

**COMPUTER AIDED DRUG DESIGN STUDY OF CYCLIN
DEPENDENT KINASE 2 (CDK2) INHIBITORS**

**SİKLİN BAĞIMLI KİNAZ 2 (CDK2) İNHİBİTÖRLERİNİN
BİLGİSAYAR DESTEKLİ İLAÇ TASARIMI ÇALIŞMASI**

ABDULİLAH ECE

Submitted to Institute of Graduate Studies in Science and Engineering
of Hacettepe University as a partial fulfillment
to the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

2011

Institute of Graduate Studies in Science and Engineering,

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of DOCTOR OF PHILOSOPHY in CHEMISTRY.

Head :
Prof. Dr. Mustafa GÜLLÜ

Advisor :
Prof. Dr. Fatma SEVİN DÜZ

Member :
Prof. Dr. Nazan TUNOĞLU

Member :
Prof. Dr. İlkey YILDIZ

Member :
Prof. Dr. Pervin CİVCİR

APPROVAL

This thesis has been certified as a thesis for the Degree of DOCTOR OF PHILOSOPHY by the above Examining Committee Members on/...../ 2011.

...../...../ 2011.

Prof. Dr. Adil DENİZLİ
Director / Institute of Graduate Studies
in Science and Engineering
Natural and Applied Sciences

This dissertation is dedicated to my family, to whom I owe so much, in every sense of the word.

COMPUTER AIDED DRUG DESIGN STUDY OF CYCLIN DEPENDENT KINASE 2 (CDK2) INHIBITORS

Abdulilah Ece

ABSTRACT

Cyclin-dependent kinases (CDKs) plays important role in the regulation of the cell division cycle. Excessive production of CDKs, or insufficient production of cyclin dependent kinase inhibitors (CKIs), can lead to a disruption of cell cycle and finally lead to cancer. Efforts have been made in discovering small molecule inhibitors of CDK2 for the purpose of restoring the cell cycle.

In the first part of this dissertation, a Quantitative Structure Activity Relationship (QSAR) study was performed so that to correlate the physicochemical properties of some pyrimidine series of CDK2 inhibitors to the biological activity of the compounds by means of a classical Hansch analysis that would lead to a meaningful model of QSAR in the pyrimidine class.

The second part focuses on identifying a number of new hit compounds with potent inhibitory activity. 3D pharmacophore models were developed based on the known inhibitors. An optimal pharmacophore model was produced and validated using external test set and Fischer's randomization method. This model was used as a 3D query for virtual screening to retrieve potential inhibitors from Life Chemicals and NCI2003 databases. The hit compounds were then subjected to molecular docking studies and finally, 11 compounds were obtained based on careful observations, analyses, comparisons and consensus scoring function.

These models give insight in designing more potent analogues against CDK2 prior to undertaking any further research including synthesis.

Keywords: CDK2 inhibitor, QSAR, 3D QSAR pharmacophore, Virtual screening, Antitumor agent, Cancer.

Advisor: Prof. Dr. Fatma SEVİN DÜZ, Hacettepe University, Department of Chemistry, Organic Chemistry Division.

SİKLİN BAĞIMLI KİNAZ 2 (CDK2) İNHİBİTÖRLERİNİN BİLGİSAYAR DESTEKLİ İLAÇ TASARIMI ÇALIŞMASI

Abdulilah Ece

ÖZ

Siklin bağlı kinazlar (CDK) hücre bölünmesi döngüsünün düzenlenmesinde önemli rol almaktadır. CDK'ların fazla üretimi veya siklin bağlı kinaz inhibitörlerinin (CDKI) yetersiz üretimi hücre döngüsünün bozulmasına ve sonuç olarak kansere yol açabilmektedir. Hücre döngüsünün tekrar onarımı için küçük CDK2 inhibitörlerinin bulunması için bir çok çalışmalar yapılmıştır.

Tezin ilk bölümünde, CDK2 inhibitörü olan bazı pirimidin serilerinin, biyolojik aktiviteleri ile fizyokimyasal özellikleri arasında ilişki kurmak için, klasik Hansch analizinden yararlanılarak, pirimidin sınıfı için Nitel Yapı Aktivite İlişkisi (QSAR) çalışması yapılmış ve QSAR modeli oluşturulmuştur.

İkinci bölümde, güçlü inhibitör aktivitesine sahip birtakım yeni başarılı bileşiklerin belirlenmesine çalışılmıştır. Bilinen inhibitörler baz alınarak üç boyutlu (3D) farmakofor modelleri geliştirilmiştir. İdeal bir farmakofor modeli oluşturulmuş ve test seti ve Fischer gelişigüzel testi ile doğrulanmıştır. Bu model, sanal taramada Life Chemicals ve NCI 2003 databanklarından potansiyel inhibitörleri bulmak için 3D sorgusu olarak kullanılmıştır. Bulunan bileşiklere moleküler kenetleme (docking) çalışmaları uygulanmış ve dikkatli gözlemler, analizler, karşılaştırmalar ve konsensus skor fonksiyonları baz alınarak 11 bileşik elde edilmiştir.

Bu bileşiklerin, ilerde sentez de dahil yapılabilecek çalışmalar için CDK2'ye karşı daha güçlü analogların tasarımında büyük katkı sağlayacağı düşünülmektedir.

Anahtar Kelimeler: CDK2 inhibitorü, QSAR, 3D QSAR farmakofor, Sanal tarama, Antitümör ajan, Kanser.

Danışman: Prof. Dr. Fatma SEVİN DÜZ, Hacettepe Üniversitesi, Kimya Bölümü, Organik Kimya Anabilim Dalı.

ACKNOWLEDGEMENTS

It is a pleasure to thank those who made this dissertation possible. Without the support, patience and guidance of the following people and institutes this thesis would not have been possible.

I would like to gratefully and sincerely thank Prof. Dr. Fatma SEVİN DÜZ for her mentorship, guidance, understanding and patience during my graduate studies at Hacettepe University. She encouraged me to not only grow as a chemist but also as an instructor and an independent thinker.

I owe my deepest gratitude to Prof. Dr. İlkey YILDIZ especially for her guideness in QSAR part.

I would also like to thank all of the members of the Sevin research group.

I also thank The Scientific and Technological Research Council of Turkey (TÜBİTAK) (Project No: 107 T 068).

I am greatly appreciative of especially ICS UNIDO, then SISSA and the support of IAEA and UNESCO for their support through training courses and conferences in Italy.

Finally, and most importantly, I would like to thank my family for their support, encouragement, quiet patience and their faith in me.

CONTENTS

ABSTRACT	i
ÖZ	ii
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	vi
LIST OF TABLES	viii
ABBREVIATIONS	ix
1. INTRODUCTION.....	1
2. GENERAL INFORMATION	4
2.1. Drug Design: A Conceptual Perspective	4
2.2. Rational Drug Design	7
2.2.1. Factors Contributing to Drug Discovery	9
2.2.2. Drug Design Theory	10
2.2.3. Role of Computers in Drug Design: Their Success and Failure	12
2.3. An Introduction to Molecular Modeling and Computer-Aided Drug Design	14
2.3.1. Molecular Modeling	14
2.4. Introduction to QSAR	15
2.4.1. Some Basic Principles	15
2.4.2. Free-Wilson Analysis	18
2.4.3. Hansch Analysis	19
2.5. Pharmacophores and Pharmacophore Searches	22
2.5.1. Pharmacophores: Historical Perspective and Viewpoint from a Medicinal Chemist.....	22
2.5.2. Definitions	22
2.5.3. Pharmacophores: the Viewpoint of a Medicinal Chemist	23
2.5.4. Criteria for a Satisfactory Pharmacophore Model	23
2.5.5. Methods Using Pharmacophore Features and Geometric Constraints	25
2.6. Docking Techniques.....	31
2.6.1. Protein Structure	32
2.6.2. Rigid Docking.....	32
2.6.3. Docking with Flexible Ligands.....	32
2.6.4. Ensembles of Ligand Conformations	33
2.7. Cell Cycle and Loss of Cell Cycle Control	33

3. WORKING PLAN	38
4. MATERIALS AND METHODS.....	39
4.1. QSAR Study.....	39
4.2. Pharmacophore Study	42
4.3. Databases Screening.....	42
4.4. Docking Study.....	44
5. RESULTS AND DISCUSSION	46
5.1. QSAR.....	46
5.1.1. Validation of the QSAR model	46
5.2. PHARMACOPHORE STUDY.....	54
5.2.1. Hypothesis Generation	54
5.2.2. Validation of Hypothesis	55
5.2.3. Databases screening	64
5.2.3. Molecular Docking Studies	65
6. CONCLUSION	69
7. REFERENCES.....	72
APPENDIX	80
CURRICULUM VITAE	88

LIST OF FIGURES

Figure 1.1.	Schematic representation of CDK2.....	1
Figure 2.1.	Drug-like molecules and druggable targets.....	6
Figure 2.2.	Enzyme-Substrate complimentary interactions	10
Figure 2.3.	Pharmacophore and receptor binding	12
Figure 2.4.	Schematic presentation of the parent structure according to the original Free–Wilson formalism (a) and in the Fujita–Ban variant of Free–Wilson analysis (b).....	19
Figure 2.5.	Hypothetical case of drug–receptor interactions.	20
Figure 2.6.	Model of the cell cycle and the cyclin/cdk complexes.....	34
Figure 2.7.	Schematic diagram of CDK2/Cyclin A	36
Figure 4.1.	Chemically diverse 16 compounds used as training set in 3D QSAR Discovery Studio/Pharmacophore generation.....	43
Figure 4.2.	Chemically diverse 14 compounds used as test set in 3D QSAR Discovery Studio/Pharmacophore generation.....	44
Figure 4.3.	Binding site in CDK2 structure	45
Figure 5.1.	Comparison of observed and predicted activities of the compounds in the training set using Eq. 1 (a) and Eq. 2 (b).....	48
Figure 5.2.	a) Experimental crystal structure of 2j bound to Thr160pCDK2/cyclin A, overall fold. (Marchetti et al., 2007) b) Calculated binding mode of 2j bound to Thr160pCDK2/cyclin A.....	51
Figure 5.3.	Comparison of observed and predicted activities of the compounds in the test set using Eq. 1 (a) and Eq. 2 (b).....	53
Figure 5.4.	The ligand-based pharmacophore models of CDK2 inhibitors reported previously by (a) Hecker, (b) Toba and (c) Vadivelan.	54
Figure 5.5.	Catalyst Hypogen pharmacophore model	58
Figure 5.6.	Best pharmacophore model Hypo1 aligned to training set compound A) Active molecule 4 (IC50 13.1 nM) B) Inactive molecule 14 (IC50 79,000 nM).....	58
Figure 5.7.	The difference in costs between HYPOGEN runs and the scrambled runs.....	60

Figure 5.8. Fisher's randomization test results. Correlation values for a particular pharmacophore across the tests correspond to a line.....	61
Figure 5.9. Graph showing the correlation (r) between experimental and predicted activities for the 14 test set molecules against the Hypo-1 model along with 16 training set molecules for CDK2 inhibitors.....	64
Figure 5.10. Binding orientation of the hit compounds NSC_649153 (A) and F5882-6930 (B) and mapping of these compounds onto the pharmacophore model Hypo1.	68
Figure A.1. Molecular structures of 11 hit compounds together with their activity values in nM.	80

LIST OF TABLES

Table 1.1.	Structures of the studied 4-cyclohexylmethoxypyrimidines used in QSAR Study.	3
Table 4.1.	Structural properties, values of the selected descriptors and biological activity (pIC50) of the training and test set 4-cyclohexylmethoxypyrimidines.	40
Table 5.1.	Observed and predicted pIC50 values of training set compounds obtained from the Eqs.(6,7).	50
Table 5.2.	Observed and predicted pIC50 values of test set compounds obtained from the Eqs.(6-7).....	52
Table 5.3.	Information of statistical significance and predictive power presented in cost values measured in bits for the top 10 hypotheses as a result of automated 3D QSAR pharmacophore generation.	57
Table 5.4.	Actual and estimated activity of the training set molecules based on the pharmacophore model Hypo1.	59
Table 5.5.	Experimental and predicted IC50 data values of 14 test set molecules against Hypo1.	63
Table 5.6.	Virtual Screening Results.....	65
Table 5.7.	Selected hit compounds and binding modes.....	67
Table A.1.	Critical values of F at 95% confidence level calculated with the Excel function.....	81
Table A.2.	Score values and molecular properties of 11 hit compounds and their 10 poses.....	82

ABBREVIATIONS

3D	Three Dimensional
ADME	Adsorption/Distribution/Metabolism/Excretion
CADD	Computer Aided Drug Design
CDK	Cyclin Dependent Kinase
CDKI	Cyclin Dependent Kinase Inhibitor
DE	Dielectric Energy
DLM	Drug Like Molecule
DS	Discovery Studio
F	Fisher F-Test of Significance
FDA	Food and Drug Administration
H	Hydrophobic
HBA	Hydrogen Bond Acceptor
HBD	Hydrogen Bond Donor
HTS	High Throughput Screening
IC50	Half Maximal Inhibitory Concentration
IUPAC	International Union of Pure and Applied Chemistry
LOO	Leave-One-Out
PDB	Protein Data Bank
PLS	Partial Least Squares
PRESS	Predictive Sum Of Squares
Q ²	Cross-Validated R ²
QSAR	Quantitative Structure-Activity Relationship
R ²	Squared Predicted Correlation Coefficient
RMSD	Root Mean Square Deviation
s	Standard Error of Prediction
SPRESS	Standard Deviation Based on PRESS

1. INTRODUCTION

The cell cycle control involves a variety of proteins called cyclins, and enzymes called cyclin-dependent kinases (CDKS) (Figure 1.1). Binding of a cyclin with its accompanying kinase is essential for the activation of the cell from one phase of the cell cycle to another. For example, CDK2/cyclin A is required for progression through the S phase. However, excessive production of CDKs, or insufficient production of cyclin dependent kinase inhibitors (CKIs), can lead to a disruption of the normal regulation controls and lead to cancer. Efforts have been made for the purpose of restoring the cancer cell cycle and these include CDK inhibition, down-regulation of cyclins, up regulation of CDK inhibitors, degradation of cyclins, or inhibition of tyrosine kinases that initiate the cell cycle activation in the first place (Patrick, 2005).

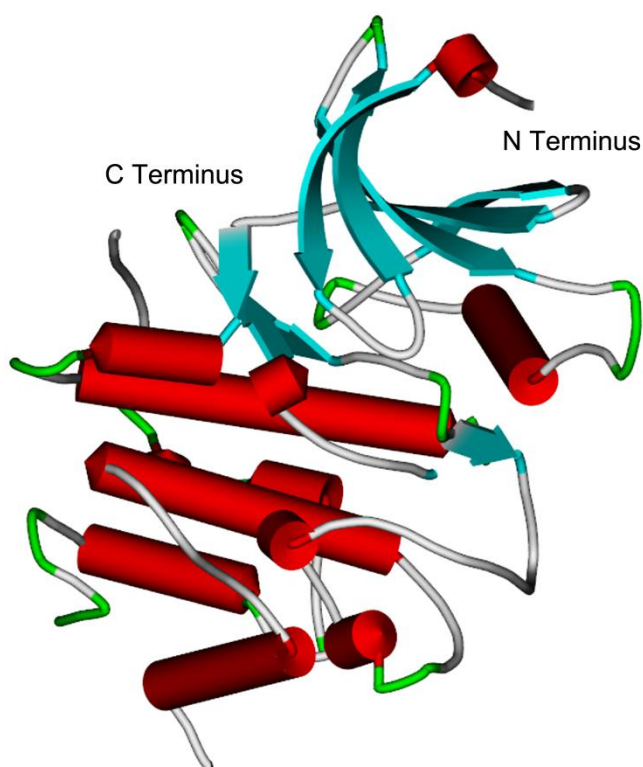


Figure 1.1. Schematic representation of CDK2 (PDB access code: 1OGU).

The developments of potent and selective small-molecule inhibitors of CDK2 have been reported and all these inhibitors are competitive with ATP. However, in order a drug to reach a market requires much effort. Because of the discovery of the long and expensive drug discovery cycle and the fundamentals of demand and supply, the drug discovery process is undergoing an extensive renovation.

Combinatorial chemistry and mass screening acquired wide acceptance. Researchers are now try to discover lead compounds in a fast and efficient manner through these techniques. Among these techniques are QSAR and pharmacophore studies.

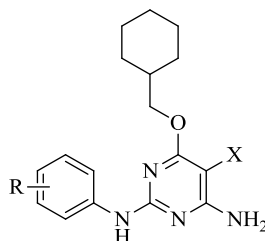
A successful QSAR model correlates physicochemical properties of a series of compounds to its biological activity. It gives information on structural requirements for designing more potent analogues.

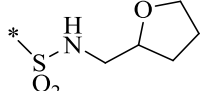
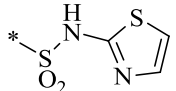
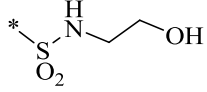
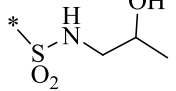
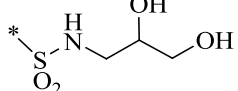
Providing that the experimentally determined high resolution 3D structure of the target is available, ligand based drug design can be performed. For instance a 3D QSAR pharmacophore study in association with molecular docking, a structure-based method, and underlying scoring functions helps to reproduce crystallographic ligand-binding modes. These methods can be combined to identify a number of new hit compounds with potent inhibitory activity and to understand the main interactions at the binding sites.

This study, as an extension of these considerations, aims to explore computer aided drug design of a series of 4-cyclohexylmethoxypyrimidines as CDK2 inhibitors (Table 1.1). Using these pyrimidine series, inhibitors of CDK2, QSAR was studied (Ece and Sevin, 2010). These CDK2 inhibitors are all competitive with ATP binding site and some of them were reported to be among the most potent and selective CDK2 inhibitors reported to date (Marchetti et al., 2007).

In addition to QSAR, a 3D QSAR pharmacophore study was investigated using a diverse set of CDK2 inhibitors along with their activities ranging over several orders to generate a good pharmacophore model. After validation of the optimum pharmacophore model, a virtual screening was performed using this model as a query in order to find out potent drug candidates through commercial databases, Life Chemicals and NCI2003. Molecular docking and consensus scoring function was used to eliminate false positive and false negative errors.

Table 1.1. Structures of the studied 4-cyclohexylmethoxypyrimidines used in QSAR Study.



Compound	R	X	Compound	R	X
1 ^{*a}	—	NO	3d ^{*b}	4-CONH ₂	H
2a ^b	4-OMe	NO	2j ^b	4-CONH ₂	NO
3a	3-Br	H	2k [*]	4-CON(CH ₃) ₂	NO
2b ^b	3-Br	NO	2l	4-CON(C ₂ H ₅) ₂	NO
2c	4-OH	NO	3e ^b	4-SO ₂ NH ₂	H
2d	3-OMe	NO	2m ^b	4-SO ₂ NH ₂	NO
3b	3-SMe	H	2n	4-SO ₂ N(C ₂ H ₅) ₂	NO
2e [*]	3-SMe	NO	2o	4- 	NO
2f ^b	4-SMe	NO	2p	4- 	NO
2g [*]	3-CH ₂ CN	NO	2q	4- 	NO
3c	4-CH ₂ CN	H	2r	4- 	NO
2h	4-CH ₂ CN	NO	2s	4- 	NO
2i	3-CH ₂ OH	NO			

*Test set compounds. ^a (Arris et al., 2000) ^b (Sayle et al., 2003).

2. GENERAL INFORMATION

2.1. Drug Design: A Conceptual Perspective

There is no doubt that drug design, although is multi-step, multidisciplinary and multi-year, has attracted many scientists from different areas of research like medicine, pharmacology, chemistry and so on. However, one should keep in mind that drug discovery is not an inevitable outcome of fundamental basic science and yet drug design is not just a technology that produces drugs for human beings on the basis of technological advances. Then, it would be for sure that more and better drugs would already be available if it were that simple.

The smallest particle of a substance that holds the chemical identity of that substance and is composed of two or more atoms held together by chemical bonds is simply defined as a molecule. Considering the fact that molecules are highly variable in the sense of structure, they may be arranged into families on the basis of specific groupings of atoms called functional groups. A functional group, regardless of the molecule in which it is located, is a particular cluster of atoms that generally reacts in the same way, hence that are responsible for the characteristic chemical reactions. For instance, the acidic property is generally given to a molecule by the carboxylic acid functional group (-COOH) in which it is inserted. It is therefore, the presence of functional groups that establishes the chemical and physical properties of a given family of molecules. Thus, a functional group is a center of reactivity in a molecule.

A drug molecule owns one or more functional groups positioned in three-dimensional space on a structural framework that holds the functional groups in a defined geometrical array that allows the molecule to bind particularly to a targeted biological macromolecule that we called the receptor. Accordingly, the structure of the drug molecule permits a desired biological response that should be beneficial (by means of inhibiting pathological processes) and that ideally prevents binding to other untargeted receptors, as a result, minimizing the probability of toxicity. The framework which holds the functional groups is typically a hydrocarbon structure like an aromatic ring or an alkyl chain and this framework is usually chemically inert in order not to participate in the binding process. Another requirement for a structural framework is that it should be relatively rigid (conformationally

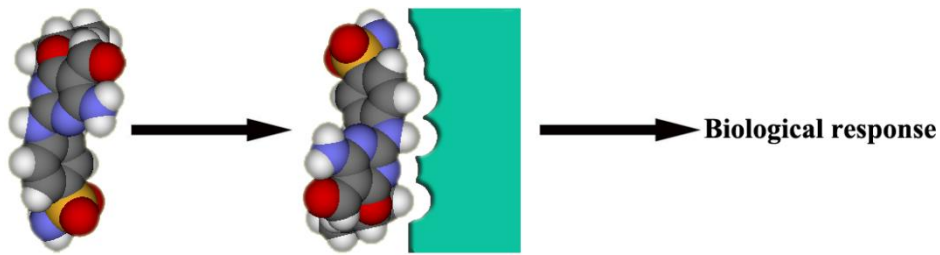
constrained). This ensures that the arrangement of functional groups is not flexible in its geometry, consequently preventing the drug from interacting with untargeted receptors by modifying its molecular shape. Nevertheless, to be successful in countering a disease process, a drug molecule must have additional properties further than the capacity to bind to a specified receptor site. It must be able to suffer the journey from its point of administration (i.e., the mouth for an orally administered drug) until it finally reaches the receptor site deep within the organism (i.e., the brain for a neurologically active drug).

A drug-like molecule (DLM) has the chemical and physical properties that will enable it to become a drug molecule should an appropriate receptor be identified (see Figure 2.1). Some specific properties allow a molecule to become a drug-like molecule and some specific properties allow a macromolecule to become a druggable target. A drug-like molecule becomes a drug molecule when it interacts with a druggable target to give a biological response and the druggable target becomes a receptor. A drug molecule becomes a useful drug molecule, when it is successfully and beneficially distributed to people with a disease.

Many scientists have asked this question: What are the properties that enable a molecule to become a drug-like molecule? Generally speaking, a molecule should be small enough to be transferred all the way through the body, hydrophilic enough to dissolve in the blood stream, and lipophilic enough to cross fat barriers within the body. Containing enough polar groups is another important requirement to enable it to bind to a receptor, but it must not have so many polar groups that it would result in being eliminated too quickly from the body through the urine to exert a therapeutic effect (Nogrody and Weaver, 2005).

Lipinski initiated the analysis of the structures of orally administered drugs, and of drug candidates (Lipinski et al., 1997). Lipinski's so called rule of five does a good job of quantifying these properties. This has so far been the primary guide to correlating physical properties with successful drug development. As stated by this rule, a drug-like molecule should have:

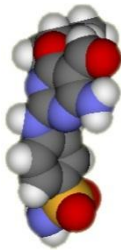
- A molecular weight no more than 500,
- A logP (logarithm of its octanol-water partition coefficient) value no more than 5,
- Not more than five hydrogen bonding donors,
- Not more than 10 hydrogen bonding acceptors.



“Drug Like Molecules”

Drugs are molecules but every molecule is not a drug!

Properties of drug-like molecules



- Low molecular weight,
- Not too hydrophilic,
- Not too lipophilic,
- Presence of functional groups.

“Druggable targets”

Receptors are macromolecules but every macromolecule is not receptor!



Properties of druggable targets

- They are usually protein,
- They lead to biological response,
- They do not cause toxicity.

Figure 2.1. Drug-like molecules and druggable targets.

In order to improve drug likeliness several improvements to Lipinski's rule of five has been reported. Ghose et al. (Ghose, 1998) reports the qualifying range together with an addition property as follows:

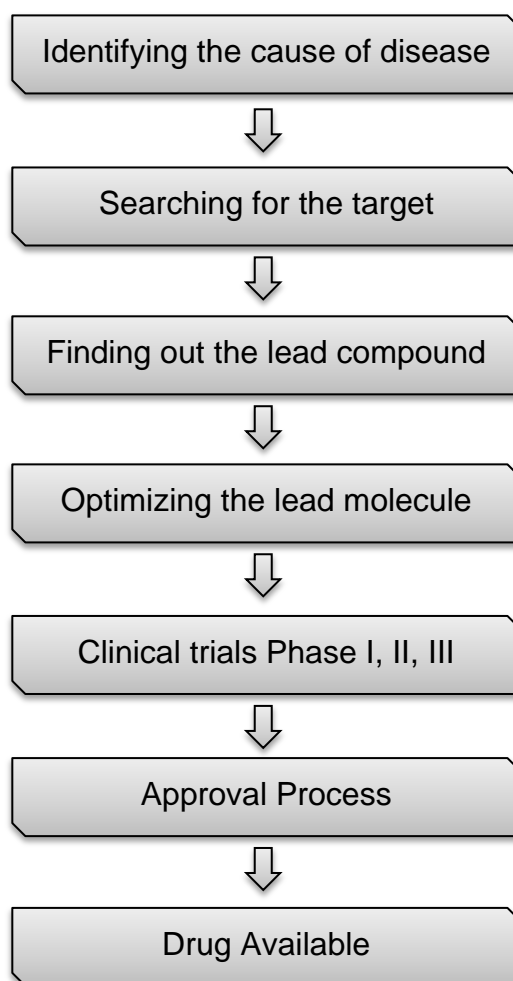
- Molecular weight between 160-480,
- log P is between -0.4 and 5.6,
- Molar refractivity between 40-130.

