hydrogen bonds in PKs in greater detail. A comprehensive study of hydrogen bond patterns was carried out between the main/side chain and ligand across the kinase sub-families. The importance of conserved residues and interactions forming hydrogen bonds in kinase sub-families are studied. Water environment in the active sites was also studied. It is hoped that these findings will help structural biologists, crystallographers and medicinal chemists to design better kinase inhibitors.

The therapeutic usages of PKs have recently opened up many research avenues [4.12–4.15]. Till date there are around 450 entries of kinases available in the Protein Data Bank (PDB) [4.16, 4.17]. The abundance of structural information for PKs provides an ideal background for structure-based drug discovery [4.15, 4.18]. Similar to various structural initiatives, online resources specific to PKs on various aspects of the kinase family are constantly appearing [4.19, 4.20]. A typical catalytic domain of kinase has 250–300 amino acids and is bilobal in nature [4.21–4.23]. This two-lobed structure can be further subdivided into 12 sub-domains (I–XI). The N-terminal lobe constituting sub-domains I–IV primarily has antiparallel  $\beta$ -sheets with the important exception of  $\alpha$ -helix C. Sub-domain V is a single polypeptide chain known as the linker region connecting the N-terminal lobe to the larger C-terminal lobe comprising sub-domains VIA–XI. The C-terminal is predominantly helical in nature. A full explanation of the structure and function of each region is available elsewhere at the PKR website [4.17]. The ATP binds in the cleft formed between the N- and C-terminal lobes of the PKs, forming several key interactions conserved across the protein kinase family [4.22, 4.24]. The adenine moiety lies in a hydrophobic region between the  $\beta$ -sheet structure of sub-domains I and II and residues from sub-domains V and VIb. A large number of protein kinase inhibitors have been observed to mimic the donor–acceptor pair of hydrogen bonds made between the protein backbone and adenine. The present study attempts to understand the nature of donor–acceptor hydrogen bonds in the expanded context of hydrogen bonding—one wherein both strong  $\text{N-H}\cdots\text{O}$  and  $\text{O-H}\cdots\text{O}$  and weaker  $\text{C-H}\cdots\text{O}$  interactions are considered.

## **4.2 Materials and methods**

### **4.2.1 Dataset**

A set of 233 X-ray structures of kinase protein–ligand complexes from the PDB was used [4.16]. This is the same dataset used as the test set in the previous chapter [4.5]. These structures have a resolution limit of 1.2 to 3.5 Å. The structures are classified into 44 sub-families of kinase. The classification of structure was carried out based on publicly available resources [4.15–4.17, 4.19]. These kinases are from both eukaryotic and prokaryotic organisms. The active site was defined by selecting all amino acid residues within a 10 Å radius of the ligand molecule. The active site also includes water molecules. The ligand selection method was adopted from an earlier study [4.4].

#### **4.2.2 Geometry optimization**

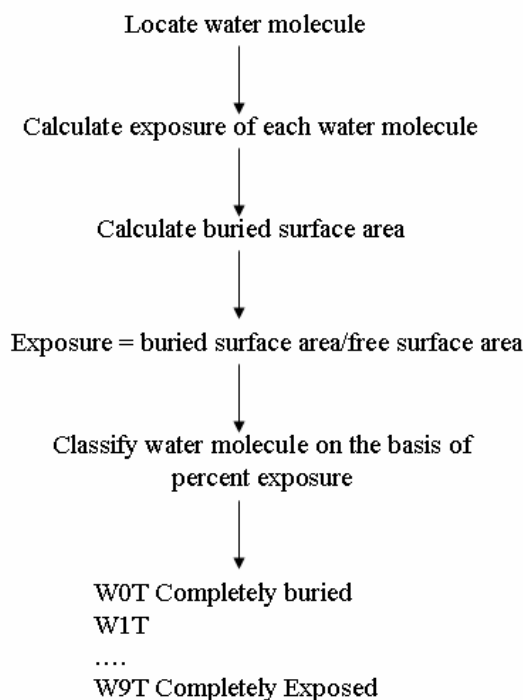
Macromolecular crystal structures rarely contain H-atom positional data with the precision required to properly evaluate hydrogen bond geometry. Therefore a method must be found to add or modify all the H-atom positions. H-atoms were added to the protein, water and ligand using the program MOE [4.25]. The H-atom positions were then refined (energy minimization) keeping the position of the non-H atoms fixed using the MMFF94x force field [4.26]. The choice of this particular force field has been discussed in the previous chapter [4.5]. All the optimized structures were exported to hydrogen bond analysis tool (HBAT) for hydrogen bond analysis [4.27].

#### **4.2.3 Hydrogen bond analysis**

Strong and weak hydrogen bonds were analyzed with the in-house developed HBAT program, which analyzes and tabulates all hydrogen bonds present in a PDB file. The output file provides distance–angle distributions across various geometry ranges while tabulation of frequencies for each residue, ligand, water and nucleic acids is done easily for any kind of interaction. HBAT is a user-friendly desktop tool, which operates both with default and user-selected parameters. The standard H-bonding criteria were set as  $d(\text{H}\cdots\text{A}) \leq 2.8 \text{ \AA}$  and  $\theta(\text{X}-\text{H}\cdots\text{A}) \geq 90^\circ$ . The hydrogen bond synthon analysis was carried out manually. A single structure from each sub-family was randomly selected for sequence alignment. The active site sequence alignment was carried out with the help of ClustalW [4.28].

#### **4.2.4 Water in the active sites of kinase**

Water molecules in the active sites were classified in 10 categories (W0T to W9T) based on solvent accessibility (exposure). This approach is very similar to the method adopted by Williams *et al.* [4.29]. The authors have described a hierarchical classification of the extent of burial of water molecules in the protein cavity. In the present context the active sites are considered as the cavity for ligand binding. This classification scheme was carried out with the help of a scientific vector language (SVL) code written in the MOE software. An overview of the method is given in the flow sheet (Scheme 4.1).



**Scheme 4.1:** Flow sheet for water classification in the active site.

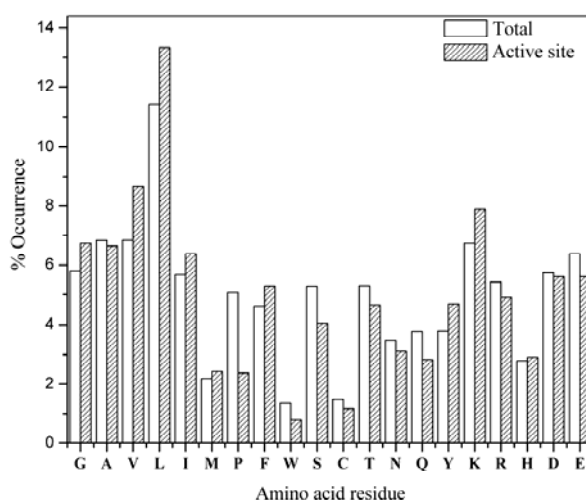
### 4.3 Results and discussion

The hydrogen bond is the key interaction for ATP and inhibitor binding in the kinase family. This is especially true of the hydrogen bonds that involve the main chain. The initial aim of the present chapter is to elucidate strong and weak hydrogen bonds geometries between the main chain, side chain and ligand. However, interactions between the main chain and the ligand are more important. This is because (a) in kinases, the hydrogen bonds geometries involving main chains are more consistent than those formed by the side chains [4.5]; (b) the majority of the hydrogen bonds in the present dataset is observed between the main chain and ligand atoms (195 out of 233 complexes); (c) main chain hydrogen bonds are crucial for proper a positioning of ligands [4.30].

#### 4.3.1 Residue frequency

The total number of residues present in these complexes is 66405, that is on average there are 285 residues present in each structure. The percentage distribution of nonpolar, charged and polar residues in the dataset is 50%, 27% and 23% respectively. The active

sites contain 18997 residues, so that each active site has around 80 residues on average. The percentage distribution for nonpolar, charged and polar residues in the active sites is 53%, 27% and 20% respectively. Among the nonpolar residues, the most frequently observed are Leu, Val, Ala and Gly. Charged residues like Lys, Glu, Asp are abundant in the active site. This finding suggests that for kinases the amino acid composition of both the entire protein and the active site is hydrophobic in nature, which is expected. However, the higher occurrence of charged residues when compared to the polar residues is uncommon [4.5]. The probable reason for this observation might be the importance of electrostatic interaction in ligand binding. In fact Bartlett *et al.* have shown in their study that, majority of catalytic residues are charged residues in 178 enzyme active sites [4.8]. Generally the charged residues are present at the surface and active sites of proteins [4.31]. The percentage occurrence of various amino acid residues in the total dataset and in the active sites is shown in Figure 4.1.



**Figure 4.1:** Distribution of amino acid residues in total and active sites of kinases.

The total number of hydrogen bonds observed between ligand and main chain is 2073. Out of the 2073 interactions, the ligand donates hydrogen bonds in 933 cases and accepts in 1140 cases (Table 4.1). Among these interactions, 53% interactions belong to nonpolar, 15% to polar and 29% to charged amino acid residues. As a donor, the frequently interacting residues are Leu (17%), Glu (20%) and His (13%). As an acceptor, ligand interacts with residues Gly (24%), and Leu (10%).

**Table 4.1:** Percent contribution of main chain donor/acceptor in ligand interaction.

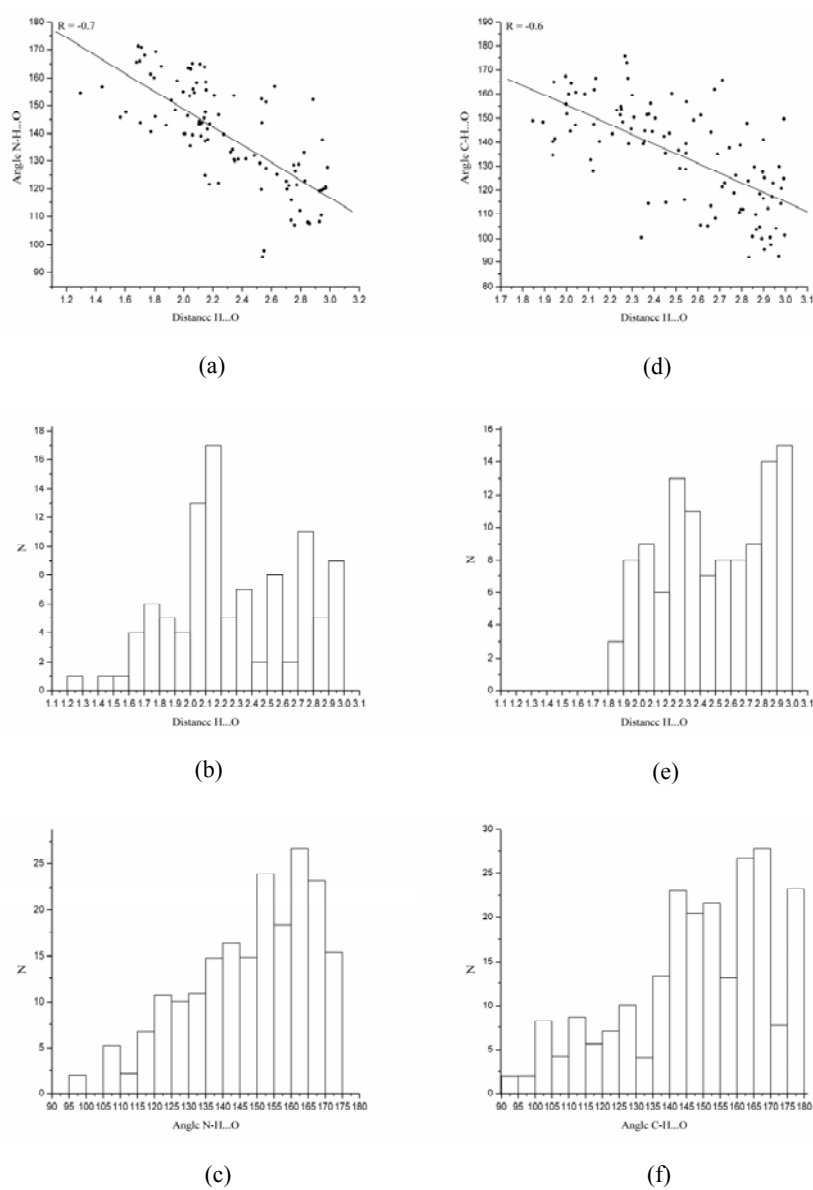
	Ligand donor main chain acceptor (in %)	Main chain donor ligand acceptor (in %)
<b>Nonpolar</b>		
Gly	4.5	24.4
Val	5.0	5.3
Leu	17.0	10.4
Ilu	5.7	1.4
Met	3.7	4.4
Pro	1.5	0.7
Phe	2.3	7.0
Trp	0.4	0.1
Ala	4.2	5.3
<b>Polar</b>		
Ser	3.2	6.0
Cys	1.9	1.8
Thr	1.4	5.7
Asn	0.4	1.4
Gln	4.9	2.0
Tyr	2.4	5.0
<b>Charged</b>		
Lys	2.6	4.5
Asp	3.3	5.0
Arg	2.9	4.1
Glu	19.9	4.2
His	13.0	1.0
Total no	933	1140

The ligand interacts with the main chain through strong and weak hydrogen bonds (Table 4.2). For reasons of clarity, the hydrogen bond abbreviation consists of three parts: hydrogen bond type, donor, and acceptor. B stands for backbone, L is ligand, W is water, D is donor, and A is acceptor. For example {NHO BD LA} signifies an N–H···O hydrogen bond involving a backbone N–H donor and a ligand O-atom acceptor. As an acceptor, ligand interacts with main chain through strong and weak hydrogen bonds, of which {NHO BD LA} 514 interactions, {CHO BD LA} 397, {NHN BD LA} 167, {CHN BD LA} 118 are noteworthy. Similarly ligand donates hydrogen bonds to the main chain, {NHO LD BA} 211, {CHO LD BA} 611, {OHO LD BA} 55 hydrogen bonds.

**Table 4.2:** Percent distribution of strong and weak hydrogen bonds among amino acid residues.

	{NHO BD LA} %	{CHO BD LA} %	{NHN BD LA} %	{CHN BD LA} %	{NHO LD BA} %	{CHO LD BA} %	{OHO LD BA} %
<b>Nonpolar</b>							
Gly	<b>20.0</b>	<b>41.4</b>	0.7	17.8	2.4	4.7	10.9
Ala	8.9	3.4	2.0	-	1.9	5.1	3.6
Val	6.1	2.4	12.2	2.5	4.7	5.0	3.6
Leu	4.3	3.4	<b>40.9</b>	20.3	27.4	14.8	5.4
Ilu	0.2	3.9	0.7	1.7	0.5	7.6	1.8
Met	1.9	0.5	26.5	-	3.3	4.0	-
Pro	-	2.1	-	-	0.9	1.5	1.8
Phe	3.4	5.3	0.7	<b>34.7</b>	0.5	2.6	5.4
Trp	0.2	-	-	-	0.5	0.4	-
<b>Polar</b>							
Ser	8.3	7.2	-	-	4.2	2.8	5.4
Cys	2.1	0.5	4.8	-	2.3	1.6	-
Thr	12.1	1.0	-	-	0.5	1.6	-
Asn	1.6	1.3	-	2.5	-	0.4	1.8
Gln	0.8	4.2	4.8	0.8	1.4	5.3	12.7
Tyr	1.9	8.2	-	14.4	0.9	2.9	3.6
<b>Charged</b>							
Lys	8.3	2.1	1.4	-	5.2	1.3	7.3
Arg	4.1	5.3	2.0	2.5	2.8	2.6	5.4
His	0.2	1.9	0.7	0.8	8.0	15.6	1.8
Asp	6.7	5.3	1.4	0.8	3.3	3.2	5.4
Glu	8.3	1.0	1.3	0.8	<b>28.9</b>	<b>16.8</b>	<b>23.6</b>
<b>TOTAL</b>	<b>514</b>	<b>397</b>	<b>167</b>	<b>118</b>	<b>211</b>	<b>611</b>	<b>55</b>

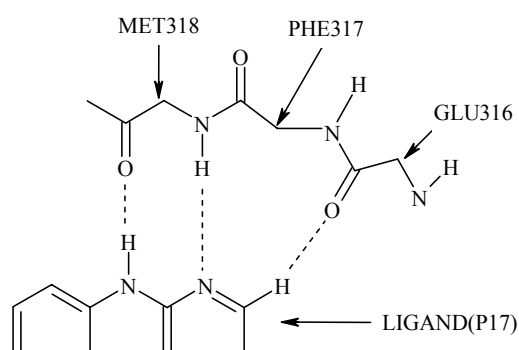
The percentage of {NHO BD LA} and {CHO BD LA} hydrogen bond interactions is highest for Gly. This reiterates the importance of Gly in the active sites of protein–ligand complexes. The possible reason of the occurrence of Gly in active sites has been discussed earlier and has to do with the small size and large conformational flexibility of this residue [4.4, 4.5]. The percentage of hydrogen bonds involving {NHN BD LA} is highest for Leu. This is because Leu is one of the frequently interacting residues in the active sites of CDK2 and also because the sub-family CDK2 represents the highest number of entries (49) in the dataset. For other types of hydrogen bonds, the percentage contribution of Leu is among the highest. Similar is the case for Glu wherein {NHO LD BA}, {CHO LD BA} and {OHO LD BA} hydrogen bonds are the most frequent. Glu is also one of the residues present at the linker region of CDK2 kinase. Apart from the CDK2 sub-family, Glu interacts commonly with ligand in other sub-families like cAMP-dependent protein kinase, PKA, FGFR1 and INSR. Phe interacts most frequently through {CHN BD LA} hydrogen bonds. In summary, various types of hydrogen bonds involving Gly, Leu and Glu are important in ligand binding in kinases. Hydrogen bond geometries for {NHO BD LA} in Gly and {CHO LD BA} in Glu are shown in Figure 4.2. For strong N–H $\cdots$ O hydrogen bonds in Gly, the median H $\cdots$ O distance,  $d$ , is less than 2.2 Å. The cone-corrected angular distributions for N–H $\cdots$ O has maxima in the range 160–165°. For the weak C–H $\cdots$ O hydrogen bond in Glu, the cone-corrected angular distributions have maxima in the range 165–170°. The inverse length-angle correlations are accordingly well-behaved in both N–H $\cdots$ O and C–H $\cdots$ O hydrogen bonds. The weak C–H $\cdots$ O hydrogen bond geometry demonstrated here is thus similar to the strong N–H $\cdots$ O hydrogen bond in its directional properties.



**Figure 4.2:** Hydrogen bond geometry for {NHO BD LA} in Gly (a)-(c) and {CHO LD BA} in Glu (d)-(f). In each case the inverse length-angle scatterplot is followed by histograms of distances and cone-corrected angular distributions.

### 4.3.2 Hydrogen bond motif and Synthons

The concept of supramolecular synthons is well-known in small molecule crystallography and crystal engineering [4.32]. Synthons are structural units within supermolecules which can be formed and/or assembled by known or conceivable synthetic operations involving intermolecular interactions. If the synthon is formed between the same functional group it is called as homosynthon, and if it is formed between two different functional groups it is referred to as a heterosynthon [4.33, 4.34]. The existence of synthons in protein–ligand complexes has been reported earlier [4.4]. In the present context, synthons are considered between various scaffolds present in the ligand molecules and the main/side chain of proteins in the active site (Table 4.3). The abbreviation for synthons is according to the short form of the hydrogen bond types. For example “CHO:NHN:NHO” represents a synthon formed by combination of  $C-H\cdots O:N-H\cdots N:N-H\cdots O$  types of hydrogen bonds (Scheme 4.2).



**Scheme 4.2:** Typical CHO:NHN:NHO synthon between main chain and ligand (PDB ID 1M52).

**Table 4.3:** Kinase protein–ligand complexes in this study. Main chain and side chain interactions are represented by M and S respectively. Hydrogen bond synthons are mentioned. Color code for various type of hydrogen bonds, (a) N–H...O red, (b) O–H...O pink, (c) N–H...N green, (d) C–H...O blue. Donor and acceptor are denoted as ‘d’ and ‘a’ respectively. Fields left blank represent complexes where other weak interaction like halogen bonds and water mediated protein–ligand interactions are observed.

Kinase sub-family		Source	PDB ID	M/S	Synthon type	Residue
1.	ABL1	Human	<i>IFPU</i>	M	NHN	MET318d
2.	ABL1	Human	<i>IEEP</i>	M	NHN	MET318d
3.	ABL1	Human	<i>IM52</i>	M	CHO:NHN:NHO	GLU136a:MET318d:MET318a
4.	Adenylate kinase	<i>Escherichia coli</i>	<i>IAKE</i>	M	NHN:CHO	VAL59d:VAL59a
5.	Adenylate kinase	<i>Saccharomyces cerevisiae</i>	<i>IAKY</i>	M	NHN:CHO	VAL63d:VAL63a
6.	Adenylate kinase	<i>Saccharomyces cerevisiae</i>	<i>IDVR</i>	M	NHO	GLN204a
7.	Adenylate kinase	<i>Zea mays</i>	<i>IZAK</i>	M	NHO	ALA195a
8.	BTK	Human	<i>IB55</i>	M	NHO	GLN15d
9.	BTK	Human	<i>IBWN</i>	M	NHO	GLN15d
10.	cAMP-dependent protein kinase	<i>Bos taurus</i>	<i>ISVE</i>	M	NHN:CHO	VAL123d:GLU121a
11.	cAMP-dependent protein kinase	<i>Bos taurus</i>	<i>IVEB</i>	M	NHN:CHO:CHO	VAL123d: VAL123a:GLU121a
12.	cAMP-dependent protein kinase	<i>Bos taurus</i>	<i>ISVH</i>	M	NHN:CHO	VAL123d:GLU121a
13.	cAMP-dependent protein kinase	<i>Bos taurus</i>	<i>ISVG</i>	M	NHN:CHO	VAL123d:GLU121a
14.	CK2	Human	<i>IOMI</i>	M	CHO	VAL116a
15.	CSK	Human	<i>IBYG</i>	M	NHO	MET269d
16.	c-Src	Human	<i>IO4A</i>	M	NHO	HIS60a
17.	c-Src	Human	<i>IO4B</i>	M	NHO	HIS60a
18.	c-Src	Human	<i>IO4D</i>	M	NHO:NHO	ARG34a:ARG34d
19.	c-Src	Human	<i>IO4E</i>	M	NHO:NHO	ILE63d:TYR61a
20.	c-Src	Human	<i>IO4F</i>	M	NHO	GLU37d
21.	c-Src	Human	<i>IO4G</i>	M	NHO	GLU37d
22.	c-Src	Human	<i>IO4H</i>	M	NHO	GLU37d

23.	c-Src	Human	IO4I	M	NHO	GLU37d
24.	c-Src	Human	IO4J	M	NHO	GLU37d
25.	c-Src	Human	IO4K	M	NHO	GLU37d
26.	c-Src	Human	IO4L	M	NHO	GLU37d
27.	c-Src	Human	IO4M	M	NHO	GLU37d
28.	c-Src	Human	IO4N	M	NHO:NHO	GLU37d:THR38d
29.	c-Src	Human	IO4O	M	NHO	GLU37d
30.	c-Src	Human	IO4P	M	NHO	GLU37d
31.	c-Src	Human	IO4Q	M	NHO	GLU37d
32.	c-Src	Human	IO4R	M	NHO:NHO	GLU37d:THR38d
33.	c-Src	Human	IO4I	M	NHO	GLU37d
34.	c-Src	Human	IO42	M	NHO:NHO	GLU37d:HIS60a
35.	c-Src	Human	IO43	M	NHO:NHO	GLU37d:HIS60a
36.	c-Src	Human	IO44	M	NHO:NHO	GLU37d:HIS60a
37.	c-Src	Human	IO45	M	NHO:NHO	GLU37d:HIS60a
38.	c-Src	Human	IO46	M	NHO:NHO	GLU37d:HIS60a
39.	c-Src	Human	IO47	M	NHO:NHO	GLU37d:HIS60a
40.	c-Src	Human	IO48	M	NHO:NHO	GLU37d:HIS60a
41.	c-Src	Human	IO49	M	NHO:NHO	GLU37d:HIS60a
42.	c-Src	Human	IY57	M	CHO:NHN:NHO	GLU39a:MET341d:MET341a
43.	SRC_RSVSA	Human	INZL	M	NHO:NHO	GLU284d:HIS307a
44.	SRC_RSVSA	Human	INZV	M	NHO:NHO	GLU284d:HIS307a
45.	CDK2	Human	IFVT	M	NHO	LEU83d
46.	CDK2	Human	IP5E	-	-	-
47.	CDK2	Human	IOGU	M	NHO:NHN:NHO	GLU81a:LEU83d:LEU83a
48.	CDK2	Human	IOI9	M	CHO:NHO:NHN	GLU81a:LEU83a:LEU83a
49.	CDK2	Human	IOIY	M	NHO:NHN	GLU81a:LEU83d
50.	CDK2	Human	IGZ8	M	NHN:NHO	LEU83d:LEU83a
51.	CDK2	Human	IJVP	M	NHN:NHO:CHO	LEU83d:LEU81a:LEU81a
52.	CDK2	Human	IHO0	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
53.	CDK2	Human	IOIT	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
54.	CDK2	Human	IHOI	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
55.	CDK2	Human	IHO8	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a

56.	CDK2	Human	IEIX	M	NHO:NHN:NHO	GLU81a:LEU83d:LEU83a
57.	CDK2	Human	IHO7	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
58.	CDK2	Human	2BHE	M	NHO:NHO:NHO	GLU81a:LEU83d:LEU83a
59.	CDK2	Human	IOIR	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
60.	CDK2	Human	IEIV	M	NHN:NHO	LEU83d:LEU83a
61.	CDK2	Human	IAQ1	M	NHO:NHO	GLU81a:LEU83d
62.	CDK2	Human	IB38	M	NHN:NHO	LEU83d:LEU83a
63.	CDK2	Human	IR78	M	NHO:NHO	GLU81a:LEU83d
64.	CDK2	Human	IGII	M	NHO:NHO	VAL83d:VAL83a
65.	CDK2	Human	IHIS	M	NHN:NHO	LEU83d:LEU83a
66.	CDK2	Human	IY8Y	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
67.	CDK2	Human	IHIR	M	NHN:NHO	LEU83d:LEU83a
68.	CDK2	Human	IPYE	M	CHO:NHO:NHN	GLU81a:LEU83a:LEU83d
69.	CDK2	Human	IKE6	M	NHO:NHO:CHO	GLU81a:LEU83d:LEU83a
70.	CDK2	Human	IKE7	M	NHO:NHO:NHO	GLU81a:LEU83d:LEU83a
71.	CDK2	Human	IKE8	M	NHO:NHO:NHO	GLU81a:LEU83d:LEU83a
72.	CDK2	Human	IKE9	M	NHO:NHO	GLU81a:LEU83d
73.	CDK2	Human	ICKP	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
74.	CDK2	Human	IB39	M	NHO:NHN	GLU81a:LEU83d
75.	CDK2	Human	IDM2	M	NHO:NHO:NHO	GLU81a:LEU83d:LEU83a
76.	CDK2	Human	IHIP	M	NHO:NHN	LEU83a:LEU83d
77.	CDK2	Human	IY9I	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
78.	CDK2	Human	IDI8	M	CHO:NHN:CHO	GLU81a:LEU83d:LEU83a
79.	CDK2	Human	IGIJ	M	NHO:NHO	VAL83d:VAL83a
80.	CDK2	Human	IW0X	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
81.	CDK2	Human	IKE5	M	NHO:NHO:NHO	GLU81a:LEU83d:LEU83a
82.	CDK2	Human	IYYZ	M	NHO:NHN:NHO	GLU81a:LEU83d:LEU83a
83.	CDK2	Human	IYIK	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
84.	CDK2	Human	IOIQ	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
85.	CDK2	Human	IP2A	M	NHO:NHO:NHO:CHO	GLU81a:LEU83d:LEU83a:HIS84a
86.	CDK2	Human	IE9H	M	NHO:NHO:NHO	GLU81a:LEU83d:LEU83a
87.	CDK2	Human	IHIQ	M	NHN:NHO	LEU83d:LEU83a
88.	CDK2	Human	IPF8	M	NHO:NHO	GLU81a:LEU83d

89.	CDK2	Human	2BHH	M	NHO:NHO:NHO	GLU81a:LEU83d:LEU83a
90.	CDK2	Human	IG5S	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
91.	CDK2	Human	IGH	M	CHO:NHO:NHO	GLU81a:LEU83d:LEU83a
92.	CDK2	Human	IURW	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
93.	CDK2	Human	IH0W	M	NHO	LEU83a
94.	CDK5	Human	IUNL	M	CHO:NHN:NHO	GLU81a:CYS83d:CYS83a
95.	CDK6	Herpivirus saimiri	IXO2	M	NHO	VAL101d
96.	CHK1	Human	INTQ	M	NHO:NHO	GLU85a:CYS87d
97.	CHK1	Human	INVR	M	NHO:NHO	GLU85a:CYS87d
98.	CHK1	Human	INVS	M	NHO:NHO	GLU85a:CYS87d
99.	C-KIT tyrosine kinase	Human	I746	M	CHO:NHN:CHO	GLU671a:CYS673d:CYS673a
100.	Death-associated protein kinase	Human	IP4F	M	CHO:CHO	GLU94a: VAL96a
101.	EGFR	Human	IM17	M	NHN:CHO	MET769d:GLN767a
102.	EGFR	Human	IXKK	M	NHN:CHO	MET793d:GLN791a
103.	FGFR1	Human	IAGW	M	NHO:NHO	ALA564d:GLU562a
104.	FGFR1	Human	IFGI	M	CHO:NHO	GLU562a:ALA564d
105.	FGFR1	Human	2FGI	M	NHO:NHN	ALA564a:ALA564d
106.	Glycerol kinase	Escherichia coli	IBO5	M	NHO	ARG83d
107.	Glycerol kinase	Escherichia coli	IBWF	M	NHO	GLY411d
108.	GSK3b	Human	IGNG		-	-
109.	GSK3b	Human	IQ3D	M	NHO	VAL135d
110.	GSK3b	Human	IQ3W	M	NHO:NHO	VAL135a:VAL135d
111.	GSK3b	Human	IQ4I	M	NHO:NHO:NHO	VAL135a:VAL135d:ASP133a
112.	GSK3b	Human	IQ4L	M	NHO:NHO	VAL135d:ASP133a
113.	GSK3b	Human	IQ5K	M	NHO:NHN:NHO	VAL135a : VAL135d: PRO136a
114.	GSK3b	Human	IR0E	M	NHO:NHO	VAL135d:ASP133a
115.	GSK3b	Human	IUV5	M	NHO:NHO:NHO	VAL135a:VAL135d: ASP133a
116.	HCK	Human	2HCK	M	NHO	MET341d
117.	Hexokinase type I	Rattus norvegicus	IBG3	S	OHO:OHO	ASP209a:ASP209a
118.	JNK3	Human	IPMN	M	NHN:CHO	MET149d:GLU147a
119.	JNK3	Human	IPMU	M	NHN:CHO	MET149d: MET149a
120.	JNK3	Human	IPMQ	M	NHN:CHO	MET149d:GLU147a
121.	JNK3	Human	IPMV	M	CHO:NHN:CHO	MET149a:MET149d:GLU147a

122.	INSR	Human	<i>IRQQ</i>	M	NHN:NHO	MET1079d:GLU1077a
123.	INSR	Human	IGAG	M	NHN:NHO	MET1079d:GLU1077a
124.	ERK2	Human	ITVO	M	NHN:NHO	MET108d:ASP106a
125.	ERK2	Human	<i>IPME</i>	M	NHN:CHO	MET108d:HIS106a
126.	ERK2	Rattus norvegicus	3ERK	M	NHN:NHO	MET106d:ASP104a
127.	ERK2	Rattus norvegicus	4ERK	M	CHO:NHN:NHO	ASP104a:MET106d:MET106a
128.	MAPK14	Human	IBL6	M	NHN:CHO	MET109d:HIS107a
129.	MAPK14	Human	IBL7	M	NHN:CHO	MET109d:HIS107a
130.	MAPK14	Human	IBMK	M	NHN:CHO	MET109d:HIS107a
131.	MAPK14	Human	IKV1	M	NHO	ASP168d
132.	MAPK14	Human	IKV2	M	NHO	ASP168d
133.	MAPK14	Human	IOUK	M	NHN:CHO	MET109d:HIS107a
134.	MAPK14	Human	IW84	M	NHN	MET109d
135.	MAPK14	Human	IOZ1	M	NHN	MET109d:HIS107a
136.	MAPK14	Human	IW82	M	NHO	ASP168d
137.	MAPK14	Human	IW83	M	NHO	ASP168d
138.	MAPK14	Human	IWBN	M	NHO	ASP168d
139.	MAPK14	Human	IWBO	S	OHO	ASP168a
140.	MAPK14	Human	IWBS	M	NHN:CHO	MET109d:HIS107a
141.	MAPK14	Human	IWBT	M	NHN:CHO	MET109d:HIS107a
142.	MAPK14	Human	IWBV	M	NHO	ASP168d
143.	MAPK14	Human	IWBW	M	NHN:NHO	MET109d:HIS107a
144.	MAPK14	Human	IYQJ	M	NHN:CHO	MET109d:HIS107a
145.	MAPK14	Human	IZZ2	S	OHO	ASP150a
146.	MAPK14	Human	<i>I49U</i>	M	CHO:NHN	HIS107a:MET109d
147.	MAPK14	Human	IDI9	M	CHO:NHN:CHO	HIS107a:MET109d:MET109a
148.	MAPK14	Human	IM7Q	M	NHO:NHO	MET109d:HIS107a
149.	MAPK14	Human	IOUY	M	NHO:NHO:NHO	HIS107a:MET109d:GLY110d
150.	MAPK14	Human	IOVE	M	NHO:NHO	MET109d:GLY110d
151.	MAPK14	Human	IW7H	M	NHN:NHO	MET109d:HIS107a
152.	MAPK14	Mus Muusculus	IYWR	M	NHN:NHO	MET109d:HIS107a
153.	MET	Human	<i>IR0P</i>	M	NHO:NHO	PRO1158a:MET1160d
154.	Nucleoside-diphosphate kinase	Dictyostelium discoideum	IBUX	S	NHO:NHO	LYS16d:HIS122d

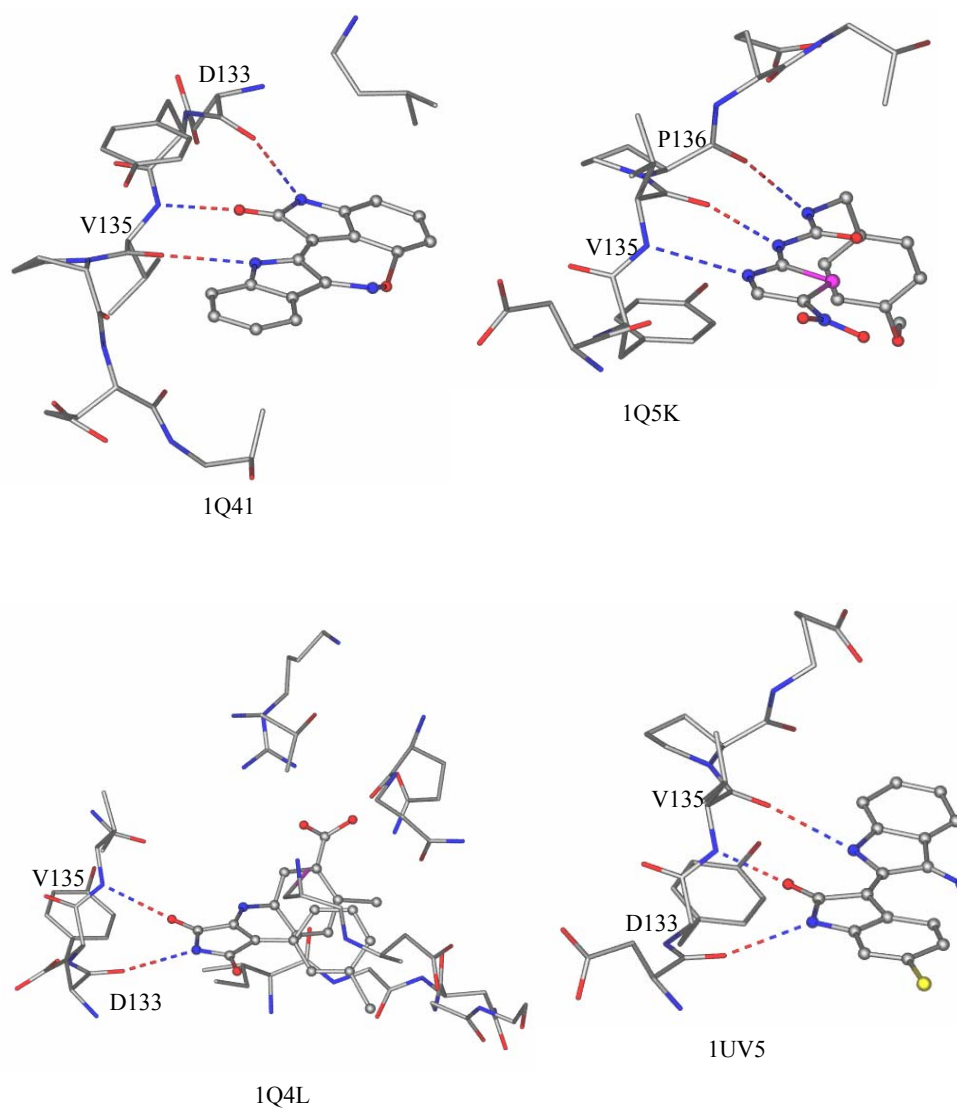
155.	P56-LCK	Human	IIJR	M	NHO:NHO:NHO	LYS60a:LYS60d:HIS58a
156.	P56-LCK	Human	IQPE	M	CHO:NHN:NHO	MET319a:MET319d:GLU317a
157.	PKD1	Human	IUU3	M	NHO:NHO	SER160a:ALA162d
158.	PKD1	Human	IUVR	M	NHO:NHO	SER160a:ALA162d
159.	PKD1	Spodoptera frugiperda	IUU9	M	NHO:NHO	SER160a:ALA162d
160.	PKD1	Human	IUU7	M	NHO:NHO	SER160a:ALA162d
161.	PKD1	Human	IUU8	M	NHO:NHO	SER160a:ALA162d
162.	PKD1	Human	IOKZ	M	NHO:NHO	SER160a:ALA162d
163.	Phosphoinositide 3 Kinase Gamma	Sus scrofa	IE8W	M	OHO:NHO	GLU880a:VAL882d
164.	Phosphoinositide 3 Kinase Gamma	Human	IE8Z	M	NHO	VAL882d
165.	Phosphoinositide 3 Kinase Gamma	Sus scrofa	IE7V	M	NHO	VAL882d
166.	Phosphoinositide 3-Kinase Gamma	Sus scrofa	IE90	M	NHO	VAL882d
167.	Phosphoenolpyruvate Carboxykinase	Human	IM51	M	NHO	PHE530d
168.	Phosphoenolpyruvate Carboxykinase	Human	INHX	M	NHO	PHE530d
169.	Phosphoglycerate kinase	Sus scrofa	IKF0	M	NHO	GLY312a
170.	Phosphoglycerate kinase	Trypanosoma brucei	I6PK	M	NHO	ALA314a
171.	PKA	Human	IBX6	M	NHO	VAL123d
172.	PKA	Human	IRE8	M	NHO	VAL123d
173.	PKA	Human	IREK	M	NHO	VAL123d
174.	PKA	Human	IBKY	M	NHN:NHO	VAL123d:GLU121a
175.	PKA	Human	IREJ	M	NHO	VAL123d
176.	PKA	Human	IFMO	M	NHN:NHO	VAL123d:GLU121a
177.	PKA	Human	IJBP	M	NHN:NHO	VAL123d:GLU121a
178.	PKA	Human	IRDQ	M	NHN:NHO	VAL123d:GLU121a
179.	PKA(alpha)	Human	ICDK	M	NHN:NHO	VAL123d:GLU121a
180.	PKA(alpha)	Human	IQ24	M	NHN:NHO	VAL123d:GLU121a
181.	PKA(alpha)	Human	ISZM	M	NHN:NHO	VAL123d:GLU121a
182.	PKA(alpha)	Human	IQ8T	M	NHO:NHO	ALA123d:GLU121a
183.	PKA(alpha)	Human	IQ8U	M	NHN:CHO	VAL123d:GLU121a
184.	PKA(alpha)	Human	IQ8W	M	NHN:CHO	VAL123d:GLU121a
185.	PKA(alpha)	Human	ISTC	M	NHN:CHO:CHO	VAL123d:GLU121a:VAL123a
186.	PKA(alpha)	Human	IYDR	M	NHN:NHO	VAL123d:GLU121a
187.	PKA(alpha)	Human	IYDS	M	NHN:CHO	VAL123d:GLU121a

188.	PKA(alpha)	Human	IYDT	M	NHN:CHO	VAL123d:GLU121a
189.	PKA(beta)	Sus scrofa	ICTP	-	-	-
190.	Protein kinase C- $\iota$	Human	IZRZ	M	NHO:NHO:CHO	GLU324a:VAL326d:VAL326a
191.	Protein kinase C- $\iota$	Human	IKPF	M	CHO	HIS42a
192.	PLK1	Human	IUMW	-	-	-
193.	PIM-1	Human	IAXWS	M	NHO	GLU121a
194.	TGFbetaR1	Human	IRW8	-	-	-
195.	TGFbetaR1	Human	IPY5	M	NHN:CHO	HIS283d:ASP281a
196.	Thymidine kinase	Herpes simplex virus	IE2K	S	NHO	GLN125d
197.	Thymidine kinase	Herpes simplex virus	IE2N	S	NHO	GLN125d
198.	Thymidine kinase	Herpes simplex virus	IE2P	S	NHO	GLN125d
199.	Thymidine kinase	Herpes simplex virus	IQHI	S	NHO:NHN	GLN125d:ARG176d
200.	Thymidine kinase	Ureaplasma urealyticum	IXMR	M	NHO:NHO	PHE128d:LYS1880a
201.	Thymidine kinase	Human herpesvirus 1	2K15	S	NHO	GLN125d
202.	Thymidine kinase	Human herpesvirus 2	IP7C	S	NHO	GLN125d
203.	Thymidine kinase	Herpes simplex virus	IE2L	S	NHO	GLN125d
204.	Thymidylate kinase	Human	IE9A	S	NHO	ARG76a
205.	Thymidylate kinase	Human	IE9B	S	NHO	ARG76a
206.	Thymidylate kinase	Human	IE9C	S	NHO	ARG76d
207.	Thymidylate kinase	Human	IE9D	S	NHO	ARG76d
208.	Thymidylate kinase	Human	IE9E	S	NHO	ARG76d
209.	Thymidylate kinase	Human	IE98	S	NHO	ARG76d
210.	Thymidylate kinase	Human	IE9F	S	NHO	ARG76d
211.	Thymidylate kinase	Saccharomyces cerevisiae	3TMK	S	NHO	ARG73d
212.	Thymidylate kinase	Escherichia coli	5TMP	S	NHO:NHO	ARG78d:GLN109d
213.	Thymidylate Kinase	Human	IE2D	S	NHO	ARG76d
214.	Thymidylate Kinase	Human	IE2E	S	NHO	ARG76d
215.	Thymidylate Kinase	Human	IE2F	S	NHO	ARG76d
216.	Thymidylate Kinase	Human	IE2G	S	NHO	ARG76d
217.	Thymidylate Kinase	Human	IE2Q	S	NHO	ARG76d
218.	Thymidylate Kinase	Mycobacterium tuberculosis	IMRN	S	NHO	ARG74d
219.	Thymidylate Kinase	Mycobacterium tuberculosis	IMRS	S	NHO	ARG74d
220.	Thymidylate Kinase	Mycobacterium tuberculosis	IG3U	S	NHO	ARG74d

221.	Thymidylate Kinase	Mycobacterium tuberculosis	1W2G	S	NHO	ARG74d
222.	Thymidylate kinase	Escherichia coli	4TMK	S	NHO	GLN109d
223.	Tyrosine kinase LCK	Human	IFBZ	M	NHO	HIS58a
224.	Uridine-cytidine kinase 2	Human	IUEI	S	NHO	ARG176d
225.	Uridine-cytidine kinase 2	Human	IUDW	S	NHO	ARG176d
226.	Uridine-cytidine kinase 2	Human	IXRJ	S	NHO	ARG176d
227.	Uridylmonophosphate/ Cytidylmonophosphate kinase	Dictyostelium discoideum	IQF9	M	NHO	VAL65d
228.	Uridylmonophosphate/ Cytidylmonophosphate kinase	Dictyostelium discoideum	3UKD	M	NHO	VAL65d
229.	Uridylmonophosphate/ Cytidylmonophosphate kinase	Dictyostelium discoideum	4UKD	M	NHO	VAL65d
230.	VEGFR2	Human	IY6A	M	CHO:NHN:NHO	GLU915a: CYS917d: CYS917a
231.	VEGFR2	Human	IY6B	M	CHO:NHN:NHO	GLU915a: CYS917d: CYS917a
232.	VEGFR2	Human	IYWN	M	NHN:NHO	CYS917d: GLU915a
233.	Wee1A kinase	Human	IX8B	M	NHN:NHO	CYS379d: GLU377a

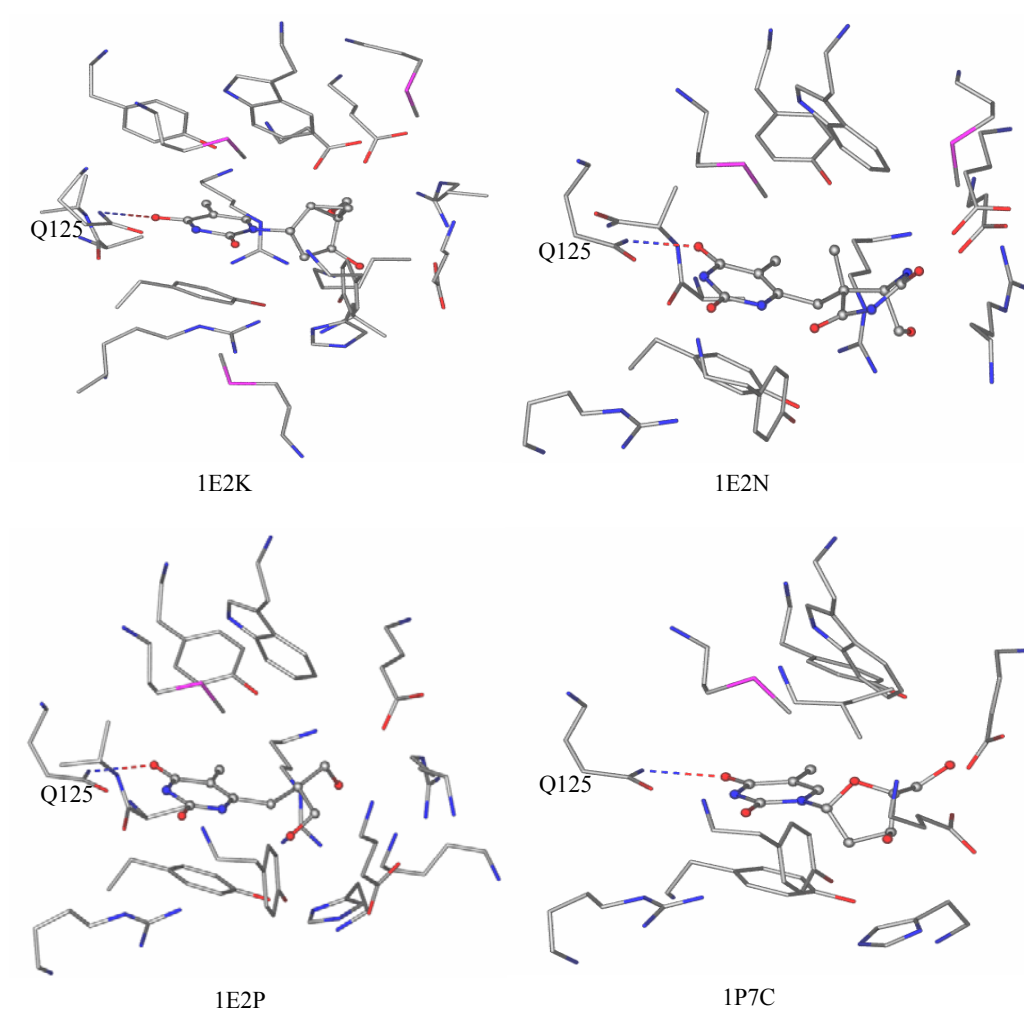
Both homosynthons and heterosynthons exist in protein–ligand complexes of kinases. The synthons are formed with a combination of N–H···O, C–H···O and N–H···N hydrogen bonds (Table 4.3). Rarely do O–H···O hydrogen bonds participate in synthon formation. The percentages of synthons formed with one hydrogen bonds (single-point) is 36%, two hydrogen bonds (two-point), 42%, three hydrogen bonds (three-point), 21% and four-point, 0.4%. Certain residues in each sub-family frequently interact with the ligand. Residues frequently interacting with ligands in different sub-families are: cAMP-dependent protein kinase sub-family, VAL123 and GLU121; CDK2 sub-family, GLU81 and LEU83; CHK1 sub-family CYS87 and GLU85; PKA sub-family, VAL123 and GLU121. These are the conserved residues present in the particular sub-family. These residues have N–H···O, C–H···O or N–H···N hydrogen bonds while interacting with various ligands within a sub-family. For example, GLU81 in the CDK2 sub-family interacts with ligands through N–H···O, C–H···O and N–H···N hydrogen bonds. Therefore in a particular kinase sub-family, conserved amino acid residues always participate in hydrogen bond interactions with a variety of ligands, so that a particular synthon is retained intact. This reflects the multifaceted hydrogen bond capability of key residues in any particular kinase sub-family. Sometimes, the type of hydrogen bonds are retained in a synthon of a particular subfamily, the best example being sub-families GSK3b, EGFR, INSR, PDK1, and a few members of PKA (Table 4.3).

Synthons in GSK3b sub-family are composed of a similar hydrogen bond patterns (Figure 4.3). VAL135 and ASP133 are observed to have NHO:NHO:NHO synthons which are retained across a variety of ligands. However, swapping of donor-acceptor pairs is observed between ligand and protein. Significantly, however, this swapping never hinders the formation of desired synthons. In summary, in each particular sub-family of kinase, various ligands bind to the key residues of the main chain with a typical synthon pattern. Of course, these inhibitors certainly have similar functional groups and scaffolds which support such interaction consistency.



**Figure 4.3:** Synthons formed by N-H...O hydrogen bonds in key residues, are shown for the GSK3b sub-family with the respective PDB ID.

The thymidine/thymidylate kinase is unusual because ligands for this sub-family generally form the NHO synthon with the side chain residue GLN125 (Figure 4.4).



**Figure 4.4:** Side chain participation in synthon formation in Thymidine kinase sub-family.

### 4.3.3 Role of conserved residue

The kinase domains across many sub-families share a great percentage of similarity in the active site residues. Hence there is a chance that residues at the same position might be involved in ligand binding. To verify this assumption an active site sequence alignment was carried out for 35 sub-families of human source, each structure representing a sub-family (Table 4.3, PDB ID with italics). Surprisingly, it was observed that residues interacting with ligands are aligned at a particular position across all sub-families with very few exceptions. The exceptions (Figure 4.5) are interactions involving side chains and adjacent residues. This finding reveals that residues determining the ligand interaction with the main chain are conserved across all the kinase sub-families. Despite the variation in amino acid substitutions across kinase sub-families, the hydrogen bond signature pattern is always retained in the conserved residues. To verify this hypothesis, sub-families of receptor tyrosine kinases are discussed here.