Veber et al. (Veber, 2002) suggest two other criteria:

- 10 or fewer rotatable bonds,
- Polar surface area equal to or less than 140 \AA^2 (or 12 or fewer H-bond donors and acceptors).

2.2. Rational Drug Design

The drug development process follows these classic steps:



Due to the fact that the traditional approach to drug discovery is really time-consuming and cost-intensive, the new approach to drug discovery goes beyond the limitations of the traditional research. It emerges that the target in the body and the potential active compound are directly related to each other.

Knowledge about the disease and previous infectious processes is necessary for designing a drug. The first step in rational drug design is to identify a molecular target that is crucial to a disease process or an infectious pathogen. The next important step which makes sense of the word rational is to determine the molecular structure of the target. The validity of rational or structure-based drug discovery depends largely on a high-resolution target structure of adequate molecule detail to permit selectivity in the screening of compounds.

German physician and immunologist Paul Ehrlich who developed the intellectual tools of medical science, such as receptors for drugs is the scientist to whom modern drug research owes its rationality. Another scientist Domagk discovered "prontosil" as an antibacterial agent. This antibacterial agent placed the foundation for the concepts of biochemical metabolites and steric analogues. This allowed medicinal chemists to employ chemical drug design. Therapeutically undesirable side effects of a lead compound forced scientists to modify the structure of the lead compound. Drug design was based on these modifications. Hansch (Hansch and Fujita, 1964) has helped medicinal chemists to target the design of drugs through quantitative structural activity relationship (QSAR) using Hammett's substituent constants (Fujita et al., 1964).

After a long process of studying about 5,000 to 10,000 compounds, only one drug arrives to the market. Each drug costs about \$156 million in the discovery phase. Food and Drug Administration (FDA) processes I, II, and III cost another \$75 million which brings the total to about \$231 million for each drug placed on the market for consumers. Then, for achieving FDA approval, a long and expensive procedure also needs to be followed (Foye, 1989).

2.2.1. Factors Contributing to Drug Discovery

Besides the long and expensive drug discovery cycle, there are several other factors that contribute to the rapidly changing landscape of the drug discovery environment:

1. Progress in molecular biology and high-throughput screening,
2. Demand fundamentals,
 - a. Aging population of the baby boomers,
 - b. Consumer demand for quality health care,
 - c. Expanded access and universal health care,
 - d. New innovation technologies,
 - e. Consumer awareness in the quality of supplements and nutrition,
3. Supplying fundamentals,
 - a. Downsizing hospitals,
 - b. Insurers reluctance to pay high reimbursements,
 - c. Transition to outpatient procedures
 - d. Disease management
 - e. Global management

Due to these factors-regulation, the cost-effectiveness of drug discovery, and the fundamentals of demand and supply-the drug discovery process is undergoing an extensive renovation. Companies that have been earning a great deal of money, thus making a fortune from the sale of drugs are expected to shift their focus to tap into information. These companies are now using intelligent software and are applying chemi information to shorten the cycle of drug discovery and therefore make the drug discovery process cost-effective (Foye, 1989; Leo et al., 1971; Wolff, 1995).

2.2.2. Drug Design Theory

Understanding the method by which the active site of a receptor selectively limits the binding of improper structures is the basic concept behind drug design. In medicinal chemistry, a *ligand* is any potential molecule that can bind to a receptor. A specific combination of atoms displaying the correct shape, size, and charge composition of the atoms is necessary for a ligand to bind and interact with a receptor. A presumed *ligand-receptor interaction* has complementary ligand-receptor shape and size, phenomenon called *steric complementarity* (Figure 2.2).

Besides the steric complementarity, electrostatic interactions also affect ligand binding by limiting the binding of improper molecules due to the fact that the ligand must contain correctly placed complementary charged atoms for interaction to occur. Nevertheless, hydrophobic interaction is the main driving force for receptor binding. Taking into account that two-thirds of our bodies are water, it is the hydrophobic nature of the ligand that provides the driving force to force the ligand to leave the water and bind to a receptor.

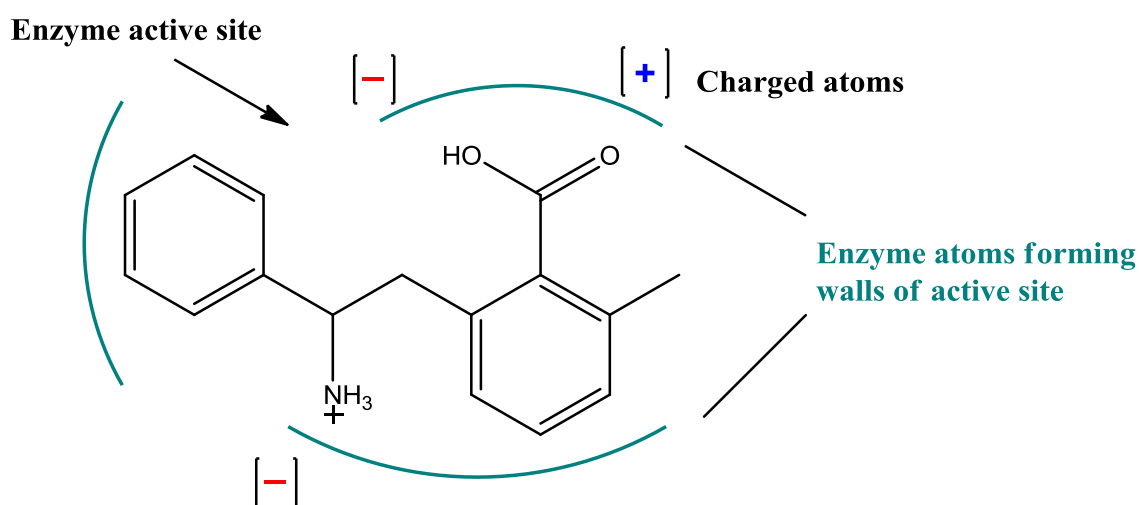


Figure 2.2. Enzyme-Substrate complimentary interactions

Among the numerous potential interactions which occur between ligands and receptors, the specific interactions that are critical for ligand recognition and binding by a receptor are called the pharmacophore.

By using a lock-and-key similarity, we can visualize a lock having numerous tumblers. Now, there may be many keys that can sterically complete the lock and

fit within the keyhole. All but the correct key will replace the wrong tumblers, yet leading to a suboptimal interaction with the lock. Only the correct key, which displays the pharmacophore to the receptor, connects the proper tumblers and interacts properly with the lock to open it. Any successful drug must incorporate the proper chemical structures and present the pharmacophore to the receptor. Hence, this is crucial to the design of pharmaceuticals (Figure 2.3).

Now to face the challenge of designing a drug that directs to a specific target receptor, the major interests are as follows:

1. Characterizing the medical condition and determining receptor targets,
2. Achieving active-site complementarity like hydrophobic, steric and electrostatic,
3. Considering biochemical mechanism for receptors,
4. Adhering to the laws of chemistry,
5. Achieving synthetic feasibility,
6. Addressing biologic considerations,
7. Developing patent considerations.

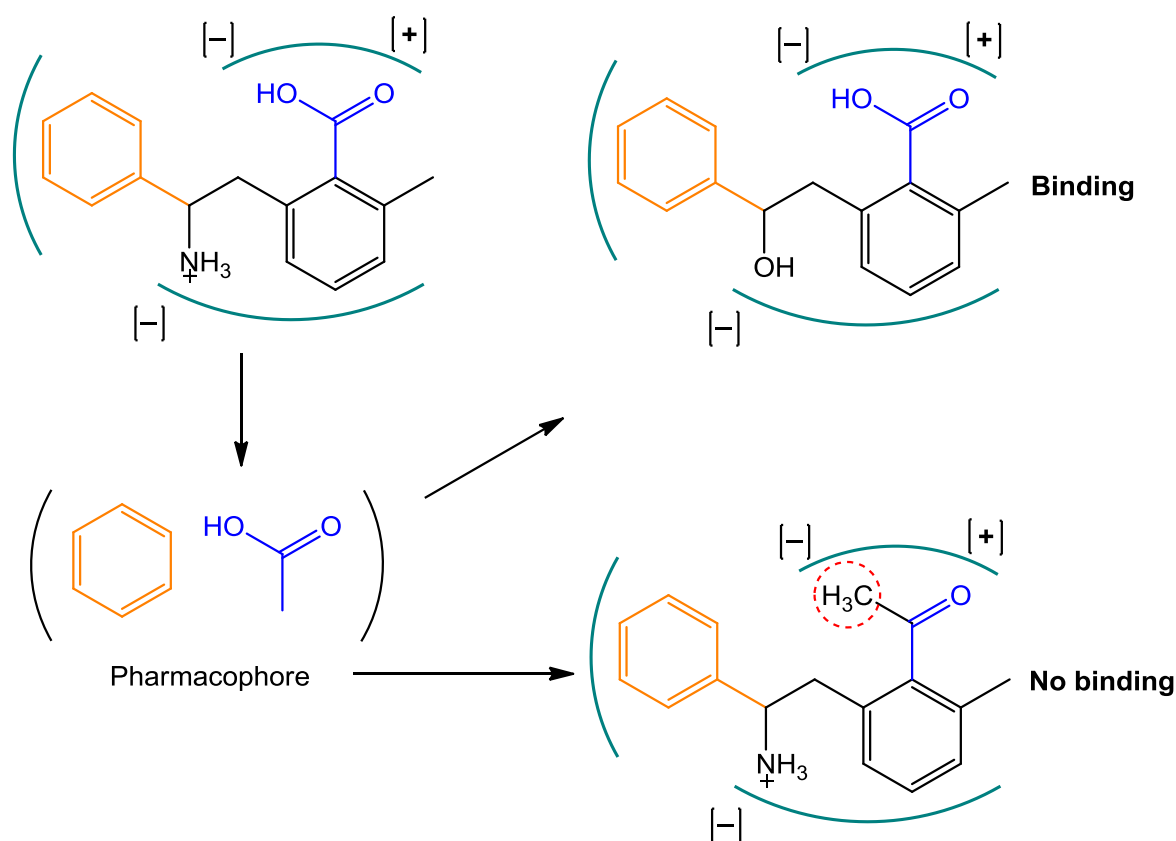


Figure 2.3. Pharmacophore and receptor binding

2.2.3. Role of Computers in Drug Design: Their Success and Failure

Computer aided drug design (CADD) was believed that it would revolutionize the way in which drugs are developed in the early 1990s and there was a great deal of optimism at that time (Nag, 2011). Although facing some difficulties, the continuing exponential increase in computing power proceeded to such a point that fundamental estimations of ligand-receptor complementarities could be performed. Scientists gained the ability to produce vector models of chemical structures and handle them in real time through computer graphics technology. By using computers, computational chemists believed that they could by-pass a considerable time and effort required for drug synthesis and testing simply by generating novel compounds with the help of computers. The concept of generating virtual lead compounds totally through the computer simulation was named de novo design.

The world's largest pharmaceutical companies spent millions of dollars on hardware and software in order to turn de novo design into a reality. Unfortunately,

apart from few cases, success was rare and de novo design proved to be a complete failure. De novo design could not prove itself to be an efficient method in discovering lead compounds. The main reasons behind this disappointment are a lack of feasible software functions and off course, limitations in computing power. Accuracy and processing time are of critical importance in scientific computing. Hence, algorithms, assumptions, approximations and other several shortcuts are required to make calculations run in a limited period of time. Later on, the calculated accuracy of any ligand-receptor interactions was greatly diminished. But again, despite the fact that chemists presumed numerous chemical structures that potentially could complement the active site based on computer simulations, the calculated binding had no correlation with reality.

This remains the most important challenge in de novo design. Even though computers have become exponentially faster, the sheer number of calculations needed to accurately predict the binding of a de novo-generated ligand to its receptor in a practical time frame still needs significant approximations. In de novo design, several attempts are being carried out to produce whole ligands from scratch and finally dock them within their receptors. The problem remains how the predicted structure acts in real life after all.

The second important problem in computer-aided de novo design is the generation of unwanted chemical structures that are useless.

In this circumstance, combinatorial chemistry and mass screening came to the fore and acquired wide acceptance. Researchers were now provided with an opportunity to discover lead compounds in a fast and efficient manner through these techniques. Whole problems associated with de novo design tools were no longer needed to generate lead structures. Now it became obvious that computational tools were indeed needed that could optimize these lead compounds into potent, useful and desired drugs.

The concept of drug optimization versus de novo design is now an important fact. In de novo ligand generation, a complete structure is created from scratch whereas in drug optimization, we begin with a lead compound whose bound

structure with the receptor has already been characterized by x-ray crystallography.

However, generation of chemical derivatives is highly influenced by computerized automation. By the help of computers, we can rapidly generate and estimate the binding of all potential derivatives by creating a list of the best potential candidates. Therefore, it's for sure that CADD software helps us to refine weakly binding lead compounds in the most effective manner (Smith et al., 1975; Hansch, 1971; Nag and Dey, 2011).

2.3. An Introduction to Molecular Modeling and Computer-Aided Drug Design

2.3.1. Molecular Modeling

Anything that is done through use of computers to describe the structure or properties of a molecule is called "molecular modeling".

The role of molecular modeling in drug design has been divided into two separate concepts:

- First one centered on the structure-activity problem. In order to design novel drugs, this paradigm seeks to justify biological activity in the absence of detailed, three-dimensional structural information about the receptor.
- The second one focused on understanding the interactions observed in receptor-ligand complexes and using the known three-dimensional structure of the therapeutic target for the purpose of designing novel drugs.

Advances in molecular biology to generate the target proteins in sufficient quantities for investigation, and the equally impressive gains in NMR (Salzmann et al., 1999; Hajduk et al., 1997; 1999; McDowell and Schaefer, 1996 and McDowell et al., 1999; Ishima and Torchia, 2000; Moore, 1999; Fejzo et al., 1999 and Dinsmore et al., 2001) and crystallography (Hajdu et al., 2000; Perrakis et al., 1999) to provide three-dimensional structures as well as identify leads, have lead a rapid increase in relevant structural information. These progressions field have stimulated the need for design tools and as a result, the molecular modeling community is rapidly developing useful approaches. However, the more common problem is one in which the receptor can only be deduced from pharmacological

studies and little, if any, structural information is available to guide through the modeling. Fortunately, useful information can be developed to guide the design and synthesis of potential novel therapeutics from an analysis of structure-activity data in the three-dimensional framework delivered by current molecular modeling techniques (Marshall and Beusen, 2003).

2.4. Introduction to QSAR

Compared with the rapid development of molecular modeling, structure-based design, and protein crystallography, traditional chemometric QSAR methods for the analysis of quantitative structure-activity relationships (QSARs) are occasionally considered to be out of fashion. Besides that, an equation is more difficult to understand than a visualized colored three-dimensional picture generated by computer graphics. Classical QSAR methods, however, still play an important role and will continue to be a practical tool in modern drug design (Martin et al., 1992; Kubinyi, 1993 and Böhm et al., 1996). Thousands of documented QSARs and success stories of QSAR predictions and QSAR-guided drug design clearly shows that they have contributed greatly in the advancement of science in medicinal chemistry attesting to their versatility (Franke and Gruska, 2003). The quantitative description of pharmacokinetic processes remains the domain of traditional QSAR techniques. This outlook and QSAR-based concepts such as “drug likeness” are achieving significant importance connected with high throughput screening (HTS) for hit to lead decisions so that to avoid the selection of compounds with unfavorable adsorption/distribution/metabolism/excretion (ADME) properties. The design of safe and selective compounds and a better understanding of toxic, carcinogenic, or mutagenic effects is another important issue.

2.4.1. Some Basic Principles

The equation below presented by Crum-Brown and Fraser in 1868 who assumed that biological activity is a function of chemical structure was probably the first general formulation of a quantitative structure–activity relationship:

$$F = f(C) \tag{1}$$

There was still a long way to go from this general formulation to the development of true QSARs. It was necessary to explain proper measures of F , appropriate mathematical formalisms for the function f , and methods to quantitatively describe chemical structure C . Modern QSAR technology appeared in 1964 starting with publications by Hansch and Fujita (Hansch and Fujita, 1964) and Free and Wilson (Free and Wilson, 1964). The first paper published led to development of the well-known and the most widely-used QSAR method Hansch analysis, also known as the extra thermodynamic or linear free-energy-related approach. The second publication resulted in development of the Free–Wilson analysis, which supplements Hansch analysis. The Free–Wilson analysis has turned out to be a very useful method for specific types of structural modifications. Both methods take advantage of multiple regression analysis as the mathematical method (f in Equation (1)) but differ in the description of chemical properties. Substituent constants and other physicochemical descriptors are used in Hansch analysis. But Free–Wilson analysis is based on chemical fragments which are directly derived from the two-dimensional structure of compounds.

A large variety of mathematical methods is ready for use today to express the f in Equation (1). The most frequently used methods are:

- Multiple regression analysis,
- Principal component and factor analysis,
- Principal component regression analysis,
- Partial least squares (PLS),
- Discriminant analysis and other classification methods, and neuronal nets.

In order to characterize chemical structure, these variety of mathematical methods has to be accompanied by a huge number of chemical descriptors. Todeschini and Consonni has presented an impressive encyclopedic guide to such descriptors in their Handbook of Molecular Descriptors (Todeschini and Consonni, 2000). But of course, not all of these descriptors have proven to be useful. They may be categorized a:

- Experimental quantities, such as $\log P$, pK_a (these quantities can also be computed), and spectroscopic data,

- Substituent constants (electronic, hydrophobic, and steric),
- Parameters derived from molecular modeling and quantum chemical computations,
- Graph theoretical indices,
- Variables describing the presence or the number of occurrences of certain substructures.

Characteristic measures of biological activity are the molar concentration C of a compound generating a certain effect derived from a dose-response curve (e.g., ED50 or IC50); binding, association, or inhibition constants; and rate constants. Reciprocal values are usually taken into account for dissociation constants and the molar-concentration-based quantities so that larger values for more active compounds are obtained. Based on kinetic or thermodynamic reasoning, such parameters can be transformed into free-energy-related quantities by logarithmic transformation. This is required for the formalism of Hansch analysis. Hence, typical expressions for F in Equation (1) are:

$$-\log C = \log 1/C = pC \text{ (examples: } pED50 \text{ or } pIC50 \text{)},$$

$\log 1/K_d$ (where K_d is a dissociation constant), and

$\log K$ (where K is an inhibition, binding, or rate constant).

By convention, the logarithmic transformation of biological measurement is used not only in methods based on linear free energy relationships (e.g. Hansch analysis) but also in all QSAR approaches applied to quantitative biological measurements. Generating better comparable results are one of the reasons. Biological measurements sometimes result in %effect data measured at a single dose. Strictly speaking, such data are not appropriate for Hansch-type and related QSAR approaches. However, experience has demonstrated that such data can still lead to meaningful QSARs after logarithmic transformation, on condition that the complete range from a few percent values to values close to 100% is covered. A logarithmic transformation according to the equation below is a good alternative for such values:

$$F = \log (\%effect (100 -\%effect)) \quad (2)$$

Another alternative is the translation of %effect data into a classification scheme which can then be analyzed by classification methods. Such methods are also necessary if biological measurements only permit a scoring of biological potency. The logarithmically transformed activity values will be designated as log BR (BR = biological response) in the following text.

2.4.2. Free-Wilson Analysis

The Free-Wilson analysis can be used for series of compounds that consist of a common (constant) parent structure and variable fragments (usually substituents) (see Figure 2.4). The fundamental assumptions of Free-Wilson analysis are:

- The parent structure and each variable fragment contribute an additive increase to the logarithm of biological response.
- There is no interaction between the fragments meaning that the increment of a given fragment is constant and independent from structural variations in other positions.

The Fujita–Ban variant of Free–Wilson analysis is used today since it is much simpler. In Fujita-Ban variant of Free–Wilson analysis, a standard substituent is specified for each position of substitution. The activity contributions of these standards are set equal to zero. Now the parent structure is defined as basic skeleton plus standard substituents (see Figure 2.4), and all activity contributions of the nonstandard substituents are calculated relative to those of the standards.

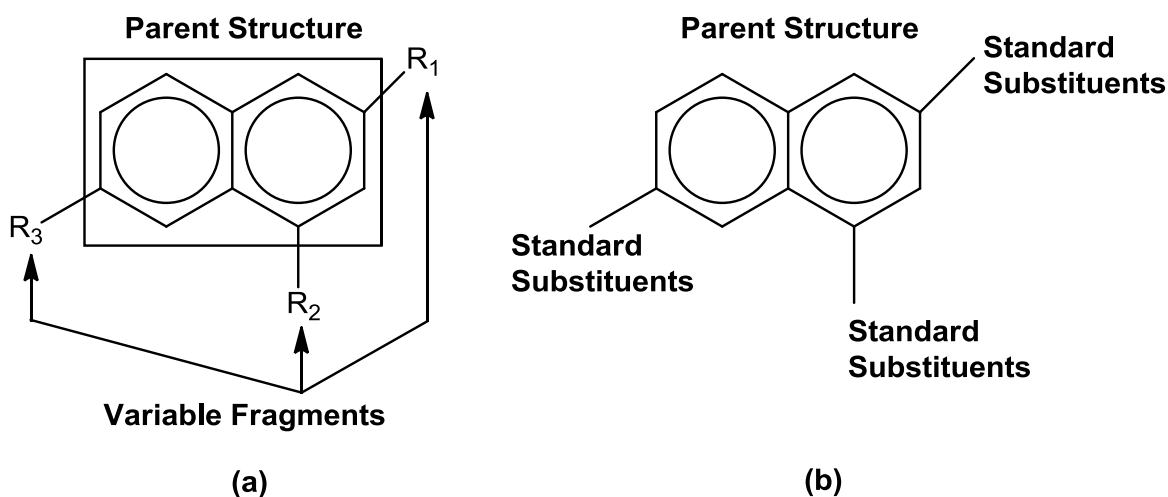


Figure 2.4. Schematic presentation of the parent structure according to the original Free–Wilson formalism (a) and in the Fujita–Ban variant of Free–Wilson analysis (b). (Figure taken from Franke and Gruska, 2003 and edited)

2.4.3. Hansch Analysis

2.4.3.1. Basic Assumptions

Basic assumptions that Hansch analysis are based on are as follows:

1. The logarithm of an appropriate biological response parameter (BR) can be considered to be related to the free energy of binding to the biological target. Thus, the logarithm of a biological response parameter can be described by the same formalisms used in physical organic chemistry to describe equilibrium or rate constants.
2. Substituents make additive and independent contributions to log BR (same assumption as in Free–Wilson analysis) in congeneric series.
3. These contributions can be resolved into factors such as electronic, hydrophobic and steric components that can be described by a linear combination of electronic (x_e), hydrophobic (x_h) and steric (x_s) parameters derived from theoretical computations or from well-defined chemical standard reactions (a_e , a_h and as .. coefficients):

$$\log \text{BR} = a_e x_e + a_h x_h + a_s x_s + \text{const.} \quad (3)$$

4. If transport processes to the site of action are included, these can be described by a parabolic or bilinear function of $\log P$ (where P is the partition coefficient in the system n-octanol/water). The following general expression results with the parabolic function:

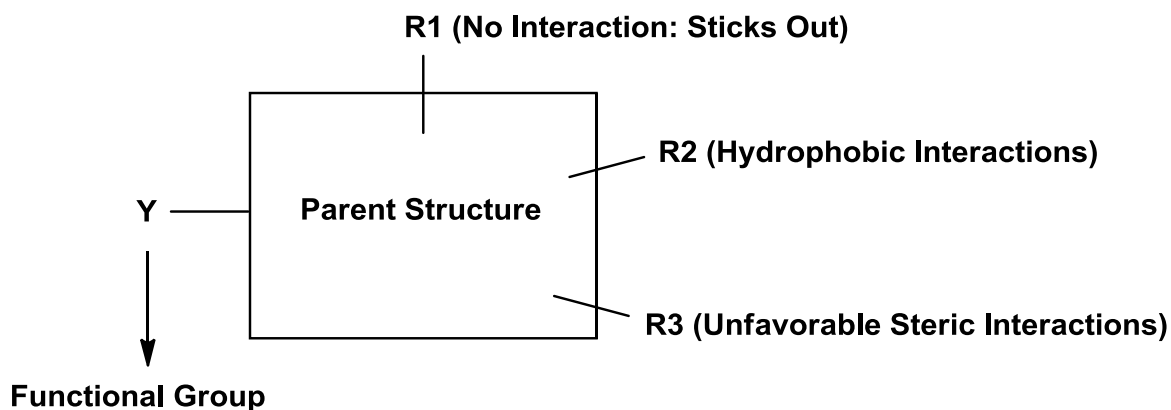


Figure 2.5. Hypothetical case of drug–receptor interactions. (Figure taken from Franke and Gruska, 2003 and edited)

$$\log BR = a_{ex}e + a_{hx}h + a_{sx}s - a_1(\log P)^2 + a_2\log P + \text{const.} \quad (4)$$

5. The real form of a Hansch equation for a given issue depends on the drug-biosystem interactions. If the hypothetical case of drug–receptor interactions shown in Figure 2.5 is taken into account, Equation (4) would turn into the following general form (including transport):

$$\log BR = a_{hx}h(R2) - a_{sx}s(R3) + a_e[x_e(R1) + x_e(R2) + x_e(R3)] - a_1(\log P)^2 + a_2\log P + \text{const.} \quad (5)$$

In this equation, $x_h(R2)$ describes the hydrophobicity of substituents in R2; $x_s(R3)$ measures steric properties (e.g., size) of substituents in R3; and electronic properties of substituents in R1, R2, and R3 are described by $x_e(R1)$, $x_e(R2)$, $x_e(R3)$. Obviously, once a Hansch equation is known, an interpretation is possible, allowing conclusions concerning the mechanism of action (Franke and Gruska, 2003).

2.4.4. PARAMETERS USED IN QSAR

Describing molecules and their properties is of critical importance in generating a good QSAR equation. Parameters are used in determining the types of intermolecular forces that underlies drug-receptor interactions. The most commonly used parameters are as follow:

2.4.4.1. Electronic Parameters

The electronic effects of various substituents will certainly have an effect on a drug's polarity or ionization. These properties may have an effect on controlling permeability of a drug through cell membranes and also determine how strongly it can bind to a receptor. Thus, it is useful to investigate the electronic effect a substituent that can have on a molecule (Patrick, 2009).

2.4.4.2. Hydrophobic Parameters

Hydrophobicity (also called lipophilicity) plays a role not only in pharmacokinetic processes but also in the interaction of drugs with many targets. It is therefore of central importance for biological potency. Generally speaking, hydrophobicity describes the tendency of molecules (or parts of molecules) to escape contact with a water environment and to move into a lipophilic environment (Franke and Gruska, 2003).

2.4.4.3. Steric Parameters

The size, bulk and shape of a drug will affect how easily it can approach and interact with a binding site. A bulky substituent may act as if it is a shield and hinder the ideal interaction between a drug and its binding site. On the other hand, a bulky substituent may help to lead a drug to place properly for maximum binding and therefore increase activity. Compared to hydrophobic or electronic properties, steric properties are more difficult to quantify (Patrick, 2009).

Apart from the common parameters defined above, there are some of the descriptors derived from entire molecular structures and many of them are completely new. For instance: polar descriptors, energetic descriptors, geometric descriptors and topological descriptors.

2.5. Pharmacophores and Pharmacophore Searches

2.5.1. Pharmacophores: Historical Perspective and Viewpoint from a Medicinal Chemist

The term “pharmacophore” has turned into one of the most popular words in medicinal chemistry since the emerging of computer-aided structure-activity studies. However, different medicinal chemistry groups assign several meanings to this term according to their scientific background and/or traditions. Thus, it is necessary to give a brief paragraph to the definition of the word pharmacophore, then a historical perspective and finally some comments from a medicinal chemistry practitioner will follow this.

2.5.2. Definitions

Many scientists use the term “pharmacophores” to explain functional or structural elements showing biological activity. However, this does not match to the official definition developed by an IUPAC working party that was published in 1998 (Wermuth et al., 1998): *A pharmacophore 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.* As a result:

- The essential, electronic and steric, function-determining points necessary for an optimal interaction with a relevant pharmacological target are described by pharmacophores.
- The pharmacophore does not represent a real molecule or a real association of functional groups. They are only an abstract concept that is responsible for the common molecular interaction capacities of a group of compounds towards their target structure.
- Pharmacophores are “pieces of molecules” (e.g. dihydropyridines, arylpiperazines) or not particular functional groups (e.g. sulfonamides).

Considering the remarks made above, a pharmacophore can be thought as the highest common denominator of a group of molecules showing a parallel

pharmacological profile and which are identified by the same site of the target protein. In spite of the official definition, many medicinal chemists still continue to call pharmacophores some particular functional groups, especially if they appear to be frequently associated with biological activity.

2.5.3. Pharmacophores: the Viewpoint of a Medicinal Chemist

Simple pharmacophores were described in the literature and considered as tools for the design of new drug molecules even before the appearance of computer-aided drug design. First structure–activity relationship considerations were accessible in the 1940s owing to the knowledge of the bond lengths and the van der Waals sizes which enabled the construction of simple two-dimensional model structures (Woods, 1940; Woods and Fildes, 1940; Dodds and Lawson, 1938 and Schueler, 1946). Although an early three-dimensional approach was made by Easson and Stedman (Easson and Stedman, 1933), access to three-dimensional models became practicable in the 1960s with the availability of X-ray analysis and conformational chemistry (Beckett, 1959; Barlow, 1964 and Belleau, 1963).

2.5.4. Criteria for a Satisfactory Pharmacophore Model (Wermuth and Langer, 1993)

A pharmacophore model has to provide acceptable information for the medicinal chemist examining structure–activity relationships in order to be recognized as a useful tool.

1. First of all, the pharmacophore model has to highlight the functional groups included in the interaction with the target, the nature of the non-covalent bonding and the various intercharge distances. This means that insignificant images of spaghetti and ribbon models (Wermuth et al., 2001), lacking indication of the molecular features of the interacting partners, have to be avoided. This is true for many redundant and opaque theoretical digressions as well. The model also has to reveal some *predictive power* and lead to the design of new, more potent compounds or, even better, of entirely novel chemical structures, not clearly deriving from the translation of structural elements from one active series into the other. A fascinating aspect of pharmacophore-based analogue design is mentioned as scaffold

hopping. This consists in the design of functional analogues via searching within wide virtual compound libraries of isofunctional structures, yet based on a different scaffold. The objective is to avoid from a patented chemical group in identifying molecules in which the central scaffold is replaced but the fundamental function-determining points are conserved and form the basis of a relevant pharmacophore (Schneider et al., 1999).

2. Distinguishing stereoisomers is the second criterion for a valid pharmacophore model that it should obey. Stereospecificity is certainly one of the main attributes of pharmacological receptors and a perfect stereochemical complementarity between the ligand and the binding-site protein is a vital criterion for high affinity and selectivity. A satisfying example of enantiomeric discrimination was observed for gamma-aminobutyric acid (GABA)-A receptor antagonists (Rognan et al., 1992).
3. The ideal model should also discriminate between agonists and antagonists. This is relatively simple for the particular category of antagonists which, according to Ariëns et al. theory (Ariëns et al., 1979), arise from the agonists basically through the addition of some additional aromatic rings which play the role of additional binding sites (e.g. the passage from GABA agonists to GABA antagonists (Rognan et al., 1992) or from muscarinic agonists to muscarinic antagonists (Wermuth, 1993)). The discrimination between the two categories becomes less obvious when the passage from agonist to antagonist depends on relatively subtle changes such as one observes for oxotremorine, glutamate and benzodiazepine antagonists.
4. A valid pharmacophore model can sometimes explain evidently *paradoxical observations*, e.g. the surprising affinity reversal found in *R*- and *S*-enantiomers of the sulpiride series on changing *N*-ethyl to *N*-benzyl derivatives (Rognan et al., 1990).
5. Eventually, it has to explain the *lack of activity* of certain analogues of the active structures. The knowledge of electronic or structural parameters

leading to poorly active or inactive compounds is a cost-lowering factor that enables the number of compounds to be synthesized to be decreased (Wermuth et al., 2006).

2.5.5. Methods Using Pharmacophore Features and Geometric Constraints

Currently there are several available pharmacophore perception methods. Only catalyst will be given here. For further details please see the reference (Poptodorov et al., 2006).

2.5.5.1 Catalyst

Catalyst® (Accelrys) was launched by BioCAD (now Accelrys) in 1992 as a tool for automated pharmacophore pattern identification in a collection of compounds on the basis of chemical features connected with three-dimensional structure and biological activity data.

Catalyst models which are called “*hypotheses*” composed of sets of abstract chemical features organized at certain positions in the three-dimensional space. The feature definitions are designed to cover various types of interactions between ligand and target, e.g. H-bond donor, H-bond acceptor, hydrophobic, positive ionizable, negative ionizable and so on. Besides in some special cases, different chemical groups that lead to the identical type of interaction, and thus to the same type of biological effect, are dealt as equivalent. The directions of the H-bonds are usually determined and are provided by vectors. Distinct chemical features in a special conformation of a compound must be placed within the tolerance constraints in order to fulfill the model. These models can also be used directly for three-dimensional database search queries in the Catalyst environment.

The pharmacophore identification procedure as implemented in the Catalyst package includes 3D structure generation, followed by conformational search and definition of the pharmacophore points in agreement with the training set.

Molecular structure editor

A 2D formula editor tool is used for the construction of molecular structures and it is provided in combination with 3D conversion. Standard potential energy minimization is carried out using the modified parameter set of the CHARM force

field (Brooks et al., 1983); the conformational models are built using Monte Carlo conformational analysis as well as poling as described in the next section.

Conformational analysis in Catalyst

(i) Overview

Many usual methods attempt to determine one global minimum energy conformation and other local minima as characteristic of the space. On the contrary, the approach to conformational analysis taken within Catalyst claims a wide coverage of bioaccessible conformational space of the molecules in the range of a user-specified energy threshold. This implies that the representative conformers generated by Catalyst are not exactly at local minima on the potential surface but are distributed widely across the space. This approach to conformational analysis is driving from the consideration that, the bound conformation of a small molecule to a receptor may not be the lowest energy conformation in many cases. Moreover, the global minimum predicted by a force field could be incorrect due to solvation effects or approximation errors in the force field.

The redundancy among conformers is a common difficulty accompanying the representation of the conformational space by sampling. Generally many hundreds or thousands of conformers are produced and then reprocessed to choose families representative of the entire space. Following a local minimization, many of these conformers may drop within the same conformation, reproduced several times. As a result, compared with methods that represent conformational space as clusters (a collection) of local minima, Catalyst focuses on the coverage of all possible bioactive conformations of a compound.

Another problem that should be mentioned briefly is the relationship between size and resolution of a conformational model especially in terms of coverage of the low-energy regions of the accessible conformational space. The coverage should in any case be in agreement with the precision of the application which uses the conformational model. The restriction is given by the tolerance of the pharmacophore query during three-dimensional pharmacophore generation for database search objectives. It has been revealed in principle that a limited number

of conformers is adequate to represent the low-energy conformational space of small- to medium-sized molecules (Smellie, 1995a; 1995b).

Conformational flexibility is provided in Catalyst by storing compounds as multiple conformers per molecule. Given that one has to generate and search through a very large number of conformers that may be indeed similar enough that can be behaved as identical when mapped on to a pharmacophore hypothesis, the need for variation with a simultaneous reduction in the number of conformers becomes obvious. The Poling algorithm of Smellie et al. (Smellie et al., 1994) implemented in Catalyst is planned to solve many of these problems.

(ii) Conformational search in Catalyst: catConf/ConFirm

There are two types of conformational search, BEST and FAST, that are employed in Catalyst. Both methods underline sufficient coverage of the conformational space, each with specific advantages. While the FAST method gives a reasonable model within a short time and is utilized for database generation purposes in the first place, the BEST method is intended to construct more precise conformational models of molecules for hypothesis generation. Both methods use Poling by default, BEST for all molecules whereas FAST depending on the size and flexibility of the compounds in investigation. Instead of Poling, the FAST method uses systematic search in the torsional space for smaller, less flexible compounds. Various aspects of conformational search parameters including Poling are user adjustable and can be turned off if required. Stereochemistry is dealt with a comprehensive manner with the options to specify explicit, relative and unknown chirality. Specified exact and relative chirality will always be conserved throughout conformational search and pharmacophore analysis, whereas for compounds with chirality marked as unknown, mirror images will be taken into account unless this is not desired by the user.

Conformational models produced by other programs can also be used for pharmacophore generation and in Catalyst databases by importing multiconformer structures stored, e.g., in SD file format.

Pharmacophore modeling with Catalyst

Catalyst has two algorithms for automated pharmacophore arrangement search: HypoGen and HipHop. While HypoGen needs biological assay data (e.g. IC50 or K_i) to derive hypotheses that can estimate quantitatively the activity of compounds, HipHop searches for a common three-dimensional configuration of chemical features shared between a set of active molecules. Regarding HypoGen, similarly to 3D QSAR, all members of the training set must own the same binding mode; the second method optionally permits automatic removal of compounds that may have a different molecular site of action. The resulting models go through a complex evaluation process by the program and the highest scoring results are reported to the user.

(i) HipHop

The HipHop algorithm (Barnum et al., 1996) tries to generate an alignment of compounds expressing specific activity against a particular target and by superposition of diverse conformations to find common three-dimensional arrangements of features shared among them. Despite the fact that HipHop does not use activity data as input, it is a good idea to pick out highly active chemically diverse compounds when composing training sets whenever possible.

HipHop determines common features by a pruned comprehensive search, starting with the simplest possible (two-feature) arrangements and extending the model to three, four, five features and so on until no more common configurations can be found. This involves a search through two large spaces - the pharmacophore domain and the conformational space of the training set. HipHop does not require a particular reference conformation. If needed, HipHop will attempt sequentially to align with each other all conformers of every training set member. Still, at least one molecule as the whole conformational model (principal compound) must be clearly described as a reference. The presentation of an exact conformer in the alignment depends on the remaining compounds and their conformational diversity and on the conditions of the run as well.

First, the program determines matches and distribution of the selected features between the training set members, followed by the alignment process. When each

of features lies within a specified distance (tolerance) from the ideal location, the features are considered superimposed, and the root mean square deviation (RMSD) for the configuration as a whole is measured at the same time. The quantitative prediction of the goodness of match between a molecule and a configuration of features (Fit) can be followed similarly to a scoring function to rank virtual screening results.

In the ideal case, superposition of all input molecules is requested. Sometimes it could be of benefit to allow some molecules, up to a specified number, to miss one, one particular or more than one of the features of a configuration so that to map all the remaining features. The advantage from such an option is that it allows one to work with compounds that may have a different binding mode or show activity in a specific assay due to experimental errors or an alternative mechanism of action.

In most cases, the result of a HipHop run will be consisting of a great number of configurations of features so there is a need to score and rank them. For instance, in many cases, the input molecules may share feature arrangements widespread among drug molecules or there may be configurations common for the training set but uncommon in general. Therefore, the ranking of the HipHop models is based on rarity (Barnum et al., 1996). Maximizing the score of a configuration will reduce the probability that the training set molecules map the model by chance, thus making the pharmacophore specific.

(ii) HypoGen

The HypoGen algorithm is intended to correlate structure and activity data for pharmacophore model generation.

HypoGen consists of three stages called phases: constructive, subtractive and optimization. Usually, the constructive phase looks like the process of the HipHop algorithm. The training set is divided into two subsets: "Active" and "inactive" compounds. First step is the identification of all pharmacophores shared between the first two highest active compounds by overlaying systematically all their conformations. Afterwards only hypotheses that fit a smallest subset of features present in the remaining active compounds are kept.

In the subtractive phase, the program examines the hypotheses previously created and dismisses those most common to the inactive part of the training set. Compounds whose activities lies 3.5 logarithmic units (this value is user adjustable) below that of the most active compound are regarded inactive.

The subtractive phase is followed by another phase called optimization phase where simulated annealing is used to improve the predictive power of the hypotheses. Few changes are made to the models and they are scored according to the accuracy in activity prediction. Finally, the simplest models that correctly predict activities are picked up (Occam's Razor) and the top N solutions are reported to the user. The method has been described in more detail in other literatures (Guner, 2000 and Kugori and Güner, 2001).

A significant assumption that is made within both HipHop and HypoGen is that more contacts to the receptor and thus more features per molecule lead to increased activity. However, in many cases, it is well known from practice that this is not true. For instance, because of unfavorable steric interactions large and feature-rich compounds may be hardly active. HypoRefine is an extension to the HypoGen algorithm which is intended to help in solving this issue by placing the exclusion volume in key locations derived from atoms of well-fitting but inactive compounds. On the other side, in case of lacking inadequate activity or when only HTS data are present, in order to produce a grid-based exclusion volume which removes false-positive HTS hits and increases enrichment rates, the HipHop Refine algorithm allows the use of "negative" information from inactive compounds matching the pharmacophore (Maynard et al., 2004 and Toba et al., 2006).

(iii) Compound databases and database searching in Catalyst

There are basically two approaches to address the problem of conformational flexibility throughout pharmacophore screening: the use of various stored conformations and on-the-fly conformer calculation (Sprague, 1997). The Fast and Best Flexible Search algorithms are the combination of both solutions offered by Catalyst. Catalyst databases composed of compounds stored as multiple conformations. When executing Fast Flexible Search, the search is carried out using only conformations already being in the database and Fast Search attempts

to find one fitting the pharmacophore among those available. The algorithm used with Best Flexible Search Databases/Spreadsheets can adjust the conformation of a molecule throughout the computation to enforce a fit within a given energy threshold.

The database search procedure begins with a fast screening process within which molecules having properties required from potential hits are sorted out from those that can be excluded likely. The screen includes substructure match followed by screens matching molecular shapes, three-dimensional pharmacophore features or text constraints (1D properties) and exclusion volumes if present in the query. All this greatly decreases the number of potential hit compounds in the database. The next step of the search process attempts a rigid fit of each conformation of each compound to the corresponding features. After the first successful mapping of all features, compounds are chosen as hits and a hit list is obtained once all compounds have passed the procedure (Kugori and Güner, 2001).

The Best database search first determines all potentially appropriate compounds by using loosened constraints, therefore including those that would fail a rigid search. Within this initial list, the algorithm tries to modify additionally the conformers in order they can fit the original query while remaining below a specific energy overflow (Kugori and Güner, 2001). The use of a Best search is justified when too small hit lists has to be dealt with.

Catalyst provides the possibility to calculate fit values that can be used for scoring once a hit list has been acquired (Poptodorov et al., 2006).

2.6. Docking Techniques

Protein–ligand docking technique is widely used in computer aided drug design and is a geometric search problem. Protein and ligand conformations, and also their relative orientations, are the relevant degrees of freedom. Although the given protein structure is reasonably well known (even though although there are many examples of conformational changes that occur upon ligand binding), the protein-bound ligand conformation is generally unknown. Hence, many docking approaches address ligand flexibility and hold the protein rigid. However,

examples of methods handling with protein flexibility will be briefly discussed. The principal concepts of docking approaches are outlined below.

2.6.1. Protein Structure

A three dimensional structure of the target protein under study at atomic resolution must be available in order to perform computational protein–ligand docking experiments. The most reliable and broadly used sources are crystal and solution structures provided by the Protein Data Bank (PDB) (Berman et al., 2000) or from in-house efforts. In the absence of experimental structures, homology models (Blundell et al., 1987 and Sander and Schneider, 1991) and pseudoreceptor models (Vedani et al., 1995) are an alternative source. It should be noted, however, that the quality of the protein structure is of critical importance for the success of subsequent docking experiments. Even small changes in structure can severely change the outcome of a computational docking experiment (Muegge, 1999). Ideally, the desired the atomic resolution of crystal structures should be below 2.5 Å (Jones et al., 1997). On the other side, the PDB provides a wealth of protein structures of many enzymes and receptors that can be used for homology modeling. It can be assumed that reasonable homology models can be constructed for many proteins coded in the human genome.

2.6.2. Rigid Docking

Because of the fact that ligand flexibility, and often also protein flexibility is crucial, rigid ligand docking is not generally appropriate for protein-ligand docking. However, such simplification is often satisfactory for the docking of small fragments or ensembles of conformations and/or molecules.

2.6.3. Docking with Flexible Ligands

In many cases, compared to the total binding affinity between ligand and target protein, energetic differences between alternative ligand conformations are small. Also, for flexible ligands it is rather common that the bioactive conformations are different from the minimum energy conformations in solution (Nicklaus et al., 1995). Small drug like molecules are generally flexible; 70% of drug like molecules involve between two and eight rotatable bonds (Oprea, 2000). Ligand flexibility is

typically treated in docking approaches by combinatorial optimization protocols like ensembles, fragmentation, simulation techniques or genetic algorithms.

2.6.4. Ensembles of Ligand Conformations

Evaluating multiple conformations of the ligands in a rigid-body docking algorithm allows ligand flexibility to be introduced. Since computing time increases directly with the number of conformations, a balance needs to be sought between computing time and coverage of conformational space (Muegge and Enyedy, 2004).

2.7. Cell Cycle and Loss of Cell Cycle Control

Proliferation is a complex process involving various subroutines that collectively causes cell division. The cell cycle lies at the heart of proliferation and consists of many processes that require to be completed in a timely and sequence specific manner. Therefore, regulation of cell cycle events is a complex affair and consists of a sequence of checks and balances that keep track of cell size, nutritional status, presence or absence of growth factors, and integrity of the genome. These cell cycle controlling routes and the signal transduction pathways that communicate with them are populated with oncogenes and tumor suppressor genes.

Cell division is divided into four stages: G1, S, G2, and M (Figure 2.6). The whole process is emphasized by two impressive events, the replication of DNA during S phase and chromosome division during mitosis or M phase. Among the four cell cycle phases, three can be allocated to replicating cells and only the G1 phase and a related inactive phase, G0, are nonreplicative in nature. Normal cycling cells that stop to proliferate enter the resting phase (G1), and their exit into the replicative phases is extremely dependent on the presence of nutrients and growth factors. Nevertheless, once the cells enter the replicative phase of the cell cycle, they become irreversibly committed to completing cell division. Thus, the situations that lead to exit from G1 and entry into S are firmly regulated and are often misregulated in neoplastic cells that show uncontrolled proliferation. Studies first directed by Arthur Pardee reported the existence of a point in G1 that restricted the passage of cells into S phase, and this was presumed to be controlled by a

labile protein factor (Pardee, 1974). Progression across this restriction point, or R point, is now known to be delicate to growth factor stimulation.

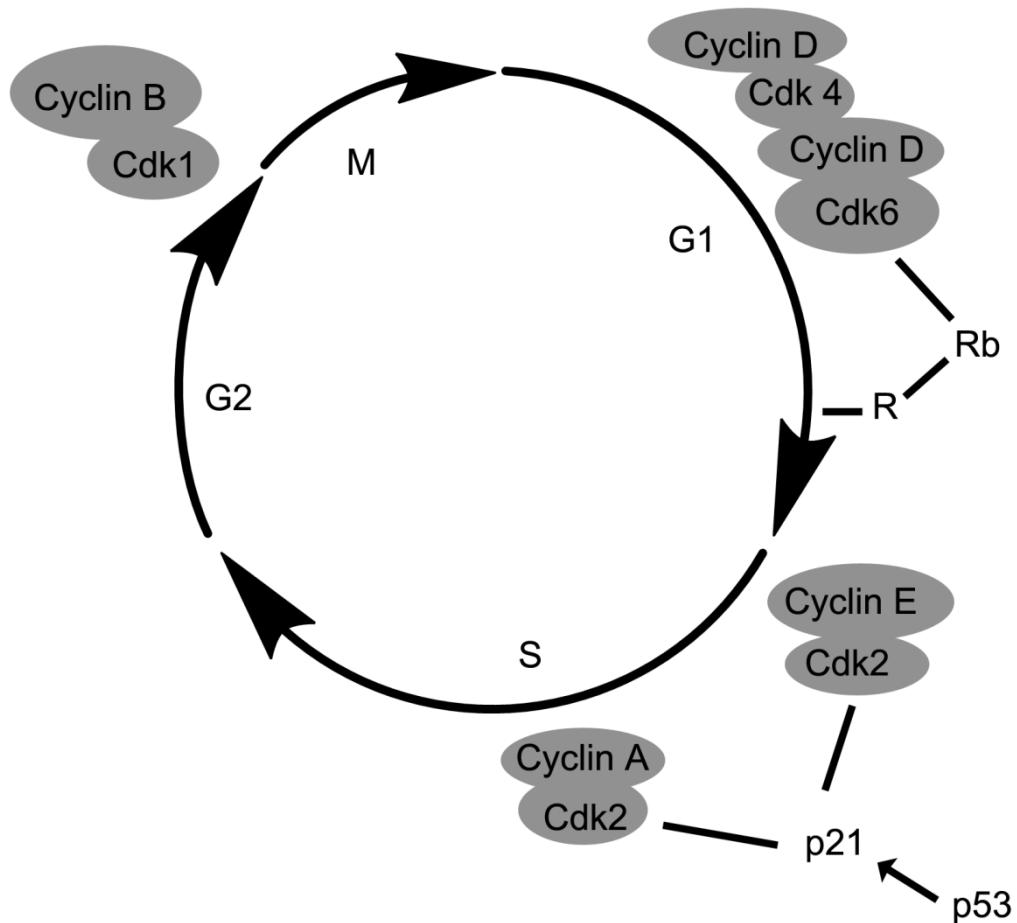


Figure 2.6. Model of the cell cycle and the cyclin/cdk complexes that are essential at each cell cycle phase. CyclinD/cdk4-6 complexes suppress Rb function by phosphorylating the protein allowing transition across the restriction R-point. **P53** suppresses cell cycle progression by stimulating the expression of the cyclin dependent kinase inhibitor p21, which binds with and inactivates a variety of cyclin/cdk complexes. (Figure taken from Martinez et al., 2003 and edited)

Progression through the cell cycle is controlled by two classes of cell cycle proteins, cyclins and cyclin dependent kinases (CDKs), which physically connect to form a protein kinase that drives the cell cycle forward (Hutchison and Glover, 1995). There are at least 15 types of cyclins and 9 types of CDKs. The accumulation of cyclin proteins occurs through cell cycle-dependent initiation of gene transcription, but removal of cyclins occurs by carefully regulated

degradation that is enabled through protein sequence tags known as destruction boxes and PEST sequences. Despite the fact that not all of the cyclin types show this oscillation in protein quantity, those cyclins that play central roles in progression through the cell cycle (cyclins A, B, and E) are most abundant during individual phases of the cell cycle. Just before the restriction point, Cyclin D1 is synthesized during G1 and plays a key role in regulation of the R point. Cyclin E is most abundant during late G1 and early S and is required for exit from G1 and progression into S phase. High levels of these two G1 cyclins can cause uncontrolled proliferation. Frankly speaking, both cyclin D1 and cyclin E are overexpressed in some tumor types, mentioning that the cyclins and other components of the cell cycle may be useful therapeutic targets (Zafonte et al., 2000).

CDK is the second component of the enzyme complex and as the name implies, needs an associated cyclin to become active (Figure 2.7). Protein kinases are numbered according to a standardized nomenclature beginning with CDK1, which for historical reasons, is most often referred to as cell division cycle 2 (*cdc2*). Different from the cyclins, abundance of the CDK proteins remains relatively constant during the cell cycle. Instead, their activity changes throughout different phases of the cell cycle in agreement with whether or not an activating cyclin is present and whether or not the kinase itself is properly phosphorylated. Both cyclins and CDKs are highly preserved from yeast to man and function in a similar manner, suggesting that the cell cycle is controlled by a universal cell cycle engine that acts through the action of conserved proteins. Thus, drug discovery studies aimed at determining agents that control the cell cycle may be performed in model organisms, such as yeast with some promise that the targeted mechanisms will also be relevant to humans.

It is now obvious that specific cyclin/cdk complexes are needed during specific phases of the cell cycle. Cyclin D1/cdk4,6 activity is necessary for passing the restriction point and crossing cells into replication. The retinoblastoma (Rb) tumor suppressor protein is an important substrate of the cyclin D1/cdk4,6 complex, which when phosphorylated by this kinase complex, is inactivated. This releases the cell from the restrictions on cell proliferation imposed by the Rb protein. It is

this event that is considered to be determinative in the stimulation of resting cells to go through proliferation. Cyclin E/cdk2 plays a role later in the cell cycle for proliferating cells by moving them from G1 into S phase. Cyclin E is overexpressed in some breast cancers where it may increase the proliferative capacity of tumor cells. Cyclin A/cdk2 supports DNA replication and is therefore needed during S phase. Cells entering mitosis up through metaphase require cyclin B/cdc2. Cyclin B is reduced at the end of metaphase and cdc2 becomes inactivated, permitting mitotic cells to progress into anaphase and to complete mitosis. Sustaining the activity of cyclin B/cdc2 causes cells to arrest in metaphase. Thus, it is the common result caused by the activation and deactivation of cyclin/cdk complexes that pushes proliferating cells through the cell cycle (Martinez et al., 2003).