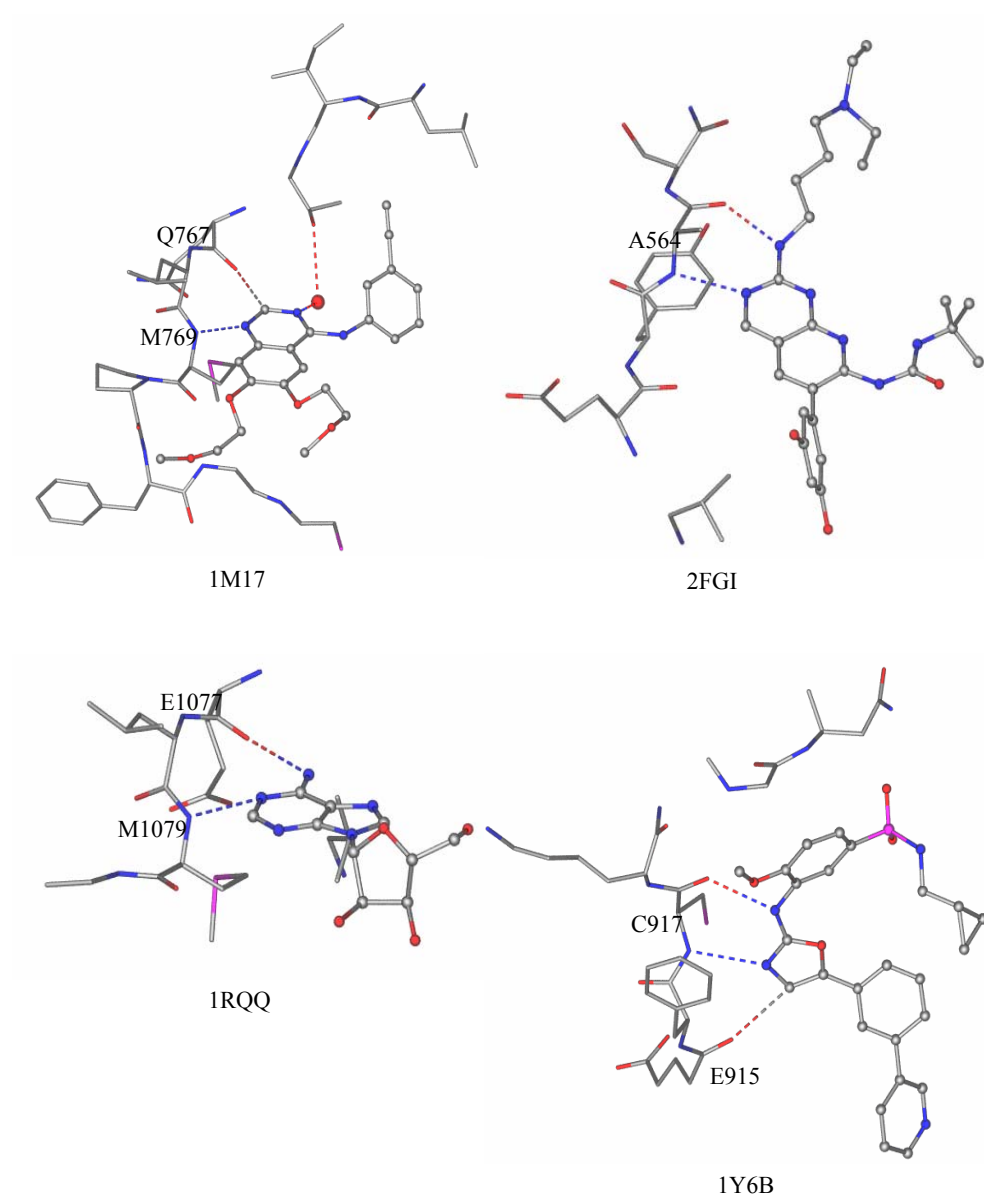
The receptor tyrosine kinases considered in the 233 datasets are EGFR, FGFR1, VEGFR2 and INSR. For these receptor tyrosine kinases, the conserved residues interact with ligands in a similar manner (Figure 4.6). The amino acid residues involved in ligand binding in these structures are aligned at a particular position (Figure 4.5). In EGFR, the interacting residues are MET769 and GLN767. For the FGFR1 sub-family the pertinent residue is ALA564. In VEGFR2, the residues are GLU915 and CYS917. In sub-family INSR, the residues are MET1079 and GLU1077. These residues interact with ligands through  $\text{N-H}\cdots\text{O}$ ,  $\text{N-H}\cdots\text{N}$  and  $\text{C-H}\cdots\text{O}$  hydrogen bonds. Therefore it is assumed that the sample receptor tyrosine kinases are similar on the basis of similar interaction patterns. These findings can also be extrapolated to other kinase sub-families. This finding has a potential role in multikinase targeting, where a single molecule can be targeted against several kinases.

```

1okz.pdb      KVPYVTRERDVMSRLDHPFFVKLYFTFQDDEKLYFGLSYAKNGELLKYIRKIGSFDETCT M
1bkx.pdb      -----ELFLVKLEF-----MVMEYVAGGEMFSHY-----M
1cdk.pdb      -----KVLQHTELFLVKLEY-----MVMEYVPGGEMFSHYR-----M
1zrz.pdb      -----VEKVFASFLVGLHS-----CLFFVIEYVNGGDLMFHMQ-----M
1om1.pdb      -----KEIILCNIVKL-----LDSLIFEYVNNT-DFKVHD-----M
1p4f.pdb      -----VELIVIT-----LHELILELVAGG-ELD-HD-----M
1nvq.pdb      -----IEIINLVVVKF-----YGYLFLEYCSGGELFDRHDI-----M
1py5.pdb      -----WREAEIY-----QILGFIAAD-----LWLVSDEYHEHGSLLFDYHDL-----M
1x8b.pdb      -----SNAREVA-----HVVRYFSAW-----MLIQNEYCNGGSLADAHD-----M
1fpu.pdb      -----LKEAAVMKEIKHNVLQLLGVC-----FYIITEFMTYGNLLDYA-----M
1qpe.pdb      -----LAEANLMK-----LVRLYAVV-----IYIITEYMEGSLVDF-----M
2hck.pdb      -----KLHA-----YIITEFMAKGSLLDFK-----M
1m17.pdb      -----DEAYVM-----AVVCRLLGIC-----VQLITQLMPFGCLLDYR-----M
1xws.pdb      -----EVL-----LSVIRLLDWF-----VLILERPEPVQDLDF-----M
1rqq.pdb      -----EEF-----EMVVRLLG-----VVMELMAHGDLKSYR-----M
1r0p.pdb      -----FMVLSLLGI-----LVVLPMKHGDLRNFN-----M
2fqi.pdb      ---DLISEMEMMKIG-KHNIINLLGACT---LYVIVEYASKGNLR---EYLQA---M
1y6b.pdb      ---ALMSELKILIV-----VNLLGACT---LMVIVEFCKFGNLS---TYLR---M
1t46.pdb      ---LMSELKVLSYLGNHNIVNLLGACT---TLVITEYCCYGDLLNFRVGMFLA---M
1byg.pdb      ---LVQLLG-----IVTEYMAKGSLV---DYLRS---M
1un1.pdb      -----EILNIVR---LHDVLTLVFEFCDDQLKKY-----M
1gz8.pdb      -----IVK---L---LDLVFEFLHQDLKK-----M
1q41.pdb      -----ELMLIVR---LR---YNLVLDPVETVY-----M
1pme.pdb      -----EIILLEI---IG---INDIIVYLVTHLMGAD-----M
1pmn.pdb      -----HRAELL---MIISLLNVFVYLVMEMLMDANLC-----M
1a9u.pdb      -----HRTELLLLKH-MVIGLLDVFVYLVTHLMGAD-----M
1o4a.pdb      -----CLSVSD-----KNVKHYKIR---YITSR-----M
1nz1.pdb      -----CLSVSDD-----KNVKHYKIRF-YITSR-----M
1fbz.pdb      -----CLSVRDF-----DVVKHYKIRNFYISPR-----M
1b55.pdb      -----LFLSYEYS-----M
1udw.pdb      -----QFNFDHPAFDVYDFVSHSRKFEGILAFVDTTRLSRRVR-----S
1e8z.pdb      -----QDMLILLRLPYGCISKIGMIEIVKDATTIAKIQ-----M
1e9a.pdb      -----VDHSVHLLFSAN-----RWEQVDRYAFSGVAFTGA-----M
1umw.pdb      -----NYMSEHLKAGAWFR-----M
1m51.pdb      -----PLSEAVNWFRKDGKFLWPGFGENSRLVLMGLHIL-----M

```

**Figure 4.5:** Active site sequence alignments for 35 entries of human kinases, each representing a subfamily. Key residues are colored red. ‘M’ stands for main chain and ‘S’ for side chain.

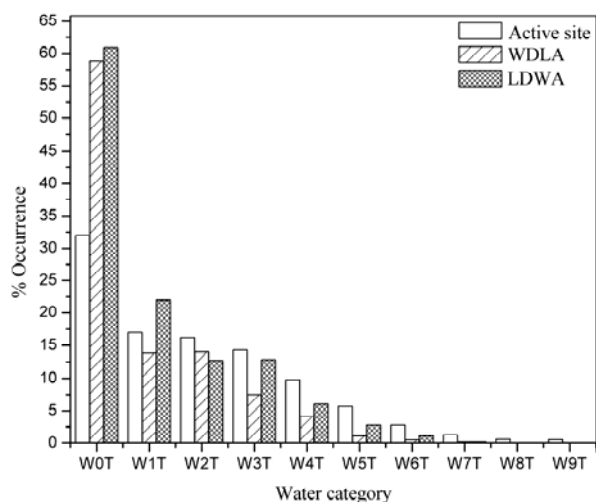


**Figure 4.6:** Conserved synthon across receptor tyrosine kinases (EGFR;1M17, FGFR1; 2FGI, INSR;1RQQ, VEGFR2;1Y6B).

#### **4.3.4 Active site solvation**

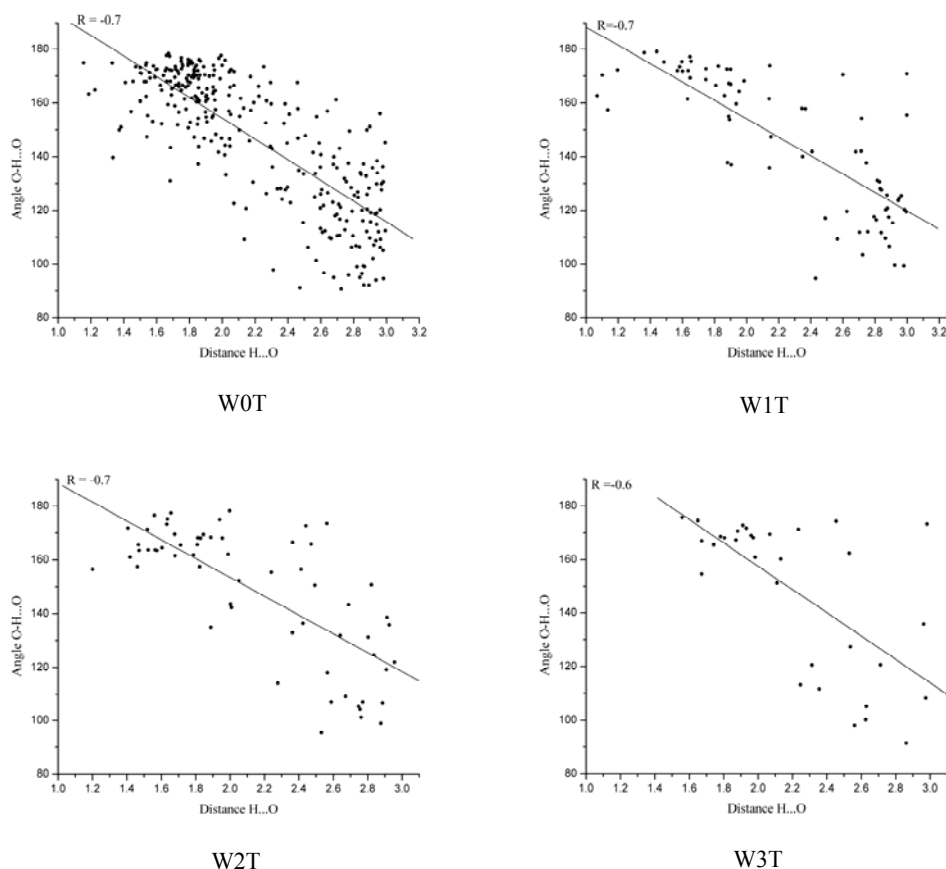
Solvation and desolvation are crucial phenomena which occur in the process of ligand receptor interaction [4.35]. During binding, the active site undergoes a rapid desolvation process to accommodate the ligand. This process finally results in the release of ordered water favoring the formation of the ligand–receptor complex. While separate, both enzyme and substrate force neighboring water molecules into an ordered shell. Binding of substrate to enzyme releases some of the ordered water, and the resulting increase in entropy provides a thermodynamic push toward formation of the ligand–receptor complex. Some of the ordered water molecules in the receptor are unaffected by this rapid desolvation process and are retained in the active site due to their structural and functional importance. These structural/functional waters usually form hydrogen bonds with neighboring residues or/and ligand and even with water molecules thus increasing the enthalpy of the final complex and all this stabilizes the ligand–protein complex. A classification of water molecules in the active sites of kinases, based on accessible surface area (exposure) was carried out. Buried water is of structural and/or functional significance while partially accessible or completely accessible water (category W1T to W9T) is less important. Structurally, the buried water molecules (W0T) are important because, very often they form hydrogen bonds, and sometimes they mediate hydrogen bonds between protein and ligand [4.36]. Functionally, these water molecules may also participate in enzyme catalysis.

The dataset contains a total of 40944 water molecules, out of which 6066 (15%) are present in the active sites. This means that, on average, there are 176 and 26 water molecules in the entire protein and in the active site respectively per structure. The percentage distribution of these water molecules in various categories based on percentage of exposure, and also the hydrogen bond between water and ligand is shown in Figure 4.7.



**Figure 4.7:** Hierarchical classification of active site water molecules. Also shown is the percentage of interacting water as an acceptor and as donor.

The majority of the active site waters are inaccessible (category W0T). This suggests that the active sites are present deep inside the proteins. Very often, the water molecules are enclosed even at the protein–ligand interface. The hydrogen bonds considered here are confined to those between ligand and water. The various types of hydrogen bonds are {OHN WD LA}, {OHO WD LA}, {CHO LD WA}, {NHO LD WA} and {OHO LD WA}. Approximately 60% of the hydrogen bonds are manifested by the W0T category of the active site water molecules. The hydrogen bonds between ligand and water are observed up to the category W6T. Beyond this an exposed water molecule rarely interacts with the ligand. The hydrogen bond geometry also depends on the water category. Hydrogen bonds geometries are better represented in the W0T category (Figure 4.8). The quality of hydrogen bonding gradually decreases from W1T to W6T. For example, the  $d$ - $\theta$  scatterplot for {OHO WD LA} is better represented in the W0T category, with the correlation becoming poorer as one moves from W1T to W3T (Figure 4.8). This suggests that buried water molecules in the active sites very often participate in hydrogen bonds with the ligand. Hence they are structurally important.



**Figure 4.8:** Inverse length-angle ( $d-\theta$ ) scatterplots for {OHO WD LA} for various categories of water (W0T to W3T). Notice the poor correlation in W3T.

#### 4.4 Conclusions

Strong (N–H...O, O–H...O, N–H...N) and weak (C–H...O, C–H...N) hydrogen bonds in the active sites of the kinase family have been studied in a dataset of 233 protein–ligand complexes. The kinase family is dominated by nonpolar and charged residues. Residues like Leu, Glu and His frequently accept hydrogen bonds from the ligand, while Gly and Leu are favored donors. Both strong and weak hydrogen bonds are of comparable importance in ligand binding. In this context, the acceptor capacity of main chain Glu is noteworthy and the geometry of C–H...O hydrogen bonds to Glu is on par with strong hydrogen bonds in the active sites. The hydrogen bonded supramolecular synthons formed between main/side chain and ligand atoms is a typical characteristic of the kinase family. These synthons are

formed by an amalgamation of strong and weak hydrogen bonds. The synthon patterns are unique to kinase sub-families. Relationships between the sub-families are established on the basis of similar synthon patterns. The similarities in synthon patterns across sub-families arise due to the conserved residues in the active sites of kinase. The active site water molecules exist in a variety of environment. Water molecules which are least exposed usually take part in hydrogen bonding with ligands. I conclude that along with strong hydrogen bonds, weak hydrogen bonds are also important in the kinase family. This information is a valuable asset for kinase inhibitor design, especially in the realm of multikinase inhibitor design.

## CHAPTER 5

---

# STRONG AND WEAK HYDROGEN BONDS IN DRUG–DNA COMPLEXES: A STATISTICAL ANALYSIS

---

### 5.1 Introduction

DNA sequences, both Watson-Crick and mismatched base pairs, can be recognized with a variety of compounds that bind in the DNA minor groove [5.1–5.5]. These compounds can be broadly classified as those that bind without a significant increase in the groove width of the free DNA, and those that bind with a broadening of the groove. The former class is exemplified by ligands containing aromatic rings and charged end-groups while the latter is typified by selective polyamide hairpin and related compounds like distamycin and netropsin [5.6]. The forces that dominate small molecule minor groove binding interactions are electrostatic, van der Waals, hydrophobic and hydrogen bonding. A crucial requirement for the ligand is a crescent shape that is complementary to the curvature of the minor groove.

Various aspects of minor groove drug–DNA recognition are revealed with the help of deposited X-ray structures in the protein databank (PDB) and the nucleic acid databank (NDB) [5.7]. Experimental, comparative structural and molecular modeling studies suggest that sequence specificity is often linked to key hydrogen bonds between the base pairs and the small molecule [5.8– 5.12].

The importance of weak C–H $\cdots$ O hydrogen bonds in macromolecules is a well-established phenomenon [5.13]. Their significance, as supporting interactions of stronger N–H $\cdots$ O and O–H $\cdots$ O bonds, in protein–ligand complexation have been described earlier [5.14, 5.15]. Based on the assumption that strong hydrogen bonding in drug–receptor interactions are thus inherently assisted by weak hydrogen bond [5.16], the aim of the present study is to analyze the importance of strong and weak hydrogen bonds in drug–DNA complexes.

The first section of the present study deals with a comparative analysis of strong and weak hydrogen bonds in a database of 70 drug–DNA complex crystal structures. The second part deals with molecular modeling applications of these results in a particular system. For the latter exercise, I have chosen a set of 26 furan derivatives targeted against Human

African Trypanosomes [5.17], *Trypanosoma brucei rhodesiense* (TBR) and *Trypanosoma brucei gambiense* (TBG). These organisms are responsible for Human African Trypanosomiasis (HAT) or sleeping sickness. The current drugs in the treatment of HAT are either toxic or difficult to use [5.18]. Only one drug for treating HAT is currently undergoing clinical trials [5.4, 5.19]. The orally available prodrug DB289 is converted systemically into another diamidine (DB75) that is active against the early-stage disease.

## 5.2 Materials and methods

### 5.2.1 Drug–DNA complexes from PDB

A set of 70 unique minor groove drug–DNA complexes was obtained from the PDB (Table 5.1). The drug molecules present in these cases are berenil (3 complexes), furans and thiophenes (8), pentamidines (5), netropsin (12), distamycin (5), Hoechst drugs (16), benzimidazoles (14), DAPI (2), polyamide (3), pyridines (2). The therapeutic uses of the above mentioned drugs are listed in Table 5.2. H-atoms were added to the drug–DNA complexes, and then subjected to minimization keeping the heavy atoms rigid. This was carried out in MOE with the MMFFx force field [5.20, 5.21]. The MOE optimized geometries were analyzed with the hydrogen bond analysis program, HBAT.

**Table 5.1:** Minor groove drug–DNA complexes in this study: (a) berenil 1-3, (b) furans 4-11, (c) pentamidines 12-16, (d) netropsin 17-28, (e) distamycin 29-33, (f) Hoechst drugs 34-49, (g) benzimidazoles 50-63, (h) DAPI 64-65, (i) polyamides 66-68, (j) pyridines 69-70.

Sl. No.	PDB ID	NDB ID	Resolution	PUBMED ID	Hetero group
1	1D63	GDL016	2.0	1640462	BRN
2	2DBE	GDL009	2.5	2323343	BRN
3	268D	GDLB42	2.0	n/a	BRN
4	289D	GDL045	2.2	8917643	D19
5	1VZK	-	1.7	n/a	D1B
6	1EEL	-	2.4	11128631	D24
7	1FMQ	DD0034	2.0	11128631	D34
8	1FMS	DD0035	1.9	11128631	D35
9	360D	GDL056	1.8	9611230	BPF
10	227D	GDL036	2.2	8639524	BGF
11	298D	GDL044	2.2	8917643	D18
12	1M6F	DD0052	1.7	12431090	CGQ
13	166D	GDL027	2.2	7813486	PET
14	1D64	GDL015	2.1	1643044	PNT

15	102D	GDL032	2.2	7608897	TNT
16	1PRP	GDL023	2.1	8268158	TNT
17	101D	GDLB31	2.2	7711020	NT
18	121D	GDL014	2.2	8395202	NT
19	195D	GDL030	2.3	n/a	NT
20	1D85	GDLB17	2.5	1332773	NT
21	1D86	GDL018	2.2	1332773	NT
22	1DNE	GDL004	2.4	2539859	NT
23	1DVL	DD0024	2.4	11914483	NT
24	261D	GDJ046	2.4	9125500	NT
25	358D	GDJB55	2.5	9826773	NT
26	375D	GDJ059	2.4	9826773	NT
27	474D	GDJB58	2.4	9826773	NT
28	6BNA	GDLB05	2.2	2991536	NT
29	267D	GDLB41	2.0	n/a	DMY
30	1JTL	DD0042	1.9	12071949	DMY
31	1K2Z	DD0046	2.4	12071949	DMY
32	2DND	GDL003	2.2	3479798	DMY
33	378D	GDH060	2.4	10089456	DMY
34	127D	GDL022	2.0	1371249	HT
35	128D	GDLB19	2.5	1371249	HT
36	129D	GDL021	2.2	1371249	HT
37	130D	GDLB20	2.5	1371249	HT
38	1D43	GDL010	2.0	1718416	HT
39	1D44	GDL011	2.0	1718416	HT
40	1D45	GDL012	1.9	1718416	HT
41	1D46	GDL013	2.0	1718416	HT
42	1DNH	GDL002	2.2	2452403	HT
43	264D	GDL026	2.4	7517864	HT
44	296D	GDL028	2.2	7515488	HT
45	8BNA	GDL006	2.2	2445998	HT
46	311D	GDL052	2.2	9162901	HT2
47	303D	GDL048	2.2	9017011	RO2
48	302D	GDL047	2.2	9017011	RO2
49	447D	DD0007	2.2	10666470	BBZ
50	263D	GDL039	2.2	8901516	TBZ
51	459D	DD0014	2.3	10373586	TBZ
52	1FTD	-	2.0	11170623	E97
53	403D	BDD001	1.4	9692982	HT1
54	443D	DD0004	1.6	10666470	IA
55	445D	DD0006	2.6	10666470	IA
56	449D	DD0009	2.1	10666470	IA
57	442D	DD0003	1.6	10666470	IB
58	444D	DD0005	2.4	10666470	IB
59	448D	DD0008	2.2	10666470	IB
60	109D	GDL033	2.0	n/a	IBB
61	1LEX	GDL037	2.2	8527438	IPL
62	1LEY	GDL038	2.2	8527438	IPL
63	269D	GDLB43	2.1	n/a	HT
64	1D30	GDL008	2.4	2627296	DAP
65	432D	DD0002	1.9	10600105	DAP
66	1CVX	DD0020	2.2	10623546	HP2
67	1CVY	DD0021	2.1	10623546	IPY
68	408D	BDD003	2.1	9756473	IPY
69	144D	GDLB24	2.2	8373768	SN6
70	328D	GDL053	2.6	9321660	SN7

**Table 5.2:** Minor groove binders (MGB) and their respective therapeutic use.

Sl. No.	Drug	Therapeutic uses
1	Berenil	Antiprotozoal agents, Intercalating agents
2	Furans	Trypanocidal agents
3	Pentamidines	Antitrypanosomal, <i>Pneumocystis carinii</i> and <i>Cryptosporidium parvum</i>
4	Netropsin	African trypanosomiasis and leishmaniasis, AIDS-related
5	Distamycin	<i>Pneumocystis carinii</i> pneumonia
6	Hoechst drugs	Filaricides, Antiviral agents, Anticancer agents
7	Benzimidazoles	Anticancer and Antiviral agents
8	DAPI	Filaricides and Antifungal agent
9	Polyamide	Antiviral
10	Pyridines	Trypanocidal Agents
		Anticancer
		Anticancer

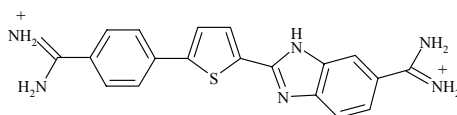
### 5.2.2 Hydrogen bond analysis tool (HBAT)

Strong and weak hydrogen bonds were analyzed with the HBAT software [5.22]. HBAT, analyzes and tabulates all hydrogen bonds present in a PDB file. The output file provides distance-angle distributions across various geometry ranges while tabulation of frequencies for each residue, ligand, water, and nucleic acids is done easily for any kind of interaction. It is a user-friendly desktop tool, which operates both with default and user-selected parameters. The standard H-bonding criteria were set as  $d(\text{H}\cdots\text{A}) \leq 2.8 \text{ \AA}$  and  $\theta(\text{X}-\text{H}\cdots\text{A}) \geq 90^\circ$ .

### 5.2.3 Docking of HAT inhibitors

Molecule building, geometry optimizations and minimizations were carried out using the MMFFx force field in the MOE software. Docking was carried out with GOLD 3.0 [5.24]. The amidinium inhibitors (26 compounds, Table 5.3) were selected from the literature [5.17]. Of these, seven (furan-based compounds) are present in the database of 70 experimental crystal structures analyzed in the first section of this chapter. Docking was performed for each of these 26 inhibitors into the DNA in eight experimental crystal structures, in other words a total of 208 docking experiments were performed. Seven of these structures correspond to the seven furans mentioned above; the eighth corresponds to a thiophene-based inhibitor in structure PDB ID-1VZK (Scheme 5.1), which has been shown to be a good choice for docking for minor groove binders [5.23]. This eighth structure is also contained in the earlier database of 70 structures. Significantly, the correlations

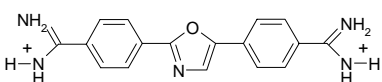
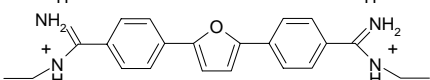
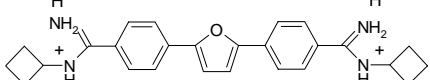
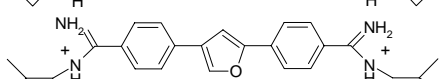
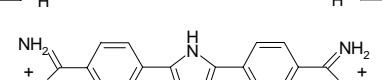
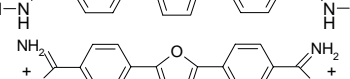
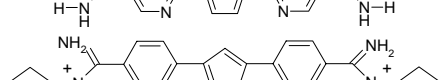
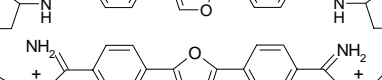
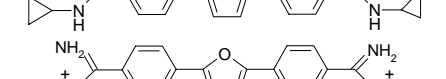
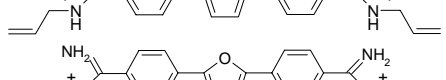
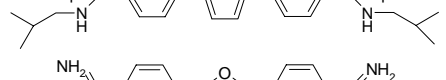
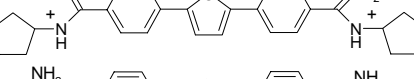
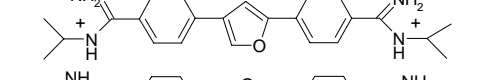
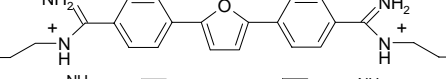
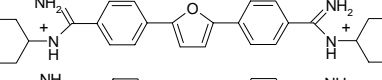
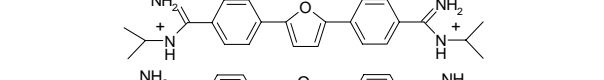
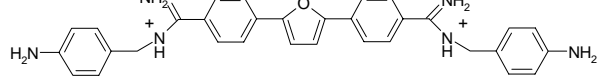
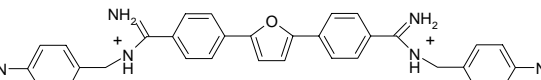
observed for the docking geometries obtained from this eighth (thiophene-based) structure were significantly better than the others, and accordingly only results from this set of 26 docking experiments are discussed further. A region of 7.0 Å radius around the ligand was defined as the active site for each drug–DNA complex. Default-set parameters were used in the docking. For each of the 10 independent Genetic Algorithm (GA) runs, with a selection pressure 1.1, 100,000 GA operations were performed on a set of 5 islands. The population size of 200 individuals was specified. Default operator weights were used for crossover, mutation, and migration of 95, 95 and 10 respectively. To further speed up the calculation, the GA docking was stopped when the top three solutions were within 1.5 Å RMSD of each other. The interaction analysis was carried out with HBAT as described above.



Scheme 5.1

**Table 5.3:** Amidinium based HAT inhibitors used in docking and their activities.

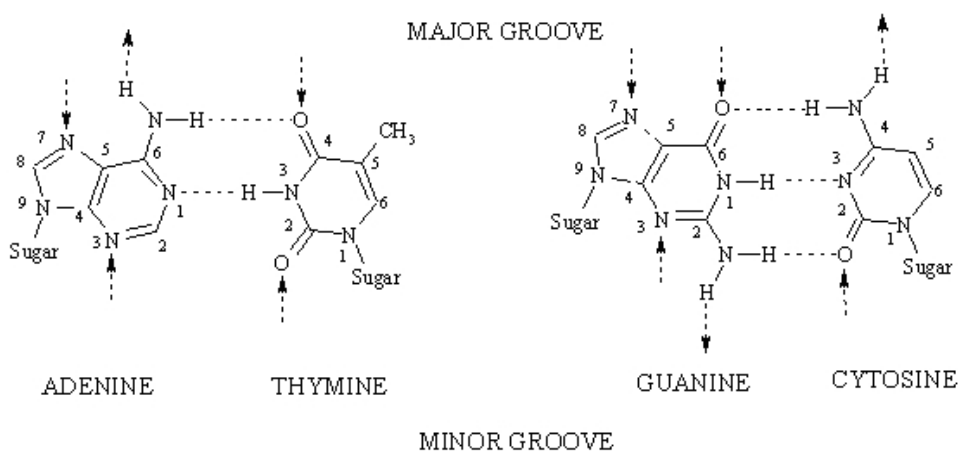
Sl. No.	Code	Structure	pIC <sub>50</sub>	IC <sub>50</sub> (in μM)
1	DB690		8.79	0.0016
2	DB867		8.69	0.0020
3	DB351		8.51	0.00302
4	DB994		8.39	0.0040
5	DB820		8.31	0.0048
6	DB75		8.29	0.0051
7	1RJL164		8.28	0.0052
8	DB417		8.18	0.0066

9	DB484		8.15	0.0070
10	DB427		8	0.0100
11	DB313		7.97	0.0107
12	DB518		7.86	0.0138
13	DB262		7.85	0.0141
14	DB829		7.77	0.0169
15	DB481		7.72	0.0190
16	DB193		7.63	0.0234
17	DB312		7.63	0.0234
18	DB235		7.45	0.0354
19	DB244		7.45	0.0354
20	DB480		7.41	0.0389
21	DB240		7.35	0.0446
22	DB249		7.17	0.0676
23	DB181		7.16	0.0691
24	DB422		7.01	0.0977
25	DB421		6.99	0.1023
26	DB568		6.77	0.1698

### 5.3 Results and discussion

#### 5.3.1 Overview of strong and weak hydrogen bond in drug–DNA complexes

DNA contains a variety of strong and weak hydrogen bond functional groups (Scheme 5.2). These groups are evenly exposed in the minor and major grooves. The ratio of the number of exposed donors to acceptors for AT and GC pairs in the major groove is 1:2, while it is 1:2 for GC and 0:2 for AT in the minor groove. The possible acceptors present in the major groove are [A{N(7)}], [G{N(7)}], [G{N(6)}] while donors belong to the amino groups in the [A{N(6)}] and [C{N(4)}] positions. In the minor groove, the acceptors are [A{N(3)}], [G{N(3)}], [T{O(2)}], [C{O(2)}] and the donor is the amino group of [G{N(2)}]. Several proteins of functional importance bind to the major groove, whereas non-covalently binding drugs bind to the minor groove. As discussed earlier, the minor groove binders (MGB) selectively bind to the AT rich region. Therefore acceptors present in the minor groove, notably N(3) of purine and C(2)=O of pyrimidine, are important sites for drug–DNA interaction. Apart from nucleotides, phosphate groups and the ribose O-atoms are also acceptors.



**Scheme 5.2:** Strong and weak hydrogen bond functional groups in DNA.

#### 5.3.2 Hydrogen bond analysis

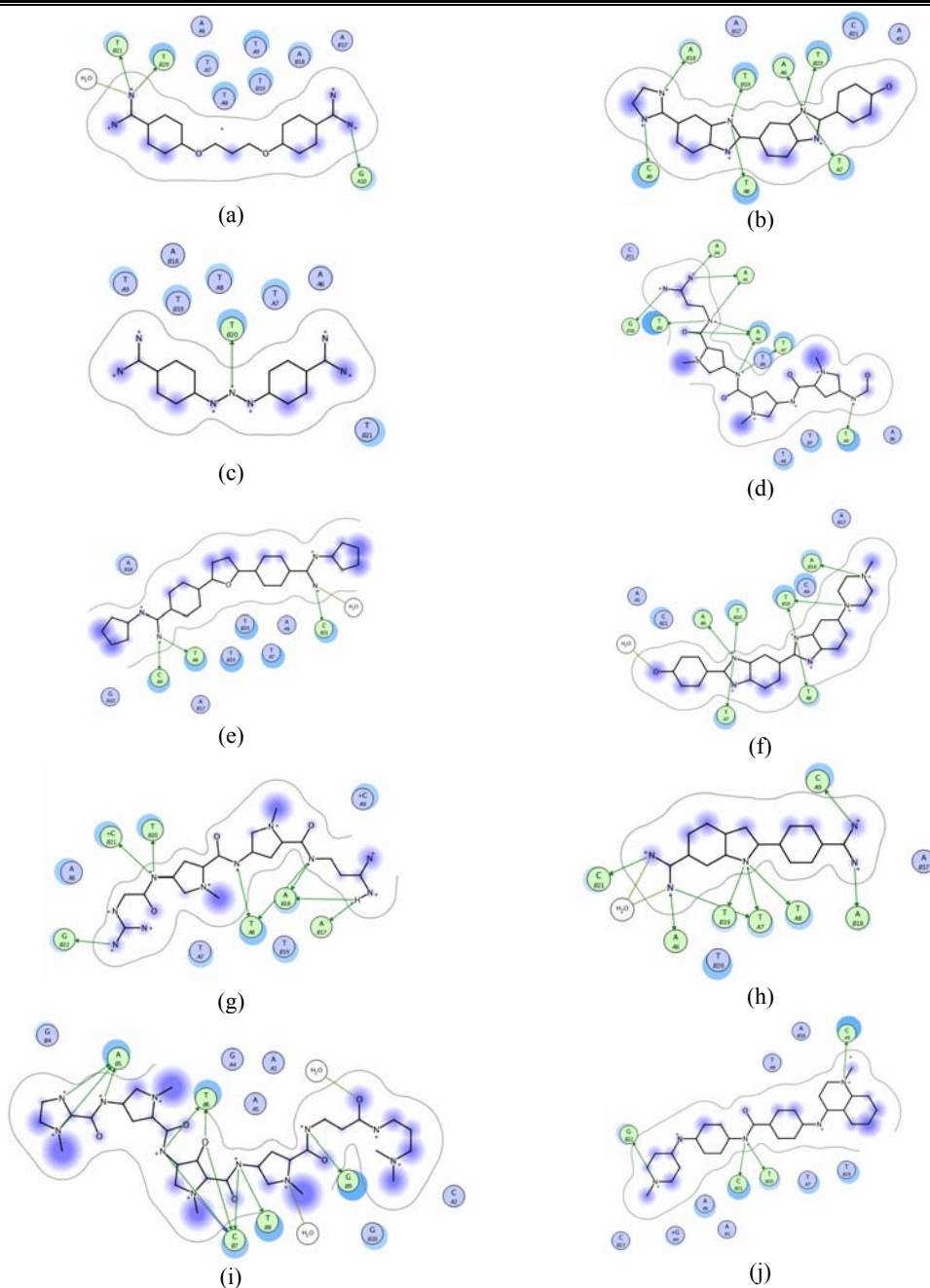
Both strong and weak hydrogen bonds are observed in the minor groove of drug–DNA complexes. As expected N(3) of purine and O(2) of pyrimidine are important (Table 5.4). The exceptions are polyamides and DAPI drugs, which interact with adenine N(3), guanine N(3), thymine O(2) and cytosine O(2). A schematic diagram for a typical inhibitor

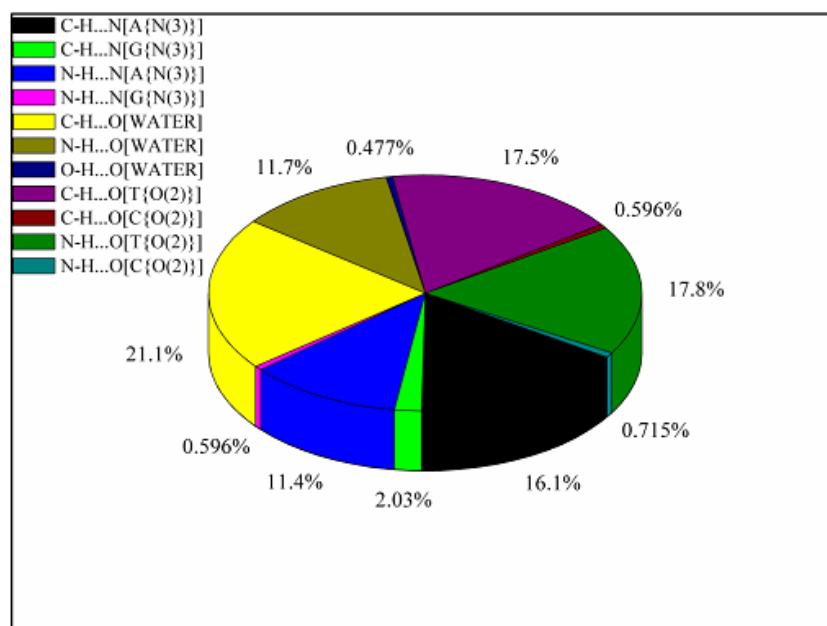
for each class is shown in Figure 5.1. The types of strong and weak hydrogen bonds observed in the case of purine is  $\text{N-H}\cdots\text{N}(3)$  and  $\text{C-H}\cdots\text{N}(3)$ , and for pyrimidine it is  $\text{N-H}\cdots\text{O}(2)$  and  $\text{C-H}\cdots\text{O}(2)$ . Also observed clearly are  $\text{N-H}\cdots\text{Ow}$  and  $\text{C-H}\cdots\text{Ow}$  interactions between the drug and water. The distribution of strong and weak hydrogen bond types in a total of 835 interactions are: (a) purine  $\text{C-H}\cdots\text{N}(3)$  18%,  $\text{N-H}\cdots\text{N}(3)$  12%, (b) pyrimidine  $\text{C-H}\cdots\text{O}(2)$  18%,  $\text{N-H}\cdots\text{O}(2)$  18%, (c)  $\text{C-H}\cdots\text{Ow}$  21%,  $\text{N-H}\cdots\text{Ow}$  12% (Figure 5.2a). The total number of  $\text{C-H}\cdots\text{N}$  and  $\text{C-H}\cdots\text{O}$  hydrogen bonds taken together is 481, while for  $\text{N-H}\cdots\text{N}$  and  $\text{N-H}\cdots\text{O}$  the combined total is 354. The present dataset contains 70 drug molecules. Thus on average, each drug molecule interacts with DNA through seven weak interactions and five strong interactions in the minor groove, which is effectively 1.4 weak hydrogen bonds for each strong hydrogen bond. Apart from nucleotides, drug molecules also interact with the sugar moiety, but they rarely interact with the phosphate group. The number of interactions observed between drug molecules and the deoxyribose sugar and phosphates taken together are 118. Among these interactions, deoxyribose sugar O4' of adenine and thymine have more hydrogen bonds (Figure 5.2b).

**Table 5.4:** Strong and weak hydrogen bonds in drug–DNA complexes.

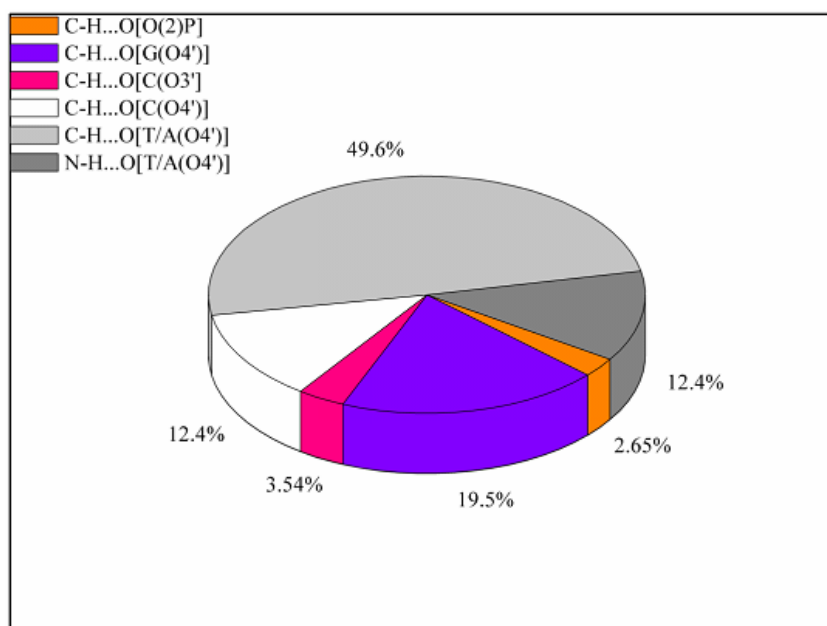
	C–H...N [A{N(3)}]	C–H...N [G{N(3)}]	N–H...N [A{N(3)}]	N–H...N [G{N(3)}]	C–H...O (WATER)	N–H...O (WATER)	O–H...O (WATER)	C–H...O [T{O(2)}]	C–H...O [C{O(2)}]	N–H...O [T{O(2)}]	N–H...O [C{O(2)}]
Berenil	5		6	3	10			5		4	
Furans	15		3	44	18			14		14	
Pentamidines	14		6	8	11			18		3	
Netropsin	12		15	12	20			17		18	
Distamycin	14		19	10	8			11		16	
Hoechst drugs	33		15	45	13		3	39		34	
Benzimidazoles	29	2	20	19	8			33		40	
DAPI	1	1	2	1	5			3		5	
Polyamides	3	14	10	28	4		1		5	13	6
Pyridines	9			7	1			7		2	
TOTAL	135	17	96	5	177	98	4	147	5	149	6

**Figure 5.1:** Schematic representation of MGBs in the minor groove. Strong hydrogen bonds are shown with green arrows, exposed ligand atoms in blue contours, and the nearest nucleotides as circles: (a) pentamidine, (b) benzimidazoles, (c) berenil, (d) distamycin, (e) furan, (f) Hoechst drugs, (g) netropsin, (h) DAPI, (i) polyamide, (j) pyridine.





(a)

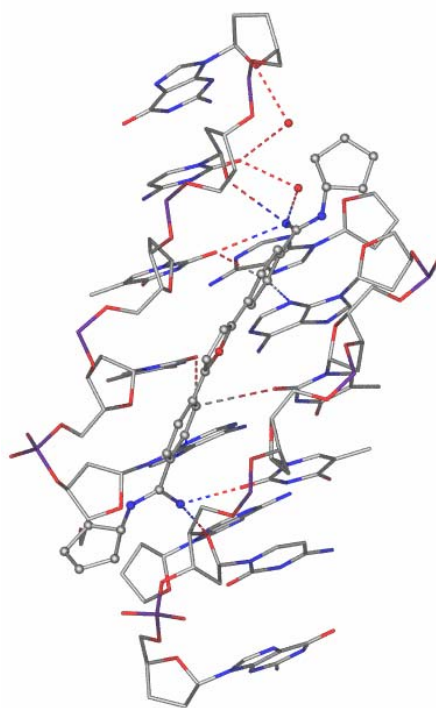


(b)

**Figure 5.2:** Distribution of hydrogen bonds in drug–DNA complexes. Interactions of drug with (a) nucleotides and water, (b) sugar and phosphate.

### 5.3.3 Donor furcation

The specificity and affinity of minor groove-binding agents are sensitive to the local width of the groove. In AT-rich sequences, the minor groove is unusually narrow (0.3-0.4 nm). Due to the narrow width, a donor present in a drug molecule possibly is hydrogen bonded to more than one acceptor on the wall of the groove. This type of situation, where a donor can interact with several acceptors simultaneously or an acceptor can interact simultaneously with many donors is termed furcation. The most frequently observed furcated interactions, in this study, involve bifurcated and trifurcated donors. Figure 5.3 shows examples of bi- and trifurcated donors in the DB244–DNA complex in PDB ID 1EEL. Typically, a single donor from the ligand is shared by two or three acceptors of the minor groove. Table 5.5 shows a distribution of strong and weak hydrogen bonds at various levels of furcation. With an increase in furcation level, the numbers of weak hydrogen bonds increase. The numbers of donor-furcated cases are: (a) bifurcated, 205; (b) trifurcated, 41; (c) tetrafurcated, 7; (d) pentafurcated, 2.



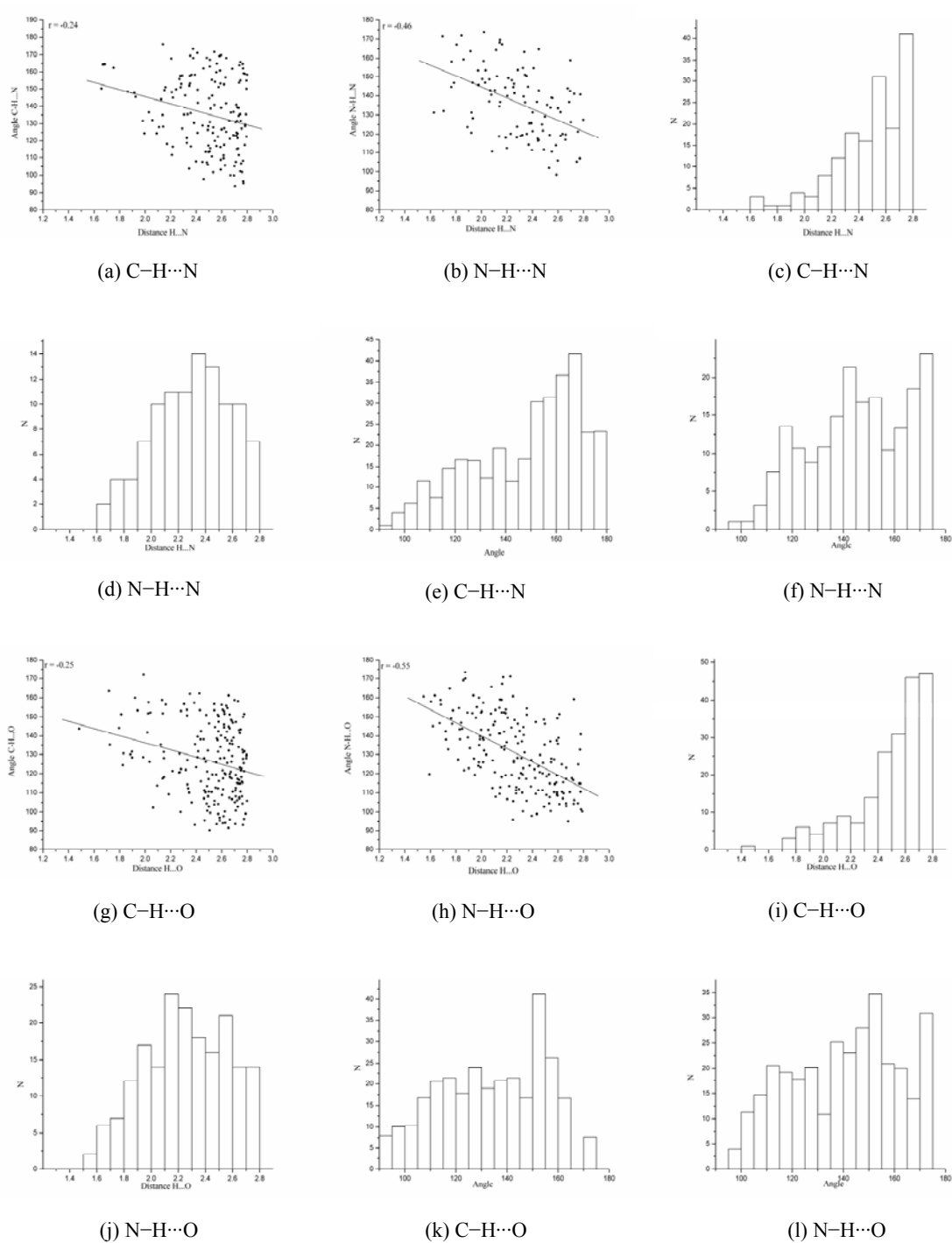
**Figure 5.3:** Bifurcated and trifurcated hydrogen bonds in DB244–DNA complex, PDB ID 1EEL.

**Table 5.5:** Strong and weak hydrogen bonds at various levels of donor furcation.

	C– <u>H</u> ···N	N– <u>H</u> ···N	C– <u>H</u> ···O	N– <u>H</u> ···O
Furcation level				
Bifurcated	84	69	216	141
Trifurcated	20	20	52	31
Tetrafurcated	6	5	10	7
Pentafurcated	6	-	4	-

### 5.3.4 Hydrogen bond geometry

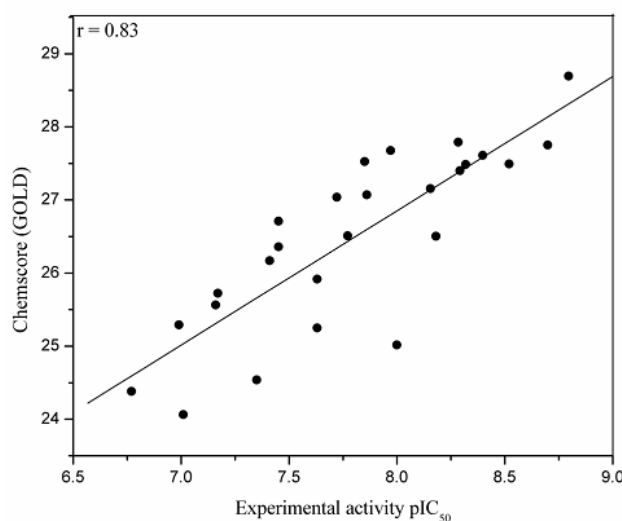
For strong hydrogen bonds of the type N–H···N and N–H···O, the median H···N/H···O distances,  $d$ , are less than 2.4 Å and 2.2 Å respectively (Figure 5.4). The cone-corrected angular distributions for N–H···N and N–H···O are similar with maxima in the range 170–175°. The inverse length-angle correlations are also well-behaved in these cases. Strong N–H···N and N–H···O hydrogen bonds shows better linearity compared to weak C–H···N and C–H···O hydrogen bonds. The weak C–H···N, C–H···O hydrogen bonds have variable geometry. The cone-corrected angular distributions of C–H···N and C–H···O interactions, have maxima at 165–170° and 150–155° respectively. To summarize, the geometries for strong and weak hydrogen bonds observed in the drug–DNA complexes are consistent and fall within acceptable limits.



**Figure 5.4:** Hydrogen bond geometry for purine (a)-(f) and pyrimidine (g)-(l) acceptors. In each case, the inverse length-angle scatterplot is followed by histograms of distances and cone-corrected angular distributions.

### 5.3.5 Human African Trypanosomiasis (Docking)

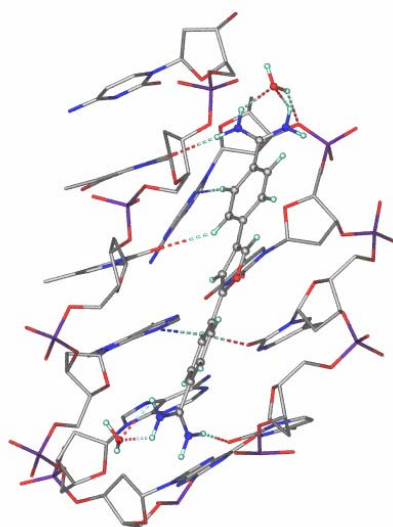
Molecular docking is a very useful technique in rational drug design. However, the purpose of using docking in the present context is different. Molecular docking was carried out to estimate the best drug–DNA geometries in cases where the crystal structures of the complex is unknown, assuming that strong and weak hydrogen bonds are optimized. Accordingly, the best geometry was selected (poses) for such complexes to obtain virtual crystal geometries for the 26 selected HAT inhibitors. The inhibitors were docked separately to each of the 8 drug–DNA complexes of the amidinium category. Each docking run was evaluated through regression analysis of experimental activities *versus* docking scores. The best correlation  $r = 0.83$  was obtained in the case of PDB ID 1VZK, where molecules with higher activity show greater docking scores (Figure 5.5) showing that the assumption regarding optimization of hydrogen bonds is a valid one. The best poses for the inhibitors were exported to HBAT for hydrogen bond analysis. The docking solutions of molecule **DB690** and **DB568** are shown in Figure 5.6.