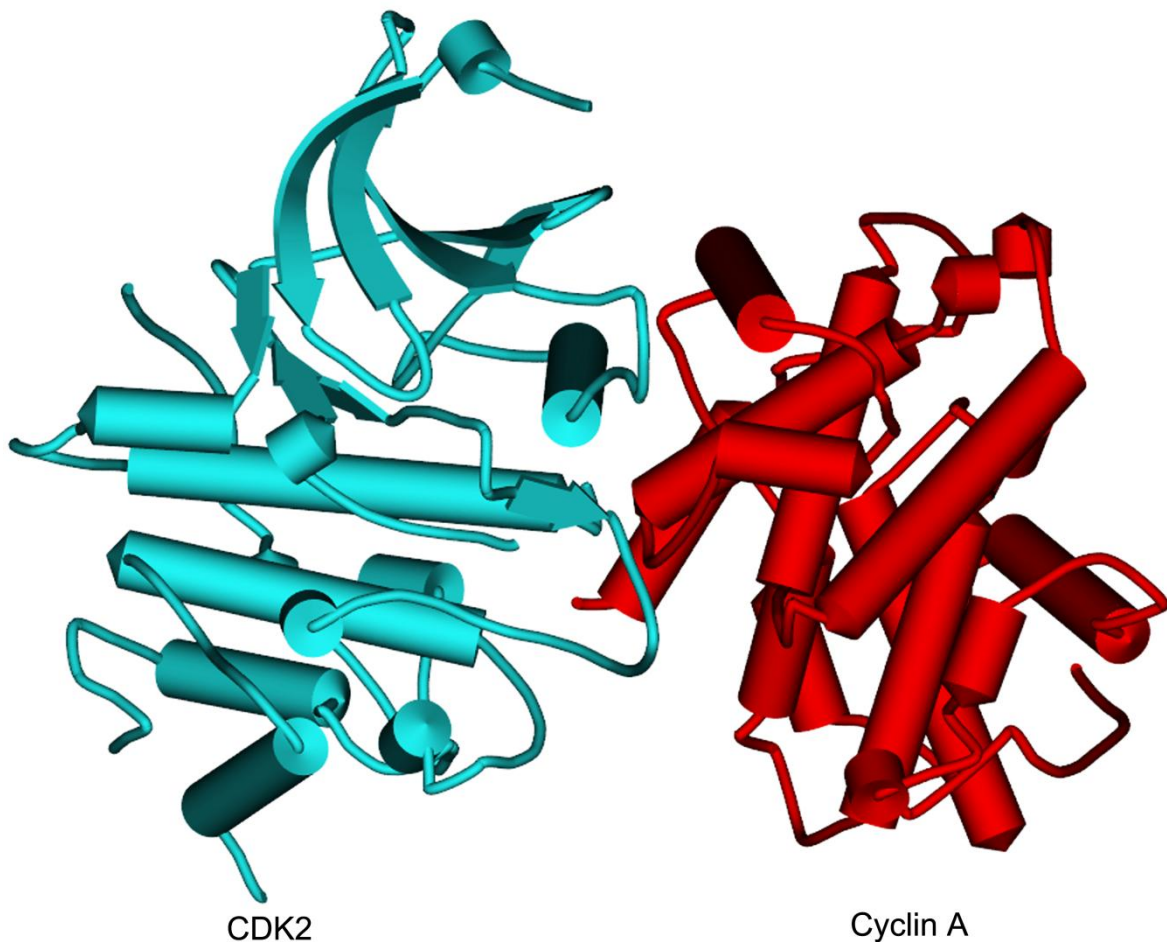


Figure 2.7. Schematic diagram of CDK2/Cyclin A (PDB access code: 1OGU).

To sum up, progress through the cell cycle is regulated by subsequent activation of cyclins and CDKs. This process can be down-regulated by the CDK inhibitors. The overall process is normally tightly controlled, such that there is a collection of a relevant cyclin-CDK complex followed by fast degradation of the complex once its task is complete.

Overactive cyclins or CDKs have been connected to several cancers. For instance, breast cancer cells frequently yield excess cyclin D and E, and skin melanoma has lost the gene that codes for the inhibitory protein p16. Half of all human tumours lack a correct functioning p53 protein, in other words the level of the inhibitory protein p21 falls. In many instances, both the pRB and p53 proteins are disabled in viral-related cervical cancers.

Oncogenic modification of cyclins, CDKs, cyclin dependent kinase inhibitors (CKIs), and other parts of the pRB pathway have been reported in 90% of human cancers, specifically in the G1 phase. Thus, excessive production of cyclins or CDKs, or inadequate production of CKIs can result in a disruption of the normal regulation controls and finally lead to cancer. By targeting molecular abnormalities, attempts have been made to identify how one can rebuild the control of the cancer cell cycle. These can involve down-regulation of cyclins, degradation of cyclins, CDK inhibition, up-regulation of CDK inhibitors, or inhibition of tyrosine kinases that initiate the cell cycle activation in the first place (Patrick, 2009).

3. WORKING PLAN

In the first part of study the structure-activity relationship of 4-cyclohexylmethoxypyrimidines, inhibitors of CDK2 as antitumor agents was explored. QSAR approach attempts to identify and quantify the physicochemical properties of a drug and to see whether any of these properties has an effect on the drug's biological activity. Using a useful QSAR equation, it becomes possible to make predictions leading to the synthesis of novel analogues. As an extension of these considerations, it was of interest to investigate whether correlating the physicochemical properties of the CDK2 inhibitors to the biological activity of the compounds by means of a classical Hansch analysis (Tute, 1990) would lead to a meaningful model of QSAR in the pyrimidine class.

The discovery of potential CDK2 inhibitors is presented in the second part. 3D QSAR pharmacophore study was explored.

Selection of an appropriate training set compounds is of crucial importance in order to establish a good pharmacophore model. The difference in the training set selection has a large influence on the final pharmacophore model. Completely different pharmacophore models of ligands interacting with the same protein target could be generated with the use of the same softwares but different training set. Different ligand-based pharmacophore models of CDK2 inhibitors were reported previously (Hecker et al., 2002; Toba et al., 2006 and Vadivelan et al., 2007).

This part of work was concentrated on generating a more accurate pharmacophore model using recent CDK2 inhibitors. A combination of pharmacophore modeling followed by validation, virtual database screening and molecular docking studies are performed.

4. MATERIALS AND METHODS

4.1. QSAR Study

Molecular properties were calculated with the QSAR protocol in the Discovery Studio environment (Accelrys). Multiple linear regression models were studied with Statgraphics Centurion XV, V.15.2.05 demo version and Bilin software (Bilin, 2005).

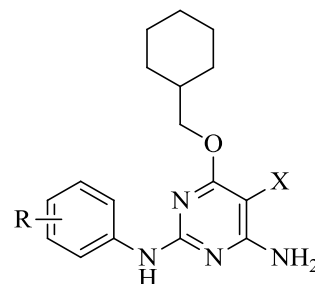
For the development of QSAR models on 4-cyclohexylmethoxypyrimidine derivatives, the CDK2 inhibitory activity was utilized. For the analysis, data of the parent compound **1** and twenty-four other derivatives (**2(a-s)**-**3(a-e)**) which have all been previously published with their CDK2 inhibitory activity (Marcetti, 2007) were employed. Of these, 5 compounds were selected to be test set (**1**, **2e**, **2g**, **2k** and **3d**). The test set molecules were selected from the most active, moderately active and less active molecules to spread out the range of activities (Golbraikh et al., 2003).

Two hundred and forty nine descriptors that included topological, charge, geometrical, aromaticity indices, constitutive properties, quantum mechanics, and thermodynamics were evaluated for each compound. In a fort step, the so-called statistical method of stepwise multiple regression procedure, based on the forward-selection and backward-elimination methods, was used for variable selection with the aim to obtain the best regression equation. For the current dataset of 20 compounds the QSAR model development was restricted to a maximum of four variables in accordance to the general accepted rule of thumb (5:1 for compounds: descriptor) during forward stepping regression.

For the internal validation, Leave-One-Out (LOO-) cross validation method was used (Tetko et al., 2001). For external validation, predictive ability of test set compounds was calculated.

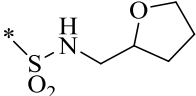
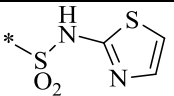
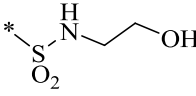
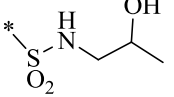
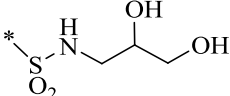
In Table 4.1, substitution pattern, values of selected descriptors and the previously published pIC₅₀ values of the training set and test set members are compiled.

Table 4.1. Structural properties, values of the selected descriptors and biological activity (pIC50) of the training and test set 4-cyclohexylmethoxypyrimidines.



Compound	R	X	Dielectric _Energy_ DMol3	Shadow _Xyfrac	Ix	N_Count	Jurs_RASA	pIC50 (in M)
1 ^{*a}	—	NO	-0.031	0.625	1	5	0.839	8.658
2a ^b	4-OMe	NO	-0.033	0.643	1	5	0.867	6.668
3a	3-Br	H	-0.034	0.618	0	4	0.911	4.585
2b ^b	3-Br	NO	-0.029	0.637	1	5	0.904	6.301
2c	4-OH	NO	-0.036	0.612	1	5	0.798	7.796
2d	3-OMe	NO	-0.031	0.637	1	5	0.868	6.469
3b	3-SMe	H	-0.036	0.637	0	4	0.916	4.222
2e [*]	3-SMe	NO	-0.031	0.668	1	5	0.91	6.398
2f ^b	4-SMe	NO	-0.033	0.63	1	5	0.907	6.921
2g [*]	3-CH ₂ CN	NO	-0.035	0.654	1	6	0.91	7.481
3c	4-CH ₂ CN	H	-0.041	0.567	0	5	0.914	4.62
2h	4-CH ₂ CN	NO	-0.037	0.574	1	6	0.908	7.194
2i	3-CH ₂ OH	NO	-0.04	0.646	1	5	0.808	7.347

Table 4.1 continues...

3d ^{*b}	4-CONH ₂	H	-0.051	0.594	0	5	0.804	4.229
2j ^b	4-CONH ₂	NO	-0.047	0.608	1	6	0.799	7.469
2k [*]	4-CON(CH ₃) ₂	NO	-0.045	0.637	1	6	0.852	7.097
2l	4-CON(C ₂ H ₅) ₂	NO	-0.047	0.645	1	6	0.859	7.699
3e ^b	4-SO ₂ NH ₂	H	-0.075	0.574	0	5	0.749	5.533
2m ^b	4-SO ₂ NH ₂	NO	-0.071	0.58	1	6	0.745	9.000
2n	4-SO ₂ N(C ₂ H ₅) ₂	NO	-0.066	0.651	1	6	0.807	7.066
2o	4- 	NO	-0.066	0.573	1	6	0.825	8.092
2p	4- 	NO	-0.074	0.601	1	7	0.784	7.721
2q	4- 	NO	-0.08	0.547	1	6	0.72	9.155
2r	4- 	NO	-0.075	0.595	1	6	0.742	8.886
2s	4- 	NO	-0.081	0.576	1	6	0.68	9.097

*Test set compounds. ^a (Arres, 2000) ^b (Sayle et al., 2003).

4.2. Pharmacophore Study

A training set of 16 compounds and test set of 14 compounds with diverse scaffold were collected from literature and tested with the same assay (FP assay) have been used in this study (Marchetti et al., 2007 and Marchetti et al., 2010). Structures and biological activities of the training set compounds are shown in Figure 4.1 and that of test set compounds are given in Figure 4.2. 3D QSAR Pharmacophore Generation module/Discovery Studio (DS) was used to construct pharmacophore model using hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), and hydrophobic (H) chemical features. It produced ten top-scored hypotheses based on the activity values of the training set molecules.

4.3. Databases Screening

The validated hypothesis generated in pharmacophore study was used to as a three-dimensional query to screen the various databases like Life Chemicals and NCI2003 which consists of 349,431 and 246,599 compounds respectively. The databases were downloaded as an SDF file and converted into a Catalyst database (i.e., a conformational model consisting of a maximum of 255 conformers was generated for each compound, so as to reproduce the flexibility of the molecule during the database search). The database (596,030 compounds, 45,603,414 conformations) search was then performed using the 'fast flexible search' method implemented in Catalyst, which retrieves compounds able to map the three-dimensional query represented by the pharmacophoric model, and finds the best fit among the conformers.

In order to filter these compounds to fit drug likeliness requirements, several molecular properties were calculated for the hit compounds returned from screening process such as molecular weight, Alogp, number of hydrogen bond acceptors and donors, molar refractivity, number of rotatable bonds and finally polar surface area.

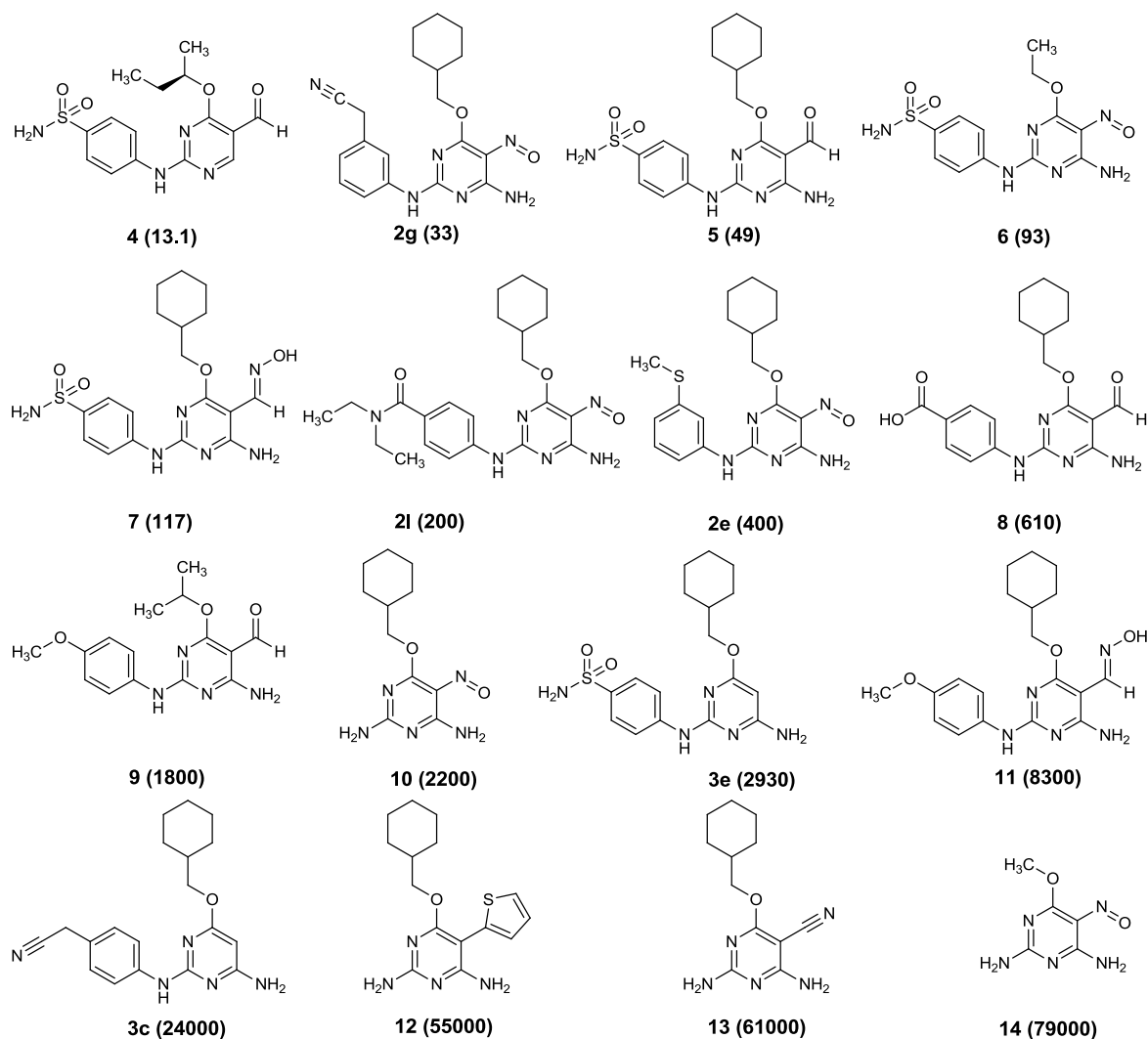


Figure 4.1. Chemically diverse 16 compounds used as training set in 3D QSAR Discovery Studio/Pharmacophore generation. IC₅₀ values are indicated in parentheses for each compound.

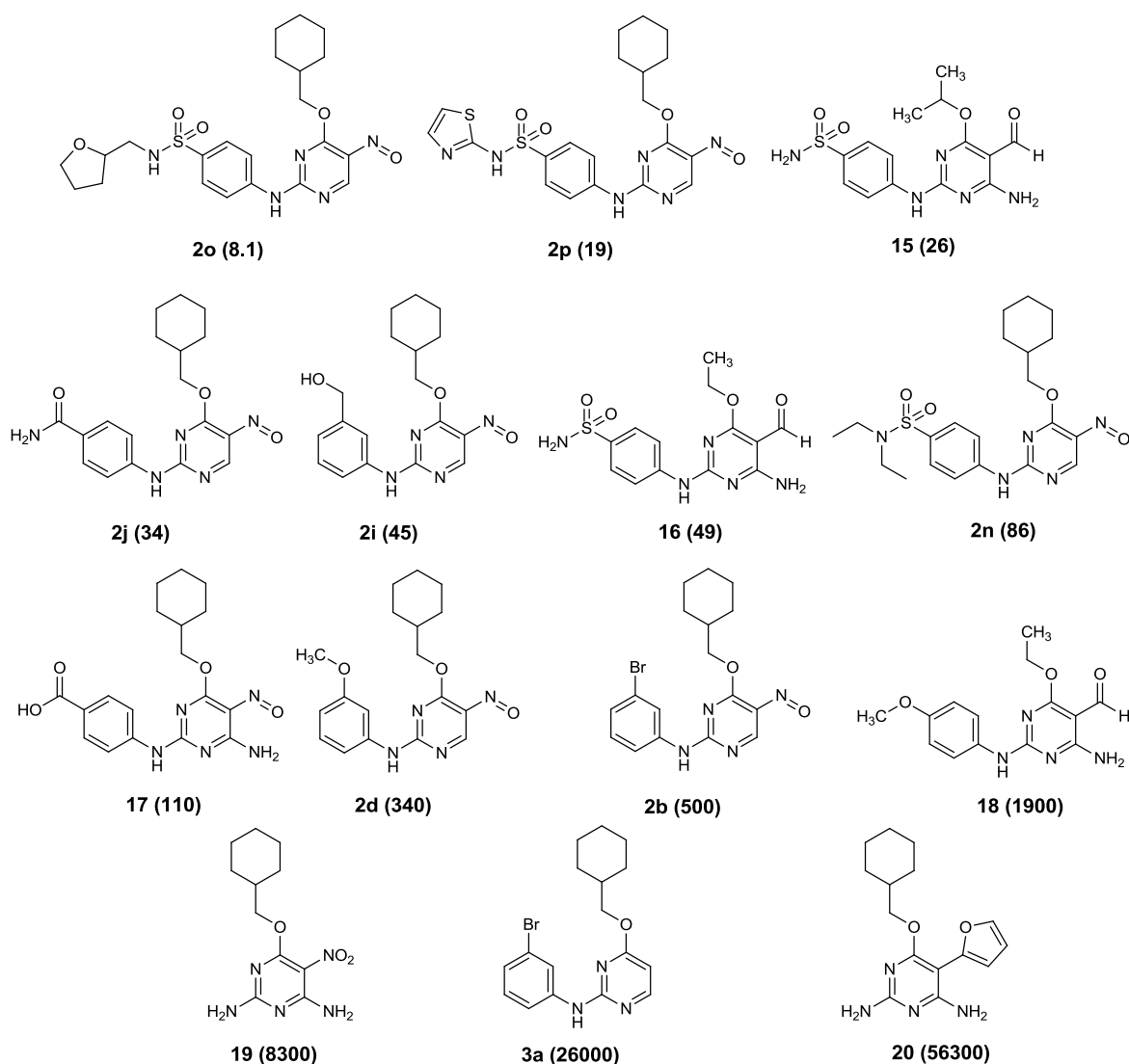


Figure 4.2. Chemically diverse 14 compounds used as test set in 3D QSAR Discovery Studio/Pharmacophore generation. IC₅₀ values are indicated in parentheses for each compound.

4.4. Docking Study

Docking studies were performed using CDOCKER protocol in Discovery Studio environment. X-ray crystal structure of carboxamide **2j** bound to the CDK2/cyclin A phosphorylated on Thr160 (T160pCDK2/cyclin A) (PDB accession code 1OGU) was obtained from protein databank and used as receptor. Binding site sphere of 8 Å radius size was determined using **2j** as reference ligand (Figure 4.3). Receptor was prepared for docking in such a way that all heteroatoms (i.e., nonreceptor atoms such as water, ions, etc.) were removed. Full potential was applied. Other

parameters were remained as their default values. The top 10 poses for each ligand were saved for comparison and analysis.

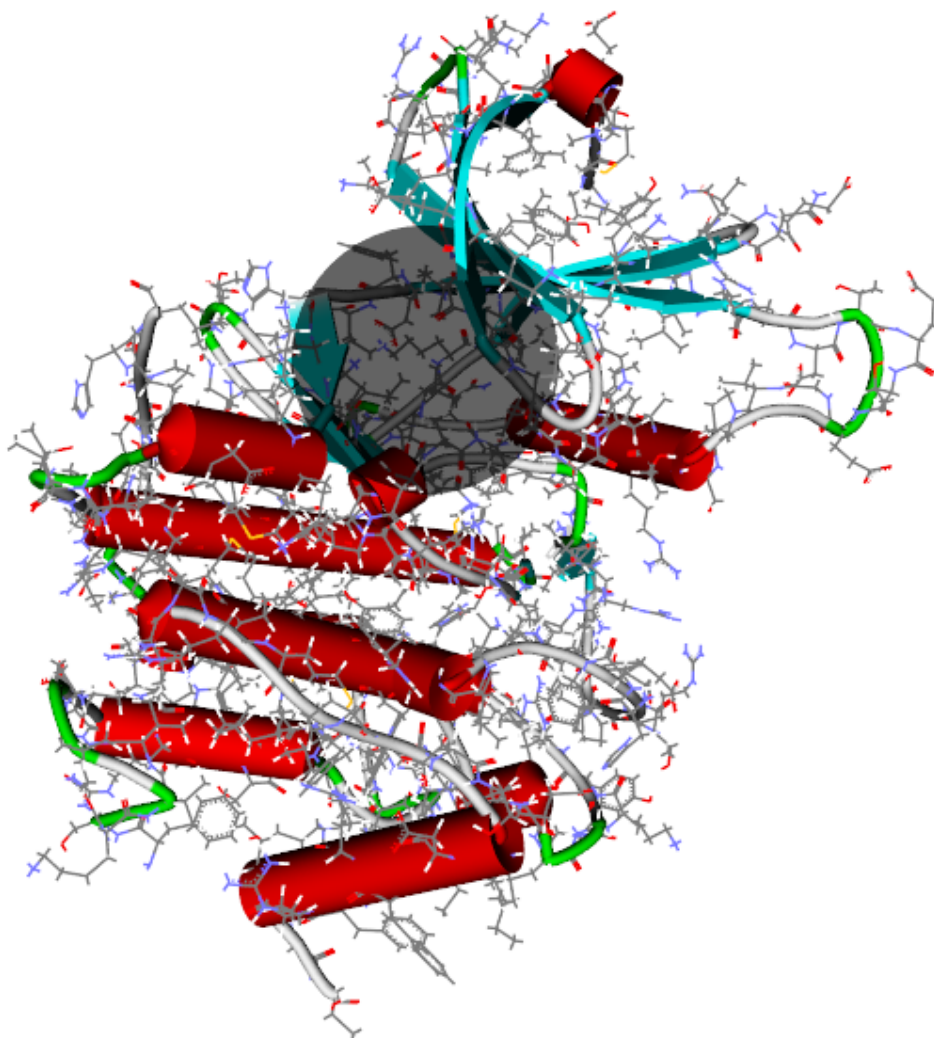


Figure 4.3. Binding site in CDK2 structure (Binding site is shown in grey circle).

5. RESULTS AND DISCUSSION

5.1. QSAR

We found the following molecular properties to be correlated with the biological activity under study:

- The dielectric energy (DE) descriptor accounts for the original charge arrangement on the surface of the molecule, and thus would be indicative of the metabolic susceptibility of a drug molecule (Turner and Maddalena, 2004).
- Shadow indices are a set of geometric descriptors to characterize the shape of the molecules (Rohrbaugh and Jurs, 1987). The descriptors are calculated by projecting the model surface on three mutually perpendicular planes, xy, yz, and xz. These descriptors depend not only on conformation but also on the orientation of the model. The shadow_XY fraction is a geometric spatial descriptor related to the breadth of a molecule and it is the fraction of area of molecular shadow in the XY plane over area of enclosing rectangle.
- Structural parameters as indicator variable, I_x was used in order to find out the role of specific substituent/substituent pattern at specific position towards the activity. For every compound that bears NO substituent, I_x was taken as 1 and for others I_x was taken as 0.
- N_Count: Number of nitrogen atoms in the molecule.
- Jurs descriptors combine shape and electronic information to characterize molecules (Stanton and Jurs, 1990). The descriptors are calculated by mapping atomic partial charges on solvent-accessible surface areas of individual atoms. Jurs_RASA is the total hydrophobic surface area divided by the total molecular solvent-accessible surface area.

5.1.1. Validation of the QSAR model

5.1.1.1. Statistical Analysis

For the internal validation, Leave-One-Out (LOO-) cross validation method was used (Tetko et al., 2001). Cross-validation is a practical and reliable method in

generating QSAR models. By using this, the predictive powers of the equations were validated. The cross-validated coefficient, q^2 , was calculated using

$$q^2 = 1 - \frac{\sum (Y_{\text{predicted}} - Y_{\text{observed}})^2}{\sum (Y_{\text{observed}} - Y_{\text{mean}})^2},$$

where $Y_{\text{predicted}}$, Y_{observed} and Y_{mean} are predicted, actual and mean values of the target property (pIC50), respectively. Here $\sum (Y_{\text{predicted}} - Y_{\text{observed}})^2$ is the predictive sum of squares (PRESS). Squared predicted correlation coefficient (R^2) cross-validated R^2 (Q^2), standard deviation based on PRESS (SPRESS) and standard error of prediction (s) were considered for the validation of these models.

The best models that were produced are as follows:

$$\begin{aligned} \text{pIC50} = & 14.543(\pm 4.75) - 41.02(\pm 13.3)\text{Dielectric_Energy_DMol3} - & (6) \\ & 0.711(\pm 0.42)\text{N_Count} + 3.612(\pm 0.61)\text{Ix} \\ & - 14.21 (\pm 6.41)\text{Shadow_XYfrac} \end{aligned}$$

(n=20; r=0.982; s=0.319; F=98.743; Q^2 =0.941; s-Press= 0.406)

$$\begin{aligned} \text{pIC50} = & 19.17(\pm 2.86) - 8.188(\pm 2.56)\text{Jurs_RASA} + 2.527(\pm 0.41)\text{Ix} - & (7) \\ & 12.16(\pm 5.55)\text{Shadow_XYfrac} \end{aligned}$$

(n=20; r=0.981; s=0.309; F=139.785; Q^2 =0.936; s-Press=0.408)

In the regression equation, n is the number of compounds considered, r is the correlation coefficient, s is the standard error of the estimate, F is the Fisher F-test of significance obtained from the ratio of explained to unexplained variance of the dependent variable, and q is the cross validated correlation coefficient derived from the predictive residual sum of squares (PRESS, leave-one-out method) (See Table A.1 in appendix for critical values for F test and calculations).

Figure 5.1 shows a comparison of the experimental and predicted activity for the 20 compounds taken in the training set with the 95% confidence regions. The R^2 (squared correlation coefficient between observed and the predicted activity) was

found to be 0.96 for both equations. Table 5.1 lists the observed and the model-predicted activity of the training set (20 compounds).

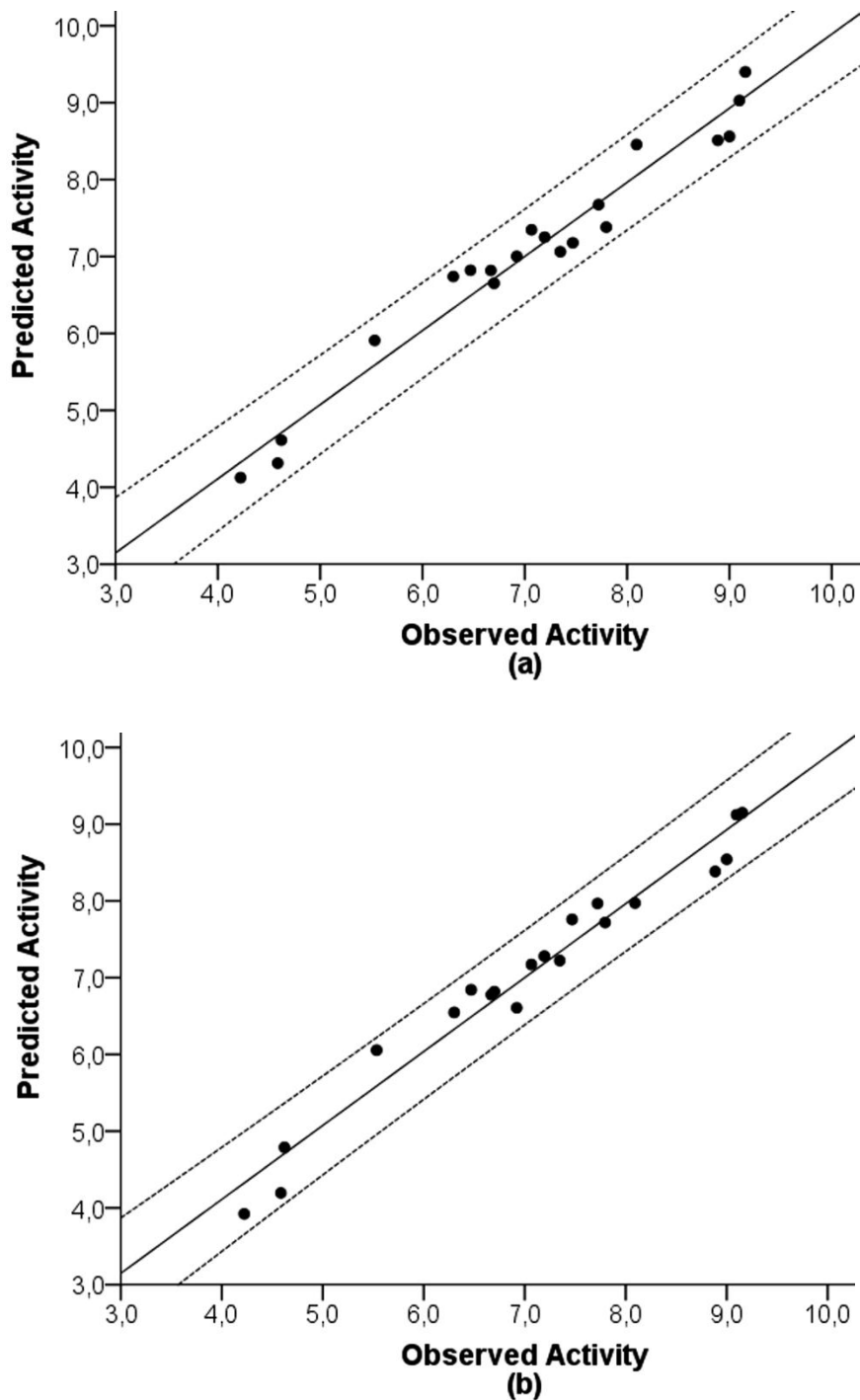


Figure 5.1. Comparison of observed and predicted activities of the compounds in the training set using Eq. 1 (a) and Eq. 2 (b) (dotted lines indicate 95% confidence limits).

Besides a good linear correlation ($r = 0.982$ for Eq. (6) and 0.981 for Eq. (7)) and an acceptable standard deviation ($s = 0.319$ for Eq. (6) and 0.309 for Eq. (7)), Eq. (6) and Eq. (7) demonstrates both high predictability ($q^2 = 0.941$ and 0.946 respectively) and significance ($F = 98.743$ and 139.785 respectively).

The negative coefficient of DE in Eq. (6) and that of JURSRASA in Eq. (7) indicates that the increase of either DE or JURSRASA (hydrophobic surface area (RASA)) is detrimental to biological activity. The negative contribution of N_Count in Eq (6) suggests that with increase in the number of N atom, CDK2 inhibitory activity will be lower. However, $|\lambda|$ values in both equations suggest that the inhibitory effect can be enhanced especially by applying NO substituent on the molecule. The observations that the 5-nitroso group forms an intramolecular hydrogen bond with the 6-amino group and make one of the amino NH bonds to interact with the backbone carbonyl of Glu 81 of CDK2 support these results (Arris et al., 2000 and Mesguiche et al., 2003). Shadow_XY fraction shows a negative contribution to biologic activity. This geometric shadow index descriptor is related to the breadth of a molecule that characterizes the shape of the molecules based on their conformation and their spatial orientation.

In an effort to see binding modes of training set compounds, docking studies were performed. **2j** was taken as reference structure due to the fact that its experimental x-ray crystal can be found in literature. In Figure 5.2, experimental x-ray crystal structure of **2j** bound to the CDK2/cyclin A phosphorylated on Thr160 (T160pCDK2/cyclin A) was compared with the final structure found after calculations. The pose with the lowest CDOCKER energy was selected. Besides some conformational changes in some regions, **2j** makes the same hydrogen bonds with the ATP binding site as observed in experimental crystal structure. Most of the active compounds were found that they shared similar binding modes, highlighting the importance of their spatial pose in the CDK2 inhibition. Large negative coefficient of Shadow_XY fraction reveals that the selectivity increased with the decreased value of Shadow_XY fraction. In other words, the smaller area of molecular shadow in the enclosing rectangle will benefit the activity.

Table 5.1. Observed and predicted pIC50 values of training set compounds obtained from the Eqs.(6,7).

Compound	Observed pIC50 (in M)	Calculated pIC50 (in M) Eq (6)	Residuals	Calculated pIC50 (in M) Eq (7)	Residuals
3a	4.585	4.313	-0.272	4.194	-0.391
3b	4.222	4.125	-0.097	3.922	-0.300
3c	4.62	4.613	-0.006	4.789	0.169
3e	5.533	5.909	0.376	6.055	0.522
2a	6.668	6.818	0.15	6.777	0.109
2b	6.301	6.739	0.438	6.547	0.246
2c	7.796	7.381	-0.415	7.719	-0.077
2d	6.469	6.821	0.352	6.842	0.373
2f	6.921	7.002	0.082	6.607	-0.313
2h	7.194	7.251	0.057	7.28	0.086
2i	7.347	7.062	-0.285	7.223	-0.123
2j	7.469	7.178	-0.29	7.759	0.291
2l	6.699	6.652	-0.046	6.818	0.119
2m	9	8.561	-0.439	8.542	-0.458
2n	7.066	7.347	0.281	7.171	0.105
2o	8.092	8.455	0.363	7.972	-0.120
2p	7.721	7.674	-0.047	7.967	0.246
2q	9.155	9.399	0.244	9.148	-0.007
2r	8.886	8.511	-0.375	8.384	-0.502
2s	9.097	9.028	-0.069	9.123	0.026

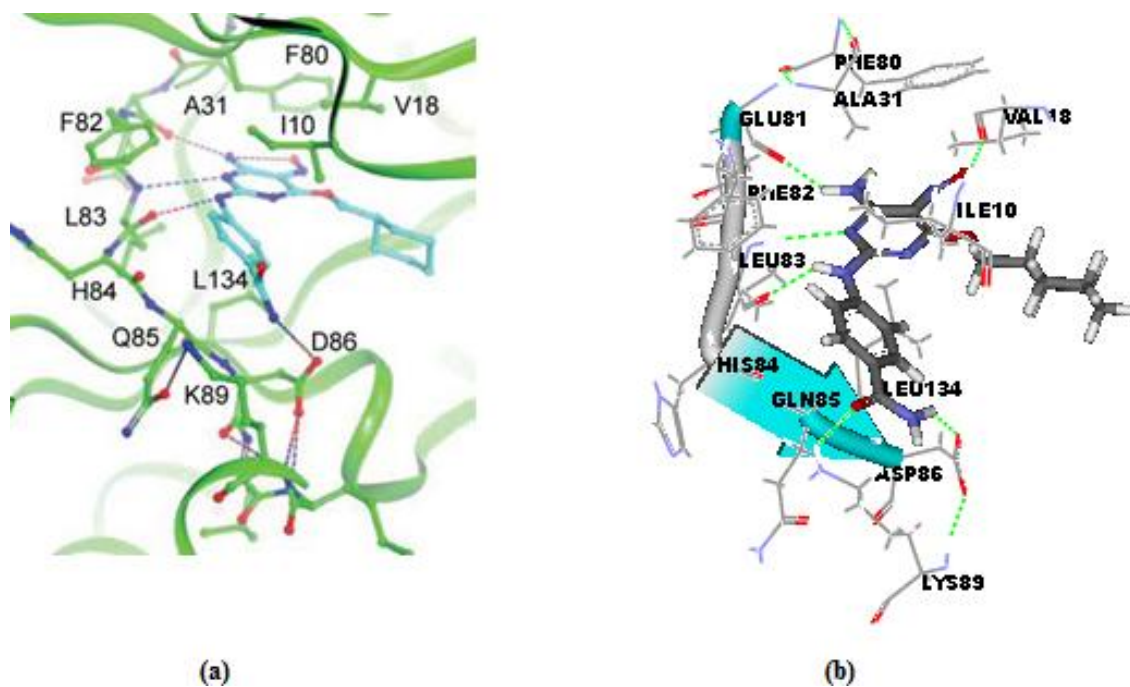


Figure 5.2. a) Experimental crystal structure of 2j bound to Thr160pCDK2/cyclin A, overall fold. (Marchetti et al., 2007) b) Calculated binding mode of 2j bound to Thr160pCDK2/cyclin A. (RMSD value was calculated as 0.5051 Å)

Prediction of the test set

Predictive r^2

The predictive ability of each QSAR model was determined from a set of five compounds that were not included in the training set. The predictive ability of the models is expressed by the predictive r^2 value, which is analogous to the cross-validated r^2 value (q^2) and is calculated using the formula:

$$r^2_{\text{pred}} = \frac{SD - \text{PRESS}}{SD},$$

where SD is the sum of the squared deviations between the biological activities of the test set molecules and PRESS is the sum of the squared deviations between the observed and the predicted activities of the test molecules.

On the basis of the models build using training set, predicted values of five compounds selected in the test were calculated. Table 5.2 lists the observed and the model-predicted activity of the test set (5 compounds). Figure 5.3 shows a

comparison of the experimental and predicted activity for the five compounds selected in the test set with the 95% confidence regions. In both cases the conventional statistical results and the predictive ability of the models for the 5 test compounds shows a good correlation between experimental and calculated pIC50 values meaning a respectable degree of predictability for the test set molecules was produced. ($R^2=0.82$, $r^2_{\text{pred}}=0.53$ for Eq. (1) and $R^2=0.83$, $r^2_{\text{pred}}=0.56$ for Eq. (2)).

Table 5.2. Observed and predicted pIC50 values of test set compounds obtained from the Eqs.(6-7).

Compound	Eq. (6)			Eq. (7)	
	pIC50 Observed (in M)	pIC50 Calculated (in M)	Residuals	pIC50 Calculated (in M)	Residuals
1	8.658	6.991	-1.666	7.225	-1.433
3d	4.229	4.64	0.411	5.362	1.132
2e	6.398	6.38	-0.018	6.121	-0.277
2g	7.481	6.032	-1.45	6.291	-1.191
2k	7.097	6.684	-0.413	6.973	-0.124

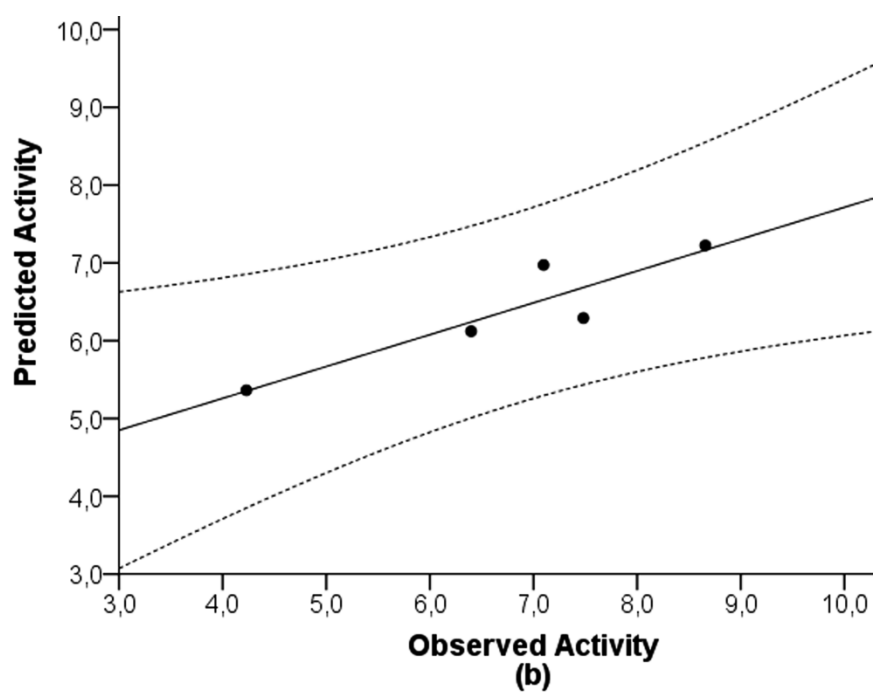
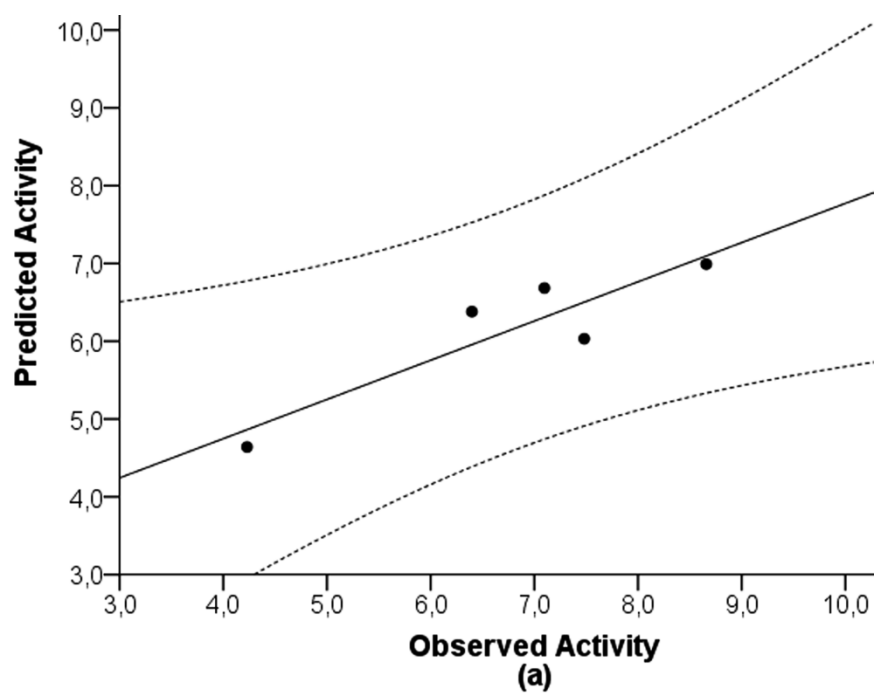


Figure 5.3. Comparison of observed and predicted activities of the compounds in the test set using Eq. 1 (a) and Eq. 2 (b) (dotted lines indicate 95% confidence limits).

5.2. PHARMACOPHORE STUDY

Specific guidelines are necessary to choose appropriate training set molecules in order to establish a good pharmacophore model. These includes the activity range and the structural diversity of the selected compounds. The difference in the training set selection has a large impact on the final pharmacophore model. A possible case is that distinctive pharmacophore models of ligands interacting with the same protein target could be generated with the use of the same program but different training set (Zou et al., 2008). For instance, three different ligand-based pharmacophore models for CDK2 inhibitors have been independently reported (Hecker et al., 2002; Toba et al., 2006 and Vadivelan et al., 2007) (Figure 5.4).

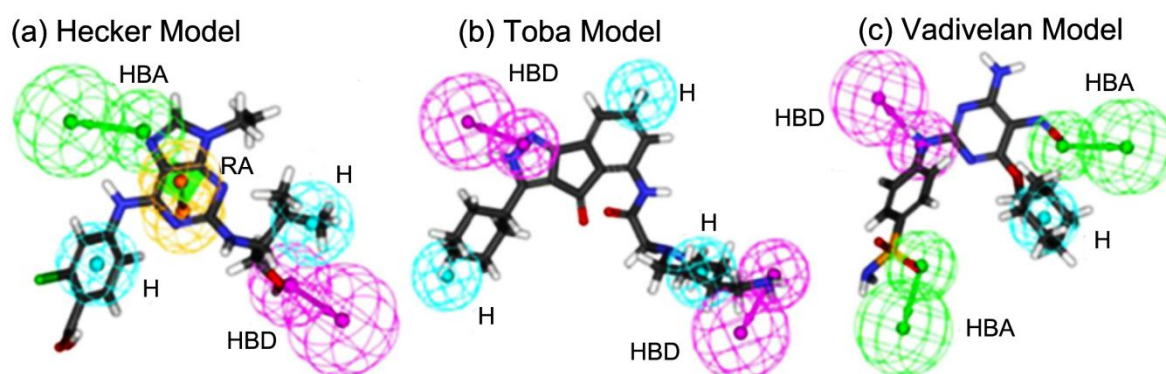


Figure 5.4. The ligand-based pharmacophore models of CDK2 inhibitors reported previously by (a) Hecker, (b) Toba and (c) Vadivelan. (HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; H, hydrophobic feature; RA, ring aromatic feature.). (Figure taken from Zou et al., 2008 and edited)

Apart from these three models, a more accurate pharmacophore model using recent CDK2 inhibitors was aimed to be generated in the following part. Results will be compared with the recent Vadivelan model that used catalyst software as well.

5.2.1. Hypothesis Generation

CATALYST allows the use of structure and activity data for a set of lead compounds to generate a hypothesis describing the activity of the lead set. In order to acquire a reliable model adequately describing the interaction of ligands with high predictability, the CATALYST procedure suggests a collection of 15-25

chemically diverse molecules with biological activity covering 4-5 orders of magnitude for the training set.

The hypotheses are described by a set of functional features like hydrogen bond donor, hydrogen bond acceptor, hydrophobic and positively and negatively ionizable sites distributed over a 3D space. Except the hydrogen bonding features that are vectors, all other functions are points (Bhattacharjee et al., 2002).

5.2.2. Validation of Hypothesis

Validating the hypothesis is one of the important methods in pharmacophore generation. There are several methods to confirm the quality of pharmacophore like preparing test set, Fischer's randomization method, correlation between the predicted and estimated values of test set etc.

The objective of pharmacophore hypothesis validation is to determine whether or not our hypothesis is powerful enough to identify active compounds from a database and able to predict their activity values accurately.

The statistical relevance of a generated hypothesis is estimated on the basis of their cost relative to the null hypothesis and their correlation coefficient. In other words, examining the cost values from the results of the pharmacophore generation process assists understanding the validity of generated pharmacophore hypotheses. Three cost values named fixed cost, total cost and null cost are generated by HypoGen during pharmacophore generation process. Every single pharmacophore generation run produces maximum of ten pharmacophore hypotheses together with their fixed and null cost values which are usually used to determine the quality of any pharmacophore hypothesis.

Debnath (Debnath, 2002) gives a detailed description regarding the quality of the HypoGen models in the sense of fixed cost, null cost, and total cost and other statistical parameters. Fixed cost (known as ideal cost as well) is the cost of the simple possible hypothesis. Fixed total cost is dependent on addition of the cost components: weight cost, error cost and configuration cost. It fits all the data perfectly. The null cost (also known as no correlation cost) is the cost predicted as mean activity. It acts as if it is a hypothesis with no features. All the generated pharmacophore hypotheses situate between fixed and null cost values.

For cost analysis, two fundamental values are used: one is the difference between the fixed and null costs and another one is the difference between null and total cost (cost difference). The fixed cost depicts a cost of the theoretical ideal hypothesis, which could definitely estimate the activity of compounds in the training set with lowest deviation, whereas null cost represents the cost of hypothesis with no features that predicts every activity to be the average activity. A best pharmacophore hypothesis should have its total cost closer to the fixed cost value and away from the null cost value. As stated by Debnath (Debnath, 2002), a large difference between the fixed cost and null cost, and a value of 40-60 bits for the unit of cost would suggest a 75-90% probability for experimental and estimated activity correlation. In order a hypothesis to be a good model, the total cost should be close to the fixed cost.

Correlation coefficient, RMSD and configuration cost are the other important parameters used to determine the quality of the pharmacophore hypothesis. Configuration cost should be smaller than 17 for a good pharmacophore hypothesis since it represents the complexity of the hypothesis while RMSD explains the quality of the correlation between the predicted and experimental activity values (Sakkiah et al., 2010).

Compared with other hypotheses, Hypo1 which consist of two HBA, one HBD and one H, establishes the highest cost difference (102.93), lower errors (52.05), best correlation coefficient (0.98) and lowest RMSD of 0.84. The fixed and the null cost values are 63.83 and 173.37, respectively (Table 5.3). Hence, Hypo1 was selected as a best hypothesis and employed for further analyses. Figure 5.5 shows, the Hypo1 chemical features with its geometric parameters. The Vadivelan pharmacophore hypothesis had cost difference of 59.55 bits, the RMSD value of 0.827 and correlation of 0.949.

Table 5.3. Information of statistical significance and predictive power presented in cost values measured in bits for the top 10 hypotheses as a result of automated 3D QSAR pharmacophore generation.

Hypo no.	Total cost	Cost difference ^a	RMSDb	Correlation	Features ^b
Hypo1	70.44	102.93	0.84	0.98	HBA, HBA, HBD, H
Hypo2	78.06	95.31	1.24	0.95	HBA, HBA, HBA
Hypo3	78.24	95.13	1.22	0.95	HBA, HBA, H
Hypo4	80.95	92.42	1.45	0.93	HBA, HBA, HBD, H
Hypo5	82.33	91.04	1.43	0.93	HBA, HBA, HBD
Hypo6	82.37	91.00	1.43	0.93	HBA, HBA, HBD
Hypo7	84.64	88.73	1.60	0.92	HBA, HBA, H
Hypo8	84.94	88.43	1.51	0.93	HBA, HBA, HBA
Hypo9	85.10	88.27	1.58	0.92	HBA, HBA, HBA
Hypo10	85.48	87.89	1.50	0.93	HBA, HBA, HBA

^a Cost difference between the null and the total cost. The null cost, the fixed cost and the configuration cost are 173.374, 63.83 and 16.2563 respectively.

^b Abbreviation used for features: RMSD, root mean square deviation; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; H, hydrophobic.

The most active and inactive compounds in the training set were aligned in Hypo1 was shown in Figure 5.6.