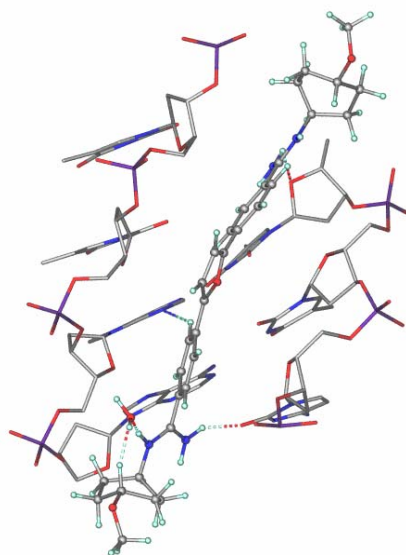
**Figure 5.5:** Fitness scores versus biological activities of HAT inhibitors.

Molecule **DB690** interacts with DNA through six weak and five strong hydrogen bonds. At one end, one amide group of the drug molecule is hydrogen bonded to T(O2). Further, this amide is also hydrogen bonded to a water molecule. The water molecule is again hydrogen bonded to the phosphate group present at the opposite strand. The other

amide group is hydrogen bonded to C(O2). Between these two ends, the weak C–H $\cdots$ N and C–H $\cdots$ O hydrogen bonds are present (Figure 6a). The inactive molecule, **DB568** binds to DNA with six weak and four strong interactions (Figure 5.6b). The interactions observed between molecule **DB690**, **DB568** and DNA are listed in Table 5.6.



(a)



(b)

**Figure 5.6:** Best docking solutions of (a) molecule **DB690** and (b) molecule **DB568** in the minor groove.

**Table 5.6:** (a) Hydrogen bonds in the molecule DB690–DNA complex.

Type	Acceptor	Acceptor atom	Residue number	$d$ (H...A)	$D$ (X...A)	$\theta$ (X–H...A)
C–H...N	A	N3	6	2.303	3.382	171.8
C–H...N	A	N3	18	2.602	3.378	127.5
C–H...O	T	O2	19	2.718	3.28	111.9
C–H...O	T	O2	19	2.449	3.284	132.4
C–H...O	T	O2	20	2.693	3.098	101.4
C–H...O	T	O2	8	2.43	3.053	114.8
N–H...O	C	O2	9	2.117	3.097	162.7
N–H...O	HOH	O	24	2.155	2.994	139.1
N–H...O	HOH	O	24	2.056	2.925	142.6
N–H...O	HOH	O	25	1.51	2.163	116.7
N–H...O	T	O2	20	2.135	3.085	155.7

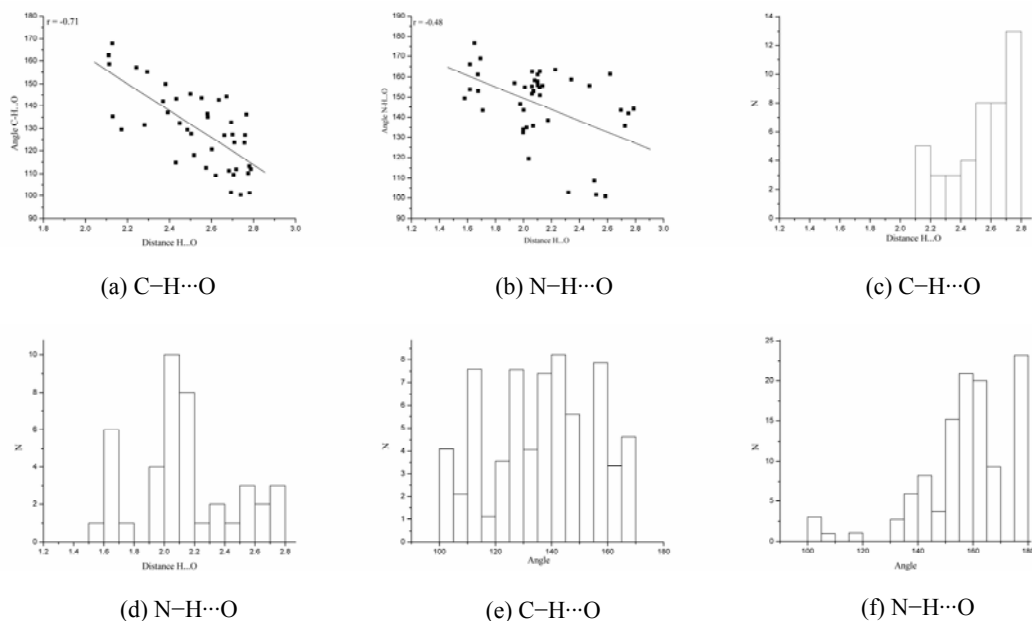
**Table 5.6:** (b) Hydrogen bonds in the molecule DB568–DNA complex.

Type	Acceptor	Acceptor atom	Residue number	$d$ (H...A)	$D$ (X...A)	$\theta$ (X–H...A)
C–H...N	A	N3	18	2.775	3.654	137.5
C–H...O	A	O1P	18	2.767	3.299	109.4
C–H...O	A	O4'	18	2.694	3.071	99.67
C–H...O	HOH	O	25	2.076	3.122	158.1
C–H...O	T	O1P	8	2.599	3.222	115.1
C–H...O	T	O2	19	2.683	3.241	111.2
N–H...N	G	N2	16	2.667	3.146	108.8
N–H...O	C	O2	9	1.709	2.592	143.4
N–H...O	HOH	O	25	1.82	2.807	163.5
N–H...O	T	O4'	7	2.403	2.836	104.6

### 5.3.6 Hydrogen bonds in HAT inhibitors

The hydrogen bond geometry for C–H...O(2) and N–H...O (2) pyrimidine acceptors are discussed here as a representatives of strong and weak hydrogen bonds. Figure 5.7 shows the C–H...O and N–H...O hydrogen bond geometries which involve DNA pyrimidine for the 26 selected HAT inhibitors. For N–H...O and C–H...O hydrogen bonds, the median H...O distances,  $d$  are less than 2.1 Å and 2.8 Å respectively. This is normal. The inverse length-angle correlations are also consistent for both types of hydrogen bonds. The cone-corrected angular distribution for N–H...O has a maximum at 175–180°. For C–H...O interactions, the maximum lies at 140–145°. This too is as expected. Similarly, the hydrogen

bond geometries obtained for both X-ray structures and the virtual complexes are consistent. Docking has produced a set of reasonable drug–DNA recognition geometries in the 19 cases where no crystal structure is available (virtual crystal structures).



**Figure 5.7:** C–H...O and N–H...O hydrogen bond geometries for 26 amidinium derivatives. In all cases, pyrimidine O(2) is the acceptor.

## 5.4 Conclusions

A statistical analysis of strong and weak hydrogen bonds in the minor groove of DNA is presented for a set of 70 drug–DNA complexes. The dataset was extracted from the PDB. The analysis was performed with the HBAT software. Both strong and weak hydrogen bonds are implicated in molecular recognition. On an average, there are 1.4 weak hydrogen bonds for every strong hydrogen bond. For both categories of interaction, the N(3) of purine and the O(2) of pyrimidine are favoured acceptors. Donor multifurcation is common with the donors generally present in the drug molecules, and being shared by hydrogen bond acceptors in the minor groove. Bifurcation and trifurcation are most commonly observed. The metrics for strong hydrogen bonds are consistent with established trends. For weak hydrogen bonds the geometries are variable. A database of recognition geometries for 26

literature amidinium-based inhibitors of HAT was generated with a docking study using seven inhibitors which occur in published crystal structures included in the list of 70 complexes mentioned above, and 19 inhibitors for which the drug–DNA complex crystal structures are unknown. The virtual geometries so generated correlate well with published activities for these 26 inhibitors.

## CHAPTER 6

---

# INTRODUCTION TO MOLECULAR MODELING AND PHARMACOPHORE MODELING

---

### 6.1 Introduction

Molecular modeling, also referred to as computational chemistry, is a set of techniques for investigating chemical problems on a computer [6.1–6.6]. Molecular modeling consists of a range of computerized techniques based on theoretical chemistry and experimental data that are used either to analyze molecules and molecular systems or to predict molecular, chemical, and biochemical properties. This chapter is intended to provide an overview of some of the techniques and methods used in molecular modeling. Later illustrated is the use of these techniques in the field of rational drug design, followed by a brief discussion on pharmacophore modeling.

The practice of molecular modeling started in the late 1960s. Since then it has been a useful technique to understand the complexity of the molecular world through the scalability and accuracy of computation. The main objective of molecular modeling is to: (a) Extract results for a particular model; (b) Compare experimental results of the system; (c) Compare theoretical predictions for the model; (d) Help to understand and interpret experimental observations; (e) Correlate microscopic details at the atomic and molecular level with macroscopic properties; (f) Provide information not available from real experiments.

### 6.2 Tools for molecular modeling

Most molecular modeling studies involve three stages. The first stage is model parameterization, wherein the intra- and inter-molecular interaction of the molecule is described on the basis of interaction energy, taking into account molecular position and/or arrangements in 3D space. The second stage is the memory crunching calculation involving energy minimization, molecular dynamics (MD), Monte Carlo simulation (MC), or a conformational search. The final step is the result analysis. Above all the accuracy of all these calculations depends on the modeler's decision to choose the correct method for

problem solving and in-depth knowledge of the “pros and cons” of the several molecular modeling tools. The most common tools used in the various stages of molecular modeling are, (a) *Ab initio* calculations, (b) Semiempirical methods, (c) Density functional theory, (d) Molecular mechanics, (e) Quantum mechanics/Molecular mechanics (QM/MM), (e) MD and MC simulations.

### 6.2.1 Quantum mechanics

Quantum mechanics (QM) is the correct mathematical description of the behavior of electrons. QM predicts exactly the properties of atoms and molecules. So far QM equations have only been solved exactly for one electron systems. QM deals with the motion of electrons under the influence of the electromagnetic force exerted by nuclear charges. The accurate understanding of the electronic behavior in the molecule helps in turn to understand molecular structure and the nature of chemical reactions. In practice, quantum chemistry is represented through *ab initio*, semiempirical methods and density functional theory. There are at least four important areas of drug design 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 [6.7].

#### 6.2.1.1 *Ab initio*

The term *ab initio* is assigned to computations that are derived directly from theoretical principles with no inclusion of experimental data [6.8–6.11]. The most common type of *ab initio* calculation is called a Hartree–Fock calculation (abbreviated HF). The disadvantage of *ab initio* methods is that they are expensive. These methods often take enormous amounts of computer CPU time, memory, and disk space.

#### 6.2.1.2 Semiempirical methods

Semiempirical methods are less strict compared to *ab initio* methods, because certain pieces of information are approximated or completely omitted [6.8, 6.11–6.14]. The core electrons are not included in the calculation and only a minimal basis set is used; further the two-electron integrals are not considered. Such parameter reduction introduces errors into the semiempirical methods. These errors are removed by further parameterization by estimating the omitted values. These values are obtained by fitting the results to

experimental data or *ab initio* calculations. The advantage of semiempirical calculations is that they are much faster than *ab initio* calculations. The disadvantage of semiempirical calculations is that the results can be erratic and fewer properties can be predicted reliably. If the molecule being computed is similar to molecules in the database used to parameterize the method, then the results may be very good. If the molecule being computed is significantly different from anything in the parameterization set, the answers may be very poor. Listed are some of the semiempirical methods practiced today, HÜCKEL, EXTENDED HÜCKEL, PPP, CNDO, MINDO, MNDO, INDO, ZINDO, SINDO1, PRDDO, AM1, PM3, PM3/TM, FENSKE–HALL, TNDO, SAM1, GAUSSIAN THEORY.

### 6.2.1.3 Density functional theory

Density functional theory (DFT) which was earlier confined to small molecules is now being introduced to macromolecules as well [6.7, 6.14, 6.15]. In DFT the energy of a molecule is determined from the electron density instead of a wave function. This formally reduces the many-body electronic problem to three coordinates only. This theory originated with a theorem by Hohenburg and Kohn to find out the ground-state electronic energy of a molecule. Later Kohn and Sham developed the practical application of this theory. Some of the more widely used functional are,  $X\alpha$ , HFS, VWN, BLYP, B3LYP, Becke3LYP, PW91, G96, P86, B96, B3P86, B3PW91.

### 6.2.2 Molecular mechanics

Molecular mechanics (MM) [6.16–6.18] is based on a mathematical model of a molecule as a collection of balls (corresponding to the atoms) held together by springs (corresponding to the bonds). The principle of Hooke's law is used to describe the ability of bonds to stretch, bend, and twist. The MM model ignores electrons. Therefore MM can be applied to molecules containing many atoms.

There are two types of interactions observed in atoms, bonded and non-bonded. The interactions are described on the basis of potential energy functions. In MM the minimum-energy geometry of a molecule is obtained through mechanical models. This is carried out by the use of a set of parameters that are empirically derived (usually termed as force-constant) and this collection of parameters is known as the force field [6.19, 6.20]. The term arises because the negative of the first derivative of the potential energy of a particle with respect to displacement along some direction is the force on the particle; a "force field"

$E(x, y, z)$  coordinates of atoms) can be differentiated to give the force on each atom. Many different force fields have been designed for various type of molecule. They have much in common, but they also have many differences. Some commonly used force fields are, AMBER, CHARMM, CFF, CHEAT, DREIDING, ECFF, EFF, GROMOS, MM1, MM2, MM3, MM4, MMFF, MOMECC, OPLS, Tripos, UFF, YETI.

### 6.2.3 QM/MM

Quantum mechanics/Molecular mechanics (QM/MM) is a hybrid method which encompasses both QM and MM [6.7, 6.20, 6.21]. This calculation is designed to give results that have very good speed when only one region needs to be modeled. The earliest QM/MM calculations were done simply by modeling different parts of the system with different techniques. For example, some crucial parts of the system could be modeled by using *ab initio* geometry-optimized calculations. The complete system could then be modeled using MM, by holding the geometry of the initial region fixed and optimizing the rest of the molecule. Sometimes the QM/MM technique uses an *ab initio* method to parameterize force field terms specific to a single system. Another variant of QM/MM is the energy subtraction method. In this method, calculations are done on various regions of the molecule with various levels of theory. Then the energies are added and subtracted to give suitable corrections. This results in computing energy for the correct number of atoms and bonds analogous to an isodesmic strategy. Bersuker and co-workers have proposed a technique called self-consistent method, whereby the atoms on the boundary between regions are included in both calculations. In this procedure, optimizations are done with each method, using the boundary atom charge from the other method, and this is repeated until the geometry is consistent between the levels of theory. The QM/MM methods are now routinely practiced for macromolecules.

## 6.3 Simulation tools

Simulation tools are very useful in elucidating biological intricacies. At present, simulations are carried out by statistical mechanics, MD or MC simulations. Statistical mechanics is the mathematical means to calculate the thermodynamic properties of bulk materials from a molecular description of the materials. Statistical mechanics provides a means for determining physical properties that are associated with the macroscopic sample of the bulk

liquid, solid, and so on. The gathering of information about possible energy levels, and about conformations of a macroscopic sample with reference to time and position is difficult through statistical mechanics. These difficulties are overcome by associating statistical mechanics with MD and MC simulations.

MD is a simulation of the time-dependent behavior of a molecular system, such as vibrational motion or Brownian motion [6.22–6.24]. It requires a way to compute the energy of the system, most often using a molecular mechanics calculation. This energy expression is used to compute the forces on the atoms for any given geometry. In MC simulation the location, orientation, and geometry of a molecule or collection of molecules are chosen according to a statistical distribution [6.25, 6.26]. MC methods are built around random sampling, which is simulated with a random-number-generating algorithm. For example, many possible conformations of a molecule could be examined by choosing the conformation angles randomly. If enough iterations are done and the results are weighted by a Boltzmann distribution, this gives a statistically valid result.

In summary, there exist several molecular modeling tools to understand the complexity and diversity of molecular system on the basis of theoretical calculations. Very often these tools are used in a variety of applications, such as material science research and rational drug design. The following section deals with some molecular modeling applications in rational drug design.

#### **6.4 Rational drug design another face of molecular modeling**

Today almost all pharmaceutical companies use molecular modeling as an integral part of their research and development programs [6.27]. According to a recent report, it costs a minimum about \$600 million to \$800 million and 12-15 years to bring a compound from the identification stage to the market [6.28]. Pharmaceutical companies are hard-pressed and so take a multitude of computer aided drug design (CADD) or rational drug design (RDD) approaches to shorten the time and reduce the cost of identifying new chemical entities (NCEs). The scalability and accuracy achieved by various molecular modeling tools mentioned above have further strengthened this process. Various avenue of RDD include quantitative structure activity relationship (QSAR), pharmacophore modeling, docking and scoring functions, *de novo* ligand design, molecular dynamics, virtual screening (VS),

ADMET predictions and chemoinformatics. In this context, pharmacophore modeling supplemented with virtual screening is an attractive research proposition in RDD.

#### **6.4.1 Quantitative structure activity relationship**

The functional groups present in a molecule determine the biological activities. For medicinal chemists, structure activity relationships (SAR) are the basis of synthesizing molecules with desired properties. Given a collection of compounds and their biological activities, it is possible to derive a statistical model which quantitatively establishes the structure–activity relationship. The importance of the functional groups present in molecules is described through chemical descriptors. QSAR is an attempt to correlate the structural features of a series of known compounds with their biological activities with the help of chemical descriptors [6.29, 6.30]. The QSAR model thus generated is used to predict the biological activity of novel inhibitors pertaining to the similar series of compounds.

#### **6.4.2 Docking and scoring functions**

The receptor–ligand complexation is the resultant of perfect complementarities of the steric and electrostatic factors between receptor and ligand. The task of predicting a favorable binding pose (lowest free energy binding mode) between receptor and ligand through computation by incorporating the essential ingredients of molecular recognition to molecular modeling techniques is the essence of docking. This is carried out through energy optimization process [6.31–6.34]. This is analogous to the docking of a ship in the harbor. In the present context the receptor is the harbor and ligand is the ship.

The scoring function used in docking is empirical and a rule based parameterized mathematical function is used to rank docking solutions based on the binding affinities of drug molecules to a receptor [6.35, 6.36]. The scoring functions are deemed to establish a linear correlation between the biological activity and the binding affinity of a diverse set of drugs for any particular target.

#### **6.4.3 De Novo ligand design**

Recent progress in structural biology has shown new light to many structural intricacies in molecular recognition [6.37]. By analyzing the active sites in terms of functional groups, the intuition of a complementary fragment very often comes to the mind of the medicinal chemist. This intuition of building complementary fragments *in silico* is the

basis of *de novo* ligand design. The first step in *de novo* design is molecular fragment fitting in active sites followed by growth of these fragments to build a complete molecule [6.38]. This exercise leads to the design of a focused virtual compounds library after several iterations. However, the virtual compounds generated in this method are often not accepted due to difficulties in their synthetic feasibilities. Therefore the problem of synthetic feasibilities is assessed by sophisticated software before attempting compound synthesis.

#### 6.4.4 Virtual screening

The method of screening large compound database *in silico* is termed as VS [6.39-6.44]. The VS method deploys a combination of known RDD techniques to generate first, a focused library of compounds and then selecting possible lead molecules *in silico* after subsequent property filtering. VS protocols include ligand based screens like 1D filters (e.g. molecular weight), 2D filters (similarity, substructure fingerprints) [6.43], 3D filters (3D-pharmacophore [6.40], 3D shape matching) and structure based screens like docking. The VS method has been implemented in many therapeutic targets [6.44].

#### 6.4.5 ADMET prediction

The final lap of the successful journey of a drug molecule depends on its Absorption, Distribution, Metabolism, Elimination, and Toxicity (ADMET) profiles. In most cases, the success of drug discovery projects depends particularly on better ADMET profile of a drug. Therefore the majority of funds in drug discovery research is diverted for designing improved ADMET profile of a drug molecule. This is the reason that ADMET property evaluation has been incorporated into the drug design strategies today [6.45, 6.46]. The ADMET model estimates physicochemical properties of compounds like ionizability ( $pK_a$ ), and lipophilicity ( $\log P$  or  $\log D$ ). These estimates provide a reasonable answer to ADMET profile of a given compound.

#### 6.4.6 Chemoinformatics

Chemoinformatics is the application of informatics methods to solve chemical problems [6.47–6.50]. Over the years chemistry has generated large amount of digital information. The prime objective of chemoinformatics is to represent chemical object digitally, to manage chemical data, elucidation of structure, the design chemical reaction and estimate synthetic feasibility. Above all, the niche of chemoinformatics is extended to

molecular modeling and RDD methods. Hence chemoinformatics is widely used in drug research.

## 6.5 Pharmacophore modeling

Pharmacophore, as defined by an IUPAC working party and published in 1998, is “*the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response*” [6.51]. As a consequence the pharmacophore: (1) describes the essential, steric and electronic, function-determining points necessary for an optimal interaction with a relevant pharmacological target. (2) does not represent a real molecule or a real association of functional groups, but it is rather a purely abstract concept that accounts for the common molecular interaction capacities of a group of compounds towards their target structure. (3) is not just a specific functional group or “pieces of molecules” [6.52, 6.53].

### 6.5.1 Historical perspectives

The structure activity relationship is the basis of pharmacophore modeling. One of the earliest to recognize structure–activity relationships was Robert Boyle in 1685, who tried to explain the specific effects of drugs in terms of mechanical philosophy [6.54]. Later, at the turn of the 20th century, the German scientist Sigmund Fränkel discussed the pharmacological action and selectivity of drug molecule in cells and tissues on the basis of the presence of certain groups in the drug molecule [6.55]. The inadequacy in the understanding of molecular worlds in the beginning of 20th century laid several controversies on the pharmacological action and the physiological activity of drugs. However, the discovery of the pharmacological receptor concept in the later part of the century confirmed the structure–activity relationships in drug molecules. The initial hypothesis on receptor–drug interaction was put forth by Langley in 1878 [6.56], who introduced the term “receptive substance” [6.57]. However, the word “receptor” was introduced later, by Paul Ehrlich [6.58, 6.59]. During the first half of the 20th century, several observations highlighted the critical features associated with the concept of receptors [6.60]. The selective interaction of drug to their respective receptor was first demonstrated by Paul Ehrlich’s through the discovery of salvarsan. This gave rise to the concept of a

chemotherapeutic “magic bullet” against specific infectious organisms. The work on enzyme and glucoside interaction by Emil Fischer later in 1894 further strengthened Ehrlich’s seminal discoveries of receptor hypothesis [6.61]. He proposed the famous “lock and key” hypothesis for enzyme substrate recognition. To extend this phenomenon in the present day context, substrate and receptor should have a greater degree of steric, electrostatic and molecular flexibility complementarities for better fit. With this background it is clear that the pharmacological action and the physiological activity of a drug molecule are hidden in the structure. Thus this provided food for thought to many to quantify and correlate structure–activity relationships in drug molecules through experiments and modeling.

#### **6.5.1.1 Pharmacophore models in the early years**

The pharmacophores were described in the literature before the advent of CADD. Due to the knowledge of the bond lengths and the van der Waals sizes in the 1940s, initial structure–activity relationships were established through the construction of simple two-dimensional model structures [6.62–6.65]. With the availability of X-ray analysis and conformational chemistry, the three-dimensional models became possible in the 1960s [6.66–6.68]. In the early 1970s, three-dimensional depictions of pharmacophore models were built for “Clonidine and Its Interaction with the  $\alpha$ -Adrenergic Receptor” [6.69–6.72]. This knowledge was later incorporated to model building using computers.

#### **6.5.1.2 Pharmacophore modeling after the use of computers**

The advent of the computer enabled scientists to quantify structure–activity more efficiently. The initial step in this direction was carried out by Korolkovas, Gund [6.73, 6.74] and by Humbelt and Marshall [6.75]. Immediately this application found its ground in the field of medicinal chemistry to correlate the effect of structure on bioactivity. These events later incorporated the ingredients of chemical descriptor to enhance the accuracy of pharmacophore modeling. With the help of computer programs the pharmacophoric patterns in molecules were identified. In this regard the program developed at Princeton University in early 1970s [6.76, 6.77] is noteworthy. This program presented the topographic patterns of molecules based on distance measurement through Dreiding models. The distances were derived from a knowledge of small molecule crystal structures. The new era of

pharmacophore modeling began thereafter with the help of many sophisticated computers [6.78, 6.79].

### **6.5.2 Pharmacophore modeling methods**

Broadly, pharmacophore methods can be classified into analog and structure based pharmacophore modeling [6.52]. The analog based pharmacophore modeling gathers information from already existing drugs/ligands that are active against target biological molecules (protein or DNA/RNA) of interest. Based on this information, a set of pharmacophore hypotheses is generated to compare biological activities between actives and in-actives and to search for lead molecules in chemical database. In this approach, the relative importance of different functionalities in drug receptor recognition is first determined; the hypothesis is next generated based on the correspondence between functional groups in congeneric series of drug molecules. Examples of analog based pharmacophore modeling package are DISCO, HipHop, GASP, HypoGen, PHASE. In the structure based approach the pharmacophore hypothesis is generated from the three-dimensional structure of the receptor. The method involves analyzing the active site to generate an interaction map of desirable features that a ligand should satisfy to adequately interact with the receptor. The features include various non-covalent interactions observed in receptor–ligand complexes. The generated hypotheses are later submitted as queries in chemical databases. The available software to carry out structure based pharmacophore modeling are structure based focusing in Cerius<sup>2</sup> and LigandScout, and pharmacophore modeling in MOE.

### **6.5.3 Application of pharmacophore modeling in rational drug design**

Pharmacophore models can be used in different ways in drug design programs: (1) as a 3D query tool in virtual screening to identify potential new compounds from 3D databases; (2) to predict the activities of a set of new compounds yet to be synthesized; (3) to filter suitable docked poses implemented in the docking exercise; (4) to understand the possible mechanisms of action; (5) to generate a focused virtual library. The pharmacophore generation approach is quite powerful and finds many applications in drug discovery research.

## **6.6 As we move on to following chapters**

The following two chapters deal with the pharmacophore modeling study on two therapeutically important targets. The studies are based on an amalgamation of techniques used in molecular modeling and drug design. These studies have a potential role in medicinal chemistry and drug design.

## CHAPTER 7

# PHARMACOPHORE MODELING ON EGFR KINASE INHIBITORS: A NOVEL STRATEGY FOR LIGAND BASED VIRTUAL SCREENING

### 7.1 Introduction

Both the reliability and utility of specific virtual screening (VS) tools grow with the knowledge available on a particular drug target and the underlying ligand–receptor interaction pattern. A widely used form of VS employs protein–ligand docking [7.1, 7.2]. Most docking approaches keep the receptor rigid and the ligand flexible with respect to conformation, orientation and position in the active site during docking. Considering protein flexibility in docking simulations drastically increases the complexity of the problem [7.3, 7.4]. Recent methods try to relax a part of the protein (side chains or in some cases even the backbone) [7.5].

In cases where there is no structural information available on the target or if it is present but difficult to model, ligand based pharmacophore models used with chemical and other filters, [7.6–7.11] can provide powerful tools for VS. There have been continuing efforts to design methods based on pharmacophore model for ligand based virtual screening (LBVS) [7.12–7.21]. The aim of the present chapter is to model a LBVS method based on pharmacophore modeling for epidermal growth factor receptor (EGFR) kinase inhibitors. To supplement the LBVS model a detailed validation of pharmacophore model was carried out by activity prediction, and mining of EGFR and non-EGFR kinase inhibitor databases. Further structure based pharmacophore modeling and docking study was carried out on eight known inhibitors. The structural basis of ligand selectivity on active and inactive form of receptor was revealed through docking study.

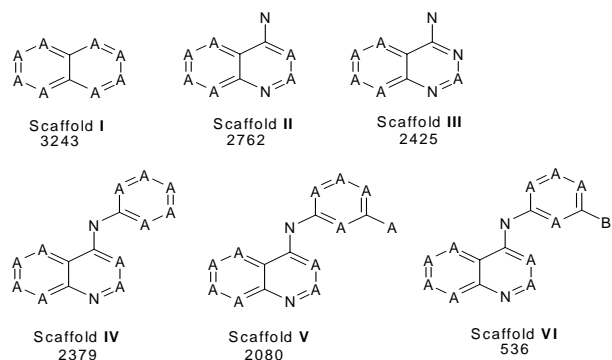
To explore the possible substitutions to the anilinoquinazoline moiety that could enhance the potency of the inhibitor, 3D-QSAR studies of this class of molecules has been done earlier from our group and by others [7.22–7.24]. Further, a study to select an appropriate combination of docking and scoring functions was carried out for EGFR kinase domain [7.25]. EGFR is frequently over-expressed in a broad range of human cancers (e.g. squamous cell carcinoma, breast, ovarian, non-small cell lung cancer etc.). EGFR is therefore a potential target for various cancer therapies [7.26–7.30]. Ligands binding EGFR are divided into two main categories, extracellular and intracellular, depending on the

receptor region targeted [7.30–7.32]. A number of compounds have shown both selectivity towards EGFR kinase domain, and potency. Many selective inhibitors of EGFR kinase are presently in the clinical trials at different phases or even marketed [7.33–7.40]. Ligands selected in this study are targeted to the kinase domain of EGFR which is intracellular.

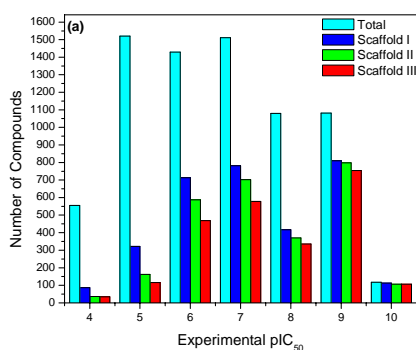
## 7.2 Materials and methods

### 7.2.1 Database profile

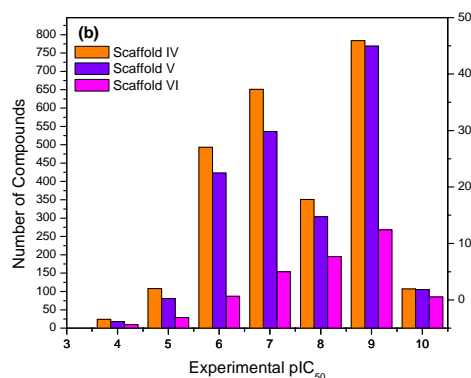
The GVK bio, kinase database contains total 1,38,805 inhibitors [7.41]. Nearly 9700 diverse molecules which have been screened for EGFR inhibitory activity are present in this maintained database, were used in this study. Among these, 3243 compounds contain scaffold **I** (Scheme 7.1). Within this set, subsets of 2762 and 2425 compounds contain scaffolds **II** and **III** respectively, that is,  $n(\text{III}) \subset n(\text{II}) \subset n(\text{I}) \subset n(\text{total})$ . Figure 7.1 is a histogram of activities for compounds containing these three scaffolds and also for all the 9490 compounds. As seen in Figure 7.1b, scaffolds **IV** and **V** consist of many active molecules. Many inhibitors belonging to scaffold **IV** and **V**, and having substitution at the 4-amino group are presently in various phases of clinical trials.



**Scheme 7.1:** Scaffolds representing database molecule.



(a)

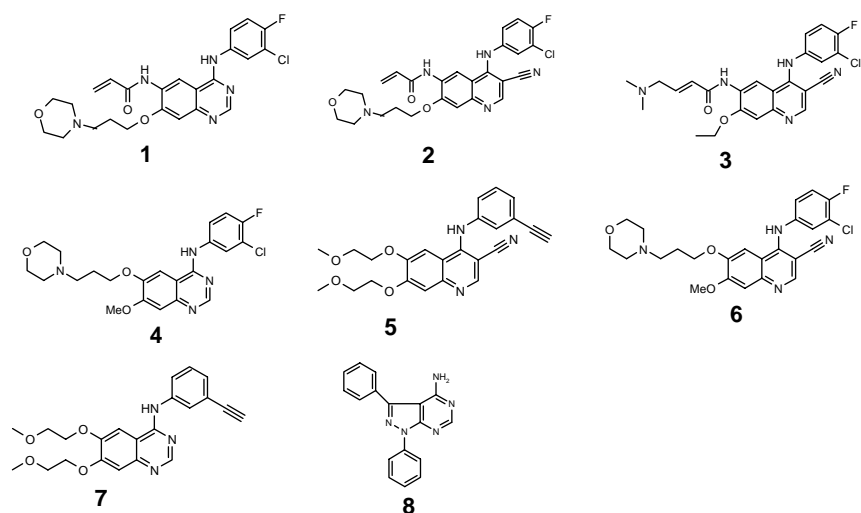


(b)

**Figure 7.1:** Frequency distribution of EGFR inhibitors (9940) based on different scaffolds, **I** to **VI** [**I–III** in (a) and **IV–VI** in (b)] The effect of phenyl substitution on the 4-amino group is evident from Figure 1b (scaffolds **IV**, **V**, **VI**). A large percentage of scaffold **VI** compounds (*meta* Br-substituted) have high activity in the nM range. To summarize,  $n(\mathbf{VI}) \subset n(\mathbf{V}) \subset n(\mathbf{IV}) \subset n(\mathbf{II})$ .

### 7.2.2 HipHop

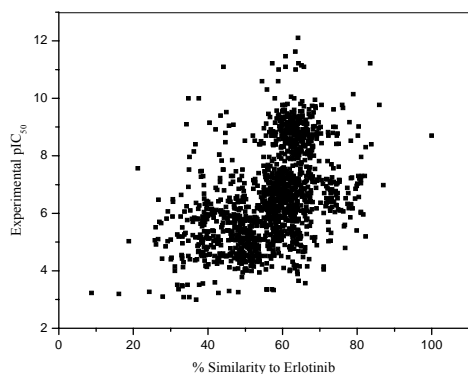
A set of eight active inhibitors **1–8** (Scheme 7.2) mostly belonging to scaffold **IV** and **V** were used to identify a qualitative pharmacophore model [7.42–7.43] using the *HipHop* module in Catalyst 4.7 [7.44]. Among these eight compounds (**1–8**), a few are in various phases of clinical trials [7.35, 7.36, 7.38]. A *HipHop* model identifies important features present in a set of active molecules. The model is built by aligning the common features present in the input training set structures, usually a small number of active compounds. *HipHop* models do not predict inhibitory activity.



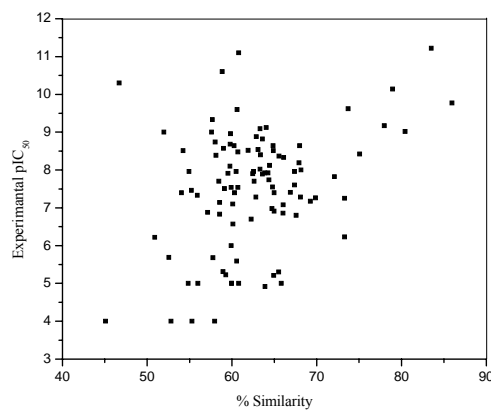
**Scheme 7.2:** Compounds considered in generating *HipHop* model and docking study.

### 7.2.3 *HypoGen*

A quantitative model was developed using *HypoGen* in Catalyst software. The *HypoGen* model was generated selecting 3800 EGFR kinase inhibitors from the GVK bio kinase database [7.41]. Ligands were selected considering diversity in structure and activities with consistent assay technique. For all these molecules ISIS structural keys [7.45, 7.46] were generated. Erlotinib (**7**), which is structurally similar to several compounds in clinical trials (Appendix II, Scheme 1), was used as the reference molecule for calculating the Tanimoto similarity coefficient. The correlation between the experimental  $\text{pIC}_{50}$  values and the Tanimoto similarity coefficients for all these inhibitors of EGFR (3800) is given in Figure 7.2.



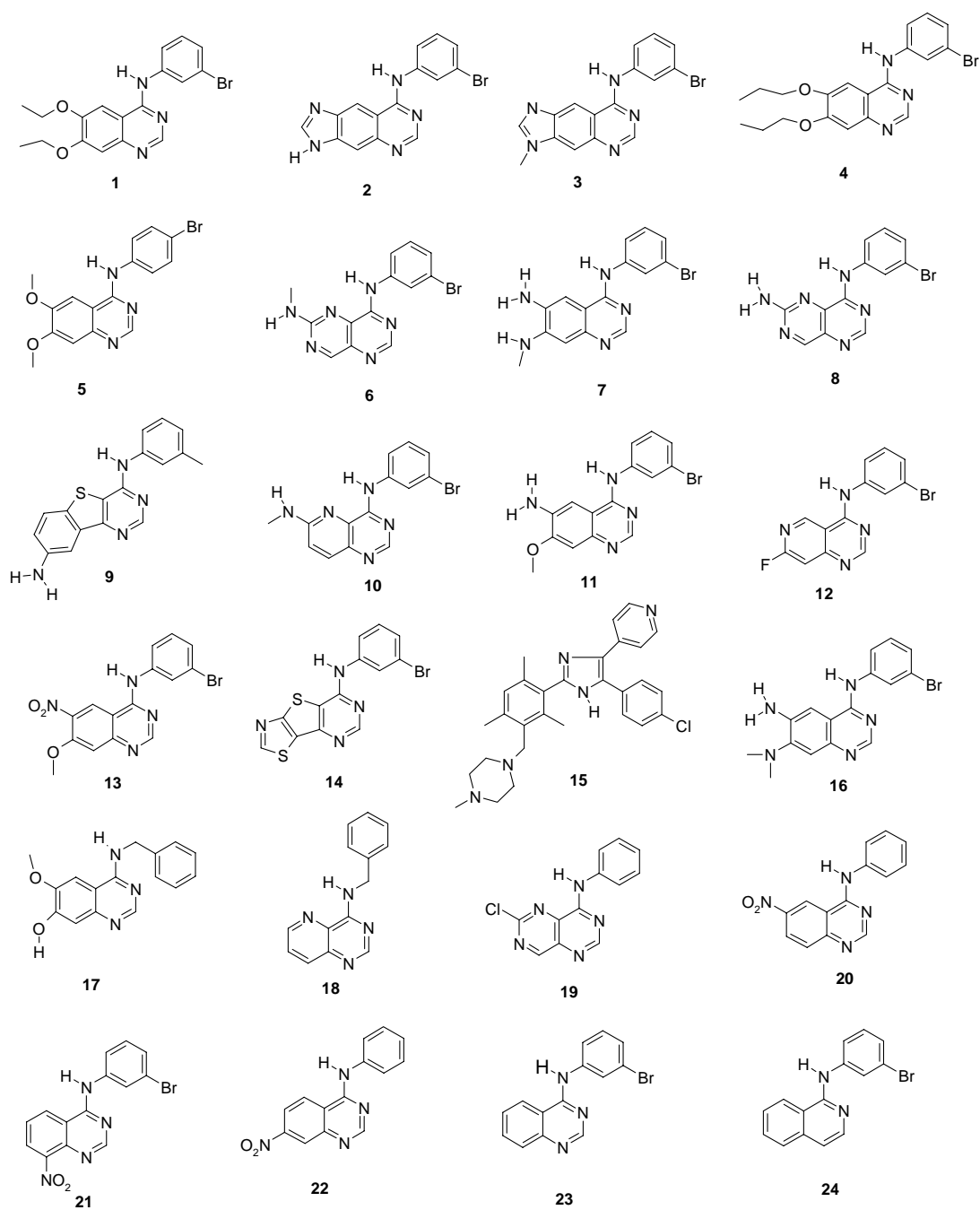
(a)



(b)

**Figure 7.2:** Percentage structural similarity with respect to erlotinib versus experimental activity for ligands (a) Diversity of the EGFR database molecules, (b) Subset 100 molecules, considered for *HypoGen* model building.

A subset of 100 molecules were selected for *HypoGen* model building by considering structural diversity and activity (Figure 7.2b). The training set comprises 24 molecules (Scheme 7.3) and the test set 76 (Appendix II, Table 1). The inhibitory activities for the training set compounds are listed in Table 7.1.

**Scheme 7.3:** Training set compounds in *HypoGen* model.

**Table 7.1:** Training set molecules considered in *HypoGen* model. Also shown are the experimental and predicted activities ( $\text{pIC}_{50}$ ).

Sl. No.	Actual $\text{pIC}_{50}$	Predicted $\text{pIC}_{50}$
1	11.22	10.55
2	11.10	10.29
3	9.77	10.20
4	10.14	9.07
5	9.00	8.43
6	8.82	8.67
7	8.51	8.76
8	8.42	7.76
9	7.89	7.76
10	8.68	8.49
11	9.12	8.69
12	10.60	10.23
13	7.82	7.79
14	7.40	8.18
15	6.24	6.14
16	6.23	7.60
17	6.22	6.40
18	6.11	5.58
19	5.59	5.95
20	5.30	5.03
21	5.00	5.06
22	4.92	5.03
23	4.00	5.00
24	4.00	5.09

In *HypoGen*, the features are correlated with the biological activity of molecules (active and inactive) and a quantitative feature activity relationship is established. The minimum number of compounds in the training set should be 16 [7.47]. The default uncertainty value of 3 was used for each compound. The uncertain factor for each compound represents the ratio range of uncertainty in the activity value based on the expected statistical straggling of biological data collection. *HypoGen* generates pharmacophore models that are a set of features in 3D space, each containing a certain tolerance and weight, that fits to the features of the training set, and further correlates to the activity. A *HypoGen* run includes three phases: (1) pharmacophore features that are common among the most active compounds are elaborated in the constructive phase; (2) those models that fit the inactive training set members are abolished in the subtractive phase and; (3) the remaining hypotheses are refined in the optimization phase. The optimization is done by random transition of features, rotation of vectored features and removal or addition of features from the preferred models. The statistical parameters like cost values also determine the significance of the model. For a statistically significant model, (i) the

difference between the fixed cost and null cost (cost difference) should be above 40 indicating an activity correlation above 75%; (ii) the configuration cost should be less than 17; (iii) the root mean square (RMS) deviation between predicted and actual activities for the training set molecules should be low and (iv) the difference between total cost and the fixed cost should be as low as possible. The configuration cost quantifies the entropy of the pharmacophore space. Any model with a configuration cost greater than 17 leads to a chance correlation of the model. The total cost is the sum of error cost, weight cost and configuration cost. The fixed cost is the lowest possible cost representing a simple model that fits all the data perfectly. The null cost represents the maximum cost of a featureless model, which estimates the activity of all the training set molecules to be the average activity.

Ligands used for pharmacophore modeling were built and minimized in Catalyst using the 3D-minimizer module with CHARMM force field parameters. The conformations were generated within Catalyst ConForm method using the “Poling” algorithm [7.48]. A maximum of 250 conformations were generated for each molecule within an energy threshold of 20.0 kcal/mol above the global energy minimum.

#### **7.2.4 LigandScout and GOLD**

The X-ray structures of human EGFR tyrosine kinase domain (1M17) and (1XKK) complexed with the inhibitor erlotinib and lapatinib respectively were obtained from the Protein Data Bank (PDB) [7.49]. Crystal structure 1M17.pdb is the active form of EGFR kinase domain, whereas 1XKK.pdb represent the inactive form [7.29, 7.50]. H-atoms were added to the proteins in the hydrogen addition option in MOE software [7.51]. Both the complexes were subjected to minimization keeping the heavy atoms rigid. This exercise was carried out in MOE with the MMFFx force field [7.52]. Structure based pharmacophore models were built for these minimized structures. This study was performed with LigandScout software [7.53]. LigandScout perform pharmacophore modeling for available protein–ligand complexes in PDB. LigandScout generates a pharmacophore model through a set of chemical features and volume constraints present in the active site. The 3-D pharmacophore models are constructed, which describe the binding mode of a ligand to the receptor.

The minimized structures were exported to GOLD for docking. Molecules considered in *HipHop* model were docked separately into the 1M17 and 1XKK crystal structures using Chemscore, GOLD [7.54]. A region of 10 Å radius around the bound

ligands, for each structure was defined as the active site. The default parameters in all the calculations were used. For each of the 10 independent Genetic Algorithm (GA) runs, with a selection pressure 1.1 and 100,000 GA operations were performed on a set of 5 islands. The population size of 200 individuals was specified. Default operator weights were used for crossover, mutation, and migration of 95, 95 and 10 respectively. To further speed up the calculation, the GA docking was stopped when the top three solutions were within 1.5 Å RMSD of each other. All the other values were set to the default. While docking, a limited flexibility is allowed for the hydrogen atoms in the –OH, and –NH<sub>3</sub><sup>+</sup> substituents of the side chains of Ser, Thr and Lys residues.

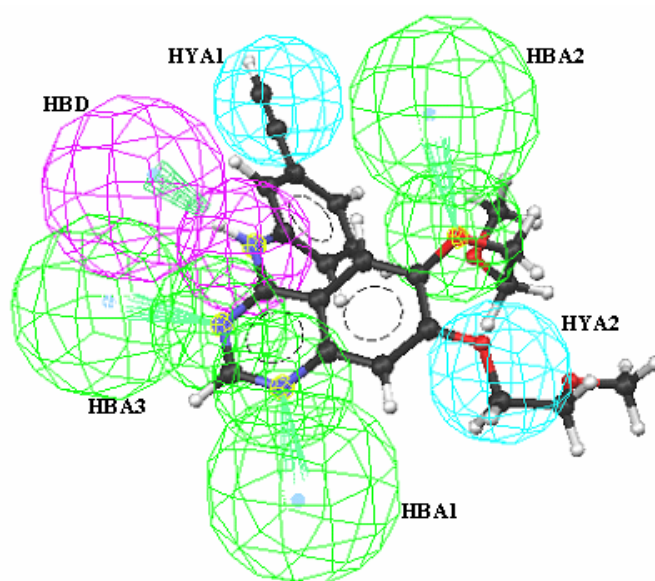
### 7.3 Results and discussion

#### 7.3.1 HipHop

A set of 10 *HipHop* models that were generated with features like hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), hydrophobic aliphatic (HYA), hydrophobic aromatic (HYR) are ranked in Table 7.2. Model 1 (shown in Figure 7.3), was used to mine EGFR kinase inhibitors in the database [7.41]. Table 7.3 shows various measures for hit quality; the underlying equations to measure the quality of hits have been mentioned in the Appendix II, Scheme 2. The results suggest poor mining performance of this model. The database mining resulted in 300 hits. However, the *HipHop* model has provided the list of important features in these inhibitors. These features were specified as input to generate a *HypoGen* model.

**Table 7.2:** Top 10 pharmacophore models generated by *HipHop*. The features corresponding to each model are listed.

Model Rank	Features	Score
1	HYA, HYA, HBD, HBA, HBA, HBA	107.169
2	HYA, HYA, HBD, HBA, HBA, HBA	103.302
3	HYA, HYA, HBD, HBA, HBA, HBA	102.478
4	HYA, HYA, HBD, HBA, HBA, HBA	101.598
5	HYA, HYA, HBD, HBA, HBA, HBA	101.396
6	HYA, HBD, HBA, HBA, HBA	88.4952
7	HYA, HBD, HBA, HBA, HBA	88.4950
8	HYA, HBD, HBA, HBA, HBA	87.3170
9	HYA, HBD, HBA, HBA, HBA	86.4055
10	HYA, HBD, HBA, HBA, HBA	86.2553



**Figure 7.3:** *HipHop* model mapping onto Erlotinib: The green contours represent the positioning of hydrogen bond acceptors (HBA 1, 2, 3) and their projected points on the receptor are indicated by conical projections. Blue contours represent hydrophobic aliphatic (HYA1, 2). Magenta contours represent the hydrogen bond donor and the projected point on the receptor is indicated by conical projections (HBD).

**Table 7.3:** Results of database mining tabulated for *HipHop* and *HypoGen* models.

Sl. No	Parameter	<i>HipHop</i>	<i>HypoGen</i>
1	Total compounds in database (D)	3800	3800
2	Total Number of actives in database (A)	1083	1083
3	Total Hits ( $H_t$ )	300	1324
4	Active Hits ( $H_a$ )	101	710
5	% Yield of Actives	33.3	50.09
6	% Ratio of actives in the Hit list	9.32	65.59
7	Enrichment or Enhancement (E)	1.18	1.88 (Maximum value of 2.87)
8	False Negatives	982	373
9	False Positives	199	614
10	GH score (Goodness of Hit list)	0.255	0.178

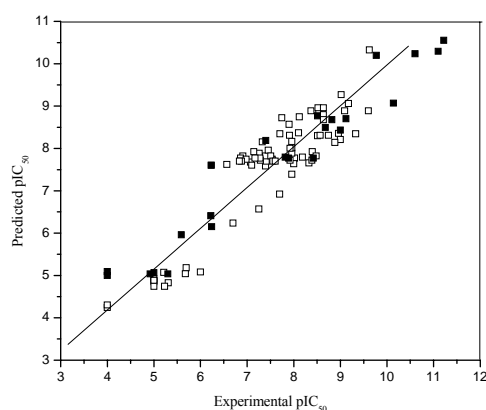
### 7.3.2 HypoGen

After 30 iterations, 10 models were selected (Table 7.4). The features found to be important are: hydrogen bond acceptor (HBA), hydrophobic aliphatic (HYA), hydrophobic aromatic (HYR). Based on various statistical analyses e.g. cost difference, RMS deviation, correlation coefficient, configuration cost, model 1 was found to be the best. It has the lowest RMSD (1.269) and highest correlation coefficient ( $r = 0.958$ ). This model was validated by predicting the activities of 76 test set compounds (Appendix II, Table 1). The correlation between experimental and predicted biological activity data is shown in Figure 7.4.

**Table 7.4:** Ranking of generated *HypoGen* models. Also mentioned are statistical values associated with each model and pharmacophore features.

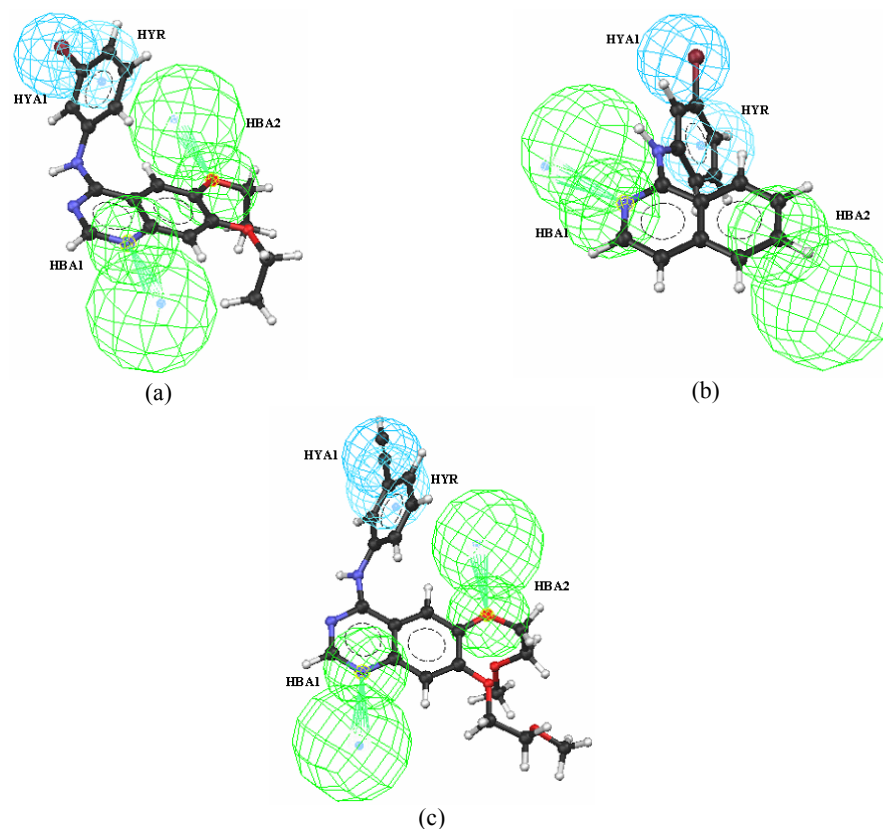
Model number	Total cost	Cost difference	RMS deviation	Correlation (r)	Features <sup>a</sup>
1	121.874	186.657	1.269	0.958	HBA, HBA, HYA, HYR
2	121.889	186.642	1.360	0.950	HBA, HBA, HYA, HYR
3	126.526	182.005	1.518	0.938	HBA, HBA, HYA, HYR
4	127.396	181.135	1.550	0.934	HBA, HBA, HYA, HYR
5	128.398	180.133	1.527	0.937	HBA, HBA, HYA, HYR
6	131.685	176.846	1.539	0.937	HBA, HBA, HYA, HYR
7	133.527	175.004	1.720	0.919	HBA, HBA, HBA, HYA
8	135.243	173.288	1.747	0.916	HBA, HBA, HYA, HYR
9	135.5	173.031	1.769	0.914	HBA, HBA, HBA, HYA
10	135.5	173.031	1.847	0.905	HBA, HBA, HYA, HYA, HYR

Null cost = 308.531, Fixed cost = 93.41, Configuration = 11.559, All cost units are in bits. <sup>a</sup>HBA hydrogen bond acceptor, HYA hydrophobic aliphatic, HYR hydrophobic aromatic.



**Figure 7.4:** Scatter plots of actual versus predicted activities of training set (■) and test set (□) molecules used in the *HypoGen* model.

The model contained four features (Figure 7.5): 1 HBA, 2 HYA and 1 HYR. The most active molecule **1** had a fitness score of 12.04 when mapped to the pharmacophore (Figure 7.5a), while the least active molecule **24** maps to a value of 6.58 (Figure 7.5b). In molecule **1** the HYA1 feature corresponds to *meta* substitution on the aniline ring. Of the two hydrogen bond acceptors, HBA1 and HBA2 correspond to the N(1) nitrogen and substitution at 6<sup>th</sup> position of the quinazoline ring respectively while HYR corresponds to the aniline ring. In the inactive molecule, HBA1 is mapped to the N(3) position of the quinazoline ring, while HYA1 and HYR are mapped to the same bromine substitution and the aniline ring. Noteworthy is the absence of the HBA2 feature at the 6<sup>th</sup> position in the inactive molecule. These features are also mapped with Erlotinib in Figure 7.5c. The distance matrix (in Å) of the generated hypotheses is provided in Table 7.5.



**Figure 7.5:** *HypoGen* model map on to: (a) most active molecule **1**, (b) least active molecule **24**, (c) Erlotinib. The green contours represent the positioning of hydrogen bond acceptors (HBA1, 2) with the projected points on the receptor indicated by conical projections. Blue contours represent the hydrophobic aliphatic (HYA1) and hydrophobic aromatic (HYR1).

**Table 7.5:** Distance matrix (in Å) of the generated quantitative model.

	HBA1 (IP)	HBA1 (PP)	HBA2 (IP)	HBA2 (PP)	HYA	HYR
HBA1 (IP)						
HBA1 (PP)	3.0					
HBA2 (IP)	5.4	7.4				
HBA2 (PP)	7.7	10.1	3.0			
HYA	7.7	5.4	9.1	12.1		
HYR	6.3	4.8	7.0	9.9	3.1	

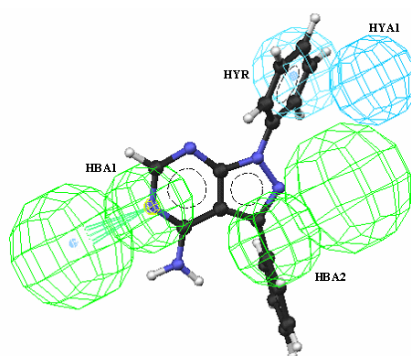
IP, initial point on ligand, PP, projected point on the receptor.

### 7.3.3 Model validation

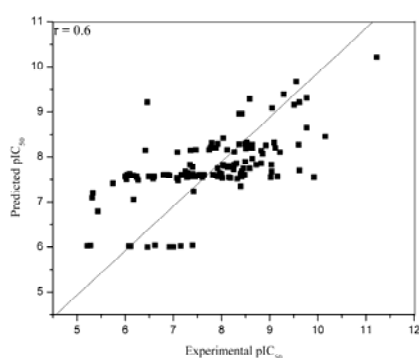
The quantitative model has mixed success in predicting the activity of the inhibitors considered in the qualitative model generation Table 7.6. The predicted activities are close to experimental activities, except for molecule **8**. This is because two of the pharmacophore features: HBA2 and HYA1 do not map properly as shown in Figure 7.6.

**Table 7.6:** Predicted activity for eight inhibitors considered in the qualitative model.

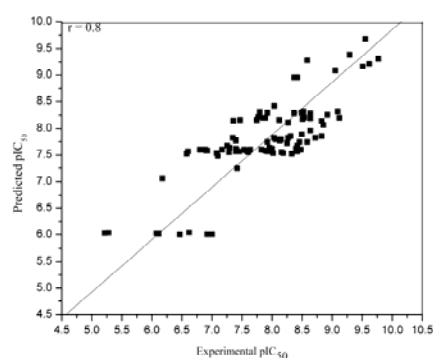
Molecule No.	Actual (pIC <sub>50</sub> )	Predicted (pIC <sub>50</sub> )	Fitness score
1	7.13	6.07	11.59
2	7.09	6.65	12.23
3	7.08	5.18	12.17
4	6.28	8.01	12.83
5	6.07	7.74	11.55
6	5.98	6.13	12.88
7	5.83	6.65	11.09
8	5.80	2.20	6.51

**Figure 7.6:** *HypoGen* model map on to molecule **8**. Notice the incomplete mapping of pharmacophore features HYA1 and HBA2.

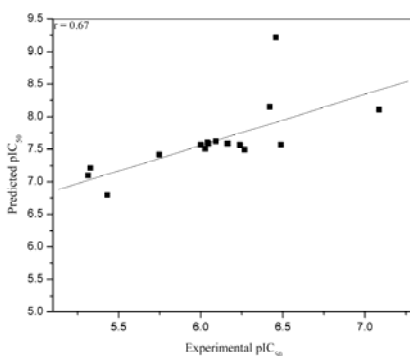
Therefore there is a large deviation between experimental activity and predicted activity. Model validation was carried out in an external dataset of 128 EGFR kinase inhibitors from the previous study ([7.25], Appendix II, Table 2). These include simple quinazolines (class **A**: 45 compounds), pyrido[3,2-d]pyrimidines (**B**: 7 compounds), pyrido[4,3-d]pyrimidines (**C**: 49 compounds), pyrido[3,4-d]pyrimidines (**D**: 5 compounds), pyrido[2,3-d]pyrimidines (**E**: 6 compounds), pyrimido[5,4-d]pyrimidines (**F**: 16 compounds). The activities were predicted for 128 ligands ( $r = 0.6$ ) shown in Figure 7.7a. Poor correlation was attributed to 32 (25 %) molecules, (Table 7.7).



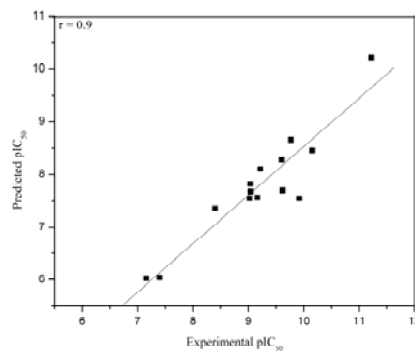
(a)



(b)



(c)



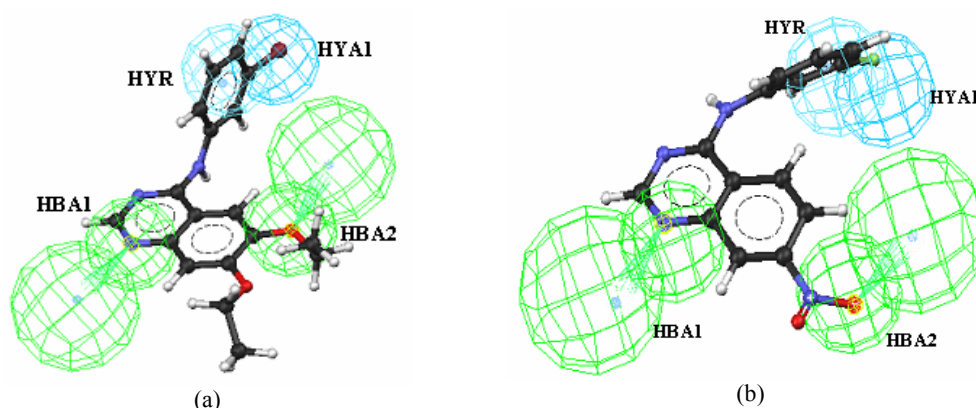
(d)

**Figure 7.7:** Scatter plot of actual versus predicted activities of 128 molecules from the external EGFR kinase inhibitor database. (a) Correlation for 128 molecules, (b) Correlation for 96 compounds, (c) False positives, (d) False negatives.

**Table 7.7:** Activity prediction for external 128 EGFR kinase inhibitors.

Compound class	Scaffold	Total Number of compounds	False positive	False negative	False prediction
quinazolines	<b>A</b>	45	8	6	31%
pyrido[3,2-d]pyrimidines	<b>B</b>	7	-	-	-
pyrido[4,3-d]pyrimidines	<b>C</b>	49	4	6	20%
pyrido[3,4-d]pyrimidines	<b>D</b>	5	-	-	-
pyrido[2,3-d]pyrimidines	<b>E</b>	6	4	-	66%
pyrimido[5,4-d]pyrimidines	<b>F</b>	16	2	2	25%
<i>Total</i>		<i>128</i>	<i>18</i>	<i>14</i>	<i>25%</i>

Among these, 18 are false positives and 14 are false negatives. The predicted activities for molecules of the pyrido[3,2-d]pyrimidine and pyrido[3,4-d]pyrimidine classes are better compared to all other compound classes. After removing the false positives and negatives the correlation between experimental and predicted activity was improved ( $r = 0.8$ ) for 96 compounds as shown in Figure 7.7*b*. Interestingly, the correlation coefficients for false positive and negatives among themselves are 0.7 and 0.9 respectively. This is shown in Figures 7.7*c* and *d*. This suggests that the model is able to rank the outliers to a reasonable limit. The outliers are usually misoriented conformations of molecules (compared to crystal structure conformation of Erlotinib/Lapatinib). Also, the pharmacophore feature HBA2 is not mapped by any of these compounds. Sometimes in these outliers the pharmacophore features HBA1, HBA2 are assigned to other functional groups instead of N(1) and substitution at 6<sup>th</sup> position. The active (reference molecule **43** in Appendix II, Table 2) and inactive (reference molecule **20** in Appendix II, Table 2) molecules collected from this data set are mapped onto the generated pharmacophore model and is shown in Figure 7.8.



**Figure 7.8:** Pharmacophore mapping of the compounds (a) most active molecule (**43**) (b) least active molecule (**20**) in the external 128 EGFR kinase inhibitors. Note the improper mapping of the hydrophobic aliphatic (HYA1) and hydrogen bond (HBA2) in the inactive molecule.

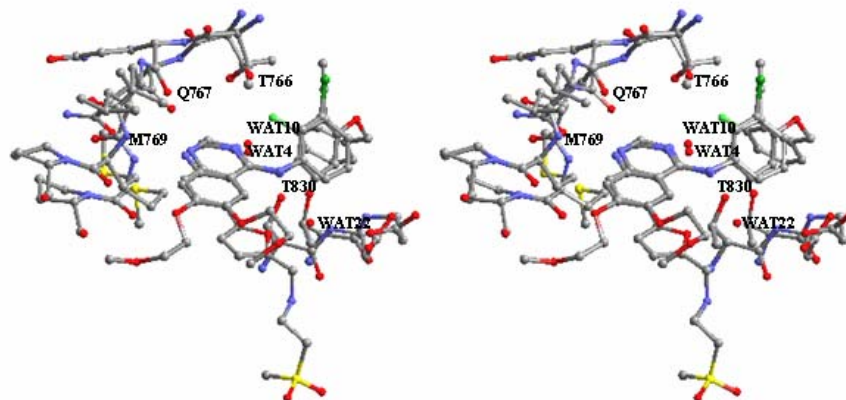
### 7.3.4 Mining

The best *HypoGen* model was used to screen the EGFR kinase inhibitor database with the best flexible search method implemented in Catalyst. The performance of *HypoGen* is better, compared to that of *HipHop* in terms of the number of hits and active hits (Table 7.3). A non-EGFR kinase inhibitor database of 4590 compounds was prepared using various kinase inhibitors. These include CDK2, FGFR kinase, IGFR kinase, MAP kinase, PDGFR kinase, PKA, Src kinase, and VEGFR kinase. The non EGFR kinase dataset molecules are spiked with 128 EGFR kinase inhibitors of the earlier study [7.25]. The database mining resulted in a smaller number of hits (129 out of 4590). Out of 129 hits, 95 belong to EGFR kinase inhibitors (i.e. 95 out of 128 EGFR kinase inhibitors were successfully retrieved). This suggests that the generated model can selectively pick EGFR kinase inhibitors from other kinase inhibitors.

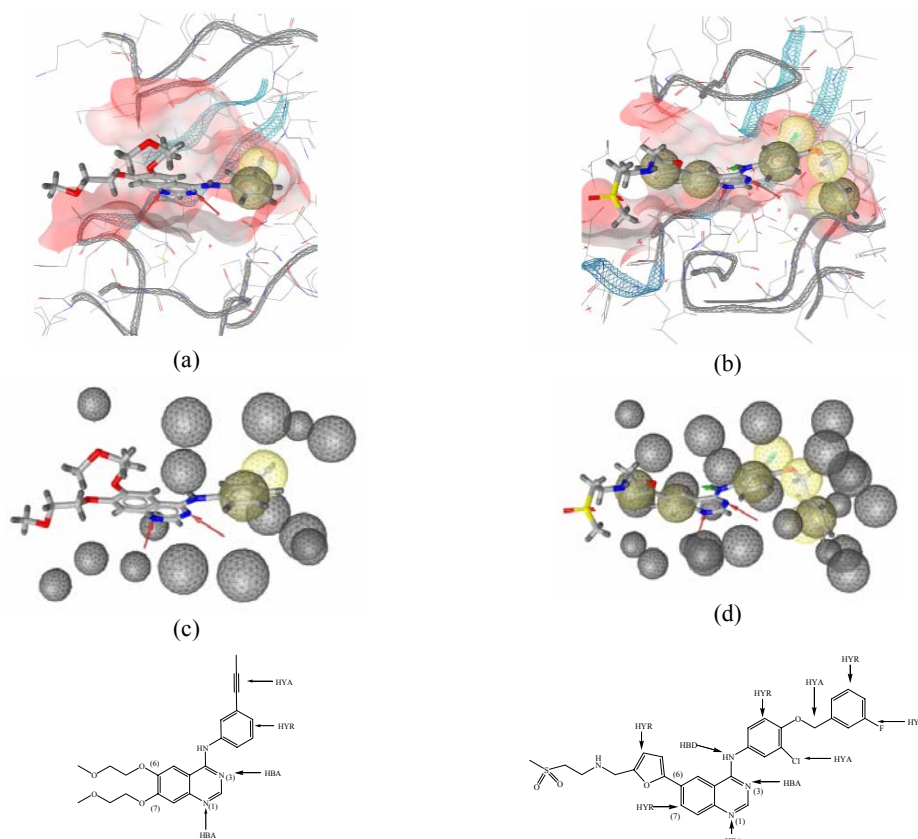
### 7.3.5 Active site and structure based pharmacophore

The X-ray structures of EGFR kinase domain active site in 1M17 and 1XKK was manually superposed (Figure 7.9). A comparison of the two structures reveals that the shape of the active site is closed and truncated in 1XKK while it is open in 1M17. This is because the A-loop and G-loop shrink towards each other in 1XKK while they are far apart in 1M17. This results in a variation in the size of the active sites,  $199 \text{ \AA}^3$  and  $145 \text{ \AA}^3$  for 1M17 and 1XKK respectively. Further 1XKK has an extra large pocket to accommodate the 3-fluorobenzyl-oxy group, whereas no such pocket is observed in the EGFR/Erlotinib complex. This pocket is where two hydrophobic aliphatic (HYA) and an aromatic (HYR) features are seen. The active site in the crystal structure of 1XKK is more hydrated than in 1M17. There are 20 water molecules within a  $10 \text{ \AA}$  radius around the bound ligand in 1XKK compared to 3 waters in 1M17. In both crystal structures, the quinazoline N(1) forms a hydrogen bond to the NH of MET769, while N(3) is hydrogen bonded to WAT 10 in 1M17 and to WAT 4 in 1XKK. However this water interacts differently with residue THR766 in 1M17 and 1XKK respectively. These hydrogen bonds are shown with HBA at N(1) and N(3) positions as pharmacophore points. Quinazoline NH is hydrogen bonded to another water molecule (WAT 22) in the EGFR/GW572016 complex. This hydrogen bond donor (HBD) is unique to 1XKK. Apart from these, two hydrophobic aromatic (HYR) pharmacophore features (quinazoline and furan substitution at 6<sup>th</sup> position) are observed in 1XKK. The orientation of the aniline ring is similar for both structures in the hydrophobic pocket, created by residues VAL702, LYS721, LEU764 (Figure 7.9, 7.10). The above

mentioned differences in active site environment are also reflected in structure based pharmacophore model generated with LigandScout software (shown in Figure 7.10).



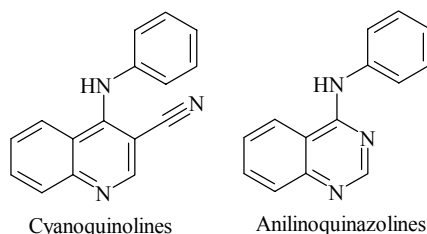
**Figure 7.9:** Stereo view of superposed crystal structures of EGFR/Erlotinib (1M17) and EGFR/GW572016 (1XKK).



**Figure 7.10:** Structure based pharmacophore model for EGFR/Erlotinib (1M17) and EGFR/GW572016 (1XKK). (a)-(b) Pharmacophore features along with polarity of binding sites. (c)-(d) Pharmacophore with excluded volume.

### 7.3.6 Docking

Docking studies were carried out for molecules in Scheme 7.2 using Chemscore in GOLD [7.54]. These molecules contain the cyanoquinoline (molecules **2**, **3**, **5**) and anilinoquinazoline (molecules **1**, **4**, **7**) fragment and are shown in Scheme 7.4.



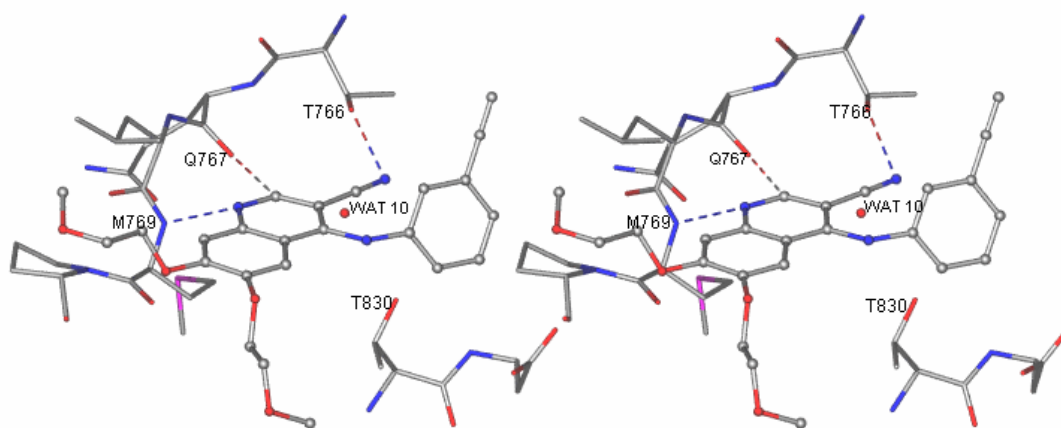
**Scheme 7.4:** Scaffold for cyanoquinoline and anilinoquinazoline

The weak interaction was modeled with a parameter which includes the weak C–H···O hydrogen bond. The initial docking runs were aimed at choosing the form of the receptor suitable to carry out the docking study. Docking poses were filtered through structure based pharmacophore criteria. Ligand poses satisfying the pharmacophore criteria were selected for further analysis. Proper docking poses and scores for molecules having the cyanoquinoline fragment were achieved only in 1M17, whereas for the anilinoquinazolines, docking poses and scores are consistent in 1M17 and 1XKK. Interestingly, the desired poses of cyanoquinoline were achieved only after removing WAT 10 from the crystal structure of 1M17. The hydrogen bonds distances are listed in Table 7.8.

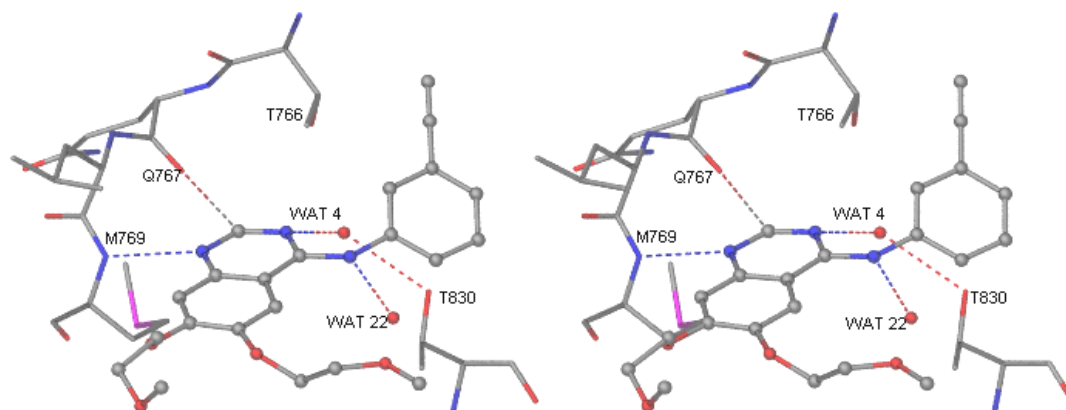
**Table 7.8:** Strong and weak hydrogen bonds for docked solutions of cyanoquinoline (in 1M17) and anilinoquinazoline (in 1XKK) inhibitors. Molecule 8 is a different class of compound.

Molecule	MET769 [N–H···N(1)]	THR766 [O–H···N]	GLN767 [C–H···O]	WAT 4 [Ow–H···N(3)]	WAT 22 [N–H···Ow]
Cyanoquinoline					
<b>2</b>	1.85	1.71	2.17		
<b>3</b>	1.85	1.97	2.29		
<b>5</b>	1.8	2.25	2.46		
<b>6</b>	1.92	1.75	2.27		
Anilinoquinazoline					
<b>1</b>	1.84	-	2.04	1.95	2.22
<b>4</b>	2.05	-	2.25	1.89	2.01
<b>7</b>	2.11	-	2.35	1.85	1.68
<b>8</b>	1.61	-	1.99	2.08	

Possible reasons for selective binding (docking scores) of cyanoquinolines to the active conformations are: (1) Open conformation of the active site in 1M17. This facilitates accessibility of the polar hydroxyl group of THR766 to the 3-cyano group in the ligand (Figure 7.11); (2) The water molecule (WAT 10 in 1M17 or WAT 4 in 1XKK) sterically hinders the correct positioning of the cyano group, (3) Crystal structure of 1XKK shows that the G-loop may also hinder interactions to THR766, The orientation of OH<sup>y</sup> of THR766 is far from the 3-cyano group. In anilinoquinazolines, water (WAT 10 in 1M17 or WAT 4 in 1XKK) forms a Ow–H···N(3) interaction as shown in Figure. 7.12 (shows docking solution of molecule **7** in 1XKK). In 1M17, WAT 10 forms a hydrogen bond with the OH<sup>y</sup> of the THR766. Similarly in 1XKK, WAT 4 interacts with THR830 through a O–H···O hydrogen bond. Therefore WAT 10 and the 3-cyano group are supramolecular isosteres in the present context. Quinazolines also show favorable binding pose in 1M17. Quinazolines form N–H···N(1) and Ow–H···N(3) interactions with MET769 and WAT 10, but the additional water (WAT 22) bonded to the quinazoline NH through a N–H···Ow hydrogen bond and the greater contact surface, discussed later in the chapter, provide extra stabilization to the quinazoline binding in 1XKK. These interactions are consistent over all the quinazolines. Sometimes water (WAT 22) interacts with the 6<sup>th</sup> position substituent with an Ow–H···O hydrogen bond.



**Figure 7.11:** Stereo view of GOLD best solution for molecule **5** (cyanoquinoline) in 1M17 crystal structure.



**Figure 7.12:** Stereo view of GOLD best solution for molecule **7** (anilinoquinazoline) in X-ray structure of 1XKK.

### 7.3.7 Ligand based pharmacophore model versus structure based model

Ligand receptor interaction is based on mutual rearrangement of both ligand and protein in order to attain complementary shapes required for binding. Ligand and protein independently may have various local or global energy minima conformations, but the complex state energy minima are not the same [7.55, 7.56]. In the present study, the ligand conformation generated in Catalyst is generally different from the conformation present in the crystal structure. This is because the crystal structure conformation is not found among the lowest energy conformation pool. For instance comparison of the crystal structure conformation of Erlotinib with that generated by Catalyst reveals that the energy of crystal structure conformation is 6 kcal/mol. This is because pharmacophore modeling in Catalyst is usually carried out on conformations generated within a threshold limit of 20 kcal/mol. The model is built on sample conformations which are local energy minima for each structure. In the present situation the sampled conformations used for model generation have energy minima well below the crystal structure conformation. In another words, the crystal structure conformation is a higher energy conformation (in vacuum) and has higher strain energy in its complexed form. To verify this hypothesis that the crystal structure conformation is in a state of higher energy, the strain energies were calculated in MOE for both ligands present in crystal structures using the E strain descriptor. For Erlotinib and Lapatinib the strain energies are 60 kcal/mol and 78 kcal/mol respectively. Similarly for eight molecules, the strain energies of the docked conformations are listed in Table. 7.9. However this strain energy is counterbalanced through enthalpy gains of various interactions like hydrogen bonds with water molecules and hydrophobic interactions.

**Table 7.9:** Docking scores and E<sub>strain</sub> energies for eight inhibitors. Interaction energies for anilinoquinazoline and cyanoquinoline are separately mentioned for active (1M17), inactive form (1XKK) of the EGFR kinase domain.

Molecule	pIC <sub>50</sub>	Fitness Score	Chemscore ΔG	E <sub>strain</sub>
1M17				
Quinoline				
2	7.09	32.3853	-34.4174	78677.91
3	7.08	32.2471	-33.3677	49342.44
5	6.07	27.4658	-32.1522	1400.799
6	5.98	27.2096	-32.1036	1242.748
Quinazoline				
1	7.13	27.7393	-31.5134	117.428
4	6.28	27.632	-30.4159	66.24467
7(Erlotinib)	5.83	27.3265	-30.1183	45.28573
8	5.80	27.1163	-27.6344	31.3697
1XKK				
Quinazoline				
1	7.13	28.518	-37.1237	59.36667
4	6.28	28.4873	-34.0155	44.67819
7(Erlotinib)	5.83	27.5014	-36.5353	29.54187
8	5.80	27.3567	-31.7945	43.93411
Quinoline				
2	7.09	22.6237	-36.9178	1086.931
3	7.08	21.913	-34.3228	3226.395
5	6.07	26.2797	-34.3154	1411.588
6	5.98	31.0039	-34.5998	1602.754

Despite differences in conformations, the ligand based pharmacophore model has generated features similar to the structure based pharmacophore model. The pharmacophore features observed in *HipHop* is similar to those in the receptor based model. HBA1 and HBA3 are shown having hydrogen bonds to MET769 and WAT 10. The docking result support HBA2 can exist, wherein WAT 22 in 1XKK is an important counterpoint to the acceptor for substitution at the 6<sup>th</sup> position. Similarly HBD is exhibited by quinazoline NH to water acceptor (WAT 22) in 1XKK. The hydrophobic aliphatic and aromatic interactions are well in agreement with the pharmacophore features of the crystal structures. The *HypoGen* model on the other hand contains features which are a subset of the features

observed in the structure based and *HipHop* models. Features like HBA1, HBA2, HYA1, HYR present in this model are able to predict well the activities of EGFR inhibitors. Presumably, features like HBA3, HBD, HYA2 which are absent in this model, negotiate the ligand strain energy by providing stabilizing interactions in the active site.

## 7.4 Conclusions

This chapter describes a method of LBVS for EGFR kinase inhibitors through pharmacophore modeling. A qualitative pharmacophore model was designed which describes the important pharmacophore features present in a set of active compounds. Another quantitative pharmacophore model was modeled through the selection of a structural similarity driven training set, taking into structure activity correlation. It was demonstrated that the quantitative pharmacophore model has better success in database mining of EGFR kinase inhibitors. This validation was carried out using both EGFR and spiked EGFR kinase inhibitors in a non-EGFR kinase inhibitor database. The predictive power of the quantitative model was validated in an external dataset of 128 EGFR kinase inhibitors. A four feature pharmacophore model, with two hydrogen bond acceptors, one hydrophobic aliphatic and a hydrophobic aromatic features, best describes the activity of EGFR kinase inhibitors.

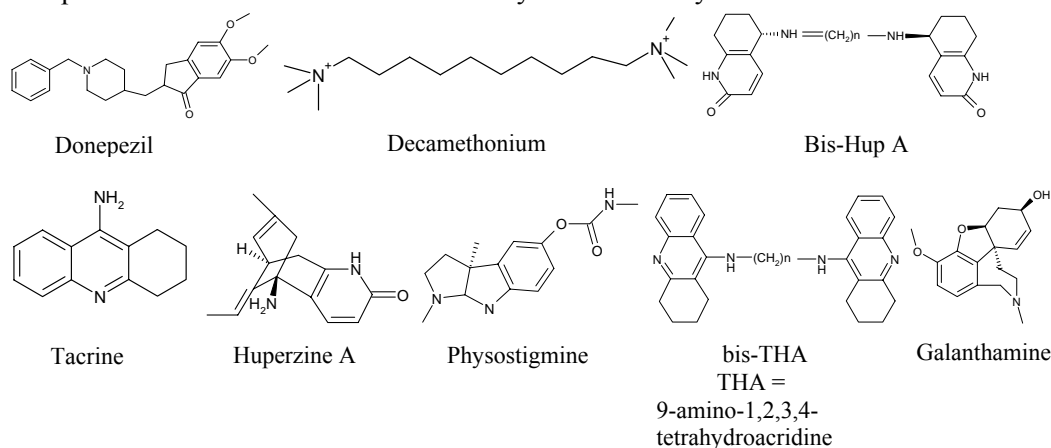
A structure based pharmacophore model was designed for two crystal structures, representing active and inactive form of the EGFR kinase domain. The ligand based pharmacophore model and the structure based pharmacophore model complement each other in various pharmacophore features. A docking study of 8 known inhibitors in the active and inactive form of the receptor reveals selectivity of the receptor toward anilinoquinazoline and cyanoquinoline. The anilinoquinazoline type of ligands have a similar binding mode irrespective of the active and inactive form of receptor, whereas cyanoquinolines have a preferred pose in the active form of the kinase domain. Ligands of the cyanoquinoline type displace a water molecule (WAT 10) to have a preferred pose in the active site. The same water molecule bridges the anilinoquinazoline ligand and THR766 (in 1M17)/THR830 (in 1XKK). In summary a VS model for large scale screening of ligand libraries against EGFR kinase has been successfully developed and validated. Additionally, the structural basis of ligand selectivity for the active and inactive forms of the EGFR kinase domain was demonstrated through docking.

## CHAPTER 8

# PHARMACOPHORE MODELING, DOCKING AND VIRTUAL SCREENING OF ACETYLCHOLINESTERASE INHIBITORS

### 8.1 Introduction

Alzheimer's disease is a progressive and fatal neurodegenerative disorder manifested by cognitive and memory deterioration, progressive impairment of activities of daily living, a variety of neuropsychiatric symptoms and behavioural disturbances [8.1]. Acetylcholinesterase (AChE) inhibitors are the only major drugs approved for the symptomatic treatment of Alzheimer's disease [8.2]. AChE is predominantly present in the central, peripheral nervous system and in skeletal muscles. It causes the termination of impulse transmission by rapid hydrolysis of acetylcholine to yield acetic acid and choline [8.3, 8.4]. Lower level hydrolysis of the neurotransmitter acetylcholine, is a feature associated with Alzheimer's disease [8.5]. There is a steady-state increase in acetylcholine resulting from cholinesterase inhibition in the brain resulting in improvement of cognitive function and mild-to-moderate cases of Alzheimer's disease [8.6]. Therefore cholinesterase inhibitors stabilize Alzheimer's disease. Hence, AChE inhibition is deemed a useful strategy in the design and development of drug candidates for the treatment of Alzheimer's disease, as exemplified by the first approved drug, tacrine. A diverse set of compounds binds to this catalytic site and elicit the catalytic process shown in Scheme 8.1 [8.7]. Two inhibitors, donepezil and tacrine have been successfully used clinically for AChE inhibition.



**Scheme 8.1:** Known AChE inhibitors.

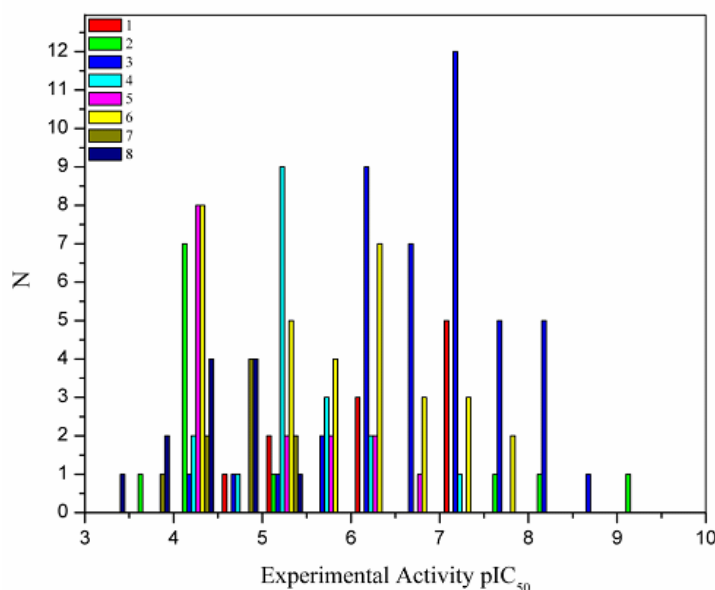
AChE is among the best studied targets with molecular modeling. Many computational studies such as theoretical [8.8], molecular dynamics simulations (MD) [8.9], quantitative structure activity relationship (QSAR) [8.10], docking [8.11] and virtual screening (VS) [8.12] have been performed on AChE. The scope of the present work lies in the determination of a virtual screening (VS) method and its application for the system of choice. The aim of the present chapter is to develop a virtual screening (VS) model that can combine both the accuracy of docking and the speed of the pharmacophore method for the AChE inhibitor.

## 8.2 Materials and methods

### 8.2.1 Structure building

A data set of 153 unique AChE inhibitors was prepared with the MOE software [8.13]. These inhibitors were collected from published papers in various journals. The inhibitors include morpholinoalkylcarbamoxyloxyseroline derivatives (11 compounds) [8.14], pyripyropene (12 compounds) [8.15], E2020 analogues as acetylcholinesterase (14 compounds) [8.16], piperidinium and pyridinium agents (30 compounds) [8.17], 11H-indeno-[1,2-b]-quinolin-10-ylamine derivatives (18 compounds) [8.18], N-benzylpiperidine aminoacid derivatives (15 compounds) [8.19], Velnacrine thiaanalogue (13 compounds) [8.20], 9-amino-1,2,3,4-tetrahydroacridine-based quaternary salts of 2-[(hydroxyimino)methyl]imidazole (9 compounds) [8.21], 1-(alkoxymethyl)-2-[(hydroxyimino)methyl]-3-methylimidazolium halides (12 compounds) [8.22], tacrine analogues (19 compounds) [8.23]. The activity profile of the above mentioned compounds is shown in Figure 8.1. The chemical structures and corresponding references of these analogues are mentioned in Appendix III, Table 1.

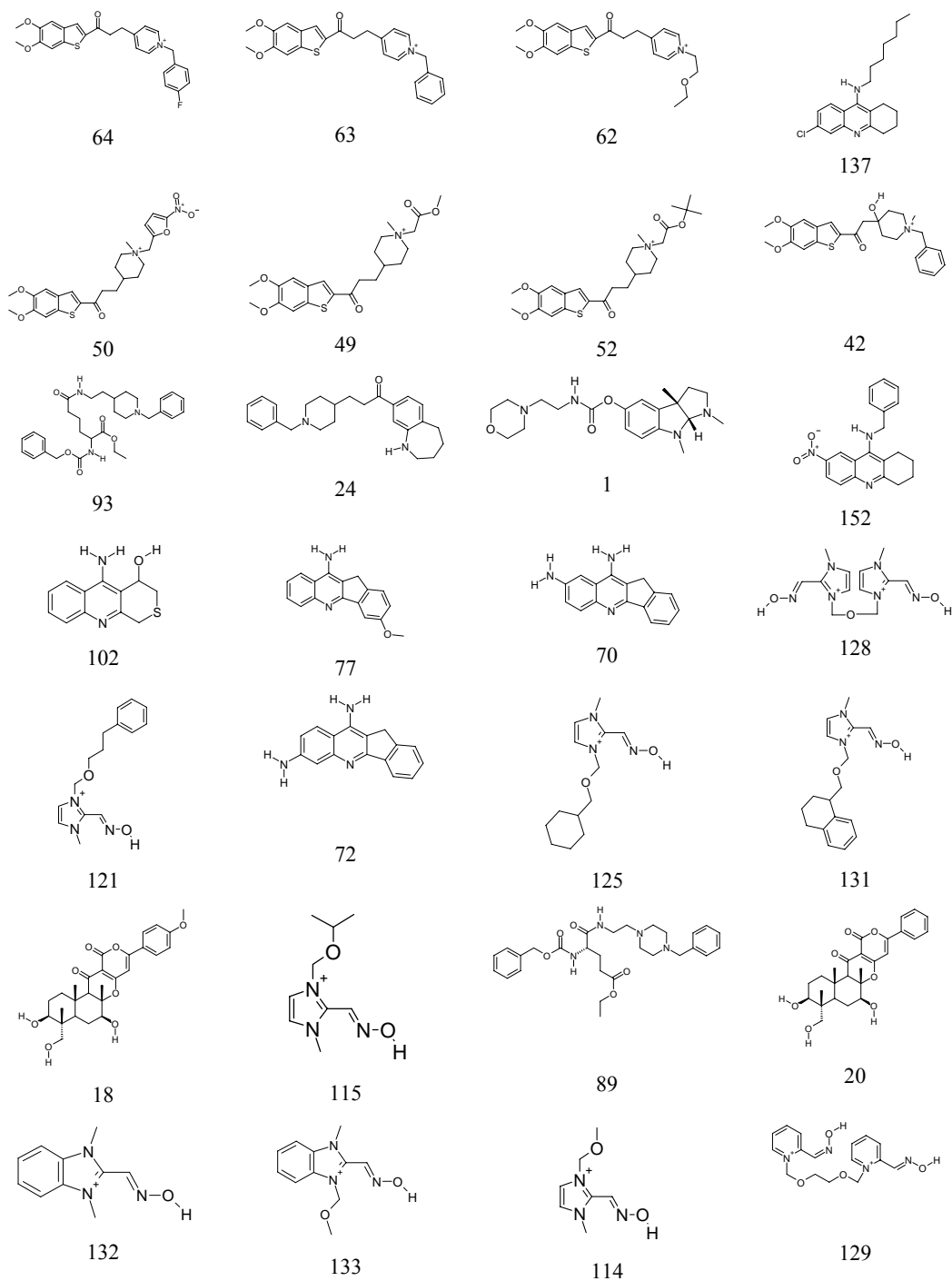
For these inhibitors, geometry optimizations and minimizations were carried out using the MMFFx force field in MOE [8.24]. The optimized ligands were exported to Catalyst 4.7 for pharmacophore generation [8.25]. The conformations for each inhibitor were generated with the best conformation option in the Catalyst ConFirm module. A maximum of 250 conformations were generated for each molecule within an energy threshold of 20.0 kcal/mol above the global energy minimum. Pharmacophore modeling in Catalyst was carried out with the *HypoGen* module and the docking study was performed with GOLD 3.0 [8.26].



**Figure 8.1:** Frequency distribution of experimental activities for inhibitors, **1:** Morpholinoalkylcarbamoyloxyseroline derivatives, **2:** pyripyropene, **3:** E2020 analogues, piperidinium and pyridinium agents, **4:** 11H-indeno-[1,2-b]-quinolin-10-ylamine derivatives, **5:** N-benzylpiperidine aminoacid derivatives, **6:** Velnacrine thiaanalogs and tacrine analogues, **7:** quaternary salts of 2-[(hydroxyimino)methyl]imidazole, **8:** 1-(alkoxymethyl)-2-[(hydroxyimino)methyl]-3-methylimidazolium halides.

### 8.2.2 HypoGen

The training set was randomly selected from the data set and is shown in Figure 8.2. The reported experimental activities were converted to pIC<sub>50</sub> values. The inhibitory activities (pIC<sub>50</sub>) of the selected compounds range from 8.5 to 3.4. The pharmacophore model was constructed using chemical features like hydrogen bond acceptor (HBA), hydrophobic aliphatic (HYA), ring aromatic (RA) and positive ionizable (PI). Pharmacophore generation was allowed to have a maximum of 5 features. The default uncertainty value of 3 was specified for each compound. *HypoGen* generates pharmacophore models that are a set of features in 3D space, each containing a certain tolerance and weight that fits to the features of the training set, and further correlates to the activity. This is carried out by comparing features which are common among the active compounds but not present in the inactive compounds. The best *HypoGen* model was used to predict the activity of 125 inhibitors present in the data set.



**Figure 8.2:** Training set molecules considered in *HypoGen* model generation. Numbers correspond to the compounds serial number.

### 8.2.3 Docking (GOLD)

The crystal structure of AChE (PDB ID 1B41), human acetylcholinesterase complexed with fasciculin-II, was obtained from the protein databank [8.27, 8.28]. H-atoms were added to the protein in the hydrogen adding option in MOE. The complexed ligand fasciculin-II was removed from the crystal structure, and then subjected to minimization by keeping the heavy atoms rigid. This was carried out in the MOE minimizer with MMFFx force field. This minimized structure was then exported to GOLD for docking.

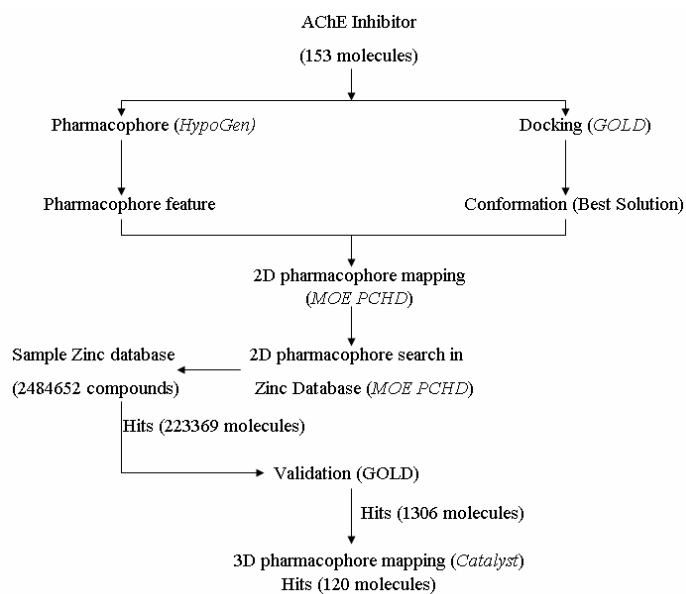
The compiled data set of 153 compounds were docked into the crystal structure using GOLD software. The active site of human AChE is characterized by a deep and narrow gorge, which penetrates half way into the enzyme; the catalytic site is present about 4 Å from its base [8.27]. A region of 10 Å radius around this gorge was selected as the active site. The default set of parameters were used during docking. For each of the 10 independent Genetic Algorithm (GA) runs, with a selection pressure 1.1, 100,000 GA operations were performed on a set of 5 islands. The population size of 200 individuals was specified. Default operator weights were used for crossover, mutation, and migration of 95, 95 and 10 respectively. To further speed up the calculation, the GA docking was stopped when the top three solutions were within 1.5 Å RMSD of each other. All other values were set to the default. While docking, a limited flexibility is allowed for the hydrogen atoms in the  $-OH$ ,  $-NH_3^+$  substituents of the side chains of Ser, Thr and Lys residues. The water molecules present in the active site were allowed for movement in due course of docking.

### 8.2.4 Database screening

*Database generation:* Compounds were collected from the ZINC database (zinc5, released Jan 2005) [8.29]. ZINC is a web-based repository database of commercially available compounds. A sample database of 2484652 compounds was built in the MOE database. This database comprises molecules provided by various vendors like NCI (161395), Ambinter (721594), ChemDiv (199997), Asinex (378794), CamBridge (599893), ChomGenex (40659), KeyOrganics (26932), LifeChemical (64660), Pubchem (57858), Maybridge (49803), Nanosyn (83067). The functional groups of each compound were mapped with the PCHD pharmacophore scheme in MOE database.

*Database screening:* To combine pharmacophore features and docked conformation, the features generated in the *HypoGen* model were assigned to the best

GOLD solution of compound **64** in the MOE pharmacophore generation module. The PCHD scheme in the pharmacophore builder was used for this purpose. The prepared ZINC database was screened with the 2-D pharmacophore in the MOE database. Generally the 2-D pharmacophore mapping is carried out through the presence or absence of functional groups corresponding to the specified pharmacophore feature. This method avoids 3-D conformation mapping of the molecule. The 3-D coordinates were built for the hit molecules and then subjected to minimization in MOE. The hits retrieved from this 2-D pharmacophore fitting were exported to GOLD for docking and interaction analysis with the receptor (1B41.pdb). For these molecules, calculations were speeded up to 7 times that of default setup. For each of the 5 independent Genetic Algorithm (GA) runs, with a selection pressure 1.1, 1000 GA operations were performed on a set of 3 islands. The population size of 100 individuals was specified. Default operator weights were used for crossover, mutation, and migration of 95, 95 and 10 respectively. The top ranked molecules in the docking were exported to Catalyst for 3-D pharmacophore fitting and screening. A conformation database was prepared for these molecules in Catalyst. The conformations were generated with the fast conformation option of the Catalyst ConForm module. A maximum of 100 conformations were generated for each molecule within an energy threshold of 20.00 kcal/mol above the global energy minimum. The *HypoGen* model was used for database screening with the “best flexible search” method in Catalyst. The overview of various steps followed in the study is shown in Figure 8.3.



**Figure 8.3:** Flowchart showing the various steps followed in the study.

All the studies were carried out in *Silicon Graphic Octane 2* and *ORIGIN 200* server. The *Octane 2* work station has a 1GB RAM and the *ORIGIN 200* server has a 512 MB RAM. Both work stations are single processor 1020MHz R12000. The sample Zinc database was prepared in a week's time in the *Octane 2* workstation. The assignment of the pharmacophore PCHD scheme and the screening was carried out in 15 days. The docking study for 223369 molecules was performed in 45 days time in the *ORIGIN 200* server. The final screening of the top score molecules of GOLD docking was carried out in 3 hours.

### 8.3 Results and discussion

#### 8.3.1 HypoGen

The best pharmacophore model was selected after several iterations of the modeling exercise. A set of ten pharmacophore models with the lowest cost differences were selected for further analysis are and shown in Table 8.1.

**Table 8.1:** Ranking of pharmacophore models generated by *HypoGen*.

Model	Total cost	Cost difference	RMSD	Training set correlation	Features <sup>a</sup>
1	123.61	53.08	0.97	0.94	HBA, HYA, PI, RA
2	128.76	50.93	1.15	0.93	HBA, HYA, PI, RA
3	130.43	46.26	1.19	0.93	HBA, HYA, HYA, PI, RA
4	130.78	45.90	1.21	0.93	HBA, HYA, PI, RA
5	131.25	45.44	1.22	0.93	HBA, HYA, PI, RA
6	132.27	44.42	1.24	0.92	HBA, HYA, PI, RA
7	133.11	43.58	1.27	0.92	HBA, HYA, HYA, PI, RA
8	133.18	43.51	1.28	0.92	HBA, HYA, PI, RA
9	133.60	43.09	1.27	0.92	HBA, HYA, PI, RA
10	134.19	42.50	1.30	0.91	HBA, HYA, PI, RA

Null cost = 176.69, Fixed cost = 110.24, Configuration = 12.48, All cost units are in bits <sup>a</sup> Hydrogen Bond Acceptor (HBA), Hydrophobic aliphatic (HYA), Positive ionizable (PI), Ring aromatic (RA).

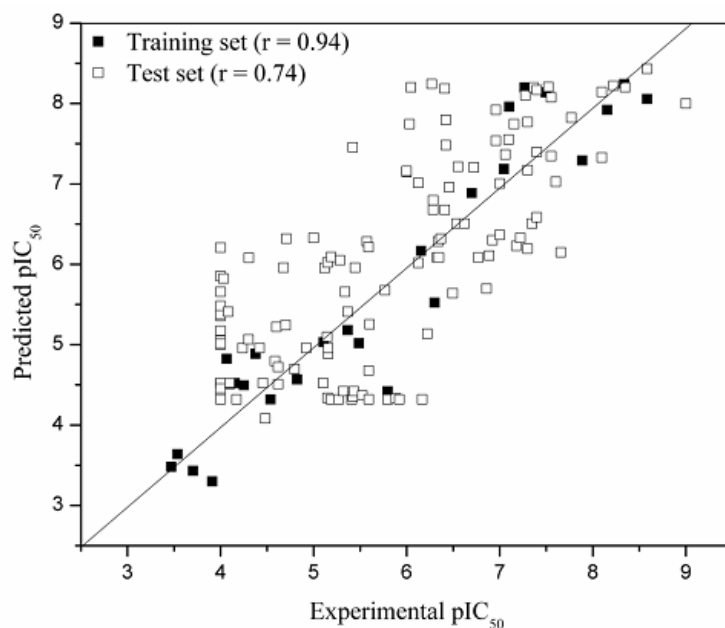
Based on various statistical parameters, model 1 is found to be the best. This model has the lowest RMSD (0.97) and highest correlation coefficient ( $r = 0.94$ ). The training set compounds are tabulated along with the experimental and predicted activity in Table 8.2. The activities were predicted for the remaining 125 test set compounds using model 1. A correlation coefficient,  $r = 0.74$  was achieved for this set. The correlation between

experimental and predicted inhibitory activities for the training set and the test set compounds is shown in Figure 8.4.

**Table 8.2:** List of compounds considered in the training set with experimental and predicted  $\text{pIC}_{50}$  values.

Sl. No.	Compound Sl. No.	Experimental $\text{pIC}_{50}$	Predicted $\text{pIC}_{50}$	Fit value <sup>a</sup>
1	64	8.58	8.05	9.57
2	63	8.33	8.24	9.69
3	62	8.15	7.92	9.24
4	137	7.88	7.29	8.51
5	50	7.49	8.13	8.40
6	49	7.26	8.2	8.49
7	52	7.10	7.95	8.85
8	42	7.04	7.18	8.68
9	93	6.69	6.88	7.71
10	24	6.30	5.52	7.48
11	1	6.15	6.16	7.36
12	152	5.79	4.42	6.55
13	102	5.48	5.01	6.31
14	77	5.36	5.18	6.33
15	70	5.22	4.31	6.30
16	128	5.1	5.03	6.27
17	121	4.82	4.56	6.26
18	72	4.53	4.31	6.31
19	125	4.37	4.88	5.91
20	131	4.25	4.49	6.33
21	18	4.15	4.52	5.75
22	115	4.09	4.5	6.33
23	89	4.06	4.82	6.44
24	20	4	4.52	6.03
25	132	3.91	3.30	4.81
26	133	3.7	3.43	4.91
27	114	3.53	3.63	4.31
28	129	3.46	3.48	4.31

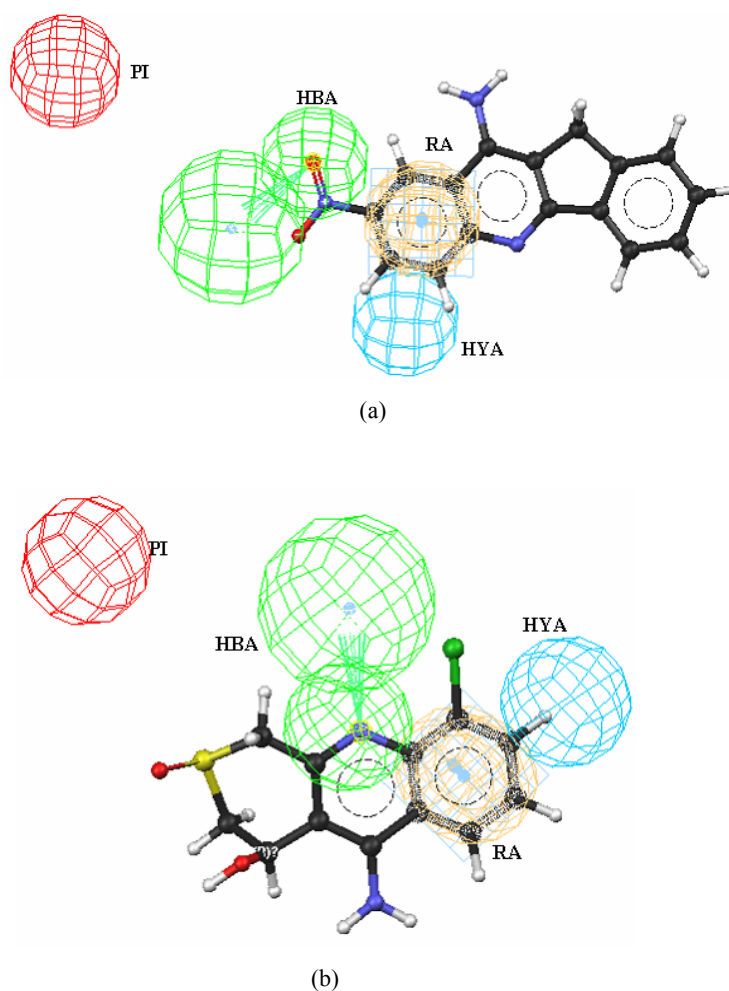
<sup>a</sup> Fit value indicates how well the features in the pharmacophore overlap with the chemical features in the molecule.