To verify the prediction accuracy of Hypo1, training set was used and the activity of each compound in training set was estimated by regression analysis. Training set compounds were classified relatively into three sets based on their activity values: $IC_{50} < 300 \text{ nM} = +++$ (highly active); $300 \text{ nM} < IC_{50} < 3000 \text{ nM} = ++$ (moderately active); $IC_{50} \geq 3000 \text{ nM} = +$ (low active). Only one inactive compound was predicted as moderate and all the remaining highly active compounds and moderately active compounds in the training set were estimated correspondingly and all inactive compounds were estimated as inactive by Hypo1. Thus Hypo1 was able to estimate the activities of compounds in their own activity ranges. The

experimental and estimated activities by Hypo1 for 16 training set compounds are shown in Table 5.4.

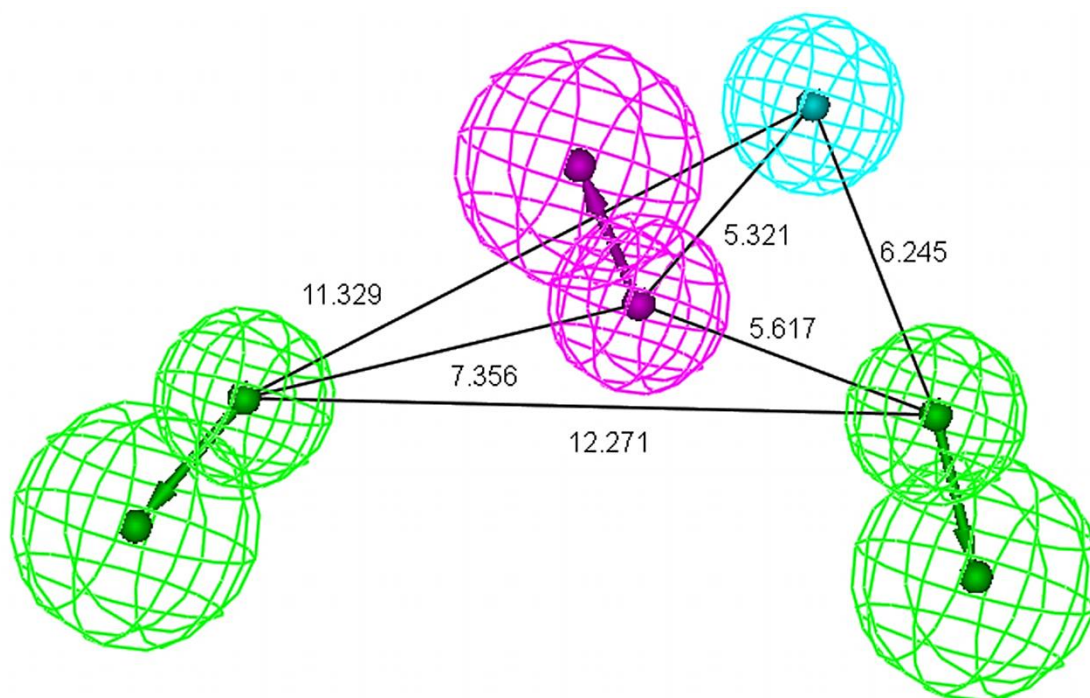


Figure 5.5. Catalyst Hypogen pharmacophore model, where H, HBA and HBD are illustrated in cyan, green and pink, respectively.

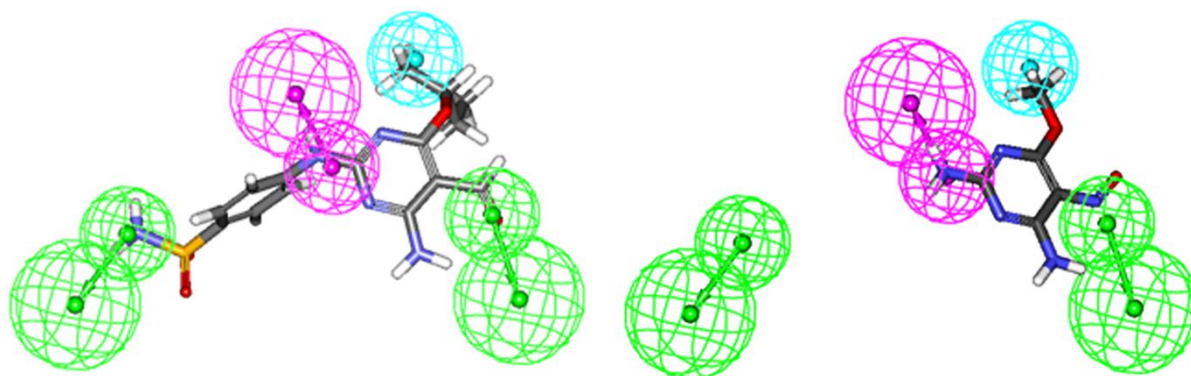


Figure 5.6. Best pharmacophore model Hypo1 aligned to training set compound A) Active molecule 4 (IC₅₀ 13.1 nM) B) Inactive molecule 14 (IC₅₀ 79,000 nM). Pharmacophore features are colour coded (H, hydrophobic, cyan, HBA, hydrogen bond acceptor, green and HBD, hydrogen bond donor, pink).

Table 5.4. Actual and estimated activity of the training set molecules based on the pharmacophore model Hypo1.

Compound No	Fit Value	Exp. IC50 (nM)	Pred. IC50 (nM)	Error	Exp. Scale	Pred. scale
4	8.73	13.1	20.0	+1.5	+++	+++
2g	8.56	33	29.7	-1.1	+++	+++
5	8.38	49	45.0	-1.1	+++	+++
6	7.98	93	111.4	+1.2	+++	+++
7	7.76	117	188.4	+1.6	+++	+++
2l	7.64	200	248.5	+1.2	+++	+++
2e	7.02	400	1,023.4	+2.6	++	++
8	7.41	610	415.9	-1.5	++	++
9	6.57	1,800	2,872.0	+1.6	++	++
10	7.05	2,200	960.6	-2.3	++	++
3e	6.60	2,930	2,713.4	-1.1	++	++
11	6.64	8,300	2,444.0	-3.4	+	++
3c	5.55	24,000	30,399.4	+1.3	+	+
12	5.53	55,000	31,757.3	-1.7	+	+
13	5.57	61,000	29,097.8	-2.1	+	+
14	4.79	79,000	174,766.0	+2.2	+	+

^a Difference between the predicted and experimental values. '+' indicates that the predicted IC50 is higher than the experimental IC50; '-' indicates that the predicted IC50 is lower than the experimental IC50;

^b Fit value indicates how well the features in the pharmacophore overlap the chemical features in the molecule. Fit Value = weight x [max(0.1 - SSE)] where $SSE = (D/T)^2$, D = displacement of the feature from the center of the location constraints and T = the radius of the location constraint sphere for the feature (tolerance) (Sakkiah et al., 2010).

^c Activity scale: IC50 < 300 nM = +++ (highly active); 300 nM < IC50 < 3000 nM = ++ (moderately active); IC50 ≥ 3000 nM = + (low active).

In addition to the cost analysis, the best pharmacophore hypothesis was also validated using external test set compounds and Fischer's validation method.

4.2.2.1. Fischer's randomization method

Fischer's randomization was used to evaluate the statistical significance of Hypo1. Validation was done by producing random spreadsheets for training set molecules, which randomly reassigned activity values to each compounds and afterwards produced the hypotheses using the same features and parameters originated for Hypo1. To obtain the confidence level of 95%, 19 random spreadsheets (random hypotheses) were generated. The significance of the hypotheses was computed using the following formula: $[1 - (1 + X)/Y] \times 100$, where X, total number of hypotheses having a total cost lower than Hypo X and Y, total number of Hypogen runs (initial + random runs). Here, $X = 0$ and $Y = (19 + 1)$, $S = [1 - ((1 + 0)/(19 + 1))] \times 100\% = 95\%$.

Figure 5.7 clearly shows that the Hypo1 hypothesis was not generated by chance, because its statistics are far more superior to all random hypotheses.

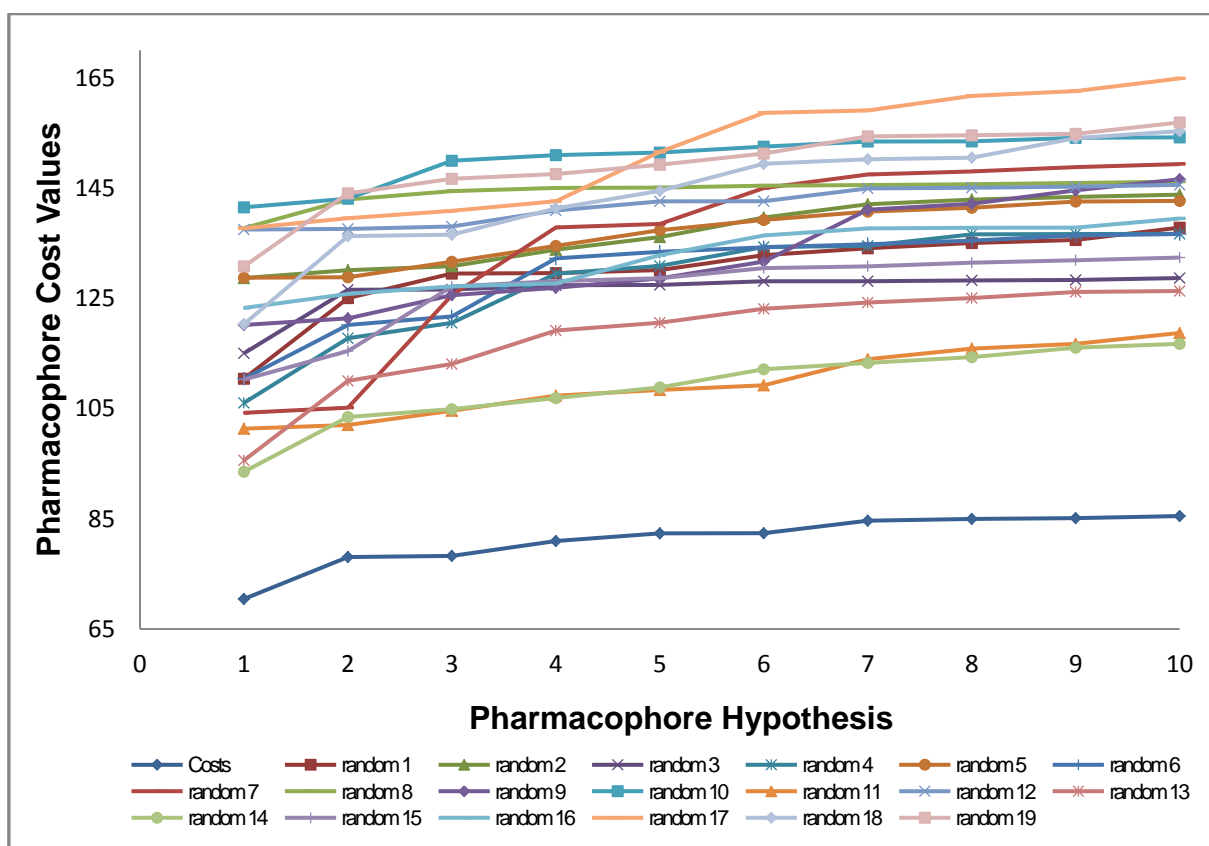


Figure 5.7. The difference in costs between HYPOGEN runs and the scrambled runs. The 95% confidence level was selected.

The correlation values of these 19 models are plotted along with those of the true hypothesis in Figure 5.8. Interestingly, none of the HypoGen models generated with randomized activity data gave higher correlation values than those of the true pharmacophore hypothesis.

As a consequence, none of these 19 new hypotheses had lower cost values and RMSD values than the true hypothesis. This test reveals that this pharmacophore hypothesis was not generated by chance, and that there is a probability of at least a 95% that it shows a valid correlation between structure and CDK2 inhibitory activity.

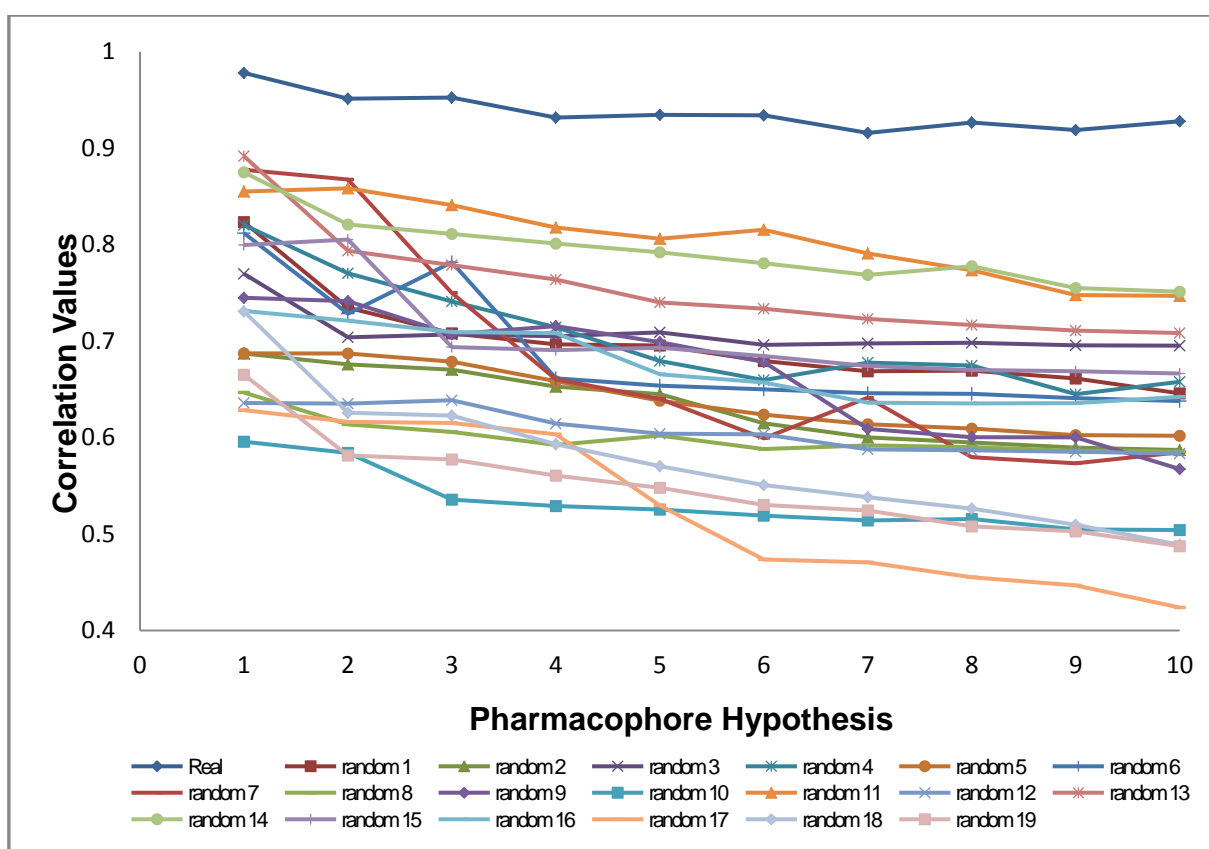


Figure 5.8. Fisher's randomization test results. Correlation values for a particular pharmacophore across the tests correspond to a line. The highest line displays the correlation values for the unscrambled pharmacophore.

4.2.2.2. Test set

Secondly, a test set was prepared using the same protocol as training set prepared and used to determine whether the hypothesis was able to predict the active compounds other than the training set molecules. Test set compounds were classified relatively into three sets based on their activity values: highly active $IC_{50} < 300$ nM = +++ (highly active); 300 nM $< IC_{50} < 3000$ nM = ++ (moderately active); $IC_{50} \geq 3000$ nM = + (low active). Except one moderately active compound that was predicted as inactive compound, all the remaining highly active compounds and moderately active compounds were estimated correspondingly and all inactive compounds were estimated as inactive by Hypo1. Thus Hypo1 was able to estimate the activities of compounds in their own activity ranges. The experimental and predicted activities of Hypo1 as applied to test set are shown in Table 5.5. This result was used for further legalization of Hypo1 suggesting that the Hypo1 not only fit for training set compounds but also for the external compounds.

The activity of the test set molecules was also scored using the best hypothesis generated, and the results gave a correlation value of 0.96. In Vadivelan model, the activity of the test set molecules was scored using the best hypothesis generated, and the results gave a correlation value of 0.912. The plot showing the correlation between the actual and predicted activity for the test set and the training set molecules is given in Figure 5.9. This indicates that the pharmacophore model generated is capable of predicting the activity of the unknown molecules with reasonable accuracy.

Table 5.5. Experimental and predicted IC50 data values of 14 test set molecules against Hypo1.

Compound No	Fit Value	Exp. IC50 (nM)	Pred. IC50 (nM)	Error	Exp. Scale	Pred. scale
2o	8.96	8.1	11.9	+1.5	+++	+++
2p	8.32	19	51.2	+2.7	+++	+++
15	8.16	26	73.9	+2.8	+++	+++
2j	8.45	34	38.2	+1.1	+++	+++
2i	7.60	45	267.8	+6	+++	+++
16	8.28	49	56.5	+1.2	+++	+++
2n	8.32	86	51.5	-1.7	+++	+++
17	8.46	110	37.1	-3	+++	+++
2d	7.09	340	881.0	+2.6	++	++
1b	6.84	500	1,552.6	+3.1	++	++
18	6.47	1,900	3,647.6	+1.9	++	+
19	5.69	8,300	21,916.0	+2.6	+	+
3a	5.57	26,000	29,191.3	+1.1	+	+
20	5.56	56,300	29,329.6	-1.9	+	+

^a Difference between the predicted and experimental values. '+' indicates that the predicted IC50 is higher than the experimental IC50; '-' indicates that the predicted IC50 is lower than the experimental IC50; a value of 1 indicates that the predicted IC50 is equal to the experimental IC50.

^b Fit value indicates how well the features in the pharmacophore overlap the chemical features in the molecule. Fit Value = weight x [max(0.1 - SSE)] where $SSE = (D/T)^2$, D = displacement of the feature from the center of the location constraints and T = the radius of the location constraint sphere for the feature (tolerance) (Sakkiah, 2010).

^c Activity scale: IC50 < 300 nM = +++ (highly active); 300 nM < IC50 < 3000 nM = ++ (moderately active); IC50 ≥ 3000 nM = + (low active).

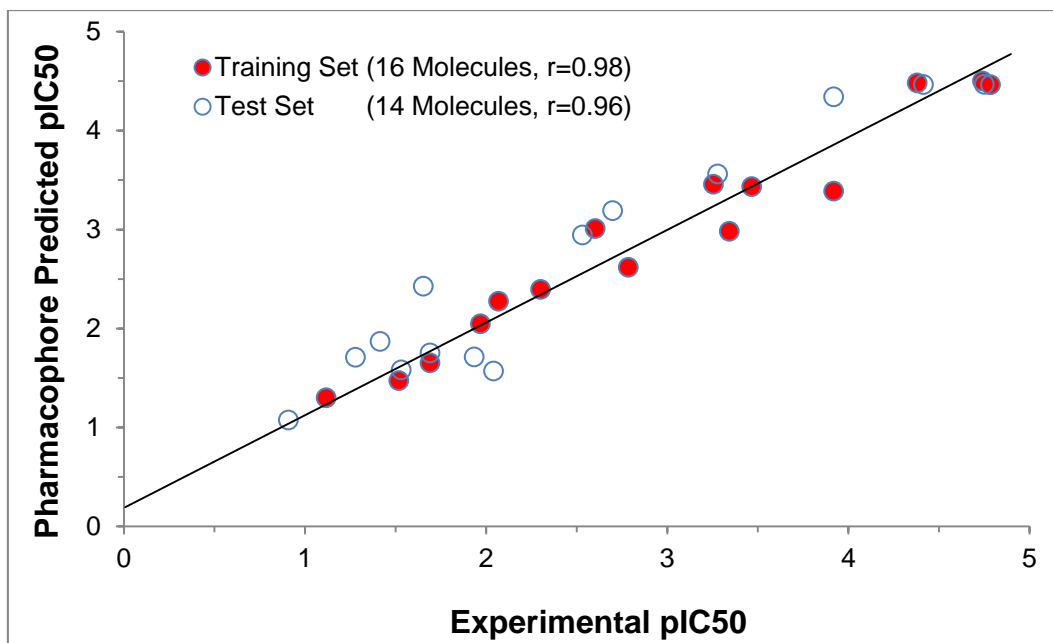


Figure 5.9. Graph showing the correlation (r) between experimental and predicted activities for the 14 test set molecules against the Hypo-1 model along with 16 training set molecules for CDK2 inhibitors.

5.2.3. Databases screening

Virtual screening of small-molecule libraries forms one side of a complex approach to drug discovery (Marcu et al., 2000). In drug discovery process virtual screening of databases is an effective option to high throughput screening (HTS). Combination of the virtual screening approach with pharmacophore modeling, molecular docking, and consensus scoring function can be used to identify and design novel CDK2 inhibitors with higher selectivity. It is believed that the molecular docking in agreement with consensus scoring functions could readily minimize false positive and false negative errors encountered by ligand-based (pharmacophore) virtual screening. Also, the complementation of molecular docking and pharmacophore can generate reliable true positive and true negative results in the following virtual screening procedure. The proper use of these methods in a drug discovery process should improve the ability to identify and optimize hits and approve their potential to serve as scaffolds for producing new therapeutic agents (Lu et al., 2011).

Three-dimensional query represented by the pharmacophoric model Hypo1 retrieved 84,040 hits from Life Chemicals database and 20,621 hits from NCI2003

database. These hits were then filtered according to the combination of Lipinski's rule of five and the improvements to this rule made by Ghose et al. (Ghose, 1998) and Veber et al. (Veber, 2002) mentioned in section 2. Finally, highly active molecules ($IC_{50} < 300nM$) were selected for further docking studies. A summary is given in Table 5.6.

Table 5.6. Virtual Screening Results

Database Name	Hit Compounds	Lipinski, Ghose and Veber Filtering	Highly Active Molecules ($IC_{50} < 300nM$)
Life Chemicals	84,040	57,523	590
NCI 2003	20,621	6,728	95

5.2.3. Molecular Docking Studies

Molecular docking is a computational method that samples conformations of small compounds in protein binding sites. Scoring functions are used to estimate which of these conformations were best complements to the protein binding site (Sakkiah et al., 2010).

Docking studies were performed following the procedure mentioned in section 3.4. 16 training set compounds together with 685 hit compounds retrieved from the databases which satisfied drug like properties were docked in the active site of CDK2.

Despite the fact that the docking program performs well in predicting and generating the correct biologically active conformation of a ligand, the present scoring functions are less successful at correctly identifying the same (Mascarenhasa and Ghoshal, 2008). It was very necessary from a virtual screening view that a methodology is adopted to filter out inactive molecules based on docking and scoring procedures. At first, the possibility of using a single scoring function for this purpose was taken into account. But since none of the investigated scoring functions had any worthwhile correlation between predicted binding affinity and the experimentally observed biological activity for the training

set compounds, another approach to address this issue was considered necessary.

4.2.3.1. Consensus scoring (Accelrys)

Instead of ranking a set of known binders, the primary objective for VS applications is to identify possible binders from available databases of compounds. Considering this objective, using of consensus scoring functions was imperative.

Ligand scoring is a technique that rapidly estimates the binding affinity of a ligand, based on a candidate ligand pose geometry docked into a target receptor structure. Scoring methods typically use empirical functions built by fitting several functional forms that characterize various aspects of the receptor-ligand interactions against binding affinity data. The use of statistical analysis of known ligand-receptor structures and the frequency of occurrence of specific receptor-ligand interactions without requiring any information about binding affinities is an alternative approach. This approach is usually referred to as a knowledge-based approach. Both types of scoring function method are available in the discovery studio that includes:

- Jain
- LigScore1
- Piecewise Linear Potential (PLP)
- Ludi
- LigScore2
- Potential of Mean Force (PMF)

The first four of the methods were built using the empirical fitting approach. The PMF function was developed using the knowledge-based statistical approach. The PLP function was at first developed as a docking function, but has been shown to correlate well with binding affinities.

The Consensus Score protocol implemented in discovery studio calculates the consensus scores of a series of docked ligands for which other scores have been computed before. Consensus scoring is a fast way to identify ligands that score high in more than one scoring function. The ligands are listed by score in descending order for each selected scoring function. The consensus score for a molecule is equal to its frequency of occurrence in the top rank percentile of each scoring function.

The very purpose of using a consensus scoring strategy is to pick up actives from huge populations of inactives, also known as “finding needles in a haystack” problem. The scaled scores were then summed up to give the final consensus scores.

In this study, eleven scoring functions were used in order to evaluate hit compounds. The 11 hits with the high binding affinities were ultimately selected after careful observations, analyses, comparisons, and consensus scores (Figure A.1 in appendix). For each ligand the pose with the highest consensus score were selected (All scoring values and critical molecular properties that are necessary for drug likeliness that belong to 11 hit compounds and their 10 poses are given in Table A.2 in appendix). As represented in Table 5.7 all hit compounds have high hit values and make critical hydrogen bonds with CDK2. Among these hit compounds, NSC_649153 and F5882-6930 have the highest activity values and they also make hydrogen bond especially with the GLU 81 and LEU 83 that are critical in inhibition of CDK2. Figure 5.10 represents the binding orientation of the hit compounds NSC_649153 and F5882-6930 also shows how well the compounds fit into Hypo1.

Table 5.7. Selected hit compounds and binding modes.