**Figure 8.4:** Scatter plot of predicted and experimental  $pIC_{50}$  value for 28 training set and 125 test set compounds.

The scatter plot suggests that the prediction for 24 molecules is poor. These poorly predicted molecules are mostly inactive. Among these, 16 are false positives and 8 are false negatives. The false positives are velnacrine thiaanalogue (compound **103**, **104**, **105**, **106**, **108**, **111**, **113**), pyripyropene derivative (**16**), huperazine (**35**), 11H-indeno-[1,2-b]-quinolin-10-ylamine derivative (**69**) and N-benzylpiperidine aminoacid derivatives (**88**, **90**, **91**, **92**, **95**, **99**). The reported inhibitory activity for these molecules is  $IC_{50} > 100$   $\mu$ M ( $pIC_{50}$  value 4). Therefore the predicted activities for these molecules are distinct from each other (ranges from 5 to 6.2  $pIC_{50}$  value). Similarly the predicted activities for the 8 compounds [1 compound from the pyridinium derivative (molecule **68**), 6 compounds of 11H-indeno-[1,2-b]-quinolin-10-ylamine derivatives (**71**, **74**, **75**, **76**, **78**, **81**), 1 velnacrine thiaanalogue (**110**)] are the same, i. e. the predicted  $pIC_{50}$  value is 4.31. Irrespective of the range of experimental activity (which ranges from  $pIC_{50}$  4 to 6.16), the predicted activity is the same. This is because of the improper mapping of the pharmacophore features in the *HypoGen* model. The two outliers compound **69** and **111** are shown in Figure 8.5. This result shows that the pharmacophore model has limited success in predicting the inhibitory activity of inactive compounds. Since these outliers belong to the inactive set of compounds, it did not hamper the predictive capacity of the pharmacophore model in the higher activity range. In fact one

of the aims of any pharmacophore modeling study is to identify active compound from a pool of inactive compounds.



**Figure 8.5:** A sample of two outlier (a) **69** and (b) **111** are shown here. Note the improper mapping of pharmacophore features **HYA** and **PI** in both the compounds.

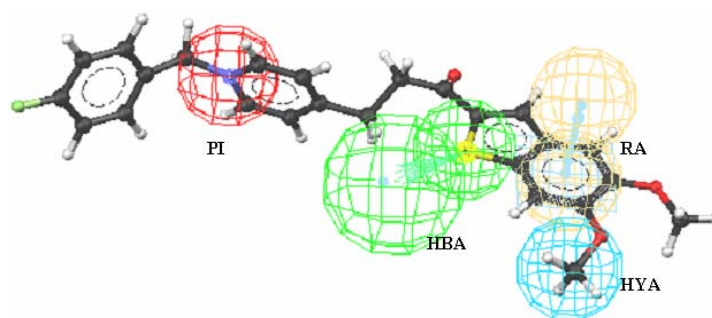
The best pharmacophore model contains five features shown in Figure 8.6. Magenta indicate positive ionizable (PI), green contours are hydrogen bond acceptor (HBA), directionalities shown as conical projections, blue contours indicate hydrophobic aliphatic (HYA), and brown contours represents ring aromatic (RA). The distance matrix for this pharmacophore model is given in Table 8.3.

**Table 8.3:** Distance matrix (in Å) of the generated *HypoGen* model.

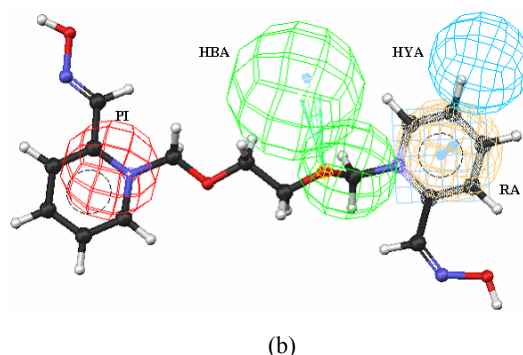
	HBA1 (IP)	HBA1 (PP)	HYA	PI	RA (IP)	RA (PP)
HBA1 (IP)						
HBA1 (PP)	3.0					
HYA	5.4	5.9				
PI	7.6	6.5	12.3			
RA (IP)	3.4	5.4	3.2	11.0		
RA (PP)	4.5	6.2	5.2	11.0	3.0	

IP, initial point on ligand, PP, projected point on the receptor.

The most active molecule **64** in the training set has a fitness score of 9.57 when mapped to the pharmacophore (Figure 8.6a), while inactive **129** maps to a value of 4.91 (Figure 8.6b). In molecule **64** the HYA feature corresponds to the *meta*-methoxy substitution on the benzothiazole ring. The HBA feature is mapped to the sulfur atom of the benzothiazole ring. RA is mapped onto the aromatic ring of the benzothiazole ring. PI corresponds to the quaternary nitrogen atom of the pyridinium ring. For inactive molecules, HBA is mapped to the methoxy group of the imidazole 3rd position. RA has been mapped to imidazole ring and HYA has been mapped to N-methyl group of imidazole ring. Noteworthy is the absence of the HYA group in molecule **129**. These results suggest a fair performance of the *HypoGen* model to predict the inhibitory activity of AChE inhibitors.



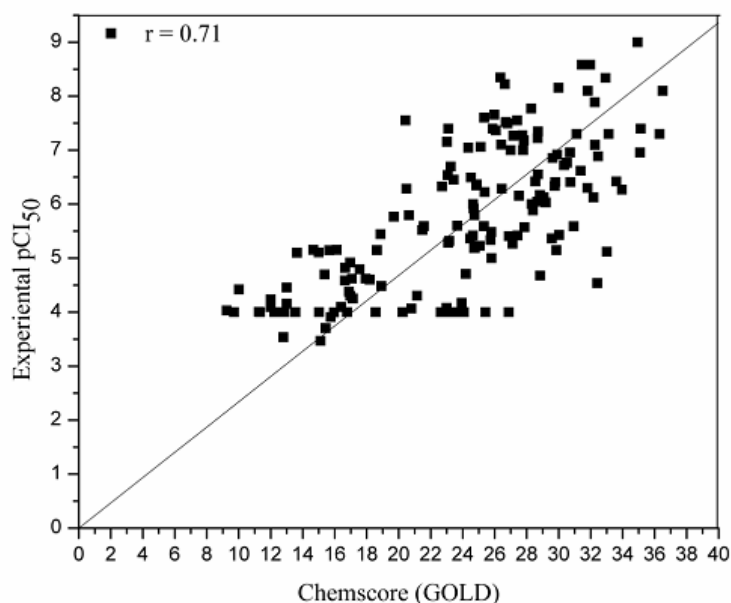
(a)



**Figure 8.6:** *HypoGen* model map on to: (a) most active molecule **64**, (b) inactive molecule **129**. The green contours represent the positioning of hydrogen bond acceptors (HBA) with their projected points on receptor indicated by conical projections. Blue contours represent the hydrophobic aliphatic (HYA) and brown color corresponds to aromatic (RA), red corresponds to positive ionizable (PI).

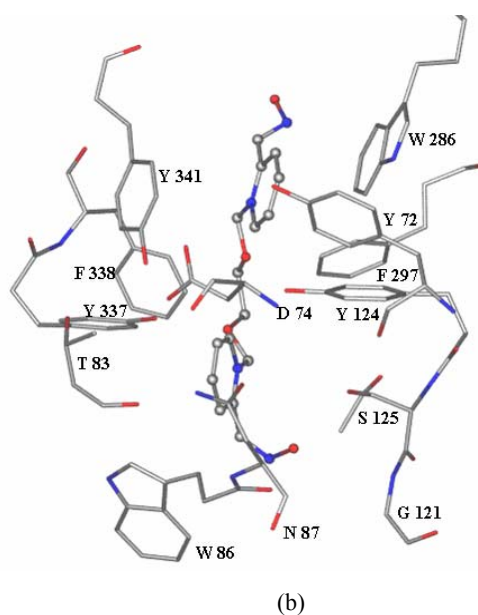
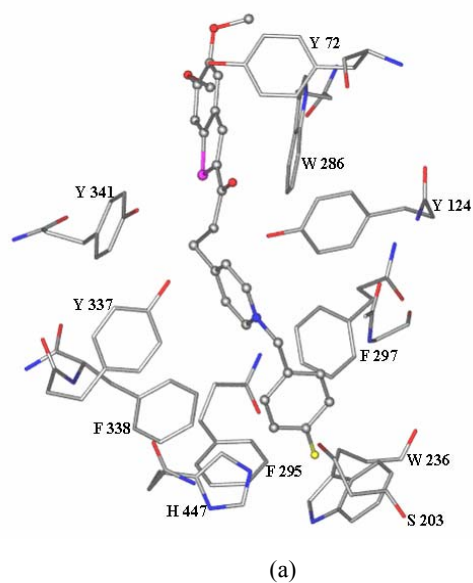
### 8.3.2 Docking analysis

The key residues present in the active site gorge are TRP286, LEU76, TYR124, PHE297, THR83, TYR337, PHE338, PHE295, TRP236, TRP86, HIS447, SER203, GLU334 [8.27, 8.30]. All the 153 ligands have been docked into 1B41, using genetic algorithm techniques of Chemscore (GOLD). The docking scores for the best possible conformation and orientation were correlated with observed biological activity (Figure 8.7). Docking results of all AChE inhibitors with 1B41 show better correlation with biological activity; molecules with higher activity exhibited better docking scores. Barring a few inactive molecules, almost all the ligand molecules bind with reasonable scores. These inactive molecules are in the same set of compounds which has experimental activity  $\text{pIC}_{50} > 4$ , and are shown to be outliers in the pharmacophore model.



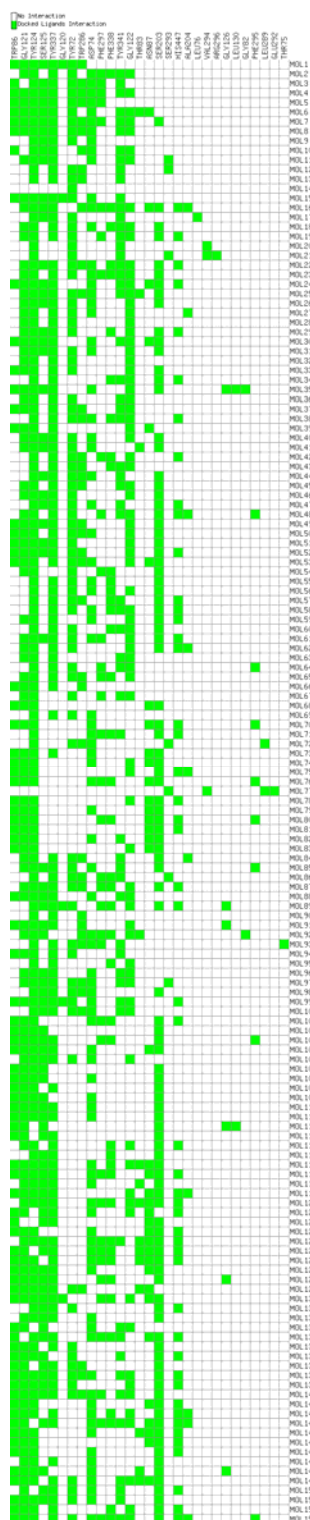
**Figure 8.7:** Scatter plot of Chemscore (GOLD) vs activity ( $pIC_{50}$ ) for 153 molecules.

The best docking solutions of molecule **64** and **129** are discussed here to address the ligand–protein interaction of the active/inactive molecule with AChE. Comparative docking scores for molecule **64** and **129** are 31 and 15 respectively, i. e. the docking score of active molecule is twice that of the inactive molecule. The orientation of molecule **64** is such that it completely occupies the active site of AChE as shown in Figure 8.8*a*. At one end, the aromatic ring of the benzothiazole ring interacts with amino acid residue TRP286 through a  $\pi$ - $\pi$  interaction. This is where the aromatic (RA) nature of pharmacophore is elucidated in the *HypoGen* model. The carbonyl oxygen of ligand interacts with  $OH^N$  of TYR124 through an O–H $\cdots$ O hydrogen bond ( $d = 1.7$ ,  $\theta = 145^\circ$ ). In the case of molecule **129**, the only hydrogen bond found is between the terminal hydroxyl group of the ligand and  $OH^V$  of SER125 ( $d = 1.9$ ,  $\theta = 150^\circ$ ) Figure 8.8*b*.

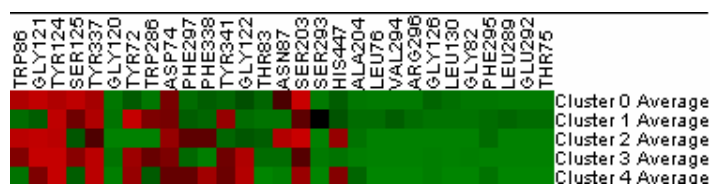


**Figure 8.8:** (a) Best GOLD (Chemscore) solutions for molecule **64** and (b) **129** are shown in the active site of 1B41.pdb.

The interactions for best solutions of 153 molecules are also shown in the interaction array through a hit map (Figure 8.9). A detailed analysis of interactions reveals that the interactions observed for these inhibitors can be clustered into 5 categories (Figure 8.10 and Table 8.4).



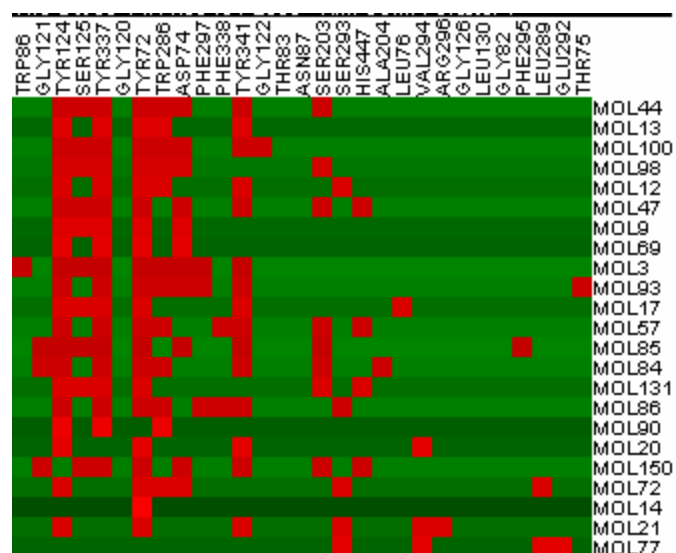
**Figure 8.9:** Hydrogen bond interaction array shown for the docking solutions of 153 AChE inhibitors. The key residues are shown in the X-axis and molecule serial numbers in the Y-axis.



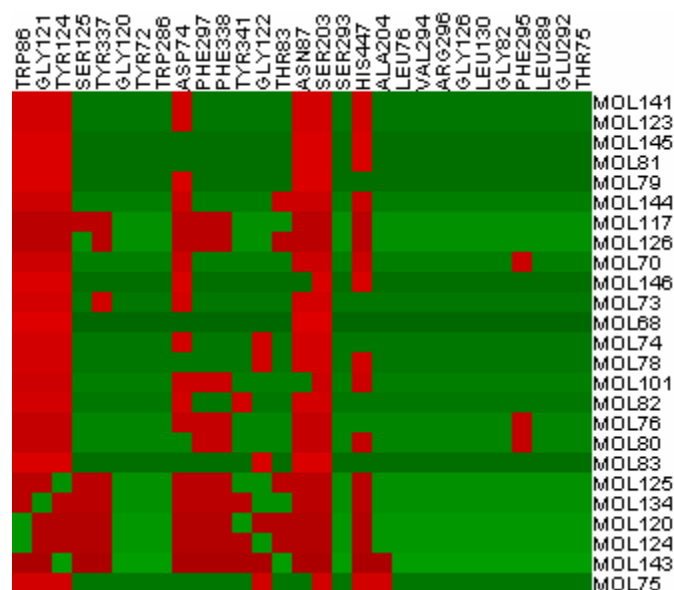
(a)



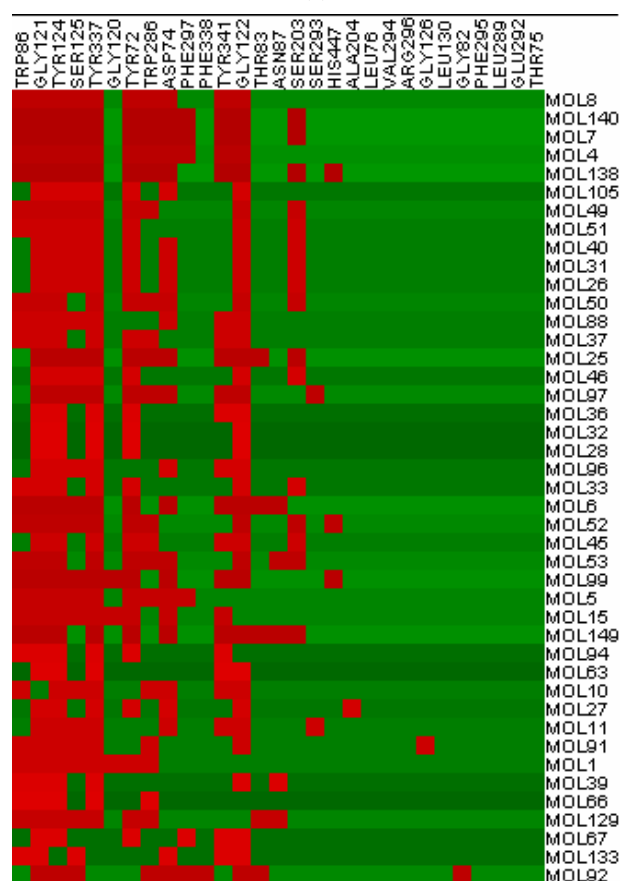
(b)



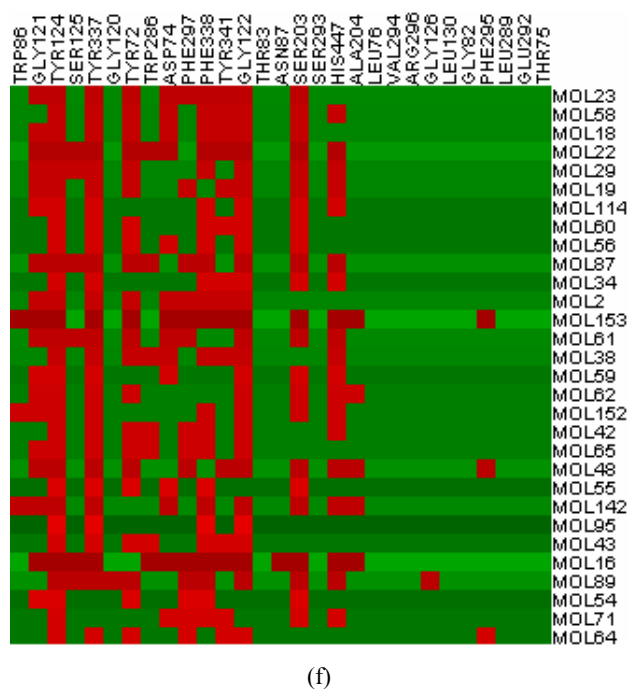
(c)



(d)



(e)



**Figure 8.10:** Docked 153 ligands are clustered in to 5 category based on hydrogen bond interaction in the active site of AChE. Average of the interaction observed for each cluster is shown in the hitmap. (a) The X-axis represents key residues of the active site while the cluster code/molecule is shown in Y. The individual member of each clusters are represented by separate hit maps. (b) Cluster 0, (c) Cluster 1, (d) Cluster 2, (e) Cluster 3, (f) Cluster 4.

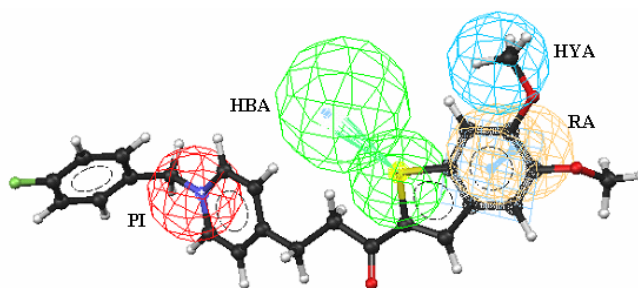
**Table 8.4:** Classification of compounds based on the hydrogen bond observed in the active site in docking study. Compound classified in each cluster are represented through majority basis, e. g. Cluster 3 and 4 contains most of the active molecule present in the dataset. Similarly cluster 0 contains mostly Tacrine and Velnacrine derivatives along with other class compounds. Also listed are the interacting residues in the active site of AChE.

Cluster code	Compound class	Key residues
0	Tacrine and Velnacrine derivatives	TRP86, GLY121, TYR124, SER125, TYR337, ASP74, ASN87, SER203
1	Pyripyropene, 11H-indeno-[1,2-b]-quinolin-10-ylamine derivatives, N-benzylpiperidine aminoacid derivatives	TYR124, SER125, TYR337, TYR72, TRP286, ASP74, TYR341, SER203, SER293
2	11H-indeno-[1,2-b]-quinolin-10-ylamine derivatives, 1-(alkoxymethyl)-2-[(hydroxyimino)methyl]-3-methylimidazolium halides, Velnacrine derivatives	TRP86, GLY121, TYR124, TYR337, ASP74, PHE297, PHE338, ASN87, SER203, HIS447
3	Morpholinoalkylcarbamoyloxyseroline, E2020 Analogues, Piperidinium and Pyridinium Agents, N-benzylpiperidine aminoacid, Tacrine derivatives	TRP86, GLY121, TYR124, SER125, TYR337, TYR72, ASP74, TYR341, GLY122, SER203
4	Pyripyropene, E2020 Analogues, Piperidinium and Pyridinium Agents, Tacrine derivatives	GLY121, TYR124, TYR337, TYR72, ASP74, PHE297, PHE338, TYR341, GLY122, SER203, HIS447

Both active and inactive compounds show hydrogen bond interactions with key amino acid residues like TRP86, GLY121, TYR124, SER125, TYR337, SER203. Among these residues TYR124 and SER125 seems to have conserved interactions with all the compounds.

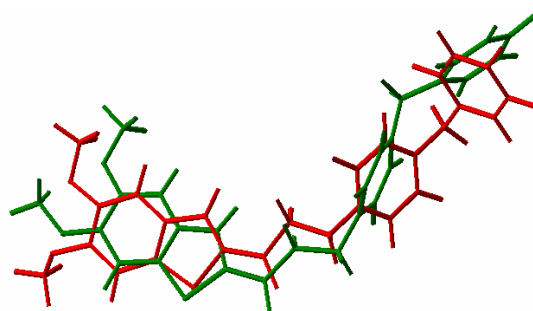
### 8.3.3 Comparative study of pharmacophore and docking

On the basis of pharmacophore features, *HypoGen* model complements the crucial interactions observed in the GOLD docked solutions. HBA is shown to have hydrogen bond interactions with TYR124 (Figure 8.11). The ring aromatic (RA) of the pharmacophore model has  $\pi$ - $\pi$  interactions with amino acid residue TRP286. The positive ionizable (PI) or cation is an important pharmacophore feature which interacts with TRP86.



**Figure 8.11:** *HypoGen* model map on to GOLD best solution for molecule **64**. Pharmacophore annotations are also shown adjacent to pharmacophore features.

The conformation of ligand and protein in 3D space is vital for proper complex formation. It has been recently reported that the bioactive conformations of the ligand depends on molecular weight, number of rotatable bonds etc [8.31, 8.32]. For any computational tool, it is a challenge to generate the bioactive conformation. However limited success has been achieved through various conformation generation tools. In this context the conformation obtained from *HypoGen* model is very similar to conformation obtained in docking solutions, e.g. the RMSD between the GOLD best solution and the *HypoGen* conformation for molecule **64** is 0.91 Å shown in Figure 8.12.



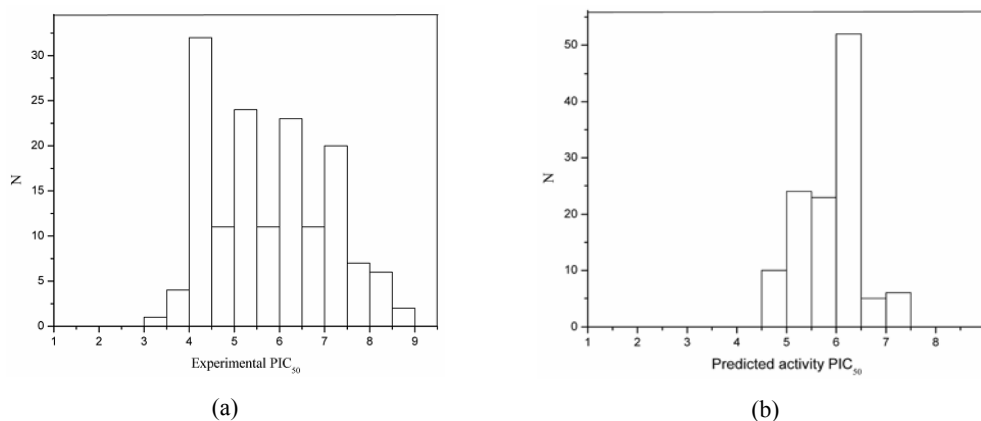
**Figure 8.12:** RMSD between GOLD best solution (red) and *HypoGen* conformation (green) for molecule **64**.

### 8.3.4 Database screening

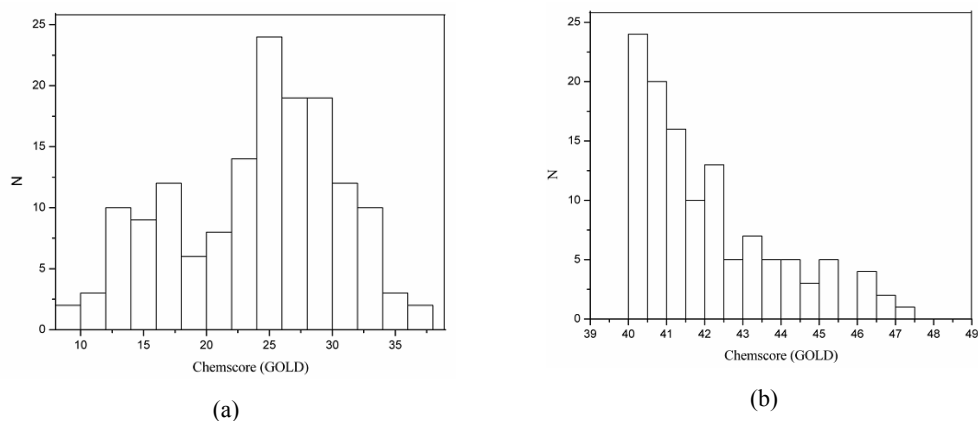
*Screen 1:* The pharmacophore model generated using PCHD pharmacophore scheme was subjected to a 2D pharmacophore search in the sample Zinc database of 2484652 compounds. This search resulted in 223369 hits. These hits are structurally very similar to known inhibitors. Therefore the complementarity between the hits and receptor were studied through docking.

*Screen 2:* The retrieved hits were subjected to docking using the GOLD software. This study further enhances the quality of hits in terms of complementarities with receptor. The docking score for these compounds ranges from  $-50$  to  $50$ . The score for top ranked hits were well above the 153 AChE inhibitors. A set of 1306 top ranked hits (docking score above 40) were found to be suitable for further interactions analysis.

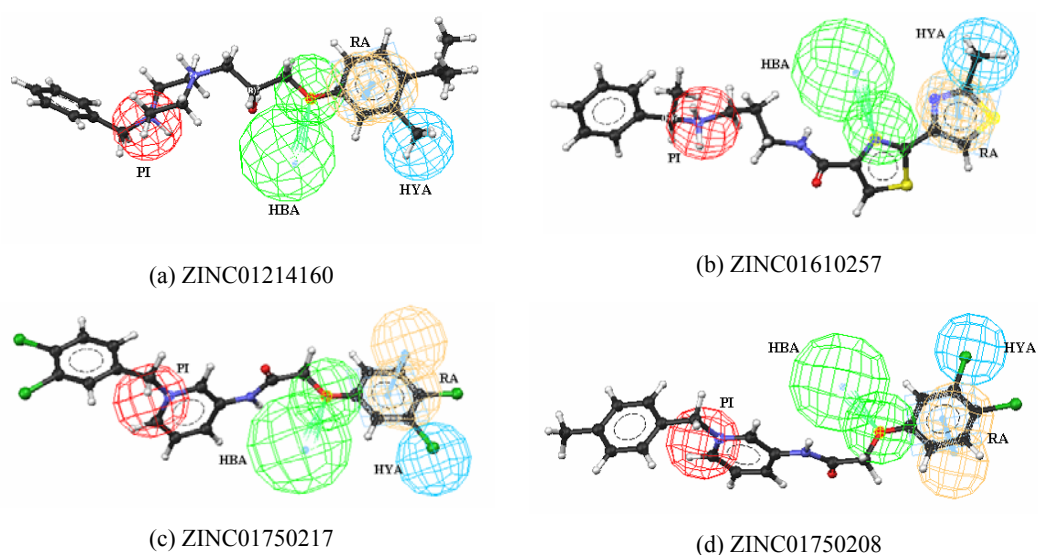
*Screen 3:* Molecules with top ranked GOLD scores (1306 molecules) were again mapped and screened with the *HypoGen* model. The number of hits found in this exercise was 120. A comparative analysis of experimental  $\text{pIC}_{50}$  values for 153 inhibitors and predicted  $\text{pIC}_{50}$  for 120 hits are shown in Figure 8.13. Similarly the docking scores for this set of compounds are shown in Figure 8.14. The predicted  $\text{pIC}_{50}$  values for the majority of hits are in the range of 6-7. About five molecules have predicted  $\text{pIC}_{50} > 7$ . This is comparable to all known inhibitors. Also the docking scores for 120 hits are well above the scores observed for the 153 AChE inhibitors. This suggests that the hits encountered in the database mining have better fitness scores compared to experimentally known 153 inhibitors and may be deemed to show inhibitory activity against AChE. As a representative of 120 hits from ZINC database, samples of four hits are shown in Figure 8.15.



**Figure 8.13:** Frequency distribution of  $\text{pIC}_{50}$  for experimental activity for (a) 153 molecules, and (b) predicted activity for 120 hits.



**Figure 8.14:** Frequency distribution of Chemscore (GOLD) for (a) known 153 AChE inhibitors, and (b) 120 hits from ZINC database.



**Figure 8.15:** HypoGen model map on to a sample of four top ranked molecules of the 120 hits from ZINC database. The respective ZINC codes are mentioned. The predicted pIC<sub>50</sub> values are as follows (a) 7.44, (b) 7.15, (c) 7.13 and (d) 7.07.

## 8.4 Conclusions

The present study describes a rational drug design approach of AChE inhibitors for the treatment of Alzheimer's disease. Various molecular modeling methods such as pharmacophore modeling, molecular docking and virtual screening were successfully implemented to identify possible inhibitors. A diverse set 153 known inhibitors were used to understand pharmacophore features required for AChE inhibitory activity. A reasonably good pharmacophore model was generated using the *HypoGen* module of Catalyst. Regression analysis between pharmacophore features and experimental activity for this set of compounds was carried out successfully. Pharmacophore model containing a hydrophobic aliphatic, ring aromatic, hydrogen bond acceptor and positive ionizable features best describe the inhibitory activity of AChE inhibitors. A molecular docking study of inhibitors with X-ray structure of human AChE revealed key hydrogen bond interactions in the active site. Each compound class seems to have specific hydrogen bond interactions with the active site. These interactions are classified into five clusters based on interacting residues. The most common residues observed in all clusters are TRP86, GLY121, TYR124, SER125, TYR337, TYR72, SER203, TRP286, ASP74, PHE297, PHE338, TYR341, GLY122, ASN87, and HIS447. The important features elucidated in the pharmacophore modeling are complemented through the docking study. Interactions between inhibitor and amino acid residues TRP286, TYR124/125, TRP86 complement ring aromatic, hydrogen bond acceptor and positive ionizable features of the pharmacophore model. The inhibitor conformations generated in the pharmacophore modeling are very similar to conformations obtained in docking. Knowledge obtained in pharmacophore modeling and docking study was used to build a composite model. Initial 2-D pharmacophore screening of composite model followed by molecular docking and 3-D pharmacophore mapping resulted in 120 plausible inhibitors from ZINC database. This short list of compounds have good docking score, reflecting better complementarity with the receptor and predicted activity comparable with the 153 known AChE inhibitors. Thus a method of rational drug design approach for AChE inhibitors was designed and validated combining pharmacophore modeling, docking and virtual screening methods.

---

---

**CONCLUSIONS AND FUTURE PROSPECTS**

---

---

**9.1 PART I**

The receptor–ligand interactions are mediated by strong and weak hydrogen bonds in the active sites of receptors. Unlike in small molecule crystallography, resolution limits are important in the structural study of hydrogen bond geometries in macromolecules. Main chain hydrogen bond geometries are better represented than the side chain. Hydrogen bond donor/acceptor furcation is very often encountered in the active sites. The amalgamation of strong and weak hydrogen bonds typically forms characteristic hydrogen bond patterns (supramolecular synthons). The geometries for strong hydrogen bonds are consistent, while the weak ones have variable geometries. Apart from hydrogen bonds there exist other weak interactions in the receptor–ligand interface. Key residues in the active sites frequently participate in hydrogen bonding with the ligand. Ligands usually maintain an optimum level of acceptor and donor ratio while interacting with the receptor. Hydrogen bonds involving water molecules in the active sites are noteworthy. With suitable computational tools, very large numbers of strong and weak intermolecular interactions in the receptor–ligand interface may be analyzed reliably. These findings have possible implication in drug discovery research.

**9.2 PART II**

Some successful implementations of molecular modeling in computer aided drug design are presented. Pharmacophore modeling, docking and virtual screening have been performed on two therapeutically important targets. The virtual screening exercise was carried out through database screening. In chapter 7 the selectivity of model was tested in the database mining of EGFR and EGFR spiked non-EGFR kinase inhibitors. In chapter 8, a composite pharmacophore model was used to identify novel inhibitors of AChE in an external database. These exercises resulted in (a) design of a selective pharmacophore model for EGFR kinase inhibitors, (b) identification of possible inhibitors for AChE. The molecular docking study on EGFR kinase domain reveals selectivity for anilinoquinazoline and cyanoquinoline type of inhibitors based on the active and inactive state of the receptor. Similar docking study in AChE suggest that, each compound class has specific hydrogen bonding interactions in the active site, and these can be categorized into

five clusters. Further, the inhibitory activities were successfully predicted for both EGFR and AChE inhibitors.

In this thesis I have tried to present some aspects of molecular recognition between receptor and ligand through a comparative crystallographic and molecular modeling study.

## REFERENCES AND NOTES

### CHAPTER 1

- [1.1] Lodish, M.; Berk, A.; Matsudaira, P.; Kaiser, C. A.; Krieger, M.; Scott, M. P.; Zipursky, S. L.; Darnel, J. (2004) *Molecular cell biology*. (Fifth edn.). W. H. Freeman & Co, New York.
- [1.2] Metzler, D. E. (2001) *Biochemistry: The chemical reaction of living cell*. (2nd edn.). Academic Press. San Diego.
- [1.3] Nelson, D. L.; Cox, M. M. (2004) *Lehninger Principles of biochemistry*. (Fourth edn.). W. H. Freeman & Co, New York.
- [1.4] Stryer, L. (2002) *Biochemistry*. (Fifth edn.). W. H. Freeman & Co, New York.
- [1.5] Garret, R. H.; Grisham, C. M. (1999) *Biochemistry*. (2nd edn.). Thomson Learning, Thomson Corporation, Stamford.
- [1.6] Felix, F., Ed. (1982) *Water a comprehensive treatise*. Plenum Press, New York.
- [1.7] Käiväräinen, A. I. (1985) *Solvent dependent flexibility of proteins and principles of their function*. D. Reidel Publishing Company, Dordrecht.
- [1.8] Arunan, E., *Current Sci.* **2007**, 92, 17-18.  
<http://www.iupac.org/projects/2004/2004-026-2-100.html>
- [1.9] Pauling, L. (1939) *The nature of the chemical bond*. Cornell University Press, Ithaca, New York.
- [1.10] Pimentel, G. C.; McClellan, A. L. (1960) *The hydrogen bond*. Freeman, San Francisco.
- [1.11] Jeffrey, G. A.; Saenger, W. (1991) *Hydrogen bonding in biological structures*. Springer-Verlag, Berlin.
- [1.12] Jeffrey, G. A. (1997) *An introduction to hydrogen bonding*. Oxford University Press, New York.
- [1.13] Scheiner, S. (1997) *Hydrogen bonding. A theoretical perspective*. Oxford University Press, Oxford.

- [1.14] Desiraju G. R.; Steiner T. (1999) *The weak hydrogen bond in structural chemistry and biology*. Oxford University Press, Oxford.
- [1.15] Nishio, M.; Hirota, M.; Umezawa, Y. (1998) *The CH/ $\pi$  interaction Evidence, Nature, and Consequences*. Wiley-VCH, Inc, New York.
- [1.16] Gerlt, J. A.; Gassman, P. G., *J. Am. Chem. Soc.* **1993**, 115, 11552-11568.
- [1.17] Cleland, W. W.; Kreevoy, M. M., *Science* **1994**, 264, 1887-1890.
- [1.18] Frey, P. A.; Whitt, S. A.; Tobin, J. B., *Science* **1994**, 264, 1927-1930.
- [1.19] Perrin, C. L.; Nielson, J. B., *Annu. Rev. Phys. Chem.* **1997**, 48, 511-544.
- [1.20] Cleland, W. W.; Frey, P. A.; Gerlt, J. A., *J. Biol. Chem.* **1998**, 25529-25532.
- [1.21] Viragh, C.; Harris, T. K.; Reddy, P. M.; Massiah, M. A.; Mildvan, A. S.; Kovach, I. M., *Biochemistry* **2000**, 39, 16200-16205.
- [1.22] Kim, K. S.; Oh, K. S.; Lee, J. Y., *Proc. Natl. Acad. Sci. USA* **2000**, 97, 6373-6378.
- [1.23] Frey, P. A., *Magn. Reson. Chem.* **2001**, 39, S190-S198.
- [1.24] Massiah, M. A.; Viragh, C.; Reddy, P. M.; Kovach, I. M.; Johnson, J.; Rosenberry, T. L.; Mildvan, A. S., *Biochemistry* **2001**, 40, 5682-5690.
- [1.25] Kim, K. S.; Kim, D.; Lee, J. Y.; Tarakeshwar, P.; Oh, K. S., *Biochemistry* **2002**, 41, 5300-5306.
- [1.26] Dean, N. E.; Miller, J. E.; Halkides, C. J.; Messina, M., *J. Chem. Inf. Model.* **2003**, 43, 554-559.
- [1.27] Poi, M. J.; Tomaszewski, J. W.; Yuan, C.; Dunlap, C. A.; Andersen, N. H.; Gelb, M. H.; Tsai, M. D., *J. Mol. Biol.* **2003**, 329, 997-1009.
- [1.28] Schutz, C. N.; Warshel, A., *Proteins* **2004**, 55, 711-723.
- [1.29] Shokhen, M.; Albeck, A., *Proteins* **2004**, 54, 468-477.
- [1.30] Anderson, S.; Crosson, S.; Moffat, K., *Acta. Crystallogr. Sect. D* **2004**, 60, 1008-1016.
- [1.31] Zhao, L.; Liao, H.; Tsai, M. D., *J. Biol. Chem.* **2004**, 279, 31995-32000.
- [1.32] Wohlfahrt, G., *Proteins* **2005**, 58, 396-406.
- [1.33] Kraut, D. A.; Sigala, P. A.; Pybus, B.; Liu, C. W.; Ringe, D.; Petsko, G. A.; Herschlag, D., *PLOS Biology* **2006**, 4, 501-519.

## References and Notes

- [1.34] Das, A.; Prashar, V.; Mahale, S; Serre, L.; Ferrer, J.-L.; Hosur, M. V., *Proc. Natl. Acad. Sci. USA* **2006**, 103, 18464-18469.
- [1.35] Baker, E. N.; Hubbard, R. E., *Progr. Biophys. Mol. Biol.* **1984**, 44, 97-179.
- [1.36] Sutor, D. J., *Journal of the Chemical Society* **1963**, 1105-1110.
- [1.37] Shefter, E.; Trueblood, K. N., *Acta Crystallographica* **1965**, 18, 1067-1077.
- [1.38] Sundaralingam, M., *Acta Crystallographica* **1966**, 21, 495-505.
- [1.39] Ramachandran, G. N.; Sasisekharan, V., *Biochim. Biophys. Acta.* **1965**, 109, 314-316.
- [1.40] Ramachandran, G. N.; Sasisekharan, V.; Ramakrishnan, C. *Biochim. Biophys. Acta.* **1966**, 112, 168-170.
- [1.41] Derewenda, Z. S.; Lee, L.; Derewenda, U., *J. Mol. Biol.* **1995**, 252, 248-262.
- [1.42] Bella, J.; Berman, H., *J. Mol. Biol.* **1996**, 264, 734-42.
- [1.43] Scheiner, S.; Gu, Y.; Kar, T., *Journal of Molecular Structure (Theochem)* **2000**, 500, 441-452.
- [1.44] Aravinda, S.; Shamala, N.; Pramanik, A.; Das, C.; Balaram, P., *Biochem. Biophys. Res. Commun.* **2000**, 273, 933-936.
- [1.45] Thakur, A. K.; Kishore, R., *Biopolymers* **2000**, 53, 447-454.
- [1.46] Senes A.; Belandia, I. U.; Engelman, D. M., *Proc. Natl. Acad. Sci. USA* **2001**, 98, 9056-9061.
- [1.47] Scheiner, S.; Kar, T.; Gu, Y., *J. Biol. Chem.* **2001**, 276, 9832-9837.
- [1.48] Weiss, M. S.; Brandl, M.; Sühnel, J.; Pal, D.; Hilgenfeld, R., *Trends Biochem. Sci.* **2001**, 26, 521-523.
- [1.49] Adams, A.; Guss, J. M.; Denny, W. A.; Wakelin, L. P., *Nucleic Acids Res.* **2002**, 30, 719-725.
- [1.50] Chamberlain, A. K.; Bowie, J. U., *J. Mol. Biol.* **2002**, 322, 497-503.
- [1.51] Jiang, L.; Lai, L., *J. Biol. Chem.* **2002**, 277, 37732-37740.
- [1.52] Babu, M. M.; Singh, S. K.; Balaram, P., *J. Mol. Biol.* **2002**, 322, 871-880.
- [1.53] Aravinda, S.; Shamala, N.; Bandyopadhyay, A.; Balaram, P., *J. Am. Chem. Soc.* **2003**, 125, 15065-15075.
- [1.54] Bhattacharyya, R.; Chakrabarti, P., *J. Mol. Biol.* **2003**, 331, 925-940.

- [1.55] Palsdottir, H.; Lojero, C. G.; Trumpower, B. L.; Hunte, C., *J. Biol. Chem.* **2003**, 278, 31303-31311.
- [1.56] Singh, S. K.; Babu, M. M.; Balaram, P., *Proteins* **2003**, 51, 167-171.
- [1.57] Wieczorek, R.; Dannenberg, J. J., *J. Am. Chem. Soc.* **2003**, 125, 8124-8129.
- [1.58] Braun, P.; Vegh, A. P.; von Jan, M.; Strohmam, B.; Hunter, C. N.; Robert, B.; Scheer, H. *Biochim. Biophys. Acta.* **2003**, 1607, 19-26.
- [1.59] Kang, B. S.; Devedjiev, Y.; Derewenda, U.; Derewenda, Z. S., *J. Mol. Biol.* **2004**, 338, 483-493.
- [1.60] Scheiner, S., *J. Phys. Chem. B Condens. Matter Mater. Surf. Interfaces Biophys.* **2006**, 110, 18670-18679.
- [1.61] Karle, I. L.; Venkateshwarlu, P.; Ranganathan, S., *Biopolymers* **2006**, 84, 502-507.
- [1.62] Thakur, A. K.; Kishore, R., *Biopolymers* **2006**, 81, 440-449.
- [1.63] Loll, B.; Raszewski, G.; Saenger, W.; Biesiadka, J., *J. Mol. Biol.* **2003**, 328, 737-747.
- [1.64] Yohannan, S.; Faham, S.; Yang, D.; Grosfeld, D.; Chamberlain, A. K.; Bowie, J. U., *J. Am. Chem. Soc.* **2004**, 126, 2284-2285.
- [1.65] Brandl, M.; Meyer, M.; Sühnel, J., *J. Biomol. Struct. Dyn.* **2001**, 18, 545-555.
- [1.66] Treger, M.; Westhof, E., *J. Mol. Recognit.* **2001**, 14, 199-214.
- [1.67] Swaminathan, C. P.; Gupta, A.; Surolia, N.; Surolia, A., *J. Biol. Chem.* **2000**, 275, 28483-28487.
- [1.68] Oku, K.; Watanabe, H.; Kubota, M.; Fukuda, S.; Kurimoto, M.; Tsujisaka, Y.; Komori, M.; Inoue, Y.; Sakurai, M., *J. Am. Chem. Soc.* **2003**, 125, 12739-12748.
- [1.69] Rajsekhar, G.; Rao, C. P.; Guionneau, P., *Carbohydr. Res.* **2003**, 338, 801-805.
- [1.70] Baddeley, T. C.; Davidson, I. G.; Glidewell, C.; Low, J. N.; Skakle, J. M.; Wardell, J. L., *Acta. Crystallogr. Sect. B* **2004**, 60, 461-471.
- [1.71] Sujatha, M. S.; Sasidhar, Y. U.; Balaji, P. V., *Biochemistry* **2005**, 44, 8554-8562.

## References and Notes

- [1.72] Yates, J. R.; Pham, T. N.; Pickard, C. J.; Mauri, F.; Amado, A. M.; Gil, A. M.; Brown, S. P., *J. Am. Chem. Soc.* **2005**, 127, 10216-10220.
- [1.73] Pandit, S. A.; Bostick, D.; Berkowitz, M. L., *Biophys. J.* **2004**, 86, 1345-1356.
- [1.74] Scheiner, S.; Kar, T., *J. Phys. Chem. B. Condens. Matter. Mater. Surf. Interfaces Biophys.* **2005**, 109, 3681-3689.
- [1.75] Steiner, T.; Koellner, G., *J. Mol. Biol.* **2001**, 305, 535-557.
- [1.76] Brandl, M.; Weiss, M. S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R., *J. Mol. Biol.* **2001**, 307, 357-377.
- [1.77] Scheiner, S.; Kar, T.; Pattanayak, J., *J. Am. Chem. Soc.* **2002**, 124, 13257-13264.
- [1.78] Harigai, M.; Kataoka, M.; Imamoto, Y., *J. Am. Chem. Soc.* **2006**, 128, 10646-10647.
- [1.79] Saraogi, I.; Vijay, V. G.; Das, S.; Sekar, K.; Guru Row T. N., *Crystal Engineering* **2003**, 6, 69-77.
- [1.80] Dougherty, D. D., *Science* **1996**, 271, 163-168.
- [1.81] Ma, J. C.; Dougherty, D. D., *Chem. Rev.* **1997**, 97, 1303-1324.
- [1.82] Zacharias, N.; Dougherty, D. A., *Trends Pharmacol. Sci.* **2002**, 23, 281-287.
- [1.83] Prajapati, R. S.; Sirajuddin, M.; Durani, V.; Sreeramulu, S.; Varadarajan, R., *Biochemistry* **2006**, 45, 15000-15010.
- [1.84] Meyer, E. A.; Castellano, R. K.; Diederich, F., *Angew. Chem. Int. Ed.* **2003**, 42, 1210-1250.
- [1.85] Ouvrard, C., Le Questel, J. -Y.; Berthelot, M.; Laurence, C., *Acta. Crystgra. Sect. B* **2003**, 59, 512-526.
- [1.86] Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S., *Proc. Natl. Acad. Sci. USA* **2004**, 101, 16789-16794.
- [1.87] Kovács, A.; Varga, Z., *Coordination Chem. Reviews* **2006**, 250, 710-727.
- [1.88] Guerra, C. F.; Bickelhaupt, F. M.; Baerends, E. J., *Chemphyschem.* **2004**, 5, 481-487.
- [1.89] Adman, E.; Watenpaugh, K. D.; Jensen, L. H., *Proc. Natl. Acad. Sci. USA* **1975**, 72, 4854-4858.

- [1.90] Ippolito, J. A.; Alexander, R. S.; Christianson, D. W., *J. Mol. Biol.* **1990**, 215, 457-471.
- [1.91] Gregoret, L. M.; Rader, S. D.; Fletterick, R. J.; Cohen, F. E., *Proteins* **1991**, 9, 99-107.
- [1.92] Rajagopal, S.; Vishveshwara, S., *FEBS J.* **2005**, 272, 1819-1832.
- [1.93] Rossmann, M. G.; Arnold, E., Eds. (**2001**) *International tables for crystallography Vol F. Crystallography of biological macromolecules*. Kluwer Academic Publishers, Dordrecht.
- [1.94] Hamilton, W. C.; Ibers, J. A. (**1968**) *Hydrogen bonding in solids*. W. A. Benjamin, New York.
- [1.95] Jeffrey, G. A. (**1992**) Accurate crystal structure analysis by neutron diffraction. In *Accurate molecular structures*. Domenicano, A.; Hargittai, I. Eds., Oxford University Press, Oxford, pp. 270-298.
- [1.96] Gutberlet, T.; Heinemann, U.; Steiner, M., *Acta. Crystallogra. Sect. D* **2000**, 57, 349-354.
- [1.97] Berstein, F., Koetzle, T., Williams, G., Meyer, E., Brice, M., Rodgers, J., Kennard, O., Shimanouchi, T.; Tasumi M., *J. Mol. Biol.* **1977**, 112, 535-542.
- [1.98] Allen, F. H.; Kennard, O., *Chem. Des. Autom. News* **1993**, 8, 31-37.  
<http://www.ccdc.cam.ac.uk/prods/csd/csd.html>
- [1.99] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., *Nucleic Acid Res.* **2000**, 28, 235-242.  
<http://www.rcsb.org/pdb>
- [1.100] Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C., *J. Mol. Biol.* **1995**, 247, 536-540.  
<http://scop.mrc-lmb.cam.ac.uk/scop>
- [1.101] Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M., *Structure* **1997**, 5, 1093-1108.  
<http://www.cathdb.info/latest/index.html>
- [1.102] Laskowski, R. A., *Nucleic Acids Res.* **2001**, 29, 221-222.  
<http://www.ebi.ac.uk/thornton-srv/databases/pdbsum>

## References and Notes

- [1.103] Chakrabarti, P.; Pal, D., *Progr. Biophys. Mol. Biol.* **2001**, 76, 1-102.
- [1.104] Porter, C. T.; Bartlett, G. J.; Thornton, J. M., *Nucleic Acid Res.* **2004**, 32, D129-D133.
- [1.105] McDonald, I. K.; Thornton, J. M., *J. Mol. Biol.* **1994**, 238, 777-793.
- [1.106] Lindauer, K.; Bendic, C.; Sühnel, J., *Comput. Appl. Biosci.* **1996**, 12, 281-289.
- [1.107] Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C., *J. Mol. Biol.* **1999**, 285, 1735-1747.
- [1.108] Collaborative computational project, number 4., *Acta. Crystallogra. Sect. D* **1994**, 50, 760-763.
- [1.109] Sobolev, V.; Sorokine, A.; Prilusky, J.; Abola, E. E.; Edelman, M., *Bioinformatics* **1999**, 15, 327-332.
- [1.110] Babu, M. M., *Nucleic Acid Res.* **2003**, 31, 3345-3348.
- [1.111] Tiwari, A.; Panigrahi, S. K.; Desiraju, G. R., *HBAT*: A complete package for analysing strong and weak hydrogen bonds in macromolecular crystal structures. (*communicated*)
- [1.112] Glusker, J. P.; Lewis, M.; Rossi, M. (**1994**) *Crystal structure analysis for chemists and biologists*. VCH, New York.
- [1.113] Böhm, H. -J.; Klebe, G., *Angew. Chem. Int. Ed.* **1996**, 35, 2588-2614.
- [1.114] Gohlke, H.; Klebe, G., *Angew. Chem. Int. Ed.* **2002**, 41, 2644-2676.
- [1.115] Klaholz, B.; Moras, D., *Structure* **2002**, 10, 1197-1204.
- [1.116] Pierce, A. C.; Sandretto, K. L.; Bemis, G. W., *Proteins* **2002**, 49, 567-576.
- [1.117] Pierce, A. C.; ter Haar, E.; Binch, H. M.; Kay, D. P.; Patel, S. R.; Li, P., *J. Med. Chem.* **2005**, 48, 1278-1281.
- [1.118] Denessiouk, K. A.; Johnson, M. S., *J. Mol. Biol.* **2003**, 333, 1025-1043.
- [1.119] Sarkhel, S.; Desiraju, G. R., *Proteins* **2004**, 54, 247-259.
- [1.120] Cashin, A. L.; Petersson, E. J.; Lester, H. A.; Dougherty, D. A., *J. Am. Chem. Soc.* **2005**, 127, 350-356.
- [1.121] Aparna, V.; Rambabu, G.; Panigrahi, S. K.; Sarma, J. A. R. P.; Desiraju, G. R., *J. Chem. Inf. Model.* **2005**, 45, 725-738.

## CHAPTER 2

- [2.1] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., *Nucleic Acid Res.* **2000**, 28, 235-242.  
<http://www.rcsb.org/pdb>
- [2.2] Desiraju, G. R.; Steiner T. (1999) *The weak hydrogen bond in structural chemistry and biology*. Oxford University Press, Oxford.
- [2.3] Steiner, T.; Koellner, G., *J. Mol. Biol.* **2001**, 305, 535-557.
- [2.4] Xantheas, S. S., *Chem. Phy.* **2000**, 258, 225-231.
- [2.5] Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S., *Proc. Natl. Acad. Sci. USA* **2004**, 101, 16789-16794.
- [2.6] McDonald, I. K.; Thornton, J. M., *J. Mol. Biol.* **1994**, 238, 777-793.
- [2.7] Lindauer, K.; Bendic, C.; Sühnel, J., *Comput. Appl. Biosci.* **1996**, 12, 281-289.
- [2.8] Collaborative computational project, number 4. *Acta. Crystallogra. Sect. D* **1994**, 50, 760-763.
- [2.9] Sobolev, V.; Sorokine, A.; Prilusky, J.; Abola, E. E.; Edelman, M., *Bioinformatics* **1999**, 15, 327-332.
- [2.10] Babu, M. M., *Nucleic Acid Res.* **2003**, 31, 3345-3348.
- [2.11] PERL ([www.perl.org](http://www.perl.org))/TK([www.tcl.tk](http://www.tcl.tk)).
- [2.12] Graphviz - Graph Visualization Software: [www.graphviz.org](http://www.graphviz.org).
- [2.13] Brandl, M.; Weiss, M. S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R., *J. Mol. Biol.* **2001**, 307, 357-377.
- [2.14] Koellner, G.; Kryger, G.; Millard, C. B.; Silman, I.; Sussman, J. L.; Steiner T., *J. Mol. Biol.* **2000**, 296, 713-735.
- [2.15] Klaholz, B.; Moras, D., *Structure* **2002**, 10, 1197-1204.
- [2.16] Sarkhel, S.; Desiraju, G. R., *Proteins* **2004**, 54, 247-259.

## CHAPTER 3

- [3.1] Glusker, J. P. (**1998**) Directional aspects of intermolecular interactions. In *Topics in current chemistry: Design of organic solids*. Weber, E., Ed., Springer-Verlag, Berlin and Heidelberg, pp. 1-56.
- [3.2] Kortemme, T.; Morozov, A. V.; Baker, D., *J. Mol. Biol.* **2003**, 326, 1239-1259.
- [3.3] Allocati, N.; Masulli, M.; Pietracupa, M.; Federici, L.; Di Ilio, C., *Biochem. J.* **2006**, 394, 11-17.
- [3.4] Shokhen, M.; Albeck, A., *Proteins* **2004**, 54, 468-477.
- [3.5] Bledsoe, R. K.; Madauss, K. P.; Holt, J. A.; Apolito, C. J.; Lambert, M. H.; Pearce, K. H.; Stanley, T. B.; Stewart, E. L.; Trump, R. P.; Willson, T. M.; Williams, S. P., *J. Biol. Chem.* **2005**, 280, 31283-31293.
- [3.6] Baker, E. N.; Hubbard, R. E., *Progr. Biophys. Mol. Biol.* **1984**, 44, 97-179.
- [3.7] Jeffrey, G. A.; Saenger, W. (**1991**) *Hydrogen bonding in biological structures*. Springer-Verlag, Berlin.
- [3.8] Glusker, J. P., *Acta. Crystallogra. Sect. D* **1995**, 51, 418-427.
- [3.9] Wahl, M. C.; Rao, S. T.; Sundaralingam, M., *Nat. Struct. Biol.* **1996**, 3, 24-31.
- [3.10] Derewenda, Z. S.; Derewenda, U.; Kobos, P. M., *J. Mol. Biol.* **1994**, 241, 83-93.
- [3.11] Fabiola, G. F.; Krishnaswamy, S.; Nagarajan, V.; Pattabhi, V., *Acta. Crystallogra. Sect. D* **1997**, 53, 316-320.
- [3.12] Thomas, A.; Benhabiles, N.; Meurisse, R.; Ngwabije, R.; Brasseur, R., *Proteins* **2001**, 43, 37-44.
- [3.13] Steiner, T.; Koellner, G., *J. Mol. Biol.* **2001**, 305, 535-557.
- [3.14] Scheiner, S.; Kar, T.; Gu, Y., *J. Biol. Chem.* **2001**, 276, 9832-9837.
- [3.15] Jiang, L.; Lai, L., *J. Biol. Chem.* **2002**, 277, 37732-37740.
- [3.16] Reddy, C. K.; Das, A.; Jayaram, B., *J. Mol. Biol.* **2001**, 314, 619-632.
- [3.17] Prasad, T.; Prathima, M. N.; Chandra, N., *Bioinformatics* **2003**, 19, 167-168.
- [3.18] Babu, M. M.; Singh, S. K.; Balaram, P., *J. Mol. Biol.* **2002**, 322, 871-880.

- [3.19] Sarkhel, S.; Desiraju, G. R., *Proteins* **2004**, 54, 247-259.
- [3.20] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., *Nucleic Acid Res.* **2000**, 28, 235-242.  
<http://www.rcsb.org/pdb>
- [3.21] Allen, F. H.; Kennard, O., *Chem. Des. Autom. News* **1993**, 8, 31-37.  
<http://www.ccdc.cam.ac.uk/prods/csd/csd.html>
- [3.22] Nissink, J. W.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R., *Proteins* **2002**, 49, 457-471.
- [3.23] Kleywegt, G., *Acta. Crystallogra. Sect. D* **2000**, 56, 249-265.
- [3.24] MOE Program, Version 2005, Chemical Computing Inc, Suite 910-1010 Sherbrooke St. W, Montreal, Quebec, Canada, H3A 2R7.
- [3.25] Allinger, N. (**1998**) Force fields: A brief introduction. In *Encyclopedia of Computational Chemistry*. Schleyer, P. v. R., Ed., John Wiley & Sons, Chicester, UK, Vol 2, pp. 1013-1015. (See also subsequent topics in force field sections).
- [3.26] Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C., *J. Mol. Biol.* **1999**, 285, 1735-1747.
- [3.27] Collaborative computational project, number 4., *Acta. Crystallogra. Sect. D* **1994**, 50, 760-763.
- [3.28] McDonald, I. K.; Thornton, J. M., *J. Mol. Biol.* **1994**, 238, 777-793.
- [3.29] Lindauer, K.; Bendic, C.; Sühnel, J., *Comput. Appl. Biosci.* **1996**, 12, 281-289.
- [3.30] Sobolev, V.; Sorokine, A.; Prilusky, J.; Abola, E. E.; Edelman, M., *Bioinformatics* **1999**, 15, 327-332.
- [3.31] Babu, M. M., *Nucleic Acid Res.* **2003**, 31, 3345-3348.
- [3.32] Brandl, M.; Weiss, M. S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R., *J. Mol. Biol.* **2001**, 307, 357-377.
- [3.33] Koellner, G.; Kryger, G.; Millard, C. B.; Silman, I.; Sussman, J. L.; Steiner, T., *J. Mol. Biol.* **2000**, 296, 713-735.
- [3.34] Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S., *Proc. Natl. Acad. Sci. USA* **2004**, 101, 16789-16794.

## References and Notes

- [3.35] Klaholz, B.; Moras, D., *Structure* **2002**, 10, 1197-1204.
- [3.36] Chakrabarti, P.; Chakrabarti, S., *J. Mol. Biol.* **1998**, 284, 867-873.
- [3.37] Kim, J.; Mao, J.; Gunner, M. R., *J. Mol. Biol.* **2005**, 348, 1283-1298.
- [3.38] Park, S.; Saven, J. G., *Proteins* **2005**, 60, 450-463.
- [3.39] Ippolito, J. A.; Alexander, R. S.; Christianson, D. W., *J. Mol. Biol.* **1990**, 215, 457-471.
- [3.40] Ragone, R., *Protein Sci.* **2001**, 10, 2075-2082.
- [3.41] Desiraju, G. R.; Steiner, T. (1999) *The weak hydrogen bond in structural chemistry and biology*. Oxford University Press, Oxford.
- [3.42] Porter, C. T.; Bartlett, G. J.; Thornton, J. M., *Nucleic Acid Res.* **2004**, 32, D129-D133.
- [3.43] Yan, B. X.; Sun, Y. Q., *J. Biol. Chem.* **1997**, 272, 3190-3194.
- [3.44] Steiner, T., *Acta. Crystallogra. Sect. D* **1995**, 51, 93-97.
- [3.45] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., *Adv. Drug Del. Rev.* **1997**, 23, 3-25.
- [3.46] Ouvrard, C.; Le Questel, J. -Y.; Berthelot, M.; Laurence, C., *Acta. Crystallogra. Sect. B* **2003**, 59, 512-526.
- [3.47] Kovács, A.; Varga, Z., *Coordination Chem. Reviews* **2006**, 250, 710-727.
- [3.48] Gregoret, L. M.; Rader, S. D.; Fletterick, R. J.; Cohen, F. E., *Proteins* **1991**, 9, 99-107.

## CHAPTER 4

- [4.1] Jeffrey, G. A.; Saenger, W. (1991) *Hydrogen bonding in biological structures*. Springer Verlag, Berlin.
- [4.2] Baker, E. N.; Hubbard, R. E., *Progr. Biophys. Mol. Biol.* **1984**, 44, 97-179.
- [4.3] Desiraju, G. R.; Steiner, T. (1999) *The weak hydrogen bond in structural chemistry and biology*. Oxford University Press, Oxford, Chapters 1 and 5.
- [4.4] Sarkhel, S.; Desiraju, G. R., *Proteins* **2004**, 54, 247-259.

- [4.5] Panigrahi, S. K.; Desiraju, G. R., *Proteins* **2007**, ASAP, DOI 10.1002/prot.21253.
- [4.6] Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. *Proc. Natl. Acad. Sci. USA* **2004**, 101, 16789-16794.
- [4.7] Glusker, J. P., *Acta. Crystallogra. Sect. D* **1995**, 51, 418-427.
- [4.8] Bartlett, G. J.; Porter, C. T.; Borkakoti, N.; Thornton, J. M., *J. Mol. Biol.* **2002**, 324, 105-121.
- [4.9] Steiner, T., *Angew. Chem. Int. Ed.* **2002**, 41, 48-76.
- [4.10] Pierce, A. C.; Sandretto, K. L.; Bemis, G. W., *Proteins* **2002**, 49, 567-576.
- [4.11] Pierce, A. C.; ter Haar, E.; Binch, H. M.; Kay, D. P.; Patel, S. R.; Li, P., *J. Med. Chem.* **2005**, 48, 1278-1281.
- [4.12] Cohen, P., *Nature Rev. Drug Discov.* **2002**, 1, 309-315.
- [4.13] Levitzki, A., *Acc. Chem. Res.* **2003**, 36, 462-469.
- [4.14] Bridges, A. J., *Chem. Rev.* **2001**, 101, 2541-2571.
- [4.15] Vieth, M.; Higgs, R. E.; Robertson, D. H.; Shapiro, M.; Gragg, E. A.; Hemmerle, H., *Biochim. Biophys. Acta* **2004**, 1697, 243-257.
- [4.16] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., *Nucleic Acid Res.* **2000**, 28, 235-242.  
<http://www.rcsb.org/pdb>
- [4.17] (a) The Protein Kinase Resource, SDSC, UC, San Diego, MC, 2004, [www.kinasenet.org/pkr](http://www.kinasenet.org/pkr)
- [4.18] Williams, D. H.; Mitchell, T., *Curr. Opin. Pharmacol.* **2002**, 2, 567-573.
- [4.19] (a) Sugan, Structure and Phylogeny of the Protein Kinases, Salk Institute, La Jolla, CA, 2002, <http://198.202.68.14/human/kinome/phylogeny.html>.  
(b) <http://kinasedb.ontology.ims.u-tokyo.ac.jp>. (c) Wang, R.; Fang, X.; Lu, Y.; Yang, C. Y.; Wang, S., *J. Med. Chem.* **2005**, 48, 4111-4119.  
[www.pdbbind.org](http://www.pdbbind.org)
- [4.20] Naumann, T.; Matter, H., *J. Med. Chem.* **2002**, 45, 2366-2378.
- [4.21] Hanks, S. K.; Hunter, T., *FASEB J.* **1995**, 9, 576-596.

## References and Notes

- [4.22] Dar, A. C.; Wybenga-Groot, L. E.; Sicheri, F. (2005) The Eukaryotic Protein Kinase Domain. In *Modular Protein Domains*, Cesareni, G.; Gimona, M.; Sudol, M.; Yaffe, M., Eds., Wiley-VCH Verlag GmbH & Co.: KGaA, Chapter 9.
- [4.23] Panigrahi, S. K.; Desiraju, G. R., *Nat. Acad. Sci. Lett. (India)* **2004**, 27, 1-11.
- [4.24] Mao, L.; Wang, Y.; Lie, Y.; Hu, X., *J. Mol. Biol.* **2004**, 336, 787-807.
- [4.25] MOE Program, Version 2006-08 Chemical Computing Inc, Sherbrooke St. W, Montreal, Quebec, Canada, H3A 2R7.
- [4.26] Halgren, T., *J. Comput. Chem.* **1996**, 17, 490-519.
- [4.27] Tiwari, A.; Panigrahi, S. K.; Desiraju, G. R. *HBAT*: A complete package for analysing strong and weak hydrogen bonds in macromolecular crystal structures. (*communicated*)
- [4.28] (a) Thompson, J. D.; Higgins, D. G.; Gibson, T. J., *Nucleic Acids Res.* **1994**, 22, 4673-4680. (b) ClustalW WWW Service at the European Bioinformatics Institute. [www.ebi.ac.uk/clustalw](http://www.ebi.ac.uk/clustalw).
- [4.29] Williams, M. A.; Goodfellow, J. M.; Thornton, J. M., *Protein Sci.* **1994**, 3, 1224-1235.
- [4.30] Aparna, V.; Rambabu, G.; Panigrahi, S. K.; Sarma, J. A. R. P.; Desiraju, G. R., *J. Chem. Inf. Model.* **2005**, 45, 725-738.
- [4.31] Gitlin, I.; Carbeck, J. D.; Whitesides, G. M., *Angew. Chem. Int. Ed.* **2006**, 45, 3022-3060.
- [4.32] Desiraju, G. R., *Angew. Chem. Int. Ed.* **1995**, 34, 2311-2327.
- [4.33] Walsh, R. D.; Bradner M. W.; Fleischman, S.; Morales, L. A.; Moulton, B.; Rodríguez-Hornedo, N.; Zaworotko, M. J., *Chem. Commun.* **2003**, 186-187.
- [4.34] Vishweshwar, P.; Nangia, A.; Lynch, V. M., *Cryst. Growth. Des.* **2003**, 3, 783-790.
- [4.35] Ladbury, J. E., *Chem. Biol.* **1996**, 3, 973-980.
- [4.36] Bottoms, C. A.; Smith, P. E.; Tanner, J. J., *Protein Sci.* **2002**, 11, 2125-2137.

## CHAPTER 5

- [5.1] Neidle, S., *Nat. Prod. Rep.* **2001**, 18, 291-309.
- [5.2] Lacy, E. R.; Madsen, E. M.; Lee, M.; Wilson, W. D. (2003) Polyamide dimer stacking in the DNA minor groove and recognition of TC mismatched base pairs in DNA. In *Small molecule DNA and RNA binders: From synthesis to nucleic acid complexes*. Demeunynck, M.; Bailly, C.; Wilson, W. D., Eds., WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, pp. 387-413.
- [5.3] Tidwell, R. R.; Boykin, D. W. (2003) Dicationic DNA minor groove binders as antimicrobial agents. In *Small molecule DNA and RNA binders: From synthesis to nucleic acid complexes*. Demeunynck, M.; Bailly, C.; Wilson, W. D., Eds., WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, pp. 414-460.
- [5.4] Wilson, W. D.; Nguyen, B.; Tanious, F. A.; Mathis, A.; Hall, J. E.; Stephens, C. E.; Boykin, D.W., *Curr. Med. Chem. -Anti-Cancer Agent* **2005**, 5, 389-408.
- [5.5] Pindur, U.; Jansen, M.; Lemster, T., *Curr. Med. Chem.* **2005**, 12, 2805-2847.
- [5.6] Dervan, P. B.; Edelson, B. S., *Curr. Opin. Struct. Biol.* **2003**, 13, 284-299.
- [5.7] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., *Nucleic Acid Res.* **2000**, 28, 235-242.  
<http://www.rcsb.org/pdb>
- [5.8] Kennard, O., *Pure & Appl. Chem.* **1993**, 65, 1213-1222.
- [5.9] Tabernero, L.; Bella, J.; Aleman, C., *Nucleic Acid Res.* **1996**, 24, 3458-3466.
- [5.10] Dervan, P. B., *Bioorg. Med. Chem.* **2001**, 9, 2215-2235.
- [5.11] Moravek, Z.; Neidle, S.; Schneider, B., *Nucleic Acid Res.* **2002**, 30, 1182-1191.
- [5.12] Rohs, R.; Bloch, I.; Sklenar, H.; Shakked, Z., *Nucleic Acid Res.* **2005**, 33, 7048-7057.
- [5.13] Desiraju, G. R.; Steiner, T. (1999) *The weak hydrogen bond in structural chemistry and biology*, Oxford University Press, Oxford.
- [5.14] Sarkhel, S.; Desiraju, G. R., *Proteins* **2004**, 54, 247-259.

## References and Notes

- [5.15] Panigrahi, S. K.; Desiraju, G. R., *Proteins* **2007**, ASAP, DOI 10.1002/prot.21253.
- [5.16] Aparna, V.; Rambabu, G.; Panigrahi, S. K.; Sarma, J. A. R. P.; Desiraju, G. R., *J. Chem. Inf. Model.* **2005**, 45, 725-738.
- [5.17] Athri, P.; Wenzler, T.; Ruiz, P.; Brun, R.; Boykin, D. W.; Tidwel, R.; Wilson, W. D., *Bioorg. Med. Chem.* **2006**, 14, 3144-3152.
- [5.18] Seed, J. R.; Boykin, D. W. (2001) Chemotherapy of African trypano-somiasis, In *World Class Parasites, Vol. 1, The African Trypanosomes*. Black, S.; Seed, J. R., Eds., Kluwer Academic Publishers, Boston, pp. 65-78.
- [5.19] Boykin, D. W., *J. Braz. Chem. Soc.* **2002**, 13, 763-771.
- [5.20] Halgren, T., *J. Comput. Chem.* **1996**, 17, 490-519.
- [5.21] MOE Program, Version 2006-08 Chemical Computing Inc, Suite 910-1010 Sherbrooke St. W, Montreal, Quebec, Canada, H3A 2R7.
- [5.22] Tiwari, A. Panigrahi, S. K.; Desiraju, G. R., *HBAT*: A complete package for analysing strong and weak hydrogen bonds in macromolecular crystal structures. (*communicated*)
- [5.23] GOLD 3.0, Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK.
- [5.24] Evans, D. A.; Neidle, S., *J. Med. Chem.* **2006**, 49, 4232-4238.

## CHAPTER 6

- [6.1] Höltje, H.-D.; Folkeis, G. (1997) *Molecular modeling: Basic principles and applications*. VCH, New York.
- [6.2] Leach, A. R. (2001) *Molecular modelling principles and applications*. (2nd edn.) Pearson Education Limited, London.
- [6.3] Young, D. C. (2001) *Computational chemistry a practical guide for applying techniques to real-world problems*. John Wiley & Sons, Inc., New York.

- [6.4] Baker, O. M.; MacKerell Jr., A. D; Roux, B.; Watanabe, M. **(2001)** *Computational biochemistry and biophysics*. Marcel Dekker, Inc. New York.
- [6.5] Tsai, C. S. **(2002)** *An introduction to computational biochemistry*. Wiley-Liss, Inc., New York.
- [6.6] Lewars, E. **(2004)** *Computational chemistry introduction to the theory and applications of molecular and quantum mechanics*. Kluwer Academic Publishers, Dordrecht.
- [6.7] Mannhold, R.; Kubinyi, H.; Folkers, G., Eds., **(2003)** *Quantum medicinal chemistry*. Wiley-VCH GmbH & Co. KGaA, Weinheim.
- [6.8] Jensen, F. **(1999)** *Introduction to computational chemistry*. John Wiley & Sons, New York.
- [6.9] Veszprémi, T.; Fehér, M. **(1999)** *Quantum chemistry: Fundamentals to applications*. Kluwer, Dordrecht.
- [6.10] Cook, D. B. **(1998)** *Handbook of computational quantum chemistry*. Oxford, Oxford.
- [6.11] Atkins, P. W.; Friedman, R. S. **(1997)** *Molecular quantum mechanics*. Oxford, Oxford.
- [6.12] Simons, J.; Nichols, J. **(1997)** *Quantum mechanics in chemistry*. Oxford, Oxford.
- [6.13] Levine, I. N. **(1991)** *Quantum chemistry*. Prentice Hall, Englewood Cliffs.
- [6.14] Jug, K.; Neumann, F. **(1998)** Configuration interaction: Semiempirical calculations. In *Encyclopedia of Computational Chemistry*. Schleyer, P. v. R., Ed., John Wiley & Sons, Chichester, UK, 1, pp. 507-513.
- [6.15] Koch, W.; Holthausen, M. C. **(2000)** *A chemist's guide to density functional theory*. Wiley-VCH, Weinheim.
- [6.16] Rappé, A. K.; Casewit, C. J. **(1997)** *Molecular mechanics across chemistry*. University Science Books, Sausalito.
- [6.17] Dobson, J. F.; Vignale, G.; Das, M. P. **(1998)** *Electronic density functional theory recent progress and new directions*. Plenum, New York.

## References and Notes

- [6.18] Schlecht, M. F. (1998) *Molecular modeling on the PC*. Wiley-VCH, New York.
- [6.19] Allinger, N. (1998) Force fields: A brief introduction. In *Encyclopedia of Computational Chemistry*. Schleyer, P. v. R., Ed., John Wiley & Sons, Chicester, UK. Vol 2, pp. 1013-1015.
- [6.20] Jalaie, M.; Lipkowitz, K. B. (2000) Published Force Field Parameters for Molecular Mechanics, Molecular Dynamics, and Monte Carlo Simulations. In *Reviews in computational chemistry*. K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 2000, Vol. 14, pp. 441-486.
- [6.21] Gao, J.; Thompson, M. (1998) *Combined quantum and molecular mechanics methods*. American Chemical Society, Washington.
- [6.22] Billing, G. D.; Mikkelsen, K. V. (1997) *Advanced molecular dynamics and chemical kinetics*. John Wiley & Sons, New York.
- [6.23] D. L. Beveridge (1998) Molecular dynamics: DNA. In *Encyclopedia of Computational Chemistry*. Schleyer, P. v. R., Ed., John Wiley & Sons, Chicester, UK, Vol 3, pp. 1620-1628
- [6.24] Balbuena, P. B.; Seminario, J. M. (1999) *Molecular dynamics*. Elsevier, Amsterdam.
- [6.25] Frenkel, D.; Smit, B. (1996) *Understanding molecular simulation*. Academic Press, San Diego.
- [6.26] Jorgensen, W. L. (1998) Monte Carlo simulations for liquids. In *Encyclopedia of Computational Chemistry*. Schleyer, P. v. R., Ed., John Wiley & Sons, Chicester, UK, Vol 3, 1754-1763.
- [6.27] Parrill, A. L.; Reddy, M. R. (1999) Overview of rational drug design. In *Rational drug design: novel methodology and practical applications*. Rami Reddy, M., Parrill, A. L., Eds., American Chemical Society, Washington DC, Vol. 719, pp. 1-11.
- [6.28] E. -L. Allan, *Drug Discovery World* **2002**, 71-75.
- [6.29] Martin, Y. C. (1998) 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, Vol. 12/13/14, pp. 3-23.
- [6.30] Abraham, D. J. **(2003)** *Burger's medicinal chemistry and drug discovery*. (Sixth edn.). John Wiley & Sons Inc., New Jersey, Vol. 1.
- [6.31] Weber, I. T.; Harrison, R. W. **(1998)** 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, Vol. 2, pp. 115-127.
- [6.32] Muegge, I.; Rarey, M. **(2001)** Small molecule docking and scoring. In *Reviews in computational chemistry*. Lipkowitz, K. B.; Boyd, D. B., Eds., John Wiley & Sons Inc., New York, Vol 17, pp. 1-46.
- [6.33] Gohlke, H.; Klebe, G., *Angew. Chem. Int. Ed.* **2002**, 41, 2644-2676.
- [6.34] Grzybowski, B. A.; Ishchenko, A. V.; Shimada, J.; Shakhnovich, E. I., *Acc. Chem. Res.* **2002**, 35, 261-269.
- [6.35] Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., *Nat. Rev. Drug Discov.* **2004**, 3, 935-949.
- [6.36] Feher, M., *Drug Discov. Today* **2006**, 11, 421-428.
- [6.37] Babine, R. E; Abdel-Meguid, S. S., Eds., **(2004)** *Protein crystallography in drug design*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
- [6.38] Zartler, E. R.; Shapiro, M. J., *Curr. Opin. Chem. Biol.* **2005**, 9, 366-370.
- [6.39] Schneider, G.; Fechner, U. *Nat. Rev. Drug Discov.* **2005**, 4, 649-663.
- [6.40] Sirois, S.; Wei, D. -Q.; Du, Q.; Chou, K. -C., *J. Chem. Inf. Model.* **2004**, 44, 1111-1122.
- [6.41] Lyne, P. D.; Kenny, P. W.; Cosgrove, D. A.; Deng, C.; Zabloudoff, S.; Wendoloski, J. J.; Ashwell, S., *J. Med. Chem.* **2004**, 47, 1962-1968.
- [6.42] Lengauer, T.; Lemmen, C.; Rarey, M.; Zimmermann, M., *Drug Discov. Today.* **2004**, 9, 27-34.
- [6.43] Willet, P., *Drug Discov. Today* **2006**, 11, 1046-1053.
- [6.44] Klebe, G., *Drug Discov. Today* **2006**, 11, 580-594.

## References and Notes

- [6.45] Ekins, S.; Swaan, P. W. (2004) Development of computational models for enzymes, transporters, channels, and receptors relevant to ADME/Tox. In *Reviews in computational chemistry*. Lipkowitz, K. B.; Larter, R.; Cundari, T. R.; Boyd, D. B., Eds., John Wiley & Sons Inc., Hoboken, New Jersey. Vol 20, pp. 333-376.
- [6.46] Tetko, I. V.; Bruneau, P.; Mewes, H. W.; Rohrer, D. C.; Poda, G. I., *Drug Discov. Today* **2006**, 11, 700-707.
- [6.47] Gasteiger, J.; Engel, T. (2003) *Chemoinformatics*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
- [6.48] Gasteiger, J. (2003) *Hand book of chemoinformatics*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
- [6.49] Bajorath, J. (2004) *Chemoinformatics: Concepts, methods and tools for drug discovery*. Methods in Molecular Biology. Vol. 275. Humana Press, New Jersey.
- [6.50] (a) Engel, T., *J. Chem. Inf. Model.* **2006**, 46, 2267-2277. (b) Chen, W. L., *J. Chem. Inf. Model.* **2006**, 46, 2230-2255. (c) Gasteiger, J., *J. Med. Chem.* **2006**, 50, 6429-6434.
- [6.51] Wermuth, C. G.; Ganellin, C. R.; Lindberg, P.; Mitscher, L. A., *Annu. Rep. Med. Chem.* **1998**, 33, 385-395.
- [6.52] Güner, O. F. (2000) *Pharmacophore: perception, development and use in drug design*. International university line: CA.
- [6.53] Langer, T.; Hoffman, R. D. (2006) *Pharmacophore and pharmacophore searches*. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.
- [6.54] Boyle, R. (1685) *Of the reconcileableness of specific medicines to the corpuscular philosophy*. Samuel Smith, London, pp. 72-75.
- [6.55] Fränkel, S. (1901) *Die Arzneimittel-Synthese auf Grundlage der Beziehungen zwischen chemischem Aufbau und Wirkung*. Julius Springer, Berlin, pp. 13-41.
- [6.56] Langley, J. N., *J. Physiol.* **1878**, 1, 339-369.
- [6.57] Langley, J. N., *J. Physiol.* **1905**, 33, 374-413.
- [6.58] Ehrlich, P.; Morgenroth, J., *Berl. Klin. Wochnschr.* **1900**, 37, 453-457.

- [6.59] Maehle, A. H.; Prull, C. R.; Halliwell, R. F., *Nat. Rev. Drug Discov.* **2002**, 1, 637-641.
- [6.60] Albert, A. (1979) *Selective Toxicity. The physicochemical basis of therapy.* Chapman and Hall, London, p. 23.
- [6.61] Fischer, E., *Ber. Dtsch. Chem. Ges.* **1894**, 27, 2985-2993.
- [6.62] Woods, D. D., *Br. J. Exp. Pathol.* **1940**, 21, 74-90.
- [6.63] Woods, D. D.; Fildes, P., *Chem. Ind.* **1940**, 59, 133-134.
- [6.64] Dodds, E. C.; Lawson, W., *Proc. R. Soc. London, Ser. B* **1938**, 125, 122-132.
- [6.65] Schueler, F. W., *Science* **1946**, 103, 221-223.
- [6.66] Easson, L. H.; Stedman, E., *Biochem. J.* **1933**, 27, 1257-1266.
- [6.67] Beckett, A. H. (1959) Stereochemical factors in biological activity. In *Fortschritte der Arzneimittel Forschung*. Birkhäuser Verlag, Basel, pp. 455-530.
- [6.68] Farmer, P. S.; Ariëns, E. J., *Trends Pharmacol. Sci.* **1982**, 3, 362-365.
- [6.69] Barlow, R. B. (1964) *Introduction to chemical pharmacology.* (2nd edn.) Methuen, London.
- [6.70] Belleau, B. (1963) An analysis of drug-receptor interactions. In *Modern concepts in the relationship between structure and pharmacological activity.* Brunings, K. J., Ed., Pergamon Press, Oxford, pp. 75-99.
- [6.71] Pullmann, B.; Coubeils, J. L.; Courrière, P.; Gervois, J. P., *J. Med. Chem.* **1972**, 15, 17-23.
- [6.72] Wermuth, C. G.; Schwartz, J.; Leclerc, G.; Garnier, J. P.; Rouot, B., *Chim. Thér.* **1973**, 1, 115-116.
- [6.73] Gund, P. (1977) Three-dimensional pharmacophoric pattern searching. In *Progress in Molecular and Subcellular Biology.* Han, F. E., Ed., Berlin, Springer-Verlag, Vol. 5, pp. 117-143.
- [6.74] Gund, P., *Ann. Reports Med. Chem.* **1979**, 14, 299-308.
- [6.75] Humblet, C.; Marshall, F. R., *Ann. Reports Med. Chem.* **1980**, 15, 267-276.
- [6.76] Gund, P.; Wipke, W. T.; Langridge, R., Computer searching of a molecular structure file for pharmacophoric patterns. In *Proc Intl Conf on Computers in*

## References and Notes

- Chem Res and Educa, Ljubljana, July 12-17, **1973**, Amsterdam, Elsevier, **1974**, 3:5/33-39.
- [6.77] Crippen G. M., *J. Comp. Phys.* **1977**, 24, 96-107.
- [6.78] Marshall, G. R. (**1993**) Binding-site modeling of unknown receptors. In *3D QSAR in Drug Design, Theory Methods and Applications*. Kubinyi, H., Ed., ESCOM, Leiden, pp. 80-116.
- [6.79] Wermuth, C. G., Langer, T. (**1993**) Pharmacophore identification. In *3D QSAR in Drug Design. Theory Methods and Applications*. Kubinyi, H., Ed., ESCOM, Leiden, pp. 117-136.

## CHAPTER 7

- [7.1] Klebe, G., *Drug Discov. Today* **2006**, 11, 580-594.
- [7.2] Schneider, G.; Böhm, H., -J., *Drug Discov. Today* **2002**, 7, 64-70.
- [7.3] Rauh, D.; Klebe, G.; Stubbs, M. T., *J. Mol. Biol.* **2004**, 335, 1325-1341.
- [7.4] Wilson, E. K., *C&EN*. **2004**, 82, 46-47.
- [7.5] Cavasotto, C. N.; Abagyan, R. A., *J. Mol. Biol.* **2004**, 337, 209-225.
- [7.6] The well-known Lipinski's rule of five may be regarded as a rudimentary form offirst generation VS. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., *Adv. Drug Deliv. Rev.* **1997**, 23, 3-25.
- [7.7] Sirois, S.; Wei, D. -Q.; Du, Q.; Chou, K. -C., *J. Chem. Inf. Model.* **2004**, 44, 1111-1122.
- [7.8] Keserü, G. M., *J. Comput. Aided Mol. Des.* **2001**, 15, 649-657.
- [7.9] Rollinger, J. M.; Hornick, A; Langer, T.; Stuppner, H.; Prast, H., *J. Med. Chem.* **2004**, 47, 6248-6254.
- [7.10] Rastelli, G.; Pacchioni, S.; Sirawaraporn, W; Sirawaraporn, R; Parenti, M. D.; Ferrari, A. M., *J. Med. Chem.* **2003**, 46, 2834-2845.
- [7.11] Muegge, I.; Enyedy, I. J., *Curr. Med. Chem.* **2004**, 11, 693-707.
- [7.12] Grüneberg, S; Stubbs, M. T.; Klebe, G., *J. Med. Chem.* **2002**, 45, 3588-3602.

- [7.13] Vieth, M.; Brooks, H. B.; Hamdouchi, C.; Mcmillen, W.; Sawyer, J. S.; Yingling, J. M.; Zhang, F., *Cell. Mol. Biol. Lett.* **2003**, 8, 566-567.
- [7.14] Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A., *J. Chem. Inf. Model.* **2006**, 46, 462-470.
- [7.15] Lengauer, T.; Lemmen, C.; Rarey, M.; Zimmermann, M., *Drug Discov. Today* **2004**, 1, 27-34.
- [7.16] Mustata, G. I.; Brigo, A.; Briggs, J. M., *Bioorg. Med. Chem. Lett.* **2004**, 14, 1447-1454.
- [7.17] Aronov, A. M.; Murko, M. A., *J. Med. Chem.* **2004**, 47, 5616-5619.
- [7.18] Parenti, M. D.; Pacchioni, S.; Ferrari, A. M.; Rastelli, G., *J. Med. Chem.* **2004**, 47, 4258-4267.
- [7.19] Renner, S.; Schneider, G., *J. Med. Chem.* **2004**, 47, 4653-4664.
- [7.20] Li, M. Y.; Tsaib, K. C.; Xia, L., *Bioorg. Med. Chem. Lett.* **2005**, 15, 657-664.
- [7.21] Krovat, E. M.; Fruhwirth, K. H.; Langer, T., *J. Chem. Inf. Model.* **2005**, 45, 146-159.
- [7.22] Vema, A.; Panigrahi, S. K.; Rambabu, G.; Gopalakrishnan, B.; Sarma, J. A. R. P.; Desiraju, G. R., *Bioorg. Med. Chem.* **2003**, 11, 4643-4653.
- [7.23] Hou, T.; Zhu, L.; Chen, L.; Xu, X., *J. Chem. Inf. Model.* **2003**, 43, 273-287.
- [7.24] Szántai-Kis, C.; Kövesdi, I.; Eros, D.; Bánhegyi, P.; Ullrich, A.; Kéri, G.; Orfi, L., *Curr. Med. Chem.* **2006**, 13, 277-287.
- [7.25] Aparna, V.; Rambabu, G.; Panigrahi, S. K.; Sarma, J. A. R. P.; Desiraju, G. R., *J. Chem. Inf. Model.* **2005**, 45, 725-738.
- [7.26] Dar, A. C.; Wybenga-Groot, L. E.; Sicheri, F. (2005) The eukaryotic protein kinase domain. In *Modular Protein Domains*, Cesareni, G.; Gimona, M.; Sudol, M.; Yaffe, M., Eds., Wiley-VCH Verlag GmbH & Co. KGaA, pp. 181-209.
- [7.27] Yang, J.; Yu, Y.; Duerksen-Hughes, P. J., *Mutat. Res.* **2003**, 543, 31-58.
- [7.28] Blume-Jensen, P.; Hunter, T., *Nature* **2001**, 411, 355-365.

## References and Notes

- [7.29] Stamos, J.; Sliwkowski, M. X.; Eigenbrot, C., *J. Biol. Chem.* **2002**, 277, 46265-46272.
- [7.30] Yarden, Y.; Ullrich, A., *Ann. Rev. Biochem.* **1988**, 57, 443-478.
- [7.31] Ullrich, A.; Schlessinger, J., *Cell* **1990**, 61, 203-212.
- [7.32] Pedersen, M. W.; Poulsen, H. S., *Sci. Med.* **2002**, 8, 206-217.
- [7.33] Traxler, P.; Bold, G.; Buchdunger, E.; Caravatti, G.; Furet, P.; Manley, P.; O'Reilly, T.; Wood, J.; Zimmermann, J., *Med. Res. Rev.* **2001**, 21, 499-512.
- [7.34] Atkins, J. H.; Gershell, L. J., *Nature Rev. Can.* **2002**, 1, 491-492.
- [7.35] Cohen, P., *Nature Rev. Drug Discov.* **2002**, 1, 309-315.
- [7.36] Dancey, J.; Sausville, E. A., *Nature Rev. Drug Discov.* **2003**, 2, 296-313.
- [7.37] Hubbard, S. R., *Curr. Opin. Struct. Biol.* **2002**, 12, 735-741.
- [7.38] Levitzki, A., *Acc. Chem. Res.* **2003**, 36, 462-469.
- [7.39] Traxler, P.; Furet, P., *Pharmacol. Ther.* **1999**, 82, 195-206.
- [7.40] Morin, M. J., *Oncogene*, **2000**, 19, 6574-6583.
- [7.41] Sarma, J. A. R. P., A kinase inhibitor database consisting of 136 678 compounds with chemical, biological, pharmacological and toxicological data was curated from journals and patents. GVK BIOSCIENCES Private Limited, August, **2004**.
- [7.42] Wissner, A.; Brawner 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.
- [7.43] Traxler, P.; Bold, G.; Frei, J.; Lang, M.; Lydon, N.; Mett, H.; Buchdunger, E.; Meyer, T.; Mueller, M.; Furet, P., *J. Med. Chem.* **1997**, 40, 3601-3616.
- [7.44] *Catalyst, Accelrys* : San Diego, CA, **1997**.
- [7.45] Xue, L.; Stahura, F. L.; Bajorath, J., *J. Chem. Inf. Model.* **2004**, 44, 2032-2039.
- [7.46] Agrafiotis, D. K.; Myslik, J. C.; Salemme, F. R., *Molecular Diversity*, **1999**, 4, 1-22.
- [7.47] Güner, O. F., Ed., **(2000)** *Pharmacophore: perception, development and use in drug design*. International university line: CA.

- [7.48] Smellie, A.; Teig, S. L.; Towbin, P., *J. Comput. Chem.* **1995**, 16, 171-187.
- [7.49] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., *Nucleic Acid Res.* **2000**, 28, 235-242.  
<http://www.rcsb.org/pdb>
- [7.50] Wood, E. R.; Truesdale, A. T.; McDonald, O. B.; Yuan, D.; Hassell, A.; Dickerson, S. H.; Ellis, B.; Pennisi, C.; Horne, E.; Lackey, K.; Alligood, K. J.; Rusnak, D. W.; Gilmer, T. M.; Shewchuk, L., *Cancer Res.* **2004**, 64, 6652-6659.
- [7.51] MOE Program, Version 2006-08 Chemical Computing Inc, Suite 910-1010 Sherbrooke St. W, Montreal, Quebec, Canada, H3A 2R7.
- [7.52] Halgren, T., *J. Comput. Chem.* **1996**, 17, 490-519.
- [7.53] Wolber, G.; Langer, T., *J. Chem. Inf. Model.* **2005**, 45, 160-169.
- [7.54] GOLD 3.0, Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK.
- [7.55] Kirchmair, J.; Laggner, C.; Wolber, G.; Langer, T., *J. Chem. Inf. Model.*, **2005**, 45, 422-430.
- [7.56] Kirchmair, J.; Wolber, G.; Laggner, C.; Langer, T., *J. Chem. Inf. Model.* **2006**, 46, 1848-1861.

## CHAPTER 8

- [8.1] Cummings, J. L., *N. Engl. J. Med.* **2004**, 351, 56-67.
- [8.2] Robichaud, A. J., *Curr. Top. Med. Chem.* **2006**, 6, 553-568.
- [8.3] Brimijoin, S., *Prog. Neurobiol.* **1983**, 21, 291-322.
- [8.4] Barnard, E. A. (1974) *The Peripheral Nervous System*. Plenum Press, New York.
- [8.5] Giacobini, E., *Neurochem. Res.* **2000**, 25, 1185-1190.
- [8.6] Giacobini, E., *Jpn. J. Pharmacol.* **1997**, 74, 225-241.
- [8.7] Carreiras, M. C.; Marco, J. L., *Curr. Pharm. Des.* **2004**, 10, 3167-3175.

## References and Notes

- [8.8] (a) Lushington, G. H.; Guo, J. X.; Hurley, M. M., *Curr. Top. Med. Chem.* **2006**, 6, 57-73. (b) Hurley, M. M.; Balboa, A.; Lushington, G. H.; Guo, J., *Chem. Biol. Interact.* **2005**, 157-158, 321-325. (c) Wlodek, S. T.; Shen, T.; McCammon, J. A., *Biopolymers* **2000**, 53, 265-271. (d) Barril, X.; Orozco, M.; Luque, F. J., *J. Med. Chem.* **1999**, 42, 5110-5119. (e) Antosiewicz, J.; Wlodek, S. T.; McCammon, J. A., *Biopolymers* **1996**, 39, 85-94. (f) Antosiewicz, J.; Gilson, M. K.; Lee, I. H.; McCammon, J. A., *Biophys. J.* **1995**, 68, 62-68.
- [8.9] (a) Niu, C.; Xu, Y.; Xu, Y.; Luo, X.; Duan, W.; Silman, I.; Sussman, J. L.; Zhu, W.; Chen, K.; Shen, J.; Jiang, H., *J. Phys. Chem. B. Condens. Matter Mater. Surf. Interfaces Biophys.* **2005**, 109, 23730-23738. (b) Senapati, S.; Bui, J. M.; McCammon, J. A., *J. Med. Chem.* **2005**, 48, 8155-8162. (c) Kousba, A. A.; Sultatos, L. G.; Poet, T. S.; Timchalk, C., *Toxicol. Sci.* **2004**, 80, 239-248. (d) Boyd, A. E.; Dunlop, C. S.; Wong, L.; Radic, Z.; Taylor, P.; Johnson, D. A., *J. Biol. Chem.* **2004**, 279, 26612-26618. (e) Xu, Y.; Shen, J.; Luo, X.; Silman, I.; Sussman, J. L.; Chen, K.; Jiang, H., *J. Am. Chem. Soc.* **2003**, 125, 11340-11349. (f) Bui, J. M.; Henchman, R. H.; McCammon, J. A., *Biophys. J.* **2003**, 85, 2267-2272. (g) Kua, J.; Zhang, Y.; McCammon, J. A., *J. Am. Chem. Soc.* **2002**, 124, 8260-8267. (h) Shen, T.; Tai, K.; Henchman, R. H.; McCammon, J. A., *Acc. Chem. Res.* **2002**, 35, 332-340. (i) Shen, T. Y.; Tai, K.; McCammon, J. A., *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2001**, 63, 041902. (j) Van Belle, D.; De Maria, L.; Iurcu, G.; Wodak, S. J., *J. Mol. Biol.* **2000**, 298, 705-726. (k) Axelsen, P. H.; Harel, M.; Silman, I.; Sussman, J. L., *Protein Sci.* **1994**, 3, 188-197.
- [8.10] (a) El Yazal, J.; Rao, S. N.; Mehl, A.; Slikker, W., Jr., *Toxicol. Sci.* **2001**, 63, 223-232. (b) Mager, P. P.; Weber, A., *Drug Des. Discov.* **2003**, 18, 127-150. (c) Zaheer-ul-Haq; Wellenzohn, B.; Tonmunpheap, S.; Khalid, A.; Choudhary, M. I.; Rode, B. M., *Bioorg. Med. Chem. Lett.* **2003**, 13, 4375-4380. (d) Kurunczi, L.; Olah, M.; Oprea, T. I.; Bologa, C.; Simon, Z., *J. Chem. Inf. Model.* **2002**, 42, 841-846. (e) Sippl, W.; Contreras, J. M.; Parrot, I.; Rival, Y. M.; Wermuth, C. G., *J. Comput. Aided Mol. Des.* **2001**, 15, 395-410.

- (f) Spassova, D. P.; Singh, A. K., *SAR QSAR Environ. Res.* **2001**, 11, 453-471.
- (g) Kaur, J.; Zhang, M. Q., *Curr. Med. Chem.* **2000**, 7, 273-294. (h) Hasegawa, K.; Kimura, T.; Funatsu, K., *J. Chem. Inf. Model.* **1999**, 39, 112-120. (i) Recanatini, M.; Cavalli, A.; Hansch, C., *Chem. Biol. Interact.* **1997**, 105, 199-228. (j) Tong, W.; Collantes, E. R.; Chen, Y.; Welsh, W. J., *J. Med. Chem.* **1996**, 39, 380-387.
- [8.11] (a) Bernard, P.; Kireev, D. B.; Chretien, J. R.; Fortier, P. L.; Coppet, L., *J. Comput. Aided Mol. Des.* **1999**, 13, 355-371. (b) Inoue, A.; Kawai, T.; Wakita, M.; Iimura, Y.; Sugimoto, H.; Kawakami, Y., *J. Med. Chem.* **1996**, 39, 4460-4470. (c) Zaheer-Ul-Haq, Z. U.; Wellenzohn, B.; Liedl, K. R.; Rode, B. M., *J. Med. Chem.* **2003**, 46, 5087-5090. (d) Alisaraie, L.; Haller, L. A.; Fels, G., *J. Chem. Inf. Model.* **2006**, 46, 1174-1187. (e) da Silva, C. H.; Campo, V. L.; Carvalho, I.; Taft, C. A., *J. Mol. Graph. Model.* **2006**, 25, 169-175. (f) Xie, Q.; Tang, Y.; Li, W.; Wang, X. H.; Qiu, Z. B., *J. Mol. Model.* **2006**, 12, 390-397. (g) Alisaraie, L.; Fels, G., *J. Mol. Model.* **2006**, 12, 348-354. (h) Graves, A. P.; Brenk, R.; Shoichet, B. K., *J. Med. Chem.* **2005**, 48, 3714-3728. (i) Guo, J.; Hurley, M. M.; Wright, J. B.; Lushington, G. H., *J. Med. Chem.* **2004**, 47, 5492-5500. (j) Pilger, C.; Bartolucci, C.; Lamba, D.; Tropsha, A.; Fels, G., *J. Mol. Graph. Model.* **2001**, 19, 374-378.
- [8.12] (a) Rollinger, J. M.; Hornick, A.; Langer, T.; Stuppner, H.; Prast, H., *J. Med. Chem.* **2004**, 47, 6248-6254. (b) Dickerson, T. J.; Beuscher, A. E. 4<sup>th</sup>; Rogers, C. J.; Hixon, M. S.; Yamamoto, N.; Xu, Y.; Olson, A. J.; Janda, K. D., *Biochemistry* **2005**, 44, 14845-14853. (c) Mizutani, M. Y.; Itai, A., *J. Med. Chem.* **2004**, 47, 4818-4828.
- [8.13] MOE Program, version 2006-08 Chemical Computing Inc, Suite 910-1010 Sherbrooke St. W, Montreal, Quebec, Canada, H3A 2R7.
- [8.14] Alisi, M. A.; Brufani, M.; Filocamo, L.; Gostoli, G.; Licandro, E.; Cesta, M. C.; Lappa, S.; Marchesini, D.; Pagella, P., *Bioorg. Med. Chem. Lett.* **1995**, 5, 2077-2080.