Name	pIC50 nM	FitValue	CScore	Hydrogen Bond
NSC_649153	75.441	8.154	8	GLU 81 LEU 83 LEU83
NSC_221631	79.75	8.13	8	GLU 81 GLN 85 ASP86
F5882-6930	112.873	7.979	9	GLU 81 LEU 83 ASP86
F5607-0191	127.196	7.927	8	LEU 83 LEU 83 HIS 84 ASP86
F5689-0078	129.083	7.92	9	LEU 83 LEU 83 HIS 84
F5382-0550	153.758	7.844	10	LEU 83 LEU 83
F3222-4739	169.63	7.802	8	LEU 83 LEU 83
F5485-0325	179.782	7.777	10	LEU 83 LEU 83
F5463-0082	231.78	7.666	9	LEU 83 LEU 83 HIS 84
F5736-0414	270.883	7.599	10	LEU 83 LEU 83
F3407-2898	286.601	7.574	10	LEU 83 LEU 83

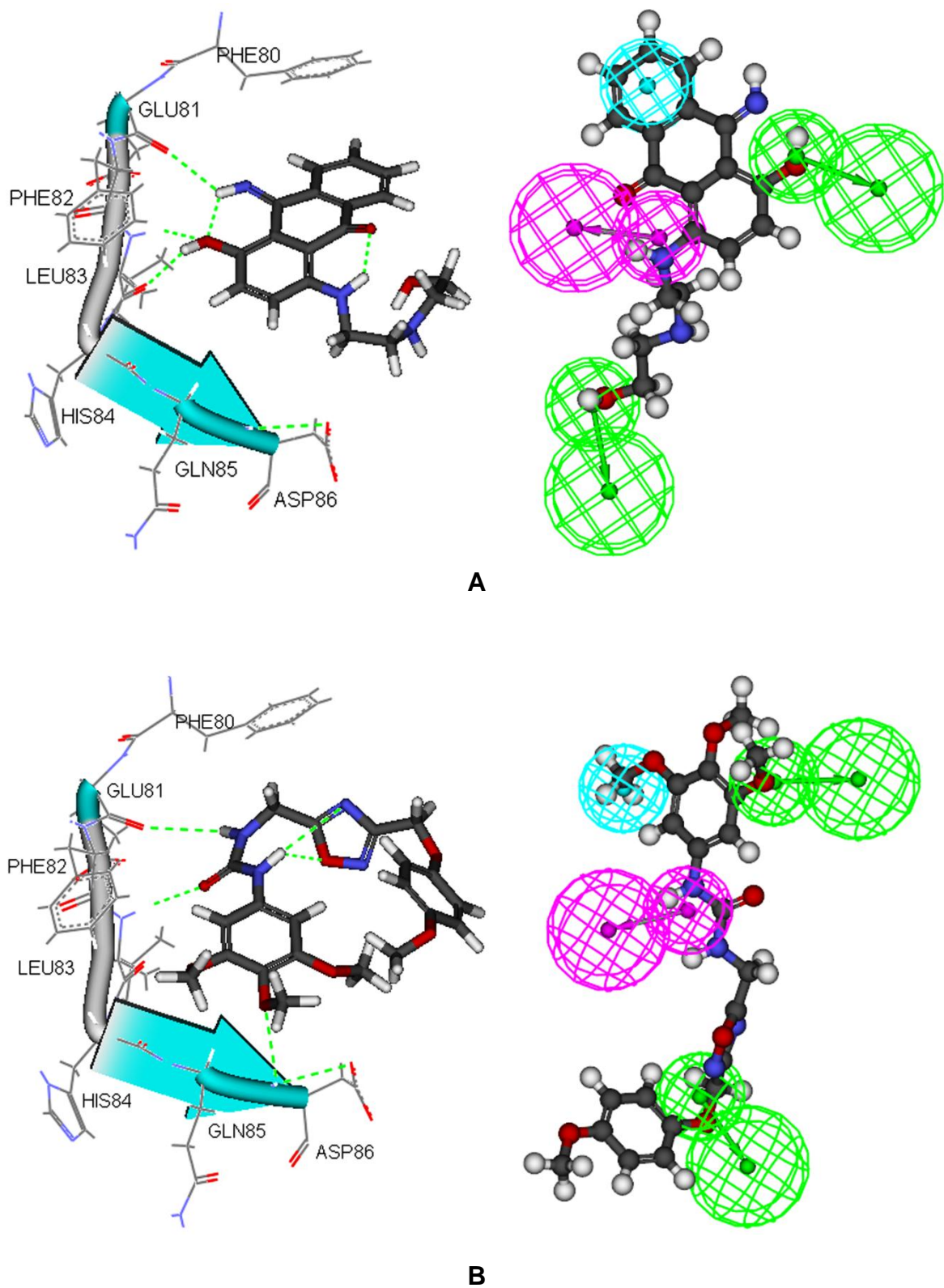


Figure 5.10. Binding orientation of the hit compounds NSC_649153 (A) and F5882-6930 (B) and mapping of these compounds onto the pharmacophore model Hypo1.

6. CONCLUSION

In this study, a systematic computer aided drug design of some CDK2 inhibitors was described using different techniques that include:

- Quantitative structure activity relationship study,
- 3D QSAR pharmacophore study:
 - Pharmacophore generation,
 - Virtual Screening,
 - Molecular Docking.

In QSAR:

Significant regression equations were obtained by multiple linear regression method for 25 pyrimidine analogue compounds according to their CDK2 inhibitory activity:

$$\text{pIC}_{50} = 14.543(\pm 4.75) - 41.02(\pm 13.3)\text{Dielectric_Energy_DMol3} - 0.711(\pm 0.42)\text{N_Count} + 3.612(\pm 0.61)\text{Ix} - 14.21(\pm 6.41)\text{Shadow_XYfrac}$$

$$(n=20; r=0.982; s=0.319; F=98.743; Q^2=0.941; s\text{-Press}=0.406)$$

$$\text{pIC}_{50} = 19.17(\pm 2.86) - 8.188(\pm 2.56)\text{Jurs_RASA} + 2.527(\pm 0.41)\text{Ix} - 12.16(\pm 5.55)\text{Shadow_XYfrac}$$

$$(n=20; r=0.981; s=0.309; F=139.785; Q^2=0.936; s\text{-Press}=0.408)$$

The QSAR models indicate that the dielectric energy, number of nitrogen atoms, NO substituent, Shadow_XYfrac and Jurs_RASA play important roles for the CDK2 receptor antagonist activities. The model developed was validated by cross validation techniques and external test set prediction. The models obtained

showed not only statistical significance but also predictive ability and revealed that minor values for DE combined with an NO substituent lead to an increasing of CDK2 inhibitory activity. Also having a large negative coefficient for Shadow_XY fraction highlights the importance of active compounds' spatial pose in the CDK2 inhibition. These models give insight on structural requirements for designing more potent analogues against CDK2.

In 3D QSAR pharmacophore Study:

Pharmacophore studies indicate that the best inhibitor model consists of (1) two hydrogen bond acceptors,(2) one hydrogen bond donor, and (3) one hydrophobic feature. The most-active molecule in the training set fits very well with the top scoring pharmacophore hypothesis. The pharmacophore model was further used to screen potential compounds from the Life Chemicals and NCI databases using virtual screening. Molecular docking and consensus scoring methods were used, as added tools for virtual screening to minimize false positive and false negative errors. Using a combination of pharmacophore modeling, virtual screening, and molecular docking, 11 putative CDK2 inhibitors were successfully identified, which can be further evaluated by in vitro and in vivo biological tests.

The pharmacophore models obtained were capable of predicting the activities over a wide variety of scaffolds and thus can be used as:

- A three-dimensional query in database searches to identify compounds with diverse structures that can function as potent inhibitors and,
- To evaluate how well any newly designed compound maps to the pharmacophore before undertaking any further study including synthesis.

Both these applications may help in the identification or design of compounds for further biological evaluation and optimization.

In Summary:

- The QSAR models obtained showed not only statistical significance but also predictive ability and highlights the active compounds' spatial pose in the CDK2 inhibition and that minor values for DE combined with an NO

substituent lead to an increasing of CDK2 inhibitory activity. The QSAR models obtained give insight on structural requirements for designing more potent analogues against CDK2.

- 3D QSAR pharmacophore study presented in this study shows that a set of CDK2 inhibitors along with their activities ranging over several orders can be used to generate a good pharmacophore model. This model can then be used as a 3D query in available database searches to determine compounds with various structures that can be effective as potent inhibitors and to assess how well newly designed compounds map onto the pharmacophore prior to undertaking any further research including synthesis. The potential hit compounds obtained from this study can be further evaluated by in vitro and in vivo biological tests.

7. REFERENCES

- Accelrys Software Inc., Discovery studio modeling environment, Accelrys Software Inc., San Diego, CA; <http://www.accelrys.com>.
- Ariens, E. J., Rodrigues de Miranda, J. F. And Simonis, A. M., 1979, The pharmacion-receptor-effector concept: a basis for understanding the transmission of information in biological systems, in *The Receptors*, O'Brien, R. D. (ed.). Plenum Press, New York, pp. 33–91.
- Arris, C.E., Boyle, F.T., Calvert, A.H., Curtin, N.J., Endicott, J.A., Garman, E.F., Gibson, A.E., Golding, B.T., Grant, S., Griffin, R.J., Jewsbury, P., Johnson, L.N., Lawrie, A.M., Newell, D.R., Noble, M.E.M., Sausville, E.A., Schultz, R. and Yu, W., 2000, Identification of Novel Purine and Pyrimidine Cyclin-Dependent Kinase Inhibitors with Distinct Molecular Interactions and Tumor Cell Growth Inhibition Profiles. *J. Med. Chem.*, 43, pp. 2797-2804.
- Barlow, R. B., 1964, *Introduction to Chemical Pharmacology*, 2nd edn. Methuen, London.
- Barnum, D., Greene, J, Smellie, A. and Sprague, P., 1996, Identification of common functional configurations among molecules, *J. Chem. Inf. Comput. Sci.*, 36, pp. 563–571.
- Bhattacharjee, A.K. ,Kyle, D.E., Vennerstrom, J.L. and Milhous, W.K., 2002, A 3D QSAR Pharmacophore Model and Quantum Chemical Structure-Activity Analysis of Chloroquine(CQ)-Resistance Reversal, *J. Chem. Inf. Comput. Sci.*, 42, pp. 1212-1220.
- Beckett, A. H., 1959, Stereochemical factors in biological activity, in *Fortschritte der Arzneimittel Forschung*, Birkhäuser Verlag, Basel, pp. 455–530.
- 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.
- Berman, HM., Westbrook, J., Feng, Z., Gilliland, G., Baht, TN., Weissig, H., Shindyalov, IN. and Bourne, PE., 2000, The Protein Data Bank, *Nucleic Acids Res*, 28, pp. 235–242.
- Bilin, 2005, Donnersbergstrasse 9, D-67256 Weisenheim am Sand, Germany.
- Blundell, TL., Sibanda, BL., Sternberg, MJE. and Thornton, JM., 1987, Knowledge-based prediction of protein structures and the design of novel molecules, *Nature*, 326, 347.
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D.J. and Swaminathan, S., 1983, *J. Comput. Chem.*, 4, 187–217.
- Böhm, H.-J., 1996, Klebe, G., and Kubinyi, H., *Wirkstoffdesign*, Spektrum Akademischer Verlag, Heidelberg.

- Debnath, A. K., 2002, Pharmacophore mapping of a series of 2,4-diamino- 5-deazapteridine inhibitors of Mycobacterium aVium complex dihydrofolate reductase, *J. Med. Chem.*, 45, pp. 41-53.
- Dinsmore, C. J., Bogusky, M. J., Culberson, J. C., Bergman, J. M., Hornick, C. F., Zartman, C. B., Mosser, S. D., Schaber, M. D., Robinson, R. G., Koblan, K. S., Huber, H. E., Graham, S. L., Hartman, G. D., Huff, J. R. and Williams, T. M., 2001, Conformational restriction of flexible ligands guided by the transferred noe experiment: potent macrocyclic inhibitors of farnesyltransferase, *J. Am. Chem. Soc.*, 123, pp. 2107-2108.
- Dodds, E. C. and Lawson, W., 1938, Molecular structure in relation to oestrogenic activity. Compounds without phenanthrene nucleus. *Proc. R. Soc. London, Ser. B*, 125, 122–132.
- Ece, A. and Sevin, F., 2010, Exploring QSAR on 4-cyclohexylmethoxypyrimidines as antitumor agents for their inhibitory activity of cdk2, *Letters in Drug Design & Discovery*, 7, pp. 625-631.
- Easson, L. H. and Stedman, E., 1933, Studies on the relationship between chemical constitution and physiological action. V. Molecular dissymmetry and physiological activity. *Biochem. J.*, 27, 1257–1266.
- Fejzo, J., Lepre, C. A., Peng, J. W., Bemis, G. W., Ajay, Murcko, M. A. and Moore, J. M., 1999, The SHAPES strategy: an NMR-based approach for lead generation in drug discovery, *Chem. Biol.*, 6, pp. 755-769.
- Foye, WO. (ed.), 1989, *Principles of Medicinal Chemistry*, 3rd ed. Philadelphia: Lea & Febiger.
- Franke, R. and Gruska, A., 2003, General introduction to QSAR, in *Quantitative structure-activity relationship (QSAR) models of mutagens and carcinogens*, Benigni, R. (eds.), CRC Press.
- Free Jr., S.M. and Wilson, J.W., 1964, A mathematical contribution to structure activity studies, *J. Med. Chem.*, 7, 395.
- Fujita, T., Iwasa, J. and Hansch, C., 1964, A new substituent constant, π , derived from partition coefficients. *J. Amer. Chem. Soc.*, 86, pp 5175–80.
- Ghose, Arup K., Viswanadhan, V. N. and Wendoloski J. J., 1998, Knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases, *J. Comb. Chem.*, 1 (1), pp. 55–68.
- Golbraikh, A., Shen, M., Xiao, Z.Y., Xiao, Y.D., Lee, K.H. and Tropsha, A., 2003, Rational selection of training and test sets for the development of validated QSAR models. *J. Comput. Aided. Mol. Des.*, 17, pp. 241-253.
- Guner, O.F. (ed.), 2000, *Pharmacophore perception, development and use in drug design*, IUL Biotechnology Series, Vol. 1. IUL, La Jolla, CA.

- Hajdu, J., Neutze, R., Sjogren, T., Edman, K., Szoke, A., Wilmouth, R. and Wilmot, C. M., 2000, Analyzing protein functions in four dimensions, *Nat. Struct. Biol.*, 7, pp. 1006-1012.
- Hajduk, P. J., Sheppard, G., Nettesheim, D. G., Olejniczak, E. T., Shuker, S. B., Meadows, R. P., Steinman, D. H., Carrera, G. M., Marcotte, P. A., Severin, J., Walter, K., Smith, H., Gubbins, E., Simmer, R., Holzman, T. F., Morgan, D. W., Davidsen, S. K., Summers, J. B. and Fesik, S. W., 1997, Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR, *J. Am. Chem. Soc.*, 119, pp. 5818-5827.
- Hajduk, P. J., Meadows, R. P. and Fesik, S. W., 1999, NMR-based screening in drug discovery, *Q. Rev. Biophys.*, 32, pp. 211-240.
- Hansch, C. and Fujita, T., 1964, ρ - σ - π Analysis: a method for the correlation of biological activity and chemical structure, *J. Am. Chem. Soc.*, 86, 1616.
- Hansch, C., 1971, Qualitative structure-activity relationship in drug design, Academic Press, Vol. 1., New York, pp. 271–342.
- Hecker, E.A., Duraiswami, C., Andrea, T.A. and Diller, D.J., 2002, Use of catalyst pharmacophore models for screening of large combinatorial libraries, *J. Chem. Inf. Comput. Sci.*, 42, pp. 1204–1211.
- Hutchison, C. and Glover, D. M., 1995, *Cell Cycle Control*, IRL Press, Oxford.
- Ishima, R. and Torchia, D. A., 2000, Protein dynamics from NMR, *Nat. Struct. Biol.*, 7, pp. 740-743.
- Kubinyi, H., 1993, *QSAR: Hansch analysis and related approaches*, VCH, Weinheim.
- Kugori, Y. and Güner, O. F., 2001, Pharmacophore modeling and three-dimensional database searching for drug design using catalyst, *Curr. Med. Chem.*, 8, 1035–1055.
- Jones, G., Willett, P., Glen, RC. and Leach, AR., 1997, Development and validation of a genetic algorithm for flexible docking, *J. Mol. Biol.*, 267, pp. 727–748.
- Leo, A., Hansch C, Elkins D., 1971, *Progress in medicinal chemistry. Chem. Rev.*, 71, pp 525.
- Lipinski, C. A., Lombardo, F., Dominy, B. W. and Feeney, P. J., 1997, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Delivery Rev.*, 23, pp. 3-25.
- Lu, S-H., Wu, J. W., Liu, H-L., Zhao, J-H., Liu, K-T., Chuang, C-K., Lin, H-Y., Tsai, W-B. and Ho, Y., 2011, The discovery of potential acetylcholinesterase inhibitors: A combination of pharmacophore modeling, virtual screening, and molecular docking studies, *Journal of Biomedical Science*, 18:8.

- Marchetti, F., Sayle, K.L., Bentley, J., Clegg, W., Curtin, N.J., Endicott, J.A., Golding, B.T., Griffin, R.J., Haggerty, K., Harrington, R.W., Mesguiche, V., Newell, D.R., Noble, M.E.M., Parsons, R.J., Pratt, D.J., Wang, L.Z. and Hardcastle, I.R., 2007, Structure-based design of 2-arylamino-4-cyclohexylmethoxy-5-nitroso-6-aminopyrimidine inhibitors of cyclin-dependent kinase, *Org. Biomol. Chem.*, 5, pp. 1577-1585.
- Marchetti, F., Cano, C., Curtin, N.J., Golding, B.T., Griffin, R.J., Haggerty, K., Newell, D.R., Parsons, R.J., Payne, S.L., Wang, L.Z. and Hardcastle, I.R., 2010, Synthesis and biological evaluation of 5-substituted O4-alkylpyrimidines as CDK2 inhibitors, *Org. Biomol. Chem.*, 8, pp. 2397–2407.
- Marcu, M.G., Chadli, A., Bouhouche, I., Catelli, N. and Neckers, L.M., 2000, The heat shock protein 90 antagonist novobiocin interacts with a previously unrecognized ATP-binding domain in the carboxyl terminus of the chaperone., *J. Biol. Chem.*, 275 (47), pp. 37181–37186.
- Marshall, G.R. and Beusen D.D., 2003, Molecular modeling in drug design in burger's medicinal chemistry and drug discovery, Sixth edition, Volume 1: Drug Discovery, Abraham D. J. (eds.), John Wiley and Sons, Inc., 78p.
- Martinez, J.D., Parker, M.T., Fultz, K.E., Ignatenko, N.A. and Gerner, E. W., 2003, Molecular biology of cancer in burger's medicinal chemistry and drug discovery, Sixth edition, Volume 5: Chemotherapeutic Agents, Abraham D. J. (eds.), John Wiley and Sons, Inc., 20p.
- Martin, Y.C., Kim, K.-H., and Bures, G.M., 1992, Computer assisted drug design in 21st century, in *Medicinal Chemistry for the 21st Century*, Wermuth, C.G. et al., Eds., Blackwell Scientific, London, chap. 20.
- Mascarenhas, N.M. and Ghoshal, N., 2008, An efficient tool for identifying inhibitors based on 3D-QSAR and docking using feature-shape pharmacophore of biologically active conformation – A case study with CDK2/CyclinA, *European Journal of Medicinal Chemistry*, Volume 43, Issue 12, pp. 2807-2818.
- Maynard, A. J., 2004, HypoGenRefine and HipHopRefine: Pharmacophore refinement using steric information from inactive compounds. Presented at the ACS National Meeting, Spring.
- McDowell, L. M. and Schaefer, J., 1996, High-resolution NMR of biological solids, *Curr. Opin. Struct. Biol.*, 6, pp. 624-629.
- McDowell, L. M., McCarrick, M. A., Studelska, D. R., Guilford, W. J., Arnaiz, D., Dallas, J. L., Light, D. R., Whitlow, M., and Schaefer, J., 1999, Conformations of trypsin-bound amidine inhibitors of blood coagulant factor Xa by double REDOR NMR and MD simulations, *J. Med. Chem.*, 42, pp. 3910-3918.

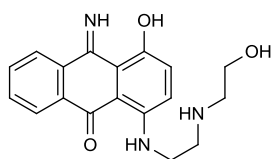
- Mesguiche, V., Parsons, R.J., Arris, C.E., Bentley, J., Boyle, F.T., Curtin, N.J., Davies, T.G., Endicott, J.A., Gibson, A.E., Golding, B.T., Griffin, R.J., Jewsbury, P., Johnson, L.N., Newell, D.R., Noble, M.E.M., Wang, L.Z. and Hardcastle, I.R., 2003, 4-Alkoxy-2,6- diaminopyrimidine derivatives: inhibitors of cyclin dependent kinases 1 and 2. *Bioorg. Med. Chem. Lett.*, 13, pp. 217-222.
- Moore, J. M., 1999, NMR techniques for characterization of ligand binding: utility for lead generation and optimization in drug discovery, *Biopolymers*, 51, 221-243.
- Muegge, I., 1999, The effect of small changes in protein structure on predicted binding modes of known inhibitors of influenza virus neuraminidase: PMF-scoring in DOCK4, *Med. Chem. Res.*, 9, pp. 490–500.
- Muegge, I. and Enyedy, I., 2004, Docking and scoring in computational medicinal chemistry for drug discovery, Bultinc, P., De Winter, H., Langenaeker, W. and Tollenaere. J.P. (eds.), Marcel Dekker Inc., New York, 406p.
- Nag, A. and Dey, B., 2011, Computer-aided drug design and delivery systems, McGraw-Hill, 1p.
- Nicklaus, MC., Wang, S., Driscoll, JS. and Milne, GWA., 1995, Conformational changes of small molecules binding to proteins, *Bioorg. Med. Chem.*, 3, pp.411–428.
- Nogray, T. and Weaver, D.F., 2005, Medicinal chemistry a molecular and biochemical approach, Oxford University Press, New York, 5p.
- Oprea, TI., 2000, Property distribution of drug-related chemical databases, *J. Comput-Aided Mol. Des.*, 14, pp. 251–264.
- Pardee, A. B., 1974, A restriction point for control of normal animal cell proliferation, *Proc. Natl. Acad. Sci., USA*, 71, pp. 1286-1290.
- Patrick, G. L., 2005, An introduction to medicinal chemistry, 3rd Edition, Oxford University Press, New York, pp. 491-493.
- Patrick, L. G., 2009, An introduction to medicinal chemistry, Oxford University Press, 382p, 384p and 522p.
- Perrakis, A., Morris, R. and Lamzin, V. S., 1999, Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol.*, 6, pp. 458-463.
- Poptodorov, K., Luu, T. and Hoffmann, R.D., 2006, Pharmacophore model generation software tools in pharmacophores and pharmacophore searches, Langer, T. and Hoffmann, R.D. (eds.), WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 27p.

- Rognan, D., Sokoloff, P., Mann, A., Martres, M.P., Schwartz, J. C., Costentin, J. And Wermuth, C. G., 1990, Optically active benzamides as predictive tools for mapping the dopamine D2 receptor. *Eur. J. Pharmacol. Mol. Pharmacol. Sect.*, 3, 59–70.
- Rognan, D., Boulanger, T., Hoffmann, R., Vercauteren, D.P., André, J. M., Durant, F. and Wermuth, C. G., 1992, Structure and molecular modeling of GABAA antagonists. *J. Med. Chem.* 35, pp. 1969–1977.
- Rohrbaugh, R.H. and Jurs, P.C., 1987, Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships. *Anal. Chim. Act.*, 199, pp. 99-109.
- Sakkiah, S., Thangapandian, S., John, S., Kwon, Y.:J: and Lee, K.W., 2010, 3D QSAR pharmacophore based virtual screening and molecular docking for identification of potential HSP90 inhibitors, *European Journal of Medicinal Chemistry*, 45, pp. 2132–2140.
- Salzmann, M., Permsh, K., Wider, G., Senn, H., and Wuthrich, K., 1999, *J. Biomol. NMR*, 14, pp. 85-88.
- Sander, C. and Schneider, R., 1991, Database of homology-derived protein structures and the structural meaning of sequence alignment, *Proteins Struct. Funct. Genet.*, 9, 56.
- Sayle, K. L., Bentley, J. F., Boyle, T. A., Calvert, H., Cheng, Y. Z., Curtin, N. J., Endicott, J. A., Golding, B. T., Hardcastle, I. R., Jewsbury, P., Mesguiche, V., Newell, D. R., Noble, M. E. M., Parsons, R. J., Pratt, D. J., Wang, L. Z. and Griffin, R. J., 2003, Structure-Based design of 2-Arylamino-4-cyclohexylmethyl-5-nitroso-6-aminopyrimidine inhibitors of cyclin-Dependent kinases 1 and 2. *Bioorg. Med. Chem. Lett.*, 2003, 13, pp. 3079-3082.
- Schneider, G., Giller, T., Neidhart, W. And Schmid, G., 1999, “Scaffold-hopping” by topological pharmacophore search: a contribution to virtual screening. *Angew. Chem. Int. Ed.*, 38, 2894–2896.
- Schueler, F.W., 1946, Sex hormonal action and chemical constitution. *Science*, 103, 221–223.
- Smellie, A., Teig, S. and Towbin P., 1994, Poling: promoting conformational variation., *J. Comput. Chem.*, 16, 171–187.
- Smellie, A., Kahn, S.D. and Teig, S. L., 1995a, Analysis of conformational coverage. 1. Validation and estimation of coverage, *J. Chem. Inf. Comput. Sci.*, 35, 285–294.
- Smellie, A., Kahn, S.D. and Teig, S. L., 1995b, Analysis of conformational coverage. 2. Applications of conformational models., *J. Chem. Inf. Comput. Sci.*, 35, 295–304.