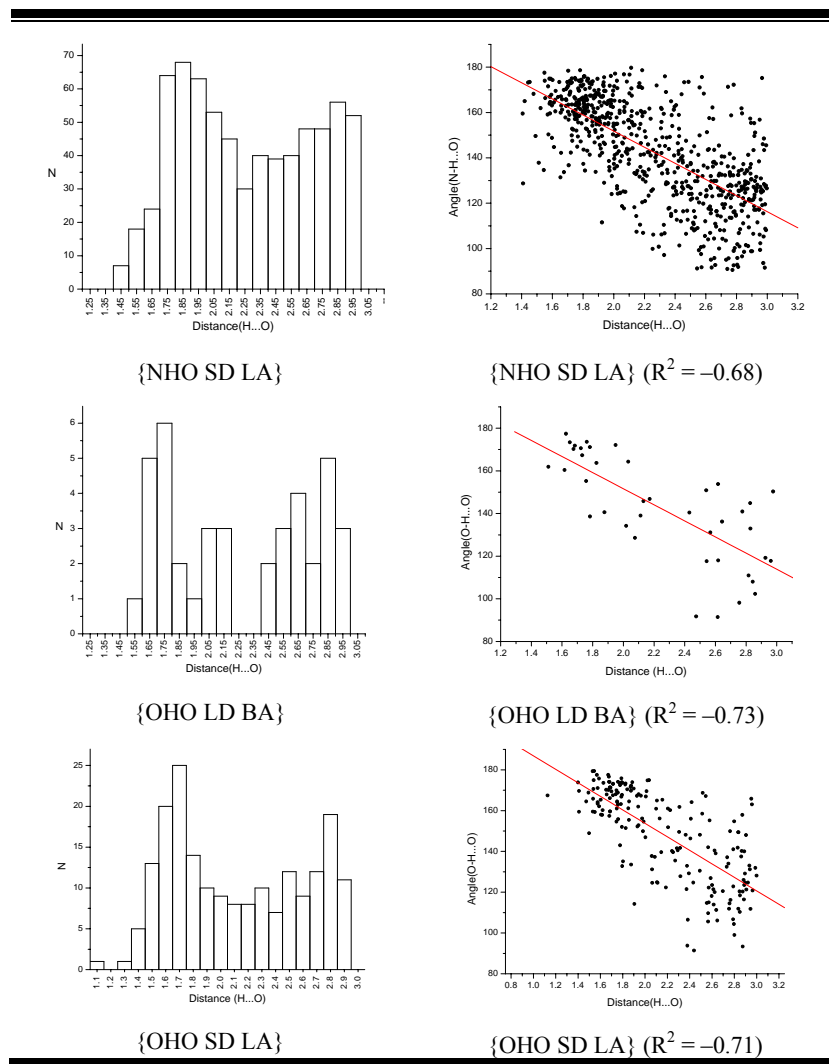
## References and Notes

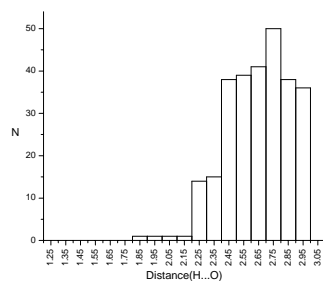
- [8.15] Obata, R.; Sunazuka, T.; Ootoguro, K.; Tomoda, H.; Harigaya, Y.; Omura, S., *Bioorg. Med. Chem. Lett.* **2000**, 10, 1315-1316.
- [8.16] Clark, J. K.; Cowley, P.; Muir, A. W.; Palin, R.; Pow, E.; Prosser, A. B.; Taylor, R.; Zhang, M. Q., *Bioorg. Med. Chem. Lett.* **2002**, 12, 2565-2568.
- [8.17] Palin, R.; Clark, J. K.; Cowley, P.; Muir, A. W.; Pow, E.; Prosser, A. B.; Taylor, R.; Zhang, M. Q., *Bioorg. Med. Chem. Lett.* **2002**, 12, 2569-2572.
- [8.18] Rampa, A.; Bisi, A.; Belluti, F.; Gobbi, S.; Valenti, P.; Andrisano, V.; Cavrini, V.; Cavalli, A.; Recanatini, M., *Bioorg. Med. Chem.* **2000**, 8, 497-506.
- [8.19] Martinez, A.; Lanot, C.; Perez, C.; Castro, A.; Lopez-Serrano, P.; Conde, S., *Bioorg. Med. Chem.* **2000**, 8, 731-738.
- [8.20] Tabarrini, O.; Cecchetti, V.; Temperini, A.; Filipponi, E.; Lamperti, M. G.; Fravolini, A., *Bioorg. Med. Chem.* **2001**, 9, 2921-2928.
- [8.21] Bedford, C. D.; Harris, R. N., 3rd; Howd, R. A.; Miller, A.; Nolen, H. W., 3rd; Kenley, R. A., *J. Med. Chem.* **1984**, 27, 1431-1438.
- [8.22] Bedford, C. D.; Harris, R. N., 3rd; Howd, R. A.; Goff, D. A.; Koolpe, G. A.; Petesch, M.; Miller, A.; Nolen, H. W., 3rd; Musallam, H. A.; Pick, R. O., *J. Med. Chem.* **1989**, 32, 493-503.
- [8.23] Recanatini, M.; Cavalli, A.; Belluti, F.; Piazzzi, L.; Rampa, A.; Bisi, A.; Gobbi, S.; Valenti, P.; Andrisano, V.; Bartolini, M.; Cavrini, V., *J. Med. Chem.* **2000**, 43, 2007-2018.
- [8.24] Halgren, T., *J. Comput. Chem.* **1996**, 17, 490-519.
- [8.25] *Catalyst* version 4.7, Accelrys: Burlington, MA.
- [8.26] GOLD 3.0, Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK.
- [8.27] Kryger, G.; Harel, M.; Giles, K.; Toker, L.; Velan, B.; Lazar, A.; Kronman, C.; Barak, D.; Ariel, N.; Shafferman, A.; Silman, I.; Sussman, J. L., *Acta. Crystallogra. Sect. D* **2000**, 56, 1385-1394.
- [8.28] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., *Nucleic Acid Res.* **2000**, 28, 235-242.  
<http://www.rcsb.org/pdb>

- [8.29] Irwin, J. J.; Shoichet, B. K., *J. Chem. Inf. Model.* **2005**, 45, 177-182.
- [8.30] Sussman, J. L.; Harel, M.; Frolow, F.; Oefner, C.; Goldman, A.; Toker, L.; Silman, I., *Science* **1991**, 253, 872-879.
- [8.31] Kirchmair, J.; Laggner, C.; Wolber, G.; Langer, T., *J. Chem. Inf. Model.* **2005**, 45, 422-430.
- [8.32] Kirchmair, J.; Wolber, G.; Laggner, C.; Langer, T., *J. Chem. Inf. Model.*, **2006**, 46, 1848-1861.

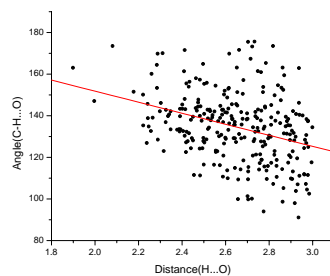
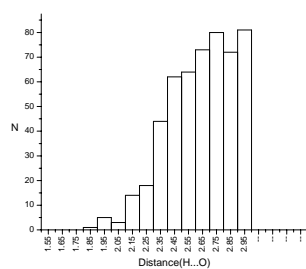
## APPENDIX I

**Figure 9:** Distance H...O histograms and  $d-\theta$  scatterplots for various hydrogen bonds in protein–ligand complexes.

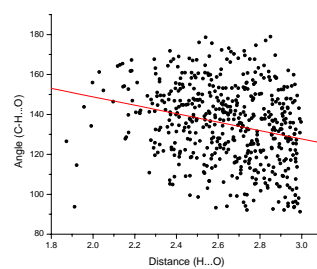
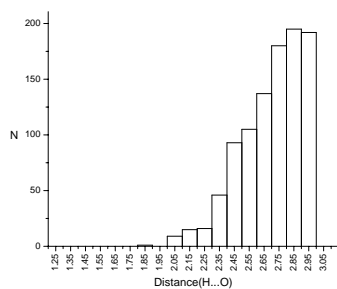




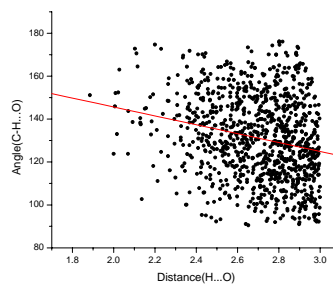
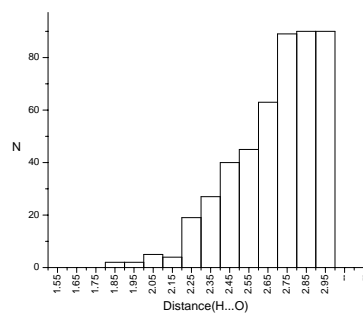
{CHO BD LA}

{CHO BD LA} ( $R^2 = -0.33$ )

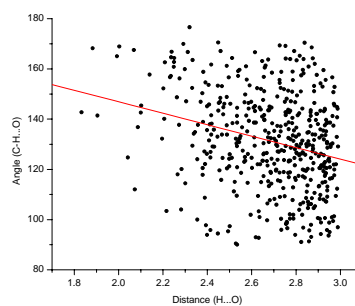
{CHO LD BA}

{CHO LD BA} ( $R^2 = -0.25$ )

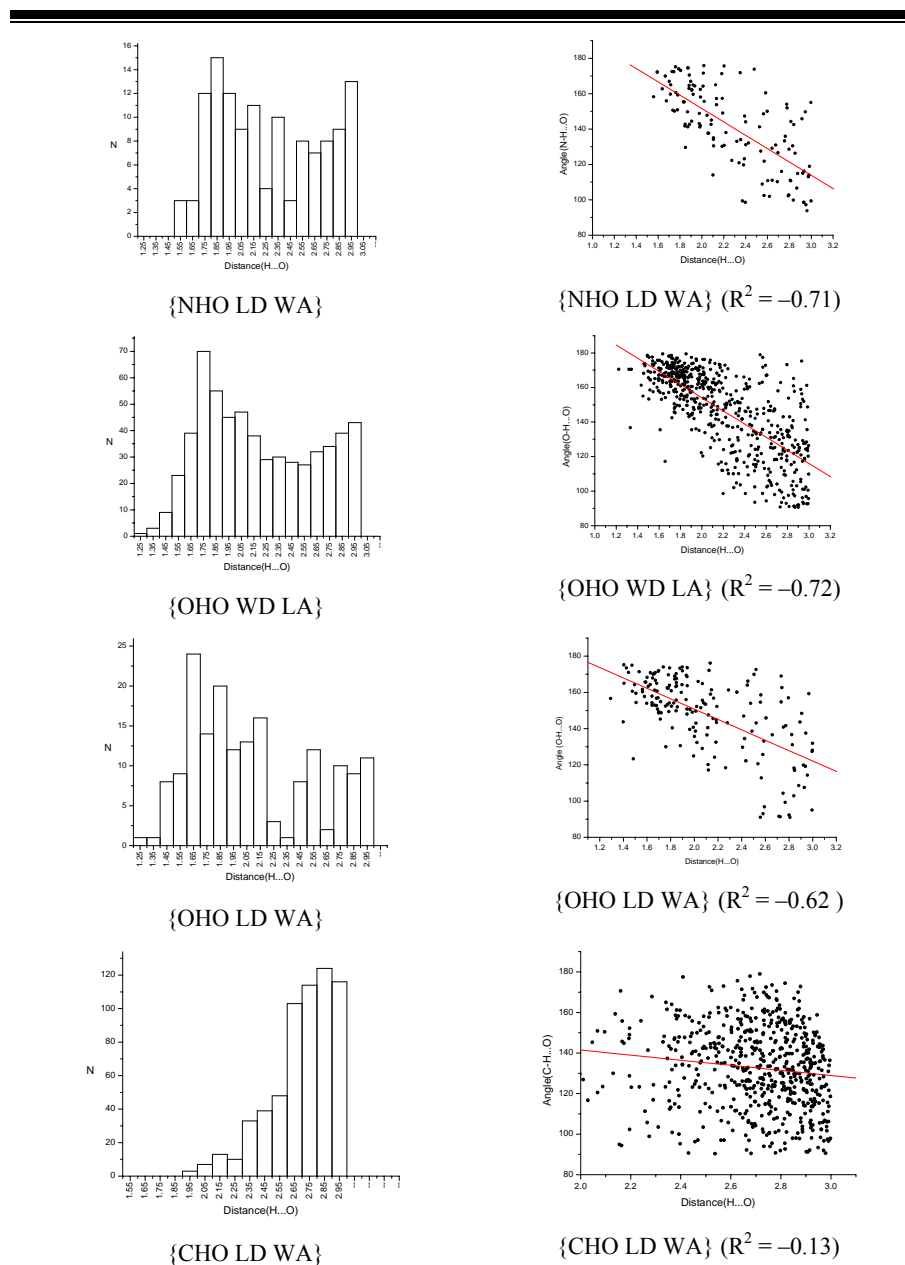
{CHO SD LA}

{CHO SD LA} ( $R^2 = -0.22$ )

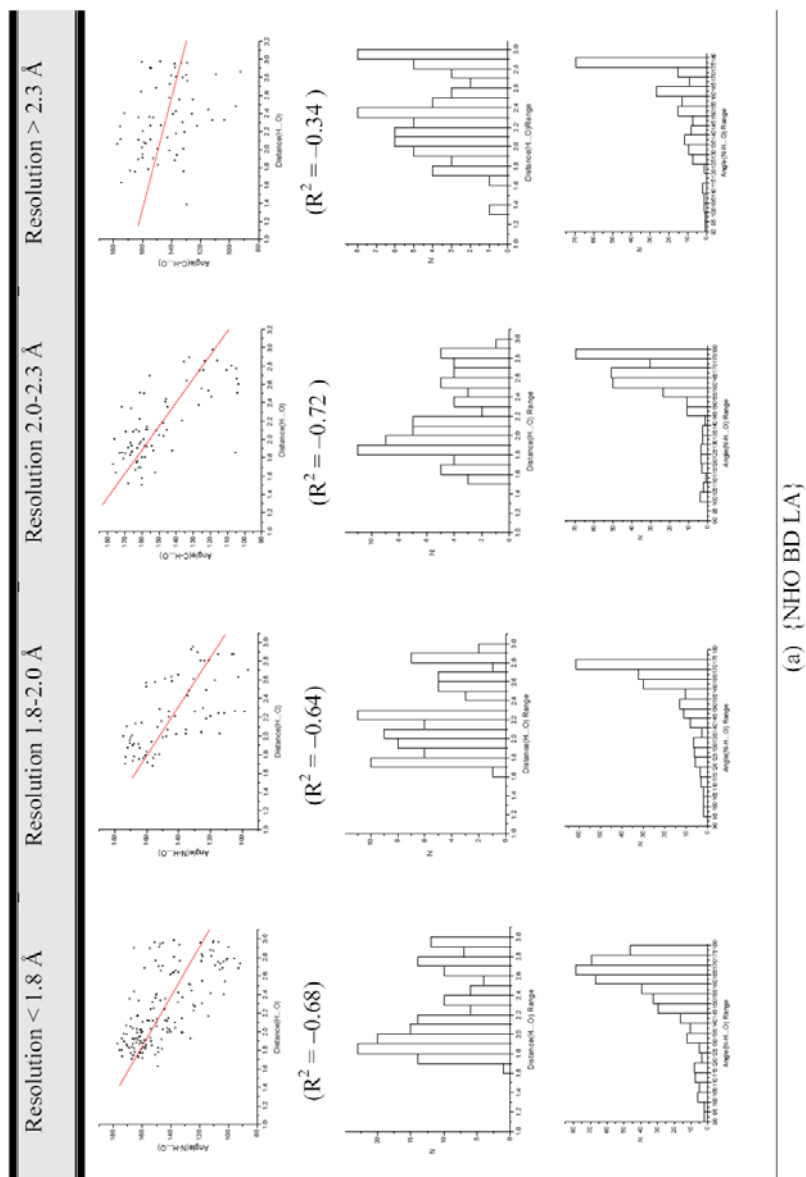
{CHO LD SA}

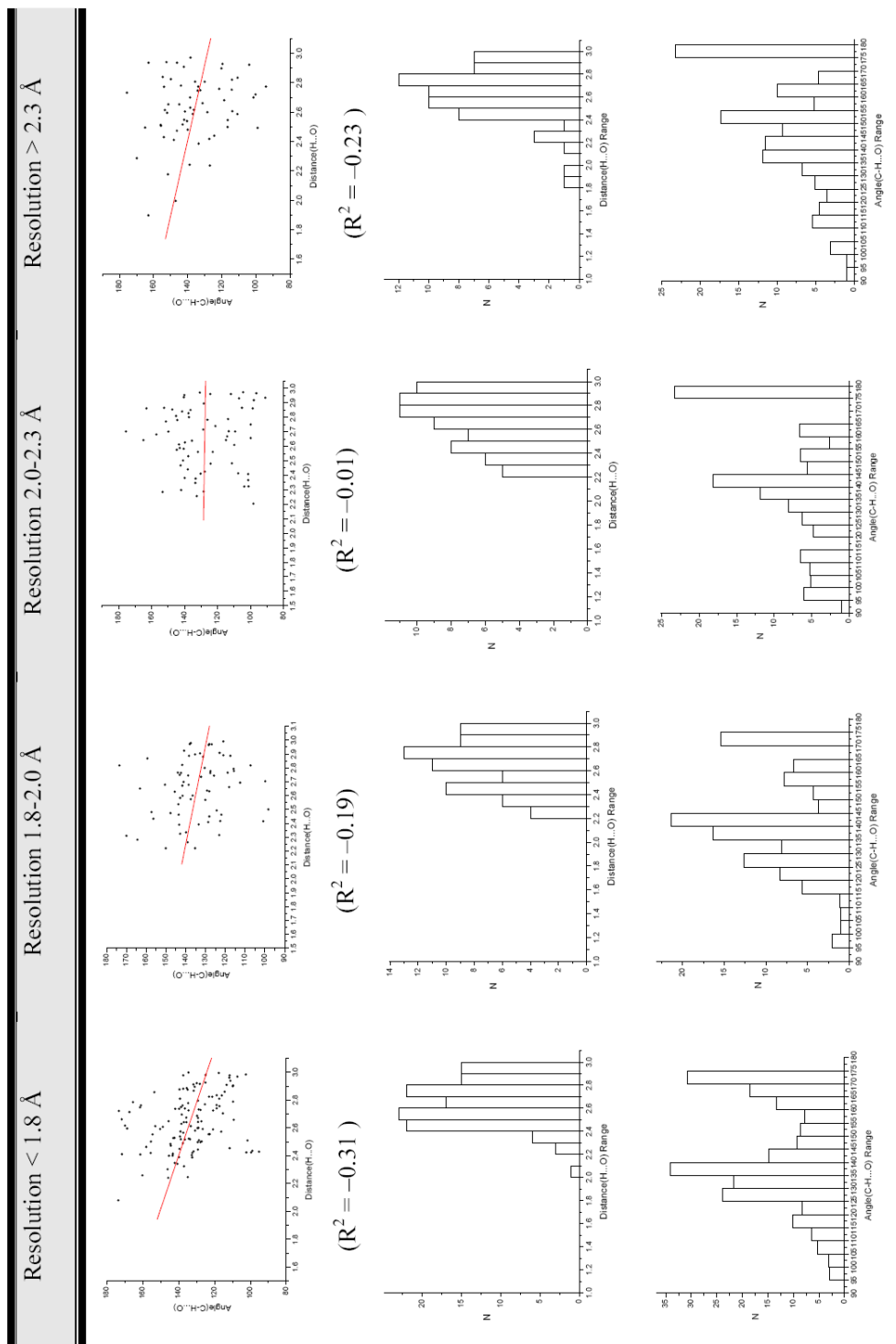
{CHO LD SA} ( $R^2 = -0.26$ )

**Figure 10:** Frequency distribution of H...O distances and respective  $d$ - $\theta$  scatterplots are shown for hydrogen bond to water as donor and/or acceptor.

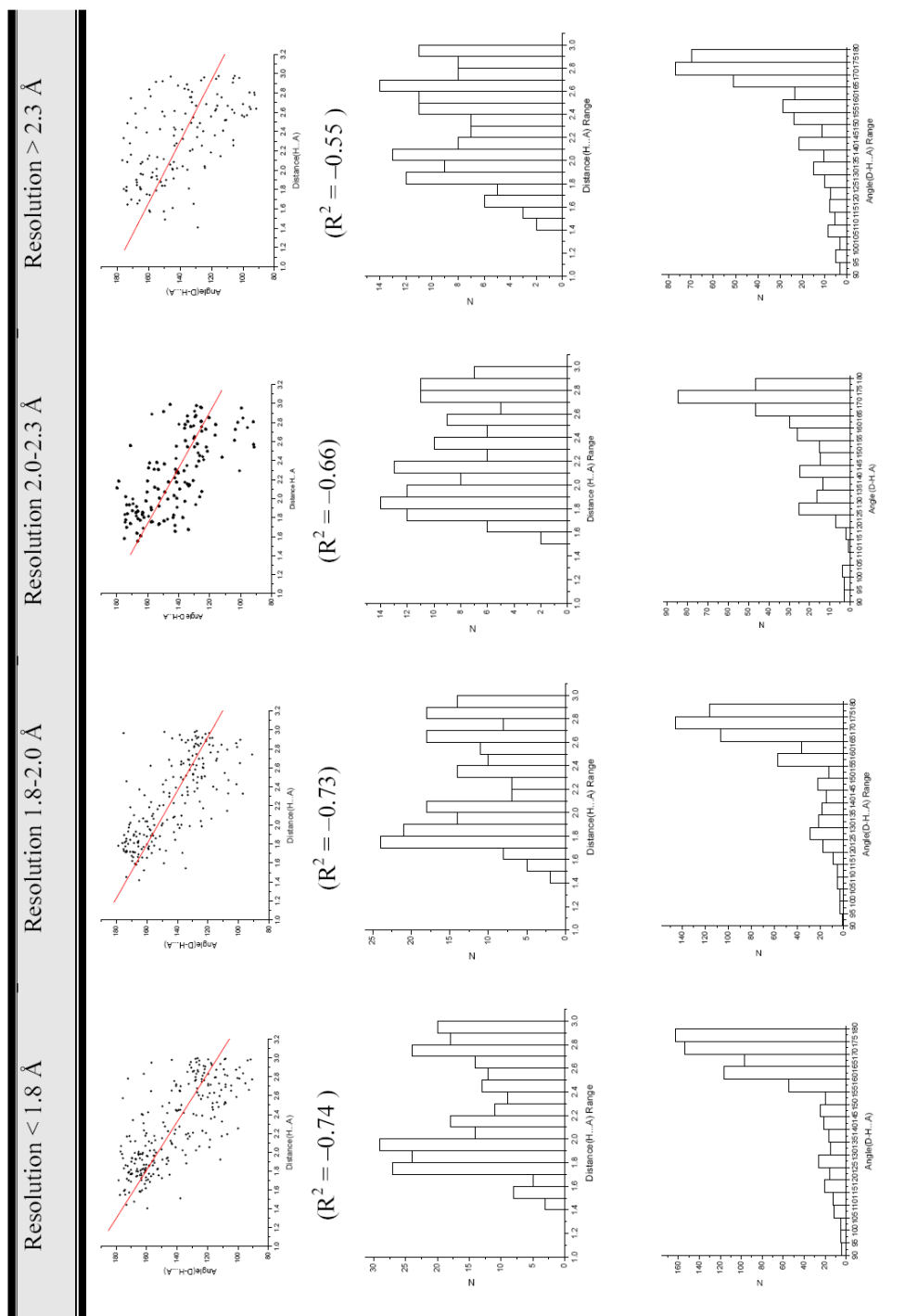


**Figure 11:** Hydrogen bond geometry as a function of resolution for {NHO BD LA}, {CHO BD LA}, {NHO SD LA} and {CHO SD LA}. For each resolution limit,  $d$ - $\theta$  plots, frequency cone corrected  $\theta$  angles are given.

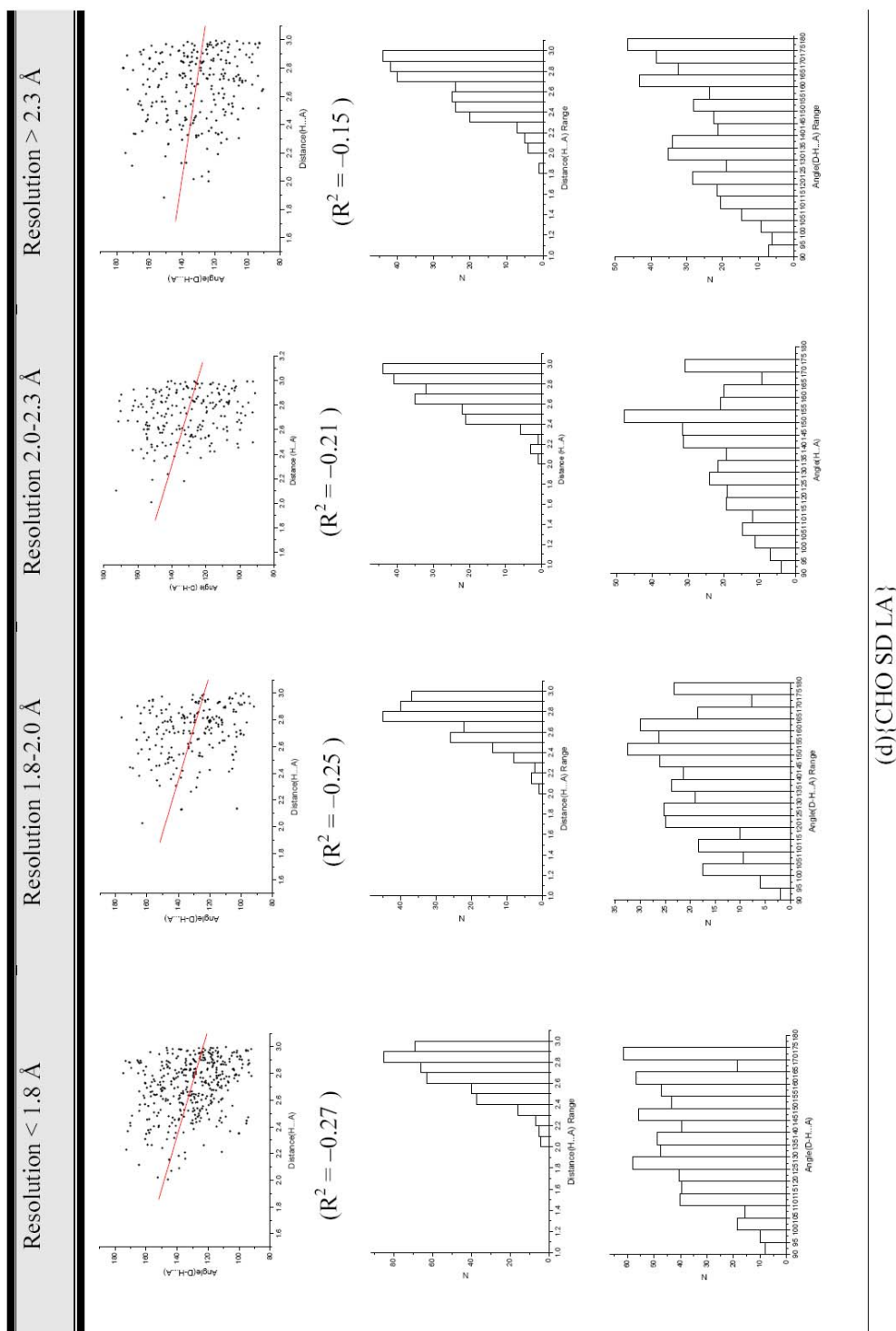




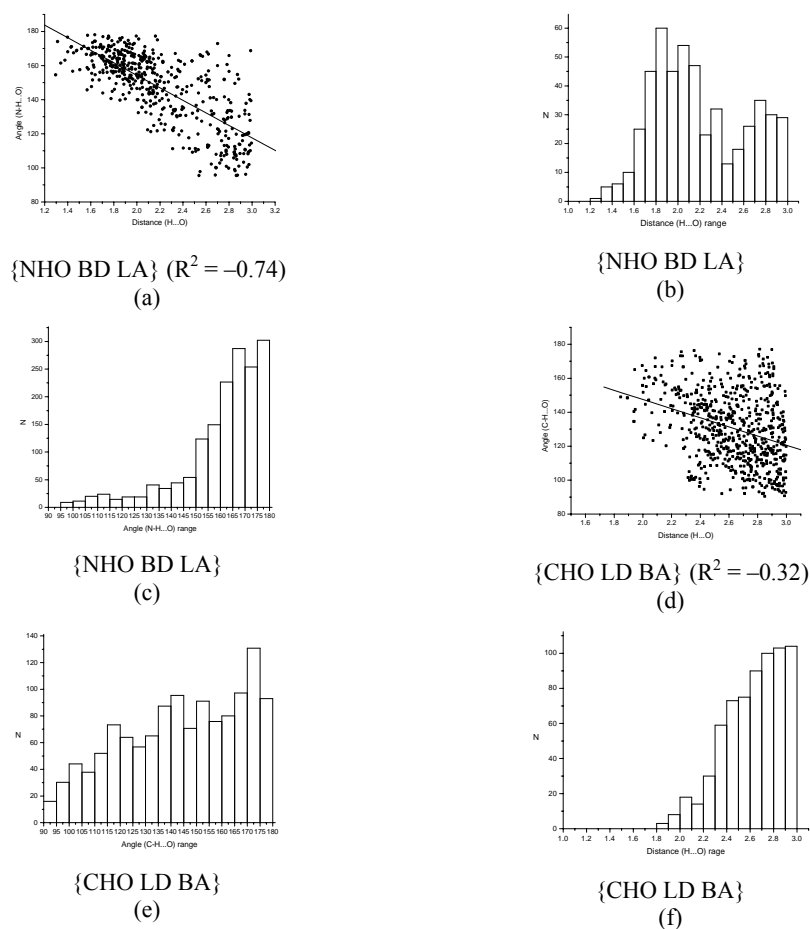
(b){CHO BD LA}

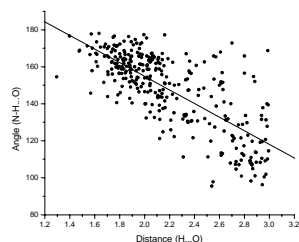


(c){NHO SD LA}

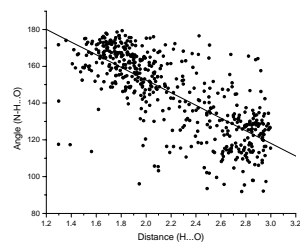


**Figure 12:** Frequency distribution of H $\cdots$ O distances and respective  $d$ - $\theta$  scatterplots and cone corrected  $\theta$  angles are shown for 233 protein–ligand complexes (kinases) for {NHO BD LA} (a)-(c) and {CHO LD BA} (d)-(f). Also shown are the  $d$ - $\theta$  scatterplots for hydrogen bond geometry as a function of resolution for {NHO BD LA} (g)-(j) and {CHO BD LA} (k)-(n).

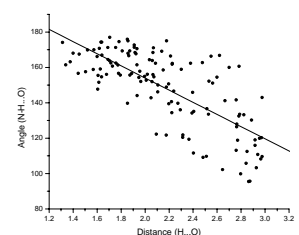




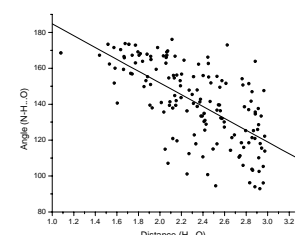
{NHO BD LA} ( $R^2 = -0.74$ )  
Resolution < 2.3 Å  
(g)



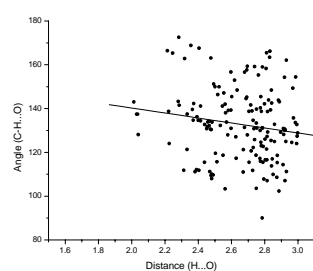
{NHO SD LA} ( $R^2 = -0.74$ )  
Resolution < 2.3 Å  
(h)



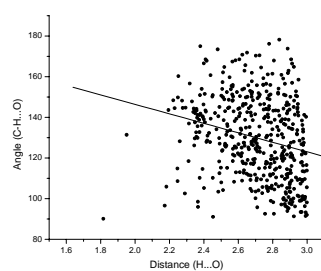
{NHO BD LA} ( $R^2 = -0.73$ )  
Resolution > 2.3 Å  
(i)



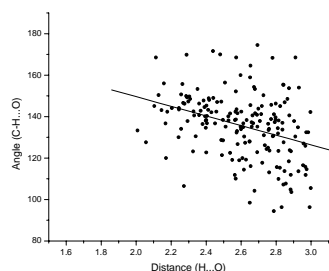
{NHO BD LA} ( $R^2 = -0.65$ )  
Resolution > 2.3 Å  
(j)



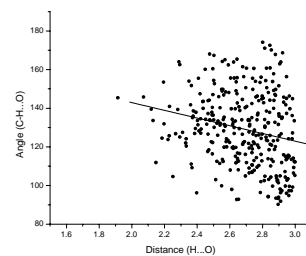
{CHO BD LA} ( $R^2 = -0.15$ )  
Resolution < 2.0 Å  
(k)



{CHO BD LA} ( $R^2 = -0.23$ )  
Resolution < 2.0 Å  
(l)

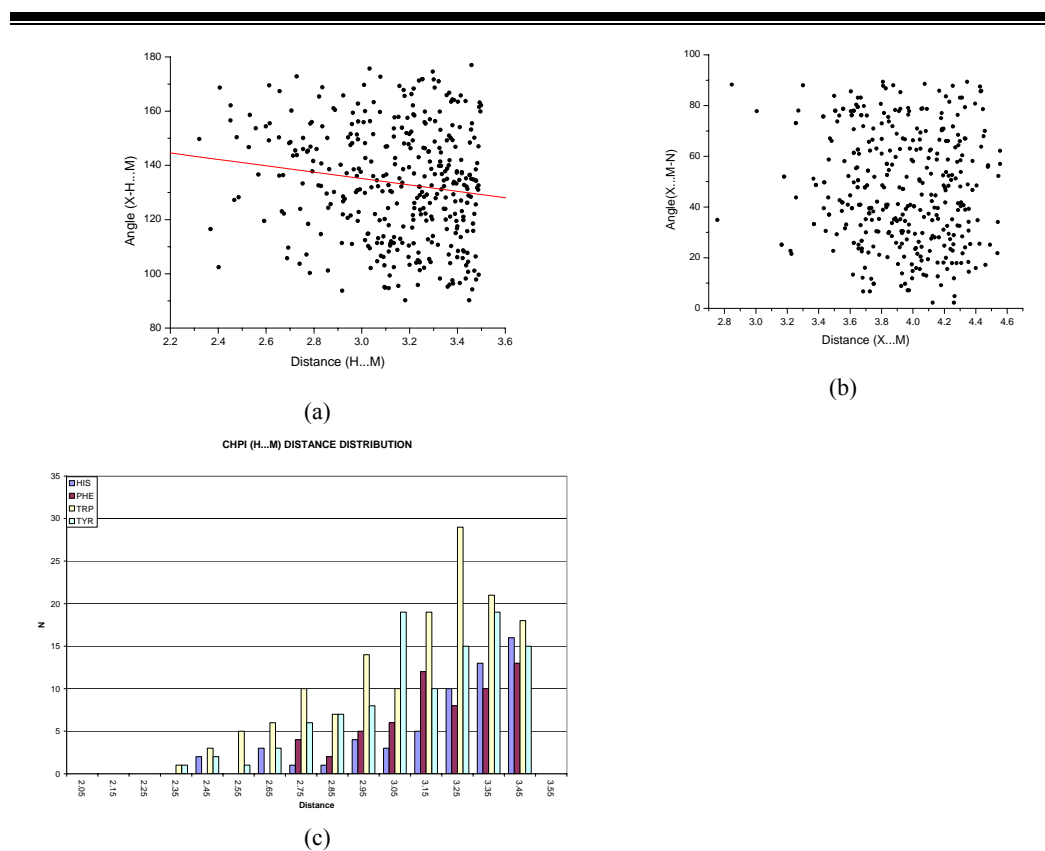


{CHO BD LA} ( $R^2 = -0.34$ )  
Resolution > 2.0 Å  
(m)



{CHO BD LA} ( $R^2 = -0.22$ )  
Resolution > 2.0 Å  
(n)

**Figure 13:** Geometry of C–H $\cdots\pi$  interactions in protein–ligand interface (a)  $d$ - $\theta$  scatterplot, (b)  $D$ - $\omega$  scatterplot (c)  $d$  (H $\cdots$ M) histogram.



**Table 6:** Percent composition of amino acid residues in 251 complexes in terms of total and active sites residues.

Nonpolar	Total %	Active site %	Polar	Total %	Active site %	Charged	Total %	Active site %
Gly	8.09	9.63	Ser	6.92	6.46	Lys	5.95	3.63
Val	7.02	7.12	Cys	1.57	1.91	Asp	5.71	5.15
Leu	8.30	8.97	Thr	6.65	6.26	Arg	4.21	4.20
Ilu	5.67	6.36	Asn	4.23	4.34	Glu	5.88	4.10
Met	2.14	2.45	Gln	3.98	3.02	His	2.23	2.83
Pro	4.76	3.86	Tyr	3.79	5.18			
Phe	3.84	5.11						
Trp	1.55	2.39						
Ala	7.41	6.80						
<b>Total %</b>	<b>48.78</b>	<b>52.69</b>		<b>27.14</b>	<b>27.17</b>		<b>23.98</b>	<b>19.91</b>

**Table 7:** Frequency distribution (%) of  $O_w-H\cdots O$  distance  $d$ , for {OHO WD LA}.

$d$ , range in Å	%
<1.21	-
1.21-1.71	13
1.71-2.21	42
2.21-2.71	26
2.71-3.21	18

**Table 8:** List of entries of 233 PDB structures (kinase) based on four resolution limits.

Resolution < 1.8 Å ( 42 Entries)								
1RDQ	1E2F	1E2D	1O49	1JVP	1O4N	1E9B	1YWN	1E2Q
1GZ8	1E9A	1O4L	1O4J	1O48	1OIT	1E9D	1O4I	1UU3
1KPF	1E9C	1OM1	1O4O	1O4G	1T46	1O4I	1H01	
1O43	1E9E	1E2G	1O4Q	1O4K	1URW	1O42	1O4R	
1O4A	1H00	1E2K	1QF9	16PK	1AKY	1O44	1O4M	
Resolution 1.8-2.0 Å ( 35 Entries)								
1H08	1X8B	1NZL	2KI5	1YWR	1O47	1O4D	1QHI	1Q5K
1NVR	1E1X	1O4P	3UKD	4TMK	1R0P	1E98	1UMW	1E1V
1NVS	1H07	1P4F	1OIR	5TMP	1WBS	1AKE	1UU7	1G3U
1O45	1O4B	1Q8U	1Y57	1UU9	1XWS	1E9F	2BHE	
Resolution 2.0-2.3 Å ( 78 Entries)								
1AQ1	1KE6	1O4E	1WBT	4UKD	1H0W	1OZ1	3ERK	1GIJ
1B38	1KE7	1O4F	1WBV	1SVG	1H1P	1P7C	1Y91	1IJR
1CDK	1KE8	1PME	1XRJ	1CKP	1IEP	1Q41	1WBO	1JBP
1E2E	1KE9	1PYE	1Y8Y	1B39	1NHX	1RE8	1DI8	1KE5
1GII	1MRS	1Q8T	1YQJ	1BWN	1NZV	1W2G	1E2N	1PMN
1H1R	1NVQ	1QPE	1ZZ2	1BX6	1OI9	1Y6A	1FMO	1PMQ
1H1S	1O46	1R78	3TMK	1DM2	1OVE	1Y6B	1FVT	1Q3D
1YDS	1O4H	1R0E	4ERK	1VYZ	1W7H	1P5E	1M51	
1Q8W	1REJ	1UNL	1W0X	1W82	1W84	1YDR		
Resolution > 2.3 Å (78 Entries)								
1PY5	1BMK	1XKK	1H1Q	1XMR	1RQQ	1GIH	1ZAK	
1Q3W	1BYG	1WBW	1KF0	2FGI	1UDW	1KV2		
1REK	1E2L	1MRN	1KV1	1OKZ	1UEI	1UV5		
1STC	1E7V	1SVE	1OUK	1PF8	2BHH	1UVR		
1SVH	1E8Z	1A9U	1OUY	1BKX	1G5S	1VEB		
1YDT	1FBZ	1BL6	1P2A	1DI9	1E90	1CTP		
1OIQ	1FPU	1BL7	1PMV	1GNG	1GAG	1XO2		
1V1K	1M7Q	1E2P	1SZM	1M17	1PMU	1BWF		
1DVR	1O1Y	1E8W	1TVO	1M52	1Q4L	1ZRZ		
1AGW	1RW8	1E9H	1UU8	1OGU	1BG3	2HCK		
1B55	1WBN	1FGI	1W83	1Q24	1BUX	1BO5		

**Table 9:** Acceptor furcations in the active sites of 233 protein–ligand complexes (kinase).

Furcation Level	Furcated acceptors in active site		
	In protein and water	In ligand	Total
Bifurcated	10211	455	10666
Trifurcated	7746	358	8104
Tetrafurcated	5645	246	5891
Pentafurcated	2512	122	2634
Hexafurcated	683	46	729
<b>Total</b>	<b>26797</b>	<b>1227</b>	<b>28024</b>
<i>Nonfurcated</i>	<i>15261</i>	<i>562</i>	<i>15823</i>

**Table 10:** Strong and weak hydrogen bonds for ligands at various levels of acceptor furcation for kinases. The total number of hydrogen bonds (O–H···O, N–H···O, C–H···O) to the 1227 furcated acceptors in Table 9 is 2942. This corresponds to an average level of furcation of 2.4 interactions to each furcated acceptor in the active site.

Furcation level	Furcated acceptor		
	O–H···O	N–H···O	C–H···O
Bifurcated	133	142	273
Trifurcated	178	290	331
Tetrafurcated	142	279	387
Pentafurcated	109	232	199
Hexafurcated	37	114	96
<i>Nonfurcated</i>	<i>84</i>	<i>76</i>	<i>107</i>

**Table 11:** N–H/O–H··· $\pi$  interactions in the protein–ligand interface.

Type	Ligand	Residue ID	Donor Atom	Acceptor	Residue ID	$d$	$D$	$\theta$	$\omega$	PDB ID
N–H··· $\pi$	TPM	400	N3	Trp	36	2.8	3.3	107.9	37.7	1F3D
N–H··· $\pi$	ACA	100	N6	Trp	62	3.2	3.7	111.2	14.1	2PK4
N–H··· $\pi$	ORN	240	NE	Tyr	14	3.2	3.8	118.3	12.6	1LAH
N–H··· $\pi$	ACA	100	N6	Trp	62	3.3	4.0	121.9	14.1	2PK4
O–H··· $\pi$	HBU	1011	O3	His	633	2.9	3.9	162.3	25.2	1HYO
O–H··· $\pi$	FLU	600	O3	Trp	101	3.3	4.2	158.4	10.1	1FLR
O–H··· $\pi$	HBU	1011	O14	His	633	3.3	3.8	112.0	32.3	1HYO

**Table 12:** (a) Interactions for halogen acceptors of type O–H $\cdots$ X. W is water, L is ligand.

Type	Donor	Acceptor	Donor	Donor Atom	Residue ID	Acceptor	Residue ID	$d$	$D$	$\theta$	PDB ID
O–H $\cdots$ Cl	W	L	HOH	O	402	CLM	221	2.3	3.2	155.7	3CLA
O–H $\cdots$ Cl	W	L	HOH	O	256	ENH	703	2.4	3.3	151.8	1CLE
O–H $\cdots$ Cl	W	L	HOH	O	267	V36	195	2.9	3.2	96.9	1VGC
O–H $\cdots$ Cl	S	L	Ser	OG	91	ENH	703	2.9	3.6	124.4	1CLE
O–H $\cdots$ F	W	L	HOH	O	409	FBA	900	2.9	3.8	161.6	1TNH

**Table 12:** (b) N–H $\cdots$ halogen interactions. B is backbone, S is side chain.

Type	Donor	Acceptor	Donor	Donor Atom	Residue ID	Acceptor	Residue ID	$d$	$D$	$\theta$	PDB ID
N–H $\cdots$ Cl	B	L	Asp	N	86	PVB	1	2.7	3.5	133.2	1CKP
N–H $\cdots$ Cl	B	L	Ala	N	100	ENH	703	2.7	3.4	127.9	1CLE
N–H $\cdots$ Cl	B	L	Thr	N	96	ENH	703	2.9	3.1	92.9	1CLE
N–H $\cdots$ Cl	S	L	Asn	ND2	35	ENH	703	2.9	3.3	107.2	1CLE
N–H $\cdots$ F	B	L	Leu	N	257	TFA	256	2.3	2.8	105.6	1BMA
N–H $\cdots$ F	B	L	Ser	N	203	TFA	256	2.6	3.1	113.4	1BMA
N–H $\cdots$ F	S	L	His	NE2	57	FPA	5	2.0	2.8	133.0	4EST
N–H $\cdots$ F	S	L	Asn	ND2	155	G2F	603	2.0	2.7	125.6	2NLR
N–H $\cdots$ F	S	L	Lys	NZ	136	FKI	404	2.2	3.0	133.0	1DY9
N–H $\cdots$ F	S	L	Arg	NH1	120	S58	701	2.3	3.3	174.5	1CX2
N–H $\cdots$ F	S	L	Arg	NH1	120	S58	701	2.4	3.0	119.4	1CX2
N–H $\cdots$ F	S	L	Asn	ND2	158	G2F	603	2.6	3.6	167.7	2NLR

**Table 12:** (c) C–H···halogen interactions.

Type	Donor	Acceptor	Donor	Donor Atom	Residue ID	Acceptor	Residue ID	<i>d</i>	<i>D</i>	$\theta$	PDB ID
C–H···Cl	B	L	Gly	CA	216	V36	195	2.5	3.4	138.7	1VGC
C–H···Cl	B	L	Thr	CA	97	ENH	703	2.6	3.3	123.1	1CLE
C–H···Cl	B	L	Arg	CA	100	ENH	703	2.6	3.5	139.9	1CLE
C–H···Cl	B	L	Thr	CA	97	ENH	703	2.9	3.9	159.6	1CLE
C–H···Cl	B	L	Arg	CA	100	ENH	703	2.9	3.8	144.0	1CLE
C–H···Cl	S	L	Leu	CB	384	IMN	701	2.4	3.1	121.2	4COX
C–H···Cl	S	L	Leu	C	384	IMN	701	2.5	3.5	161.5	4COX
C–H···Cl	S	L	Arg	CG	100	ENH	703	2.5	3.2	122.6	1CLE
C–H···Cl	S	L	Pro	CG	96	ENH	703	2.6	3.6	154.3	1CLE
C–H···Cl	S	L	Lys	CD	273	PP2	1904	2.7	3.8	163.5	1QPE
C–H···Cl	S	L	Leu	CD2	47	ENH	703	2.8	3.5	126.5	1CLE
C–H···Cl	S	L	Ala	CB	105	CLM	221	2.8	3.6	128.3	3CLA
C–H···Cl	S	L	Ser	CB	91	ENH	703	2.8	3.3	106.9	1CLE
C–H···Cl	S	L	Ile	CG2	314	PP2	1904	2.9	3.5	114.9	1QPE
C–H···Cl	S	L	Arg	CB	100	ENH	703	2.9	3.6	122.4	1CLE
C–H···F	B	L	Val	CA	4	FPA	5	2.2	2.7	104.3	4EST
C–H···F	B	L	Gln	CA	192	FBA	900	2.5	3.5	148.3	1TNH
C–H···F	B	L	Ala	CA	234	961	1	2.5	3.3	123.4	4LBD
C–H···F	B	L	Gly	CA	34	CHF	1	2.7	3.6	139.3	1EPO
C–H···F	S	L	Tyr	CE2	355	S58	701	2.1	3.1	150.2	1CX2
C–H···F	S	L	Thr	CB	106	SB2	800	2.3	3.2	149.7	1A9U
C–H···F	S	L	Leu	CG	359	A26	169	2.3	3.3	152.3	1D3H
C–H···F	S	L	Thr	CG2	221	TFA	256	2.4	3.4	155.7	1BMA
C–H···F	S	L	Arg	CD	120	S58	701	2.4	3.4	141.7	1CX2
C–H···F	S	L	Leu	CB	104	SB4	800	2.5	3.1	119.3	1B17
C–H···F	S	L	Leu	CD1	531	S58	701	2.5	3.5	166.6	1CX2
C–H···F	S	L	Val	CG1	224	TFA	256	2.5	3.4	142.2	1BMA
C–H···F	S	L	Leu	CB	104	SB2	800	2.5	3.4	135.6	1A9U
C–H···F	S	L	Val	CG1	349	S58	701	2.5	3.2	119.1	1CX2
C–H···F	S	L	Lys	CE	136	FK1	404	2.6	3.1	107.9	1DY9
C–H···F	S	L	Leu	CD2	86	SB2	800	2.6	3.5	135.6	1A9U
C–H···F	S	L	Pro	CD	364	A26	169	2.6	3.3	119.4	1D3H
C–H···F	S	L	Gln	CG	192	FPA	5	2.6	3.6	142.5	4EST
C–H···F	S	L	Ala	CB	234	961	1	2.7	3.3	114.1	4LBD
C–H···F	S	L	Val	CG2	224	TFA	256	2.7	3.6	142.3	1BMA

**Table 13:** (a) O–H···S interactions.

Type	Donor	Acceptor	Donor	Donor Atom	Residue ID	Acceptor	Acceptor Atom	Residue ID	<i>d</i>	<i>D</i>	$\theta$	PDB ID
O–H···S	L	S	BAR	O1	212	Cys	SG	75	2.4	3.3	144.3	2CHT
O–H···S	L	S	DES	OP3	800	Met	SD	343	2.6	3.6	165.4	3ERD
O–H···S	S	L	Tyr	OH	7	GTB	SG2	1	2.2	3.1	160.5	1GLQ
O–H···S	S	L	Tyr	OH	7	GTS	SG2	1	2.6	3.5	171.1	1GLP
O–H···S	S	L	Tyr	OH	367	COM	S1	9800	2.6	3.2	119.6	1MRO
O–H···S	S	L	Tyr	OH	128	HDS	S1	132	2.6	3.4	140.8	1LIC
O–H···S	S	L	Thr	OG1	92	MES	S	6001	2.9	3.5	122.4	3CHB
O–H···S	W	L	HOH	O	710	ACV	S17	351	2.3	3.3	163.3	1BK0
O–H···S	W	L	HOH	O	713	ACV	S17	351	2.4	3.3	149.9	1BK0
O–H···S	W	L	HOH	O	201	478	S1	200	2.6	3.5	153.9	1HPV
O–H···S	W	L	HOH	O	1178	BLG	S	401	2.7	3.3	126.5	1D01
O–H···S	W	L	HOH	O	7563	GL3	S	445	2.7	3.4	126.8	1MRO

**Table 13:** (b) N–H···S interactions.

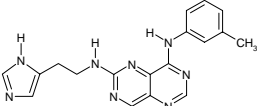
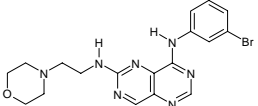
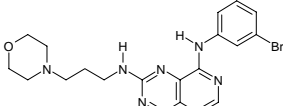
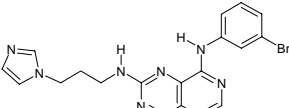
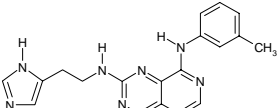
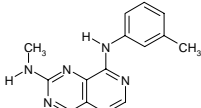
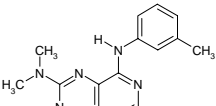
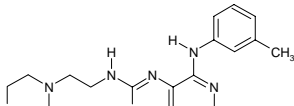
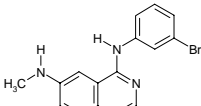
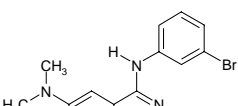
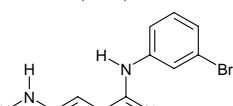
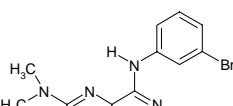
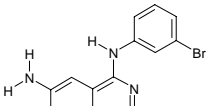
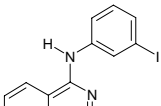
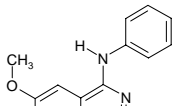
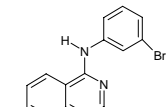
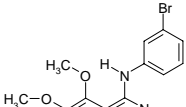
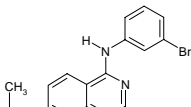
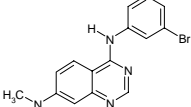
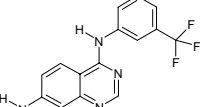
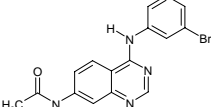
Type	Donor	Acceptor	Donor	Donor Atom	Residue ID	Acceptor	Acceptor Atom	Residue ID	<i>d</i>	<i>D</i>	$\theta$	PDB ID
N–H···S	B	L	Leu	N	170	SLE	S	301	2.1	3.1	161.2	1ATL
N–H···S	B	L	Leu	N	448	GL3	S	445	2.5	3.5	161.2	1MRO
N–H···S	B	L	Val	N	482	TP7	S7	9500	2.5	3.5	159.8	1MRO
N–H···S	B	L	Thr	N	199	BZO	S1	555	2.9	3.9	172.1	1A42
N–H···S	L	S	PLU	N	500	Met	SD	270	2.8	3.8	174.0	1LCP
N–H···S	L	S	BPP	N15	400	Cys	SG	267	2.9	3.9	171.8	1D4P
N–H···S	S	L	ARG	NH1	341	MNO	S1	601	2.4	3.3	136.8	1BMQ
N–H···S	S	L	ASN	ND2	481	TP7	S7	9500	2.6	3.6	169.8	1MRO
N–H···S	S	L	His	NE2	214	ACV	S17	351	2.6	3.3	127.7	1BK0
N–H···S	S	L	Lys	NZ	43	MES	S	6002	2.7	3.6	143.1	3CHB
N–H···S	S	L	Arg	NE	126	HDS	S1	132	2.7	3.6	149.9	1LIC
N–H···S	S	L	His	ND1	237	MNO	S1	601	2.7	3.7	165.4	1BMQ
N–H···S	S	L	Gln	NE2	19	I10	S39	300	2.8	3.6	144.5	1BGO
N–H···S	S	L	Arg	NH1	94	Met	S	6001	2.8	3.6	133.6	3CHB
N–H···S	S	L	Arg	NH2	96	TK4	S1	1001	2.8	3.8	148.1	1C5C

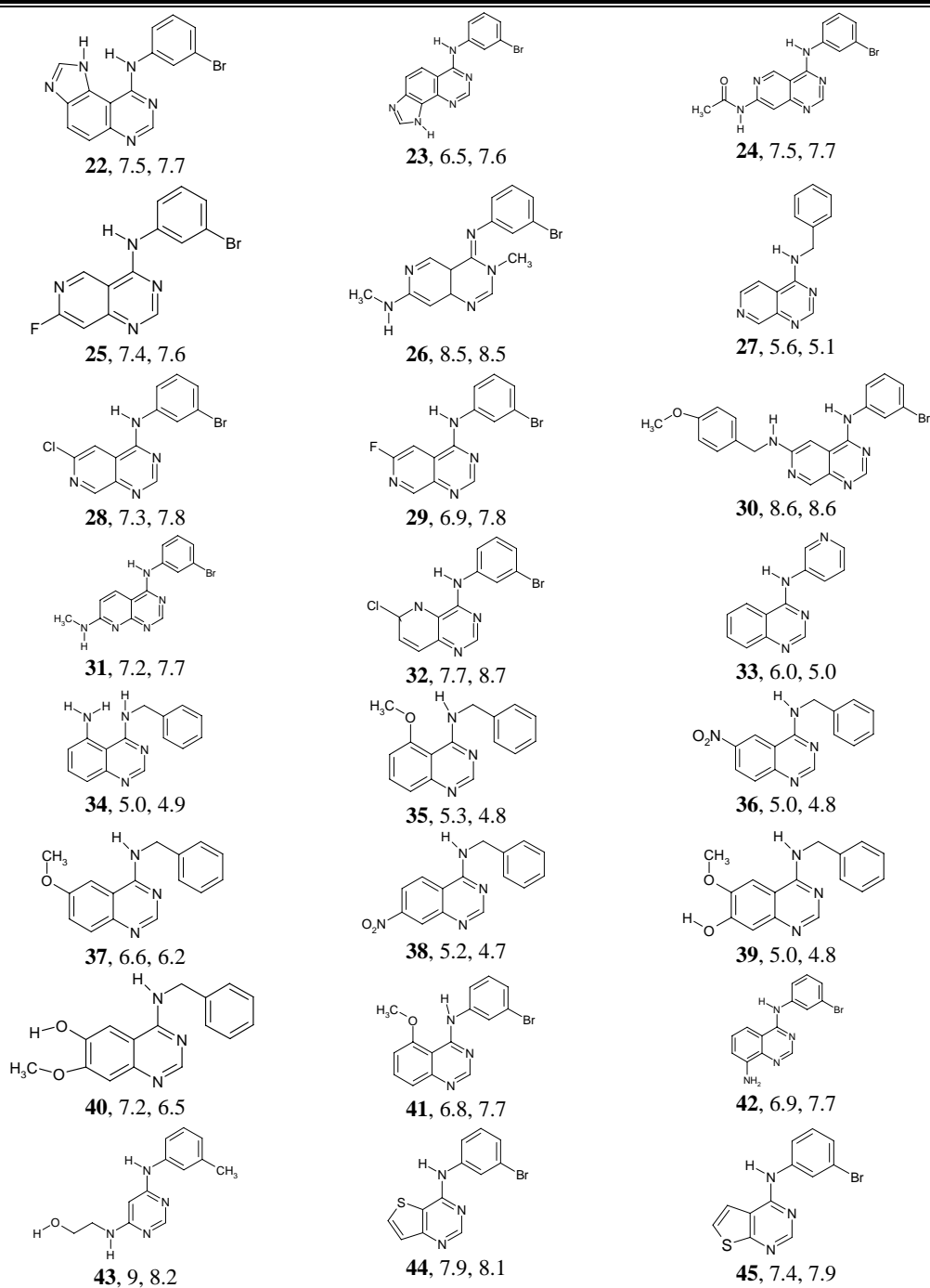
**Table 13:** (c) C–H···S interactions.

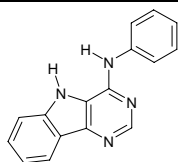
Type	Donor	Acceptor	Donor	Donor Atom	Residue ID	Acceptor	Acceptor Atom	Residue ID	$d$	$D$	$\theta$	PDB ID
C–H···S	B	L	Gly	CA	169	SLE	S	301	2.5	3.4	136.9	1ATL
C–H···S	B	L	Asn	CA	481	TP7	S7	9500	2.6	3.6	156.6	1MRO
C–H···S	L	S	HEC	C3B	601	Cys	SG	65	2.5	2.8	93.7	1QKS
C–H···S	L	S	F43	CHB	800	Met	SD	233	2.7	3.7	161.8	1MRO
C–H···S	L	S	STR	C19	1	Met	SD	759	2.7	3.5	130.7	1A28
C–H···S	L	S	AZE	C5	178	Met	SD	88	2.7	3.8	177.8	1FEN
C–H···S	L	S	MPD	C6	1	Cys	SG	37	2.8	3.6	138.2	1NCO
C–H···S	L	S	DES	CP4	800	Met	SD	343	2.8	3.7	142.0	3ERD
C–H···S	L	S	MPD	C3	1	Cys	SG	37	2.8	3.7	135.7	1NCO
C–H···S	L	S	CBO	C25	301	Met	SD	184	2.8	3.9	170.4	1HDC
C–H···S	L	S	AMP	C2	338	Met	SD	177	2.9	3.4	107.9	4FBP
C–H···S	L	S	HEC	CMB	601	Cys	SG	65	2.9	3.5	118.8	1QKS
C–H···S	L	S	HDS	C8	132	Met	SD	20	2.9	3.9	157.2	1LIC
C–H···S	L	S	IQB	C3	1	Met	SD	120	2.9	3.8	137.7	1YDT
C–H···S	S	L	Pro	CG	81	PY2	S	5	2.5	3.6	168.9	1IDA
C–H···S	S	L	Leu	CD2	198	BZO	S2	555	2.6	3.5	132.9	1A42
C–H···S	S	L	Val	CG2	523	S58	S1	701	2.7	3.5	139.1	1CX2
C–H···S	S	L	Phe	CZ	443	TP7	S7	9500	2.8	3.2	106.9	1MRO
C–H···S	S	L	Leu	CD2	198	ETS	S1	262	2.8	3.8	147.8	1CIL
C–H···S	S	L	Arg	CG	179	MNO	S1	601	2.8	3.8	154.2	1BMQ
C–H···S	S	L	His	CE1	35	TK4	S1	1001	2.8	3.5	121.8	1C5C
C–H···S	S	L	Val	CG2	121	AZM	S2	264	2.8	3.7	140.5	2H4N
C–H···S	S	L	Val	CG2	121	BZO	S2	555	2.9	3.5	117.3	1A42
C–H···S	L	S	UNK	CD	188	Cys	SG	243	2.9	3.8	144.4	2YHX

## APPENDIX II

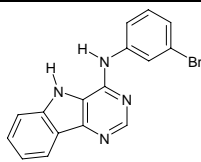
**Table 1:** Test set molecules for quantitative (*HypoGen*) model. Molecule serial number (in bold) followed by actual and predicted pIC<sub>50</sub> values.

 <b>1</b> , 9.6, 8.8	 <b>2</b> , 9.0, 8.8	 <b>3</b> , 8.5, 8.8
 <b>4</b> , 8.6, 8.7	 <b>5</b> , 8.5, 8.9	 <b>6</b> , 8.3, 8.8
 <b>7</b> , 8.3, 7.9	 <b>8</b> , 8.6, 8.9	 <b>9</b> , 7.9, 7.3
 <b>10</b> , 7.0, 7.6	 <b>11</b> , 8.1, 8.7	 <b>12</b> , 8.0, 7.7
 <b>13</b> , 8.1, 7.7	 <b>14</b> , 7.0, 7.6	 <b>15</b> , 7.2, 7.8
 <b>16</b> , 8.0, 7.6	 <b>17</b> , 9.1, 9.0	 <b>18</b> , 7.9, 7.7
 <b>19</b> , 7.9, 7.8	 <b>20</b> , 8.4, 7.8	 <b>21</b> , 7.3, 7.5

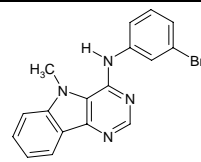




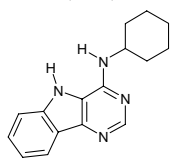
**46**, 5.6, 5



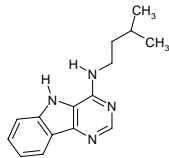
**47**, 7.1, 7.9



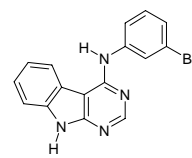
**48**, 6.8, 7.6



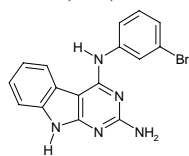
**49**, 4.0, 4.2



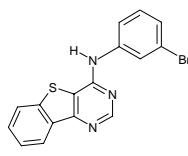
**50**, 4.0, 4.3



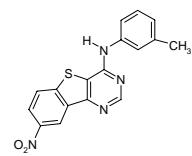
**51**, 7.5, 7.8



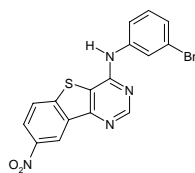
**52**, 6.8, 7.6



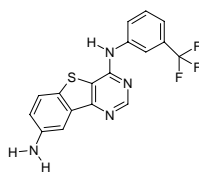
**53**, 8.7, 8.3



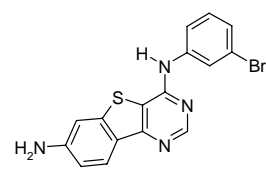
**54**, 7.9, 8.0



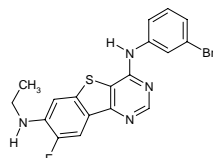
**55**, 7.9, 8.0



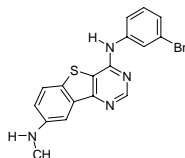
**56**, 7.3, 8.1



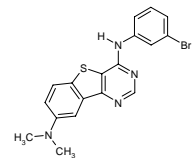
**57**, 9.3, 8.3



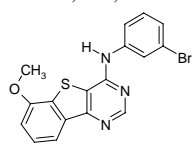
**58**, 8.1, 8.3



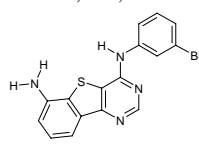
**59**, 8.9, 8.3



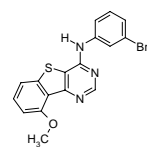
**60**, 7.7, 8.3



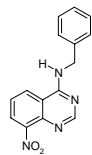
**61**, 7.8, 8.5



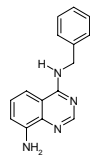
**62**, 8.5, 8.3



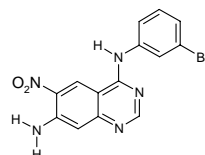
**63**, 7.6, 6.9



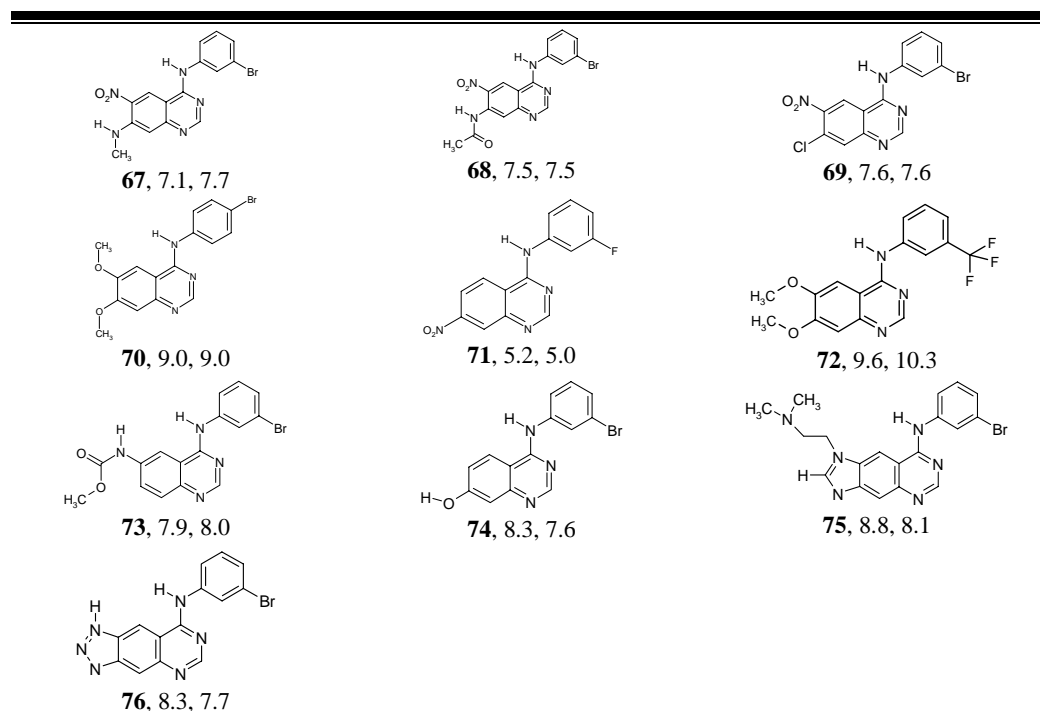
**64**, 5.0, 4.7



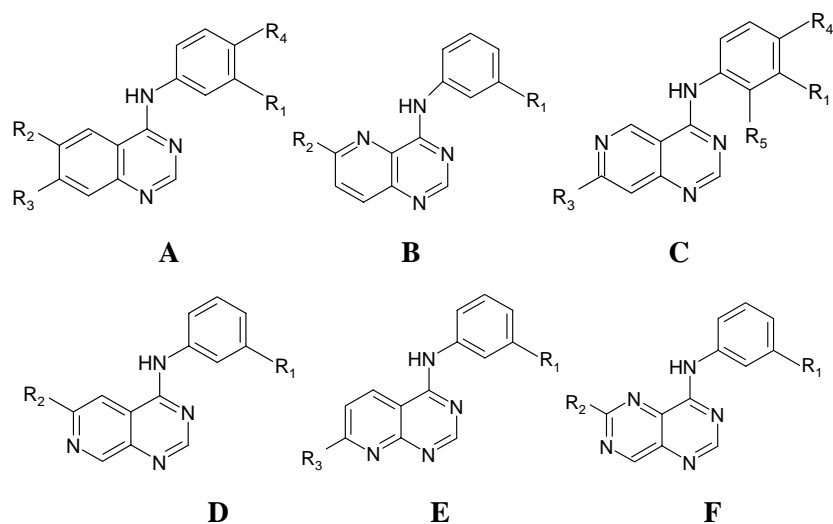
**65**, 5.0, 4.8



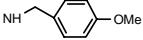
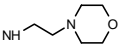
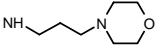
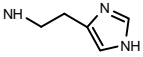
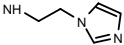
**66**, 7.2, 7.7

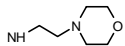
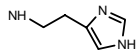
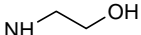
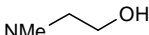
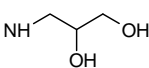
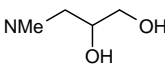

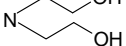
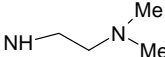
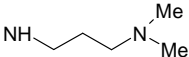
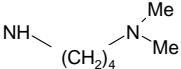
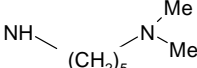
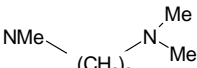
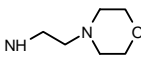
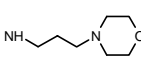
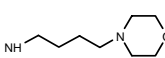
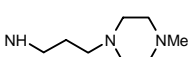
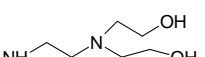
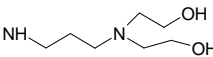
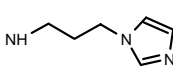


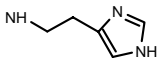
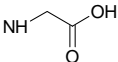
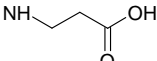
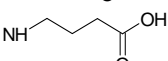
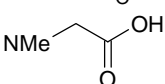
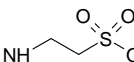
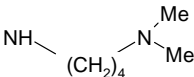
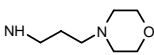
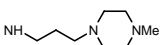
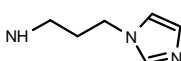
**Table 2:** List of 128 EGFR kinase inhibitors from earlier study. False positives are marked in red while negatives are in blue.



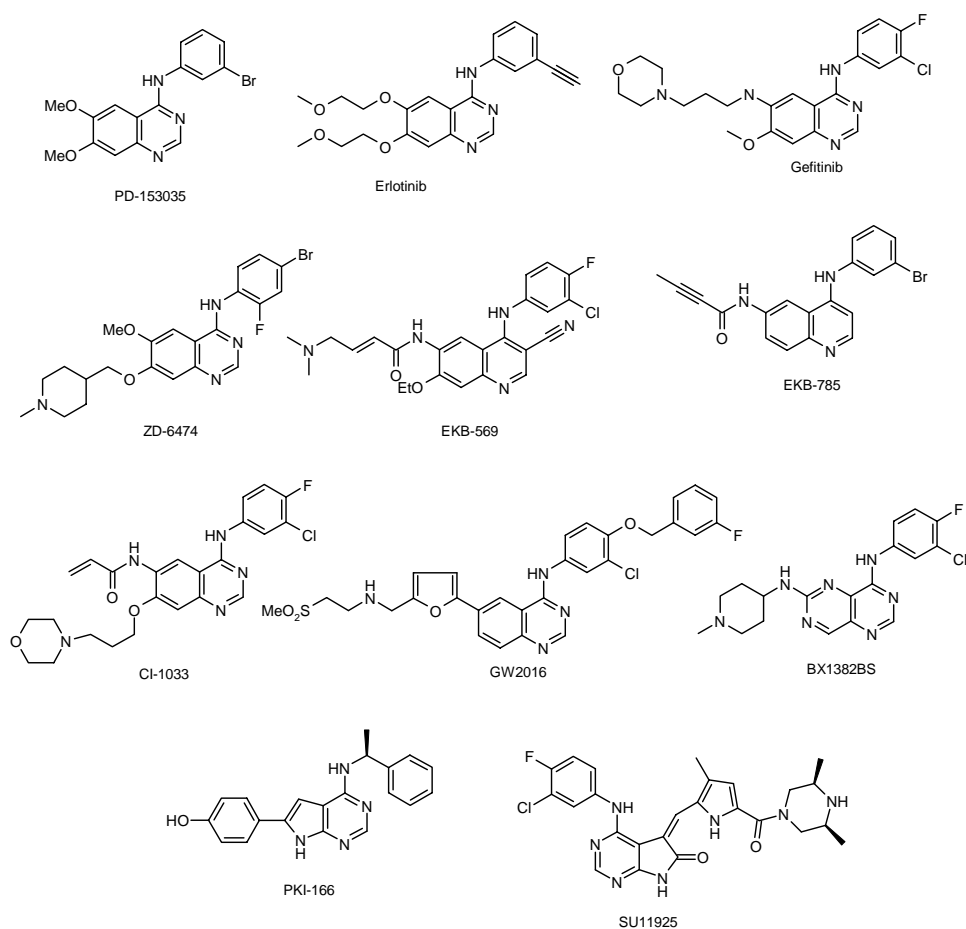
Sl. No.	Class	Substitution					Activity pIC <sub>50</sub>
		R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	
1	A	-	-	-	-	-	6.46
2	A	Me	-	-	-	-	6.04
3	A	Cl	-	-	-	-	7.63
4	A	Br	-	-	-	-	7.56
5	A	I	-	-	-	-	7.09
6	A	CF <sub>3</sub>	-	-	-	-	6.23
7	A	Br	NO <sub>2</sub>	-	-	-	6.04
8	A	Br	OMe	-	-	-	6.45
9	A	Br	-	NO <sub>2</sub>	-	-	6.0
10	A	Br	-	OMe	-	-	8.0
11	A	Br	OH	OH	-	-	9.76
12	A	Br	NH <sub>2</sub>	NH <sub>2</sub>	-	-	9.92
13	A	F	-	-	-	-	7.25
14	A	-	OMe	-	-	-	7.25
15	A	-	NH <sub>2</sub>	-	-	-	6.11
16	A	CF <sub>3</sub>	NH <sub>2</sub>	-	-	-	6.24
17	A	-	OMe	-	-	-	6.92
18	A	-	-	NH <sub>2</sub>	-	-	7.0
19	A	CF <sub>3</sub>	-	NH <sub>2</sub>	-	-	8.48
20	A	F	-	NO <sub>2</sub>	-	-	5.21
21	A	Cl	-	NO <sub>2</sub>	-	-	6.09
22	A	I	-	NO <sub>2</sub>	-	-	6.26
23	A	-	OMe	OMe	-	-	7.53
24	A	F	OMe	OMe	-	-	8.42
25	A	Cl	OMe	OMe	-	-	9.5
26	A	I	OMe	OMe	-	-	9.05
27	A	CF <sub>3</sub>	OMe	OMe	-	-	9.61
28	A	Br	NHMe	-	-	-	8.39
29	A	Br	NMe <sub>2</sub>	-	-	-	7.07
30	A	Br	NHCOOMe	-	-	-	7.92
31	A	Br	-	OH	-	-	8.32
32	A	Br	-	NHAc	-	-	7.39
33	A	Br	-	NHMe	-	-	8.15
34	A	Br	-	NHEt	-	-	7.92
35	A	Br	-	NMe <sub>2</sub>	-	-	7.95
36	A	Br	NH <sub>2</sub>	NHMe	-	-	9.16
37	A	Br	NH <sub>2</sub>	NMe <sub>2</sub>	-	-	6.79
38	A	Br	NH <sub>2</sub>	OMe	-	-	8.42
39	A	Br	NH <sub>2</sub>	Cl	-	-	8.18
40	A	Br	NO <sub>2</sub>	NHMe	-	-	7.16
41	A	Br	NO <sub>2</sub>	OMe	-	-	7.82
42	A	Br	NO <sub>2</sub>	Cl	-	-	7.6
43	A	Br	OEt	OEt	-	-	11.22
44	A	Br	O-n-Pr	O-n-Pr	-	-	9.76
45	A	H	OMe	OMe	Br	-	10.14
46	B	Br	-	-	-	-	7.46
47	B	Br	NH <sub>2</sub>	-	-	-	8.11
48	B	Br	Cl	-	-	-	7.74
49	B	Br	F	-	-	-	7.35
50	B	Br	NHMe	-	-	-	8.5

51	B	Br	NMe <sub>2</sub>	-	-	-	8.01
52	B	Br	OMe	-	-	-	8.36
53	C	Br	-	-	-	-	7.45
54	C	Br	-	NHAc	-	-	7.53
55	C	Br	-	F	-	-	7.88
56	C	Br	-	OMe	-	-	7.40
57	C	-	-	NH <sub>2</sub>	-	-	6.60
58	C	NO <sub>2</sub>	-	NH <sub>2</sub>	-	-	7.39
59	C	-	-	NH <sub>2</sub>	-	Br	6.61
60	C	Br	-	NH <sub>2</sub>	-	-	8.00
61	C	-	-	NH <sub>2</sub>	Br	-	7.59
62	C	-	-	NH <sub>2</sub>	CF <sub>3</sub>	-	5.32
63	C	-	-	NH <sub>2</sub>	-	OMe	5.43
64	C	OMe	-	NH <sub>2</sub>	-	-	6.88
65	C	-	-	NH <sub>2</sub>	OMe	-	6.17
66	C	-	-	NH <sub>2</sub>	-	NH <sub>2</sub>	5.27
67	C	NMe <sub>2</sub>	-	NH <sub>2</sub>	-	-	5.74
68	C	-	-	NH <sub>2</sub>	NMe <sub>2</sub>	-	5.31
69	C	F	-	NH <sub>2</sub>	-	-	6.07
70	C	Cl	-	NH <sub>2</sub>	-	-	6.92
71	C	OH	-	NH <sub>2</sub>	-	-	7.15
72	C	Me	-	NH <sub>2</sub>	-	-	7.39
73	D	Br	-	-	-	-	7.29
74	D	Br	Cl	-	-	-	7.39
75	D	Br	F	-	-	-	6.9
76	D	Br	OMe	-	-	-	8.58
77	D	Br		-	-	-	8.63
78	E	Br	-	-	-	-	6.16
79	E	Br	-	NH <sub>2</sub>	-	-	6.02
80	E	Br	-	F	-	-	6.16
81	E	Br	-	NHMe	-	-	7.28
82	E	Br	-	NMe <sub>2</sub>	-	-	6.48
83	E	Br	-	OMe	-	-	6.58
84	F	H	NHMe	-	-	-	7.88
85	F	Br	Cl	-	-	-	7.08
86	F	Br	NH <sub>2</sub>	-	-	-	8.82
87	F	Br	NHMe	-	-	-	9.11
88	F	Br	NMe <sub>2</sub>	-	-	-	9.02
89	F	Br	OMe	-	-	-	8.42
90	F	Br		-	-	-	9.09
91	F	Br		-	-	-	8.53
92	F	Br		-	-	-	9.6
93	F	Br		-	-	-	8.63
94	F	Me	Cl	-	-	-	6.42
95	F	Me	NH <sub>2</sub>	-	-	-	7.76
96	F	Me	NHMe	-	-	-	8.36

97	F	Me	NMe <sub>2</sub>	-	-	-	8.39
98	F	Me		-	-	-	8.63
99	F	Me		-	-	-	8.52
100	C	Br	-		-	-	9.61
101	C	Br	-		-	-	8.58
102	C	Br	-		-	-	9.03
103	C	Br	-		-	-	8.49
104	C	Br	-		-	-	7.85
105	C	Br	-		-	-	7.92
106	C	Br	-		-	-	7.34
107	C	Br	-		-	-	8.05
108	C	Br	-		-	-	8.13
109	C	Br	-		-	-	8.07
110	C	Br	-		-	-	7.39
111	C	Br	-		-	-	8.49
112	C	Br	-		-	-	8.72
113	C	Br	-		-	-	8.26
114	C	Br	-		-	-	8.30
115	C	Br	-		-	-	8.03
116	C	Br	-		-	-	8.92
117	C	Br	-	NHNH <sub>2</sub>	-	-	8.14
118	C	Br	-		-	-	9.29

<b>119</b>	C	Br	-		-	-	9.04
<b>120</b>	C	Br	-		-	-	8.82
<b>121</b>	C	Br	-		-	-	9.21
<b>122</b>	C	Br	-		-	-	9.55
<b>123</b>	C	Br	-		-	-	7.79
<b>124</b>	C	Br	-		-	-	8.85
<b>125</b>	C	Me	-		-	-	8.26
<b>126</b>	C	Me	-		-	-	8.03
<b>127</b>	C	Me	-		-	-	8.25
<b>128</b>	C	Me	-		-	-	8.45

**Scheme 1:** Some inhibitors and their trade name which are in clinical trials. PD-153035, Erlotinib (Tarceva, CP-358774, OSI-774), Gefitinib (Iressa, ZD-1839), ZD-6474, EKB-569, EKB-785 (CL-387785), Canertinib (CI-1033, PD-183805), GW-2016 (GW-572016), BX-1382BS are all based on scaffold **V**, while PKI-166 and SU11925 have a five membered ring (rather than a six membered ring) fused to the 4-aminopyrimidine.



**Scheme 2.** Equations used in quality check in *HipHop* and *HypoGen*.

Percent yield of actives:  $\%Y = H_a / H_t \times 100$

Percent ratio of actives:  $\%A = H_a / A \times 100$

$$\text{Enrichment (enhancement)} \quad E = \frac{H_a / H_t}{A / D} = \frac{H_a \times D}{H_t \times A}$$

False negative =  $A - H_a$

False positive =  $H_t - H_a$

$$\text{Goodness of fit} = \left( \frac{H_a (3A + H_t)}{4 H_t A} \right) \times \left( 1 - \frac{H_t - H_a}{D - A} \right)$$

Where

$H_a$  = Number of active hits

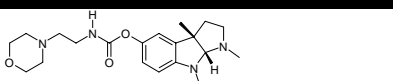
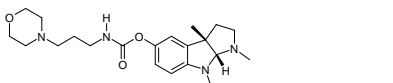
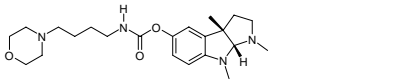
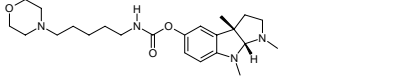
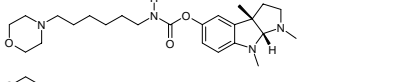
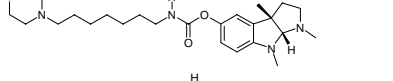
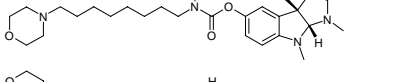
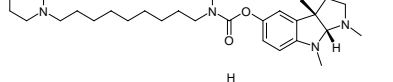
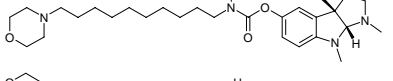
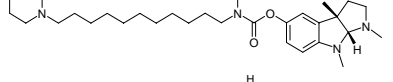
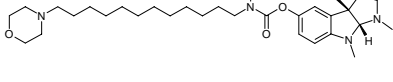
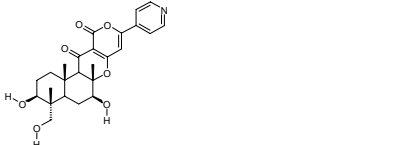
$H_t$  = Total number of hits

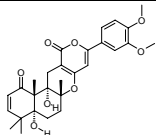
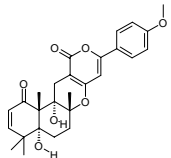
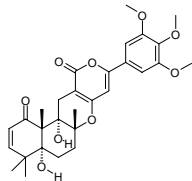
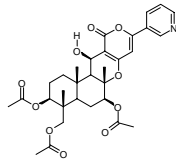
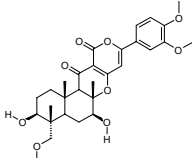
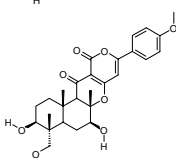
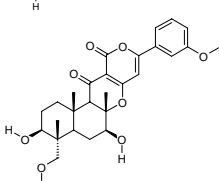
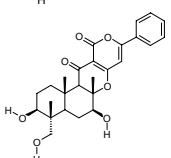
$D$  = Total number of molecules present in the database

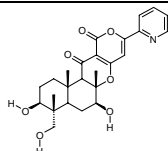
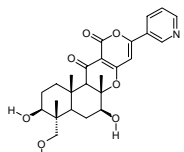
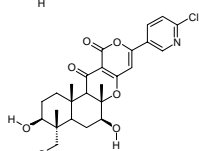
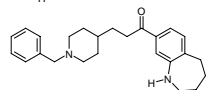
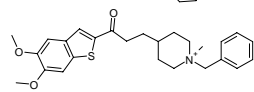
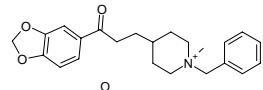
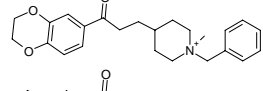
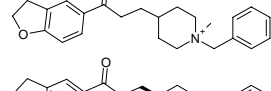
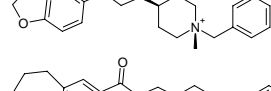
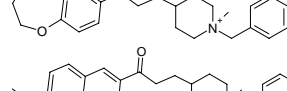
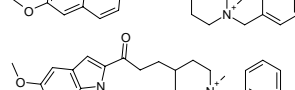
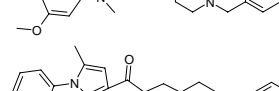

$A$  = Total number of active molecules present

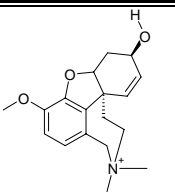
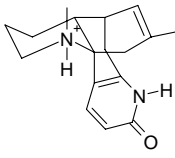
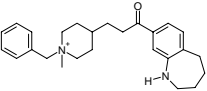
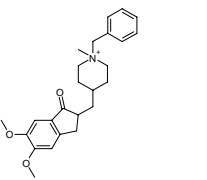
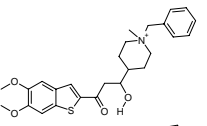
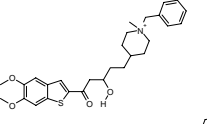
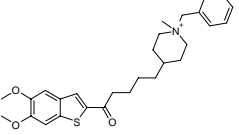
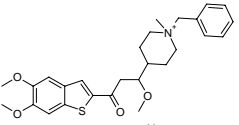
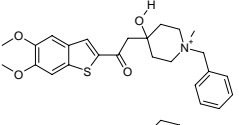
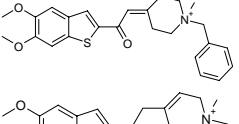
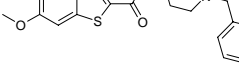
## APPENDIX III

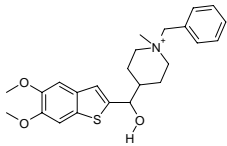
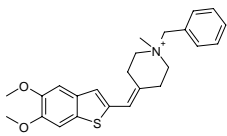
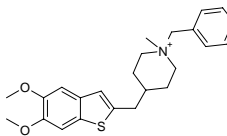
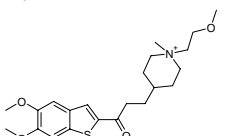
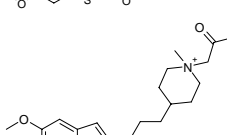
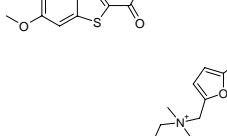
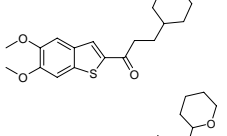
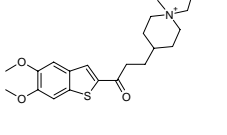
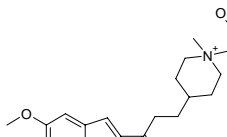
**Table 1:** List of compounds used in the study. The experimental and *HypoGen* predicted activities are listed along with the docking scores.

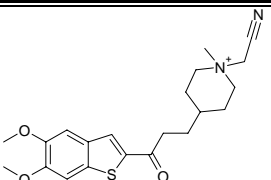
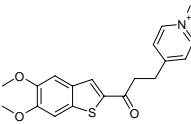
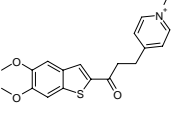
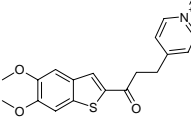
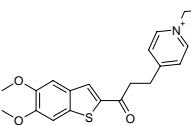
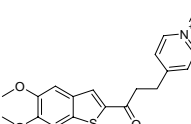
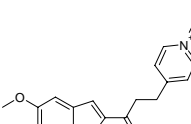
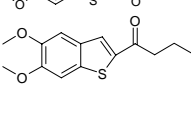
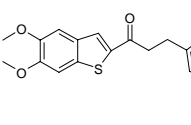
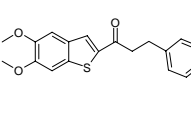
Molecule No.	Structure	Original compound from reference	Reference	Experimental (pIC <sub>50</sub> )	Predicted (pIC <sub>50</sub> )	Chemscore
1		7a	[8.14]	6.2	6.2	27.5
2		7b	[8.14]	5.4	7.5	24.6
3		7c	[8.14]	4.7	6.0	28.8
4		7d	[8.14]	5.1	6.0	33.0
5		7e	[8.14]	6.0	7.7	29.2
6		7f	[8.14]	6.4	7.5	28.5
7		7g	[8.14]	7.1	7.6	32.3
8		7h	[8.14]	7.2	6.3	28.7
9		7i	[8.14]	7.4	6.6	25.9
10		7l	[8.14]	7.4	7.4	23.1
11		7m	[8.14]	7.2	7.7	23.0
12		13a	[8.15]	5.1	4.5	13.6

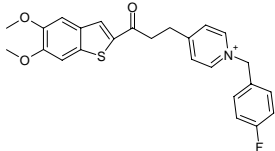
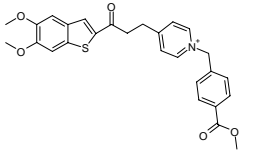
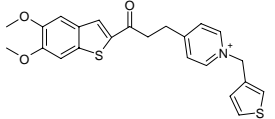
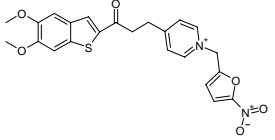
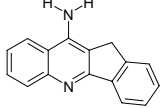
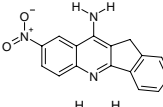
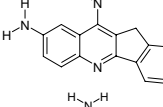
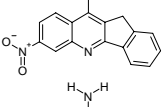
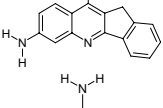
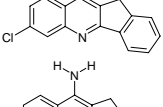
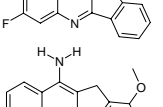
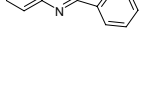
13		1 (arisugasin A)	[8.15]	9.0	8.0	34.9
14		2 (arisugasin B)	[8.15]	7.6	6.4	31.5
15		3 (Territrem B)	[8.15]	8.1	7.3	31.8
16		4 (pyripyropene)	[8.15]	4.0	4.5	9.7
17		6a	[8.15]	4.2	5.0	12.0
18		7a	[8.15]	4.2	4.5	13.0
19		8a	[8.15]	4.0	5.8	9.3
20		9a	[8.15]	4.0	4.5	11.3

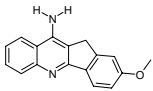
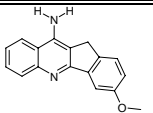
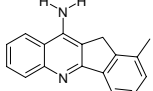
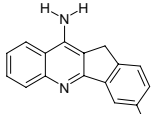
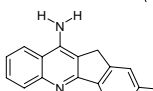
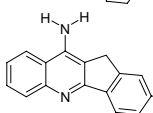
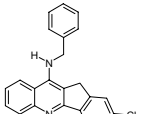
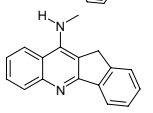
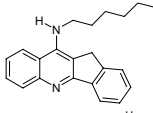
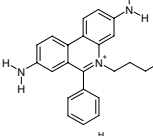
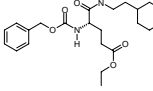
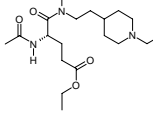
21		10a	[8.15]	4.0	4.5	16.0
22		11a	[8.15]	4.5	4.5	13.0
23		12a	[8.15]	4.1	4.5	12.0
24		TAK 147	[8.16]	6.3	5.5	31.8
25		8	[8.16]	8.1	8.1	36.5
26		9	[8.16]	6.6	6.5	31.4
27		10	[8.16]	7.0	6.4	27.8
28		11	[8.16]	7.8	7.8	28.3
29		12	[8.16]	7.2	6.2	27.8
30		13	[8.16]	6.9	5.7	29.6
31		14	[8.16]	7.3	7.2	36.3
32		15	[8.16]	6.9	6.3	29.9
33		16	[8.16]	7.7	6.1	26.0

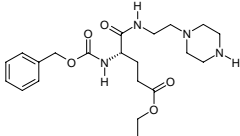
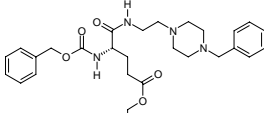
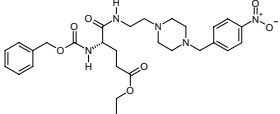
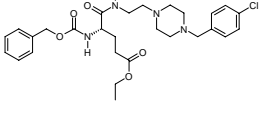
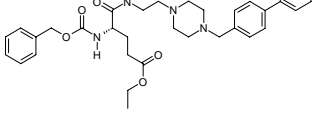
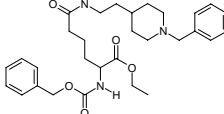
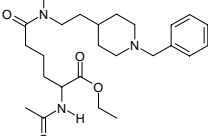
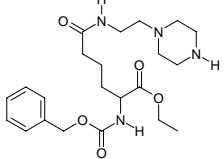
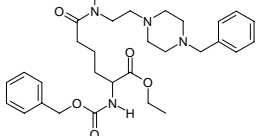
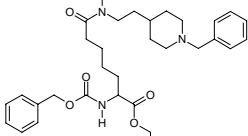
34		Gallanthamine	[8.16]	7.3	6.2	33.1
35		Huperzine	[8.16]	4.0	4.5	11.3
36		TAK 147	[8.16]	7.3	7.8	31.1
37		E2020	[8.16]	7.6	7.0	25.3
38		6	[8.17]	7.4	8.2	26.1
39		7	[8.17]	6.4	7.8	33.6
40		8	[8.17]	7.0	7.5	35.1
41		10	[8.17]	6.7	7.2	30.3
42		14	[8.17]	7.0	7.2	24.4
43		17	[8.17]	6.3	6.7	26.4
44		18	[8.17]	6.1	7.0	32.2

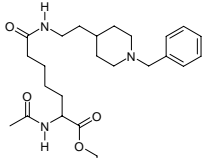
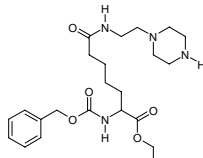
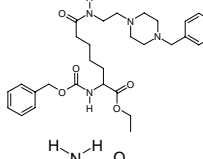
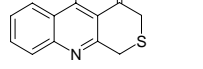
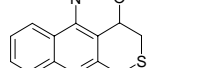
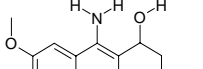
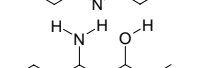
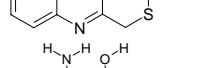
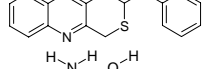
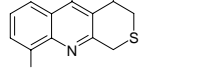
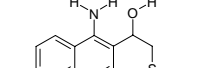
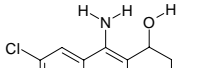
45		21	[8.17]	4.7	6.3	24.2
						
46		23	[8.17]	5.6	6.3	27.9
						
47		25	[8.17]	5.0	6.3	25.8
						
48		26	[8.17]	7.3	8.1	22.7
						
49		27	[8.17]	7.3	8.2	27.7
						
50		28	[8.17]	7.5	8.1	27.2
						
51		29	[8.17]	7.6	8.1	26.8
						
52		30	[8.17]	7.1	8.0	27.4
						
53		31	[8.17]	6.4	8.2	26.4
						

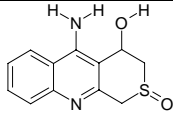
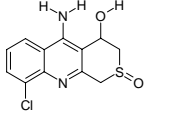
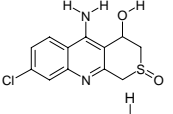
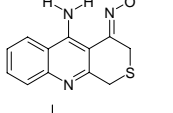
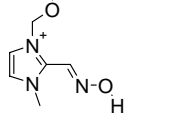
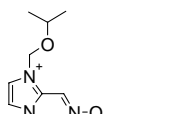
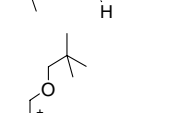
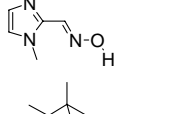
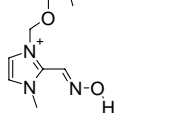
54		32	[8.17]	6.0	7.1	29.8
						
55		36	[8.17]	6.0	8.2	28.3
56		37	[8.17]	6.6	7.2	28.7
57		38	[8.17]	6.3	8.2	28.7
58		39	[8.17]	5.6	6.2	34.0
59		40	[8.17]	7.0	7.9	30.9
60		41	[8.17]	7.4	8.2	30.7
61		42	[8.17]	7.5	8.2	35.1
62		43	[8.17]	8.2	7.9	26.7
63		44	[8.17]	8.3	8.2	30.0

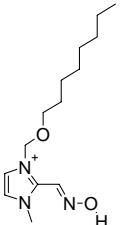
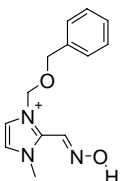
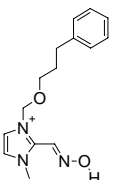
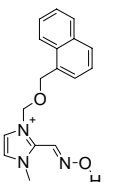
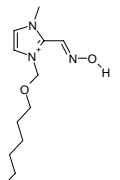
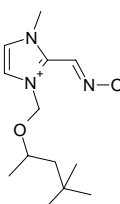
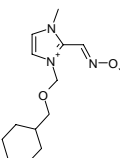
64		45	[8.17]	8.6	8.1	32.9
65		46	[8.17]	6.0	7.2	32.0
66		47	[8.17]	8.2	8.2	24.6
67		48	[8.17]	8.3	8.2	26.6
68		1a	[8.18]	6.2	4.3	26.4
69		2a	[8.18]	4.0	4.3	28.8
70		2b	[8.18]	5.2	4.3	26.9
71		2c	[8.18]	4.2	4.3	25.1
72		2d	[8.18]	4.5	4.3	23.9
73		2e	[8.18]	5.2	6.1	32.4
74		2f	[8.18]	5.9	4.3	24.8
75		2g	[8.18]	5.8	4.3	24.7

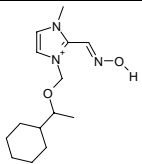
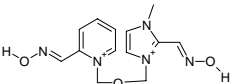
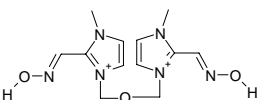
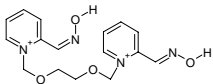
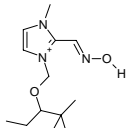
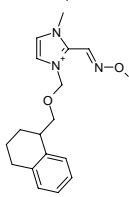
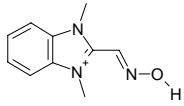
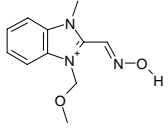
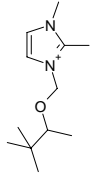
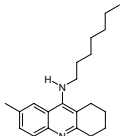
76		2h	[8.18]	5.2	4.3	24.7
77		2i	[8.18]	5.4	5.2	24.7
78		2j	[8.18]	5.4	4.3	24.5
79		2k	[8.18]	5.3	5.7	26.9
80		2l	[8.18]	6.4	6.3	25.7
81		2m	[8.18]	5.3	4.3	24.9
82		2n	[8.18]	5.1	4.3	27.1
83		2o	[8.18]	5.9	4.3	29.9
84		2p	[8.18]	5.4	5.4	28.4
85		propidium	[8.18]	7.1	5.1	29.6
86		1	[8.19]	5.1	6.0	15.7
87		2	[8.19]	4.1	5.4	18.6

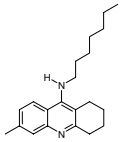
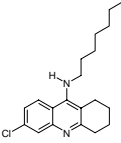
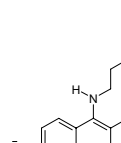
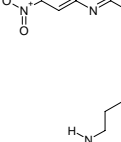
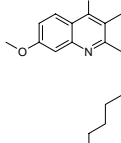
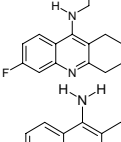
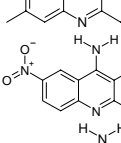
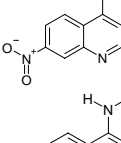
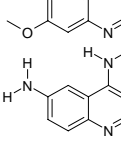
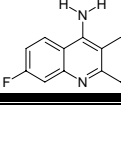

88		3	[8.19]	4.0	5.4	23.0
89		4	[8.19]	4.1	4.8	12.6
90		5	[8.19]	4.0	5.7	20.8
91		6	[8.19]	4.0	5.4	12.9
92		7	[8.19]	4.0	5.9	15.0
93		8	[8.19]	6.7	6.9	13.5
94		9	[8.19]	6.3	6.8	23.3
95		10	[8.19]	4.0	5.5	20.5
96		11	[8.19]	5.6	5.3	16.8
97		12	[8.19]	6.2	5.1	23.7

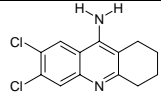
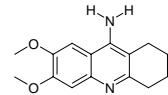
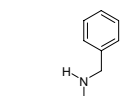
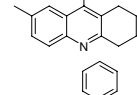
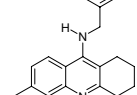
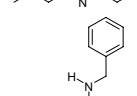
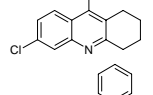
98		13	[8.19]	5.8	5.7	25.4
99		14	[8.19]	4.0	6.2	19.7
100		15	[8.19]	5.4	6.0	12.2
101		4ag	[8.20]	5.6	4.7	18.9
102		5ag	[8.20]	5.5	5.0	25.3
103		5fg	[8.20]	4.0	5.0	25.8
104		5ah	[8.20]	4.0	5.1	22.6
105		5aiI	[8.20]	4.0	5.2	24.1
106		5bg	[8.20]	4.0	5.0	18.6
107		5cg	[8.20]	6.3	6.1	25.4
108		5dg	[8.20]	4.0	5.0	24.8
109		5eg	[8.20]	6.5	5.6	23.8

110		6ag	[8.20]	5.6	4.3	24.5
111		6bgI	[8.20]	4.0	4.3	21.6
112		6cgI	[8.20]	4.3	6.1	23.2
113		7ag	[8.20]	4.0	4.4	21.2
114		2a	[8.21]	3.5	3.6	20.3
115		2b	[8.21]	4.1	4.5	12.8
116		2c	[8.21]	4.9	5.0	16.4
117		2d	[8.21]	4.8	4.7	17.0
118		2e	[8.21]	4.7	5.2	17.6

119		2f	[8.21]	5.2	4.9	15.4
120		2g	[8.21]	4.3	5.1	16.1
121		2h	[8.21]	4.8	4.6	17.0
122		2i	[8.21]	5.2	5.0	16.7
123		2d	[8.22]	4.6	5.2	14.7
124		2j	[8.22]	4.6	4.7	18.2
125		2w	[8.22]	4.4	4.9	17.1

126		2x	[8.22]	4.6	4.8	16.9
127		5	[8.22]	4.4	5.0	16.6
128		6	[8.22]	5.1	5.0	10.0
129		7	[8.22]	3.5	3.5	15.0
130		2i	[8.22]	4.6	4.5	15.1
131		2ab	[8.22]	4.3	4.5	17.9
132		3a	[8.22]	3.9	3.3	17.1
133		3b	[8.22]	3.7	3.4	15.8
134		4	[8.22]	4.5	4.1	15.4
135		17	[8.23]	6.4	6.7	18.9

136		18	[8.23]	6.9	6.1	30.7
137		19	[8.23]	7.9	7.3	32.5
138		20	[8.23]	6.5	6.5	32.3
139		21	[8.23]	6.3	6.3	23.1
140		22	[8.23]	7.3	6.5	29.8
141		2	[8.23]	7.0	7.0	28.7
142		5	[8.23]	5.5	4.4	27.0
143		6	[8.23]	7.6	7.3	21.5
144		7	[8.23]	6.5	7.0	20.4
145		8	[8.23]	5.4	4.4	23.4
146		9	[8.23]	7.1	7.4	27.4

147		10	[8.23]	6.3	6.1	25.1
148		11	[8.23]	5.3	6.1	23.1
149		12	[8.23]	5.4	4.4	30.0
150		13	[8.23]	6.1	6.0	29.1
151		14	[8.23]	6.8	6.1	30.5
152		15	[8.23]	5.8	4.4	20.6
153		16	[8.23]	5.3	4.4	23.2

## **ABOUT THE AUTHOR**

Sunil Kumar Panigrahi, son of Jogeswar Panigrahi and Kamala, was born in Kantamal, a village in the Boudh district of Orissa, India, in 1977. He received his primary education in Kantamal U. G. M. E. school. He completed his secondary school and senior school certificate education in Jawahar Navodaya Vidyalaya, Phulbani (now Kandhamal) and Sambalpur respectively. Later he completed his B. Sc. from Gangadhar Meher College, Sambalpur. After the completion of his M. Sc. from Sambalpur University, he joined the School of Chemistry, University of Hyderabad to pursue the Ph. D. degree in 2002. He was awarded research fellowship by Council of Scientific and Industrial Research (JRF and SRF) for 2002-2007.

## LIST OF PUBLICATIONS

- [1] Aparna, V.; **Panigrahi, S. K.**; Rambabu, G.; Gopalakrishnan, B.; Sarma, J. A. R. P.; Desiraju G. R. Design of EGFR kinase inhibitors: A ligand-based approach and its confirmation with structure-based studies. *Bioorg. Med. Chem.* **2003**, 11, 4643-4653.
- [2] **Panigrahi, S. K.**; Desiraju, G. R. Homology modelling in protein structure prediction: Epidermal Growth Factor Receptor kinase domain. *Nat. Acad. Sci. Lett. (India)* **2004**, 27, 1-11.
- [3] Aparna, V.; Rambabu, G.; **Panigrahi, S. K.**; Sarma, J. A. R. P. ; Desiraju G. R. Virtual screening of 4-anilinoquinazoline analogues as EGFR kinase inhibitors: Importance of hydrogen bonds in the evaluation of poses and scoring functions. *J. Chem. Inf. Model.* **2005**, 45, 725-738.
- [4] **Panigrahi, S. K.**; Desiraju, G. R. Strong and weak hydrogen bonds in the protein–ligand interface. *Proteins* **2007**, ASAP, DOI 10.1002/prot.21253.
- [5] Tiwari, A.; **Panigrahi, S. K.**; Desiraju, G. R. *HBAT*: A complete package for analysing strong and weak hydrogen bonds in macromolecular crystal structures. (*communicated*)
- [6] **Panigrahi, S. K.**; Desiraju, G. R. Strong and weak hydrogen bonds in protein–ligand complexes of kinases: A comparative study. (*communicated*)
- [7] **Panigrahi, S. K.**; Desiraju, G. R. Strong and weak hydrogen bonds in drug–DNA complexes: A statistical analysis. (*communicated*)
- [8] **Panigrahi, S. K.**; Desiraju, G. R. Pharmacophore modeling on EGFR kinase inhibitors: A novel strategy for ligand based virtual screening. (*in preparation*)
- [9] **Panigrahi, S. K.**; Desiraju, G. R. Pharmacophore modeling, docking and virtual screening of Acetylcholinesterase inhibitors. (*in preparation*)