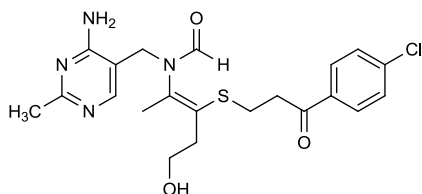
- Smith, R.N., Hansch, C. and Ames, M.M., 1975, Selection of a reference partitioning system for drug design work. *J. Pharm. Sci.*, 64, pp 599-606.
- Sprague, P., Hoffmann, R., 1997, in *Computer-assisted lead finding and optimization*, B.T. H. van de Waterbeemd, G. Folkers (eds.). Wiley-VCH, Weinheim, pp. 225–240.
- Stanton, D.T. and Jurs P.C., 1990, Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure–property relationship studies. *Anal. Chem.*, 62, pp. 2323-2329.
- Şener, E. A. and Yalçın, İ., 2003, *Farmasötik ve Medisinal Kimya’da İlaç Etken Madde Tasarım Yöntemleri -1, Kantitatif Yapı-Etki İlişkileri Analizleri (QSAR)*, Ankara University, Faculty of Pharmacy Publications, No:86, pp. 174.
- Tetko, I.V., Tanchuk, V.Y. and Villa, A.E.P., 2001, Prediction of octanol/water partition coefficients from physprop database using artificial neural networks and E-state indices. *J. Chem. Inf. Comput. Sci.*, 41, pp. 1407-1421.
- Toba, S., Srinivasan, J., Maynard, A. J. and Sutter, J., 2006, Using pharmacophore models to gain insight into structural binding and virtual screening: an application study with CDK2 and human DHFR, *J. Chem. Inf. Model.*, 46, 728–735.
- Todeschini, R. and Consonni, V. , 2000, *Handbook of molecular descriptors*, Wiley-VCH, Weinheim.
- Turner, J.V. and Maddalena, D.J., 2004, Agatonovic-Kustrin S. Bioavailability prediction based on molecular structure for a diverse series of drugs. *Pharm. Res.*, 21, pp. 68-82.
- Tute, M. S., 1990, In: *Comprehensive Medicinal Chemistry Quantitative Drug Design*, Hansch, C., Sammes, P. G.. Taylor, J. B. (Eds.), Pergamon Press, Oxford, 4, pp. 1-31.
- Vadivelan, S., Sinha, B.N., Irudayam, S.J. and Jagarlapudi, S.A.R.P., 2007, Virtual screening studies to design potent CDK2-Cyclin A inhibitors, *J. Chem. Inf. Model.*, 47, pp. 1526-1535.
- Veber, D. F., Johnson, S. R., Cheng, H-Y, 2002, Smith, B.R., Ward, K. W. and Kopple, K. D., Molecular properties that influence the oral bioavailability of drug candidates, *J. Med. Chem.*, 45, pp. 2615-2623.
- Vedani, A., Zbinden, P., Snyder, J.P. and Greenidge, P.A., 1995, Pseudoreceptor modeling: the construction of three-dimensional receptor surrogates, *J. Am. Chem. Soc.*, 117, pp. 4987–4994.
- Wermuth, C. G., 1993, Aminopyridazines—an alternative route to potent muscarinic agonists with no cholinergic syndrome. *Farmaco*, 48, 253–274.

- Wermuth, C. G. and Langer, T., 1993, Pharmacophore identification. In 3D QSAR in Drug Design. Theory Methods and Applications, Kubinyi, H. (ed.). ESCOM, Leiden, pp. 117–136.
- Wermuth, C. G., Ganellin, C. R., Lindberg, P. and Mitscher, L. A., 1998, Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1997). *Annu. Rep. Med. Chem.*, 33, 385–395.
- Wermuth, C. G., 2001, The impact of QSAR and CADD methods in drug design. In *Rational Approaches to Drug Design*, Hoeltje, H.D., Sippl, W. (eds.). Prous Science, Barcelona, pp. 3–20.
- Wermuth, D.C., 2006, Pharmacophores: Historical Perspective and Viewpoint from a Medicinal Chemist in *Pharmacophores and Pharmacophore Searches*, Langer, T. and Hoffmann, R.D. (eds.), WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 3p.
- Wolff, ME., 1995, *Burger's Medicinal Chemistry and Drug Discovery*, Wiley, New York.
- Woods, D.D., 1940, The relation of p-aminobenzoic acid to the mechanism of the action of sulphonamide. *Br. J. Exp. Pathol.*, 21, 74–90.
- Woods, D.D. and Fildes, P., 1940, The anti-sulphanilamide activity (in vitro) of p-aminobenzoic acid and related compounds. *Chem. Ind.*, 59, 133–134.
- Zafonte, B. T., Hult, J., Amanatullah, D. F., Albanese, C., Wang, C., Rosen, E., Reutens, A., Sparano, J. A., Lisanti, M. P. and Pestell, R. G., 2000, Cell-cycle dysregulation in breast cancer: breast cancer therapies targeting the cell cycle, *Front Biosci.*, 5, D938-D961.
- Zou, J., Xie, H-Z., Yang, S-Y, Chen, J.J., Ren, J-X and Wein, Y-Q., 2008, Towards more accurate pharmacophore modeling: Multicomplex-based comprehensive pharmacophore map and most-frequent-feature pharmacophore model of CDK2, *Journal of Molecular Graphics and Modelling*, 27, pp. 430–438.

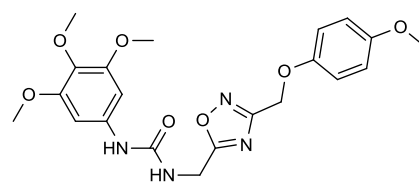
APPENDIX



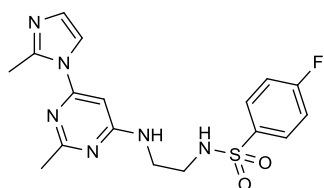
NSC_649153 (75.441)



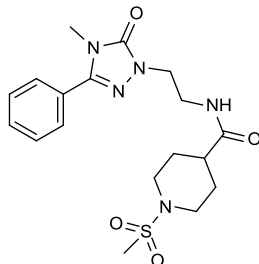
NSC_221631 (79.75)



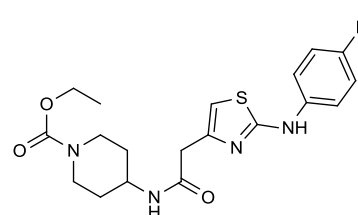
F5882-6930 (112.873)



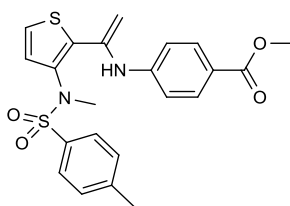
F5607-0191 (127.196)



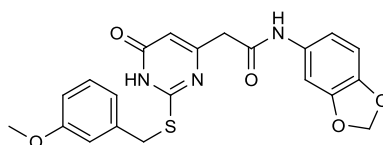
F5689-0078 (129.083)



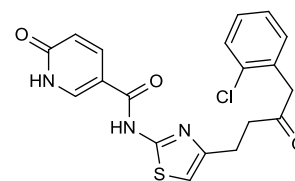
F5382-0550 (153.758)



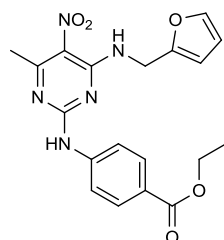
F3222-4739 (169.63)



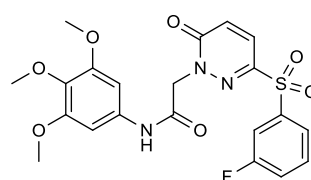
F5485-0325 (179.782)



F5463-0082 (231.78)



F5736-0414 (270.883)



F3407-2898 (286.601)

Figure A.1. Molecular structures of 11 hit compounds together with their activity values in nM.

Table A.1. Critical values of F at 95% confidence level calculated with the Excel function FINV(probability,deg_freedom1,deg_freedom2).

DF*	Number of Parameters in QSAR Equation (k)						
	1	2	3	4	5	6	7
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51

*DF= Degrees of freedom= $n-k-1$, where n is number of compounds, k is number of parameters (Sener and Yalçın; 2003).

F test value should be greater than 3.06 for Eq. (1) (DF= $(20-4-1)=15$; $k=4$) and 3.24 for Eq. (2) (DF= $(20-4-1)=16$; $k=3$).

Table A.2. Score values and molecular properties of 11 hit compounds and their 10 poses.

Name	POSE_NUM	Mol_Weight	Num_H_Acceptors	Num_H_Donors	Num_RotBonds	ALogP	ALogP_MR	Mol_PolSurfArea	LigScore1_Dreiding	LigScore2_Dreiding	-PLP1	-PLP2	Jain	-PMF	-PMF4	-CDOCKER_EN	Ludi_1	Ludi_2	Ludi_3	Consensus	FitValue	Estimate
NSC_649153	1.0	325.4	6.0	5.0	6.0	1.5	94.1	105.4	4.2	6.3	88.5	89.0	3.7	93.8	39.6	37.8	532.0	421.0	418.0	8.0	8.2	75.4
NSC_649153	9.0	325.4	6.0	5.0	6.0	1.5	94.1	105.4	4.6	6.2	82.3	80.9	2.2	65.1	24.3	35.1	409.0	349.0	270.0	4.0	8.2	75.4
NSC_649153	2.0	325.4	6.0	5.0	6.0	1.5	94.1	105.4	3.9	5.7	78.8	75.1	2.5	101.7	34.2	36.1	361.0	328.0	327.0	2.0	8.2	75.4
NSC_649153	4.0	325.4	6.0	5.0	6.0	1.5	94.1	105.4	3.9	5.7	80.1	75.3	2.1	101.0	34.8	35.9	387.0	334.0	340.0	2.0	8.2	75.4
NSC_649153	3.0	325.4	6.0	5.0	6.0	1.5	94.1	105.4	2.4	5.3	76.8	72.6	1.7	54.6	16.0	36.0	370.0	321.0	248.0	1.0	8.2	75.4
NSC_649153	5.0	325.4	6.0	5.0	6.0	1.5	94.1	105.4	2.2	5.2	73.9	68.9	1.3	53.7	12.4	35.5	355.0	329.0	265.0	1.0	8.2	75.4
NSC_649153	6.0	325.4	6.0	5.0	6.0	1.5	94.1	105.4	3.5	5.6	73.3	68.2	2.3	64.2	20.1	35.5	348.0	307.0	249.0	1.0	8.2	75.4
NSC_649153	7.0	325.4	6.0	5.0	6.0	1.5	94.1	105.4	3.7	5.7	73.8	68.6	2.5	66.9	21.7	35.4	297.0	266.0	216.0	1.0	8.2	75.4
NSC_649153	8.0	325.4	6.0	5.0	6.0	1.5	94.1	105.4	3.5	5.5	75.5	71.4	2.9	89.1	27.2	35.3	408.0	353.0	344.0	1.0	8.2	75.4
NSC_649153	10.0	325.4	6.0	5.0	6.0	1.5	94.1	105.4	2.5	5.3	75.5	69.2	0.7	61.2	17.4	35.1	352.0	327.0	267.0	1.0	8.2	75.4
NSC_221631	7.0	449.0	7.0	2.0	10.0	1.4	122.6	134.7	4.8	6.2	104.6	96.4	5.4	79.6	14.1	24.4	557.0	488.0	477.0	8.0	8.1	79.8
NSC_221631	8.0	449.0	7.0	2.0	10.0	1.4	122.6	134.7	5.5	6.3	102.1	97.0	3.0	55.2	5.6	23.9	399.0	413.0	391.0	5.0	8.1	79.8
NSC_221631	1.0	449.0	7.0	2.0	10.0	1.4	122.6	134.7	5.1	6.3	86.0	79.7	3.0	72.8	20.6	28.8	500.0	432.0	357.0	4.0	8.1	79.8
NSC_221631	4.0	449.0	7.0	2.0	10.0	1.4	122.6	134.7	5.3	6.3	105.5	91.3	2.3	87.9	29.5	26.8	424.0	397.0	383.0	4.0	8.1	79.8
NSC_221631	5.0	449.0	7.0	2.0	10.0	1.4	122.6	134.7	3.9	6.1	97.5	83.1	4.4	75.6	14.0	25.5	339.0	365.0	343.0	4.0	8.1	79.8
NSC_221631	2.0	449.0	7.0	2.0	10.0	1.4	122.6	134.7	4.0	5.2	79.6	69.2	3.5	52.8	10.3	28.0	336.0	342.0	307.0	2.0	8.1	79.8
NSC_221631	6.0	449.0	7.0	2.0	10.0	1.4	122.6	134.7	3.7	5.9	92.9	83.4	2.2	77.7	19.5	25.1	384.0	371.0	355.0	2.0	8.1	79.8
NSC_221631	3.0	449.0	7.0	2.0	10.0	1.4	122.6	134.7	3.7	5.5	80.6	75.9	2.6	58.8	14.8	27.3	347.0	335.0	274.0	0.0	8.1	79.8
NSC_221631	9.0	449.0	7.0	2.0	10.0	1.4	122.6	134.7	3.3	5.5	67.4	59.2	0.1	60.9	23.5	22.8	341.0	340.0	254.0	0.0	8.1	79.8
NSC_221631	10.0	449.0	7.0	2.0	10.0	1.4	122.6	134.7	3.9	5.9	79.0	72.1	2.3	64.2	27.5	22.6	417.0	370.0	375.0	0.0	8.1	79.8

Table A.2. continues...

Name	POSE_NUM	Mol_Weight	Num_H_Acceptors	Num_H_Donors	Num_RotBonds	ALogP	ALogP_MR	Mol_PolSurfArea	LigScore1_Dreiding	LigScore2_Dreiding	-PLP1	-PLP2	Jain	-PMF	-PMF4	-CDOCKER_EN	Ludi_1	Ludi_2	Ludi_3	Consensus	FitValue	Estimate
F5882-6930	1.0	444.4	8.0	2.0	10.0	1.9	111.2	126.2	3.6	6.2	115.8	98.4	3.3	102.8	35.9	28.9	577.0	474.0	522.0	9.0	8.0	112.9
F5882-6930	8.0	444.4	8.0	2.0	10.0	1.9	111.2	126.2	4.1	6.3	108.7	96.5	3.3	94.8	33.2	26.2	575.0	469.0	523.0	9.0	8.0	112.9
F5882-6930	2.0	444.4	8.0	2.0	10.0	1.9	111.2	126.2	3.6	6.2	117.0	98.1	3.2	100.4	34.2	28.8	598.0	470.0	511.0	8.0	8.0	112.9
F5882-6930	3.0	444.4	8.0	2.0	10.0	1.9	111.2	126.2	4.8	6.4	96.5	79.6	1.4	110.9	48.3	27.4	645.0	487.0	400.0	8.0	8.0	112.9
F5882-6930	5.0	444.4	8.0	2.0	10.0	1.9	111.2	126.2	4.9	6.5	100.2	80.5	1.7	120.5	53.1	27.2	629.0	497.0	401.0	8.0	8.0	112.9
F5882-6930	6.0	444.4	8.0	2.0	10.0	1.9	111.2	126.2	4.9	6.7	97.0	79.4	1.2	99.8	47.4	26.6	659.0	489.0	416.0	7.0	8.0	112.9
F5882-6930	10.0	444.4	8.0	2.0	10.0	1.9	111.2	126.2	4.8	6.4	96.7	76.9	1.8	101.1	40.3	25.2	588.0	447.0	364.0	7.0	8.0	112.9
F5882-6930	7.0	444.4	8.0	2.0	10.0	1.9	111.2	126.2	4.6	6.5	89.6	75.5	1.1	92.6	42.9	26.4	649.0	451.0	380.0	6.0	8.0	112.9
F5882-6930	9.0	444.4	8.0	2.0	10.0	1.9	111.2	126.2	4.6	6.3	83.5	67.6	0.6	103.5	46.6	25.4	653.0	480.0	409.0	6.0	8.0	112.9
F5882-6930	4.0	444.4	8.0	2.0	10.0	1.9	111.2	126.2	2.9	6.1	106.7	87.4	2.0	97.3	32.1	27.3	502.0	407.0	429.0	5.0	8.0	112.9
F5607-0191	8.0	390.4	6.0	2.0	7.0	1.1	97.3	110.2	4.1	6.1	84.5	76.3	4.4	89.0	41.0	39.3	579.0	502.0	503.0	8.0	7.9	127.2
F5607-0191	5.0	390.4	6.0	2.0	7.0	1.1	97.3	110.2	4.0	6.1	96.2	79.3	3.2	107.0	52.4	39.8	411.0	389.0	385.0	6.0	7.9	127.2
F5607-0191	1.0	390.4	6.0	2.0	7.0	1.1	97.3	110.2	3.2	5.9	79.9	69.2	3.1	111.3	57.5	41.3	399.0	369.0	304.0	4.0	7.9	127.2
F5607-0191	2.0	390.4	6.0	2.0	7.0	1.1	97.3	110.2	3.5	6.0	86.3	74.3	2.8	78.1	44.4	41.3	426.0	413.0	401.0	2.0	7.9	127.2
F5607-0191	4.0	390.4	6.0	2.0	7.0	1.1	97.3	110.2	3.4	6.0	82.8	71.1	2.0	79.0	40.6	40.1	420.0	410.0	395.0	2.0	7.9	127.2
F5607-0191	6.0	390.4	6.0	2.0	7.0	1.1	97.3	110.2	3.4	6.0	81.7	70.2	2.0	81.6	42.9	39.8	454.0	413.0	389.0	2.0	7.9	127.2
F5607-0191	9.0	390.4	6.0	2.0	7.0	1.1	97.3	110.2	3.5	6.0	79.6	68.1	1.9	76.1	42.6	38.7	430.0	383.0	372.0	2.0	7.9	127.2
F5607-0191	10.0	390.4	6.0	2.0	7.0	1.1	97.3	110.2	3.7	5.8	77.1	62.1	2.1	78.8	38.3	38.4	328.0	312.0	318.0	2.0	7.9	127.2
F5607-0191	3.0	390.4	6.0	2.0	7.0	1.1	97.3	110.2	3.5	5.9	87.6	77.6	2.8	66.7	31.0	40.1	435.0	406.0	401.0	1.0	7.9	127.2
F5607-0191	7.0	390.4	6.0	2.0	7.0	1.1	97.3	110.2	3.3	5.9	88.9	75.5	2.2	60.9	26.6	39.7	390.0	377.0	300.0	1.0	7.9	127.2

Table A.2. continues...

Name	POSE_NUM	Mol_Weight	Num_H_Acceptors	Num_H_Donors	Num_RotBonds	ALogP	ALogP_MR	Mol_PolSurfArea	LigScore1_Dreiding	LigScore2_Dreiding	-PLP1	-PLP2	Jain	-PMF	-PMF4	-CDOCKER_EN	Ludi_1	Ludi_2	Ludi_3	Consensus	FitValue	Estimate
F5689-0078	10.0	407.5	5.0	1.0	6.0	0.2	104.5	110.8	4.8	6.2	92.1	77.0	4.3	114.7	57.9	24.7	607.0	515.0	488.0	9.0	7.9	129.1
F5689-0078	9.0	407.5	5.0	1.0	6.0	0.2	104.5	110.8	3.8	6.2	93.8	76.0	4.1	97.7	46.8	25.3	488.0	447.0	444.0	7.0	7.9	129.1
F5689-0078	2.0	407.5	5.0	1.0	6.0	0.2	104.5	110.8	4.1	5.9	97.0	80.2	4.0	75.0	35.0	28.4	532.0	443.0	444.0	6.0	7.9	129.1
F5689-0078	1.0	407.5	5.0	1.0	6.0	0.2	104.5	110.8	4.9	6.3	94.2	78.0	3.9	69.6	28.8	30.1	484.0	447.0	383.0	5.0	7.9	129.1
F5689-0078	3.0	407.5	5.0	1.0	6.0	0.2	104.5	110.8	4.1	5.9	94.6	79.4	3.9	73.4	32.8	27.6	509.0	432.0	433.0	4.0	7.9	129.1
F5689-0078	5.0	407.5	5.0	1.0	6.0	0.2	104.5	110.8	3.5	5.6	90.5	81.8	3.4	73.6	30.3	26.1	525.0	478.0	412.0	4.0	7.9	129.1
F5689-0078	8.0	407.5	5.0	1.0	6.0	0.2	104.5	110.8	4.3	6.2	82.0	71.5	2.1	82.6	35.4	25.5	408.0	408.0	373.0	3.0	7.9	129.1
F5689-0078	4.0	407.5	5.0	1.0	6.0	0.2	104.5	110.8	4.0	5.8	87.2	73.9	3.7	68.4	25.0	26.3	450.0	413.0	359.0	1.0	7.9	129.1
F5689-0078	6.0	407.5	5.0	1.0	6.0	0.2	104.5	110.8	4.3	6.0	88.0	70.1	1.6	71.5	22.8	25.8	331.0	350.0	349.0	1.0	7.9	129.1
F5689-0078	7.0	407.5	5.0	1.0	6.0	0.2	104.5	110.8	4.0	5.8	83.8	69.1	1.9	64.4	23.1	25.6	378.0	377.0	368.0	1.0	7.9	129.1
F5382-0550	9.0	406.5	5.0	2.0	7.0	2.0	104.8	111.8	4.0	6.7	98.4	83.5	4.1	81.7	37.9	45.0	522.0	479.0	470.0	10.0	7.8	153.8
F5382-0550	5.0	406.5	5.0	2.0	7.0	2.0	104.8	111.8	4.1	6.4	94.2	84.1	4.1	78.8	31.0	46.9	492.0	451.0	484.0	8.0	7.8	153.8
F5382-0550	7.0	406.5	5.0	2.0	7.0	2.0	104.8	111.8	4.1	6.4	95.6	84.8	3.9	82.6	31.5	46.4	498.0	458.0	444.0	8.0	7.8	153.8
F5382-0550	8.0	406.5	5.0	2.0	7.0	2.0	104.8	111.8	4.0	6.3	94.4	84.4	4.8	80.2	30.6	45.2	504.0	478.0	503.0	8.0	7.8	153.8
F5382-0550	1.0	406.5	5.0	2.0	7.0	2.0	104.8	111.8	3.5	6.3	99.1	88.0	4.7	81.9	33.0	49.2	507.0	470.0	512.0	7.0	7.8	153.8
F5382-0550	2.0	406.5	5.0	2.0	7.0	2.0	104.8	111.8	3.5	6.3	96.9	86.5	4.6	85.0	33.4	49.0	495.0	464.0	454.0	7.0	7.8	153.8
F5382-0550	3.0	406.5	5.0	2.0	7.0	2.0	104.8	111.8	3.5	6.3	98.3	86.8	4.2	85.5	32.9	48.1	498.0	464.0	450.0	7.0	7.8	153.8
F5382-0550	4.0	406.5	5.0	2.0	7.0	2.0	104.8	111.8	3.8	6.4	96.8	85.4	4.1	83.6	30.1	47.5	492.0	462.0	448.0	7.0	7.8	153.8
F5382-0550	10.0	406.5	5.0	2.0	7.0	2.0	104.8	111.8	3.5	6.3	94.2	77.0	4.2	77.3	31.6	43.8	516.0	444.0	435.0	5.0	7.8	153.8
F5382-0550	6.0	406.5	5.0	2.0	7.0	2.0	104.8	111.8	2.8	5.6	86.5	74.4	3.0	75.1	30.3	46.4	439.0	410.0	390.0	2.0	7.8	153.8

Table A.2. continues...

Name	POSE_NUM	Mol_Weight	Num_H_Acceptors	Num_H_Donors	Num_RotBonds	AlogP	AlogP_MR	Mol_PolSurfArea	LigScore1_Dreiding	LigScore2_Dreiding	-PLP1	-PLP2	Jain	-PMF	-PMF4	-CDOCKER_EN	Ludi_1	Ludi_2	Ludi_3	Consensus	FitValue	Estimate
F3222-4739	8.0	444.5	5.0	1.0	7.0	3.6	116.4	129.4	4.1	6.4	95.8	93.6	4.0	88.8	25.8	17.8	579.0	506.0	494.0	8.0	7.8	169.6
F3222-4739	10.0	444.5	5.0	1.0	7.0	3.6	116.4	129.4	4.1	6.4	95.0	92.1	3.7	89.0	27.0	17.5	549.0	488.0	481.0	8.0	7.8	169.6
F3222-4739	1.0	444.5	5.0	1.0	7.0	3.6	116.4	129.4	3.9	6.5	98.4	86.2	2.8	96.6	32.6	22.1	451.0	438.0	477.0	5.0	7.8	169.6
F3222-4739	2.0	444.5	5.0	1.0	7.0	3.6	116.4	129.4	3.7	6.4	97.0	84.8	2.8	97.9	31.4	22.1	451.0	436.0	473.0	5.0	7.8	169.6
F3222-4739	3.0	444.5	5.0	1.0	7.0	3.6	116.4	129.4	3.5	6.3	92.0	80.7	2.2	89.8	31.2	21.7	463.0	429.0	483.0	4.0	7.8	169.6
F3222-4739	4.0	444.5	5.0	1.0	7.0	3.6	116.4	129.4	3.7	6.4	92.5	80.9	2.1	88.9	32.6	21.6	475.0	423.0	479.0	4.0	7.8	169.6
F3222-4739	7.0	444.5	5.0	1.0	7.0	3.6	116.4	129.4	3.7	5.6	79.3	68.5	2.5	70.9	14.9	17.9	529.0	480.0	452.0	3.0	7.8	169.6
F3222-4739	5.0	444.5	5.0	1.0	7.0	3.6	116.4	129.4	3.9	5.9	83.6	70.7	1.0	96.0	36.6	19.8	506.0	442.0	367.0	2.0	7.8	169.6
F3222-4739	9.0	444.5	5.0	1.0	7.0	3.6	116.4	129.4	3.7	5.9	74.2	59.2	0.0	90.9	35.7	17.5	434.0	371.0	290.0	1.0	7.8	169.6
F3222-4739	6.0	444.5	5.0	1.0	7.0	3.6	116.4	129.4	3.8	5.9	77.5	67.3	1.1	85.2	32.0	19.2	473.0	432.0	337.0	0.0	7.8	169.6
F5485-0325	4.0	439.5	7.0	2.0	8.0	2.8	117.5	123.6	4.9	6.6	107.4	97.5	3.9	106.2	38.1	21.7	610.0	494.0	521.0	10.0	7.8	179.8
F5485-0325	3.0	439.5	7.0	2.0	8.0	2.8	117.5	123.6	5.5	6.6	118.8	108.4	6.7	98.0	35.1	22.0	603.0	551.0	617.0	9.0	7.8	179.8
F5485-0325	6.0	439.5	7.0	2.0	8.0	2.8	117.5	123.6	4.9	6.5	102.2	93.4	3.5	103.6	36.1	20.8	581.0	474.0	401.0	9.0	7.8	179.8
F5485-0325	5.0	439.5	7.0	2.0	8.0	2.8	117.5	123.6	5.1	6.6	103.5	95.5	3.7	102.7	35.0	21.4	565.0	466.0	389.0	8.0	7.8	179.8
F5485-0325	7.0	439.5	7.0	2.0	8.0	2.8	117.5	123.6	4.6	6.4	98.9	87.6	3.5	78.8	25.3	20.3	502.0	446.0	434.0	6.0	7.8	179.8
F5485-0325	8.0	439.5	7.0	2.0	8.0	2.8	117.5	123.6	4.2	6.1	104.3	87.5	2.3	108.0	31.8	20.0	439.0	405.0	463.0	6.0	7.8	179.8
F5485-0325	9.0	439.5	7.0	2.0	8.0	2.8	117.5	123.6	4.2	6.1	106.4	89.6	2.4	112.4	32.3	19.8	458.0	417.0	477.0	6.0	7.8	179.8
F5485-0325	2.0	439.5	7.0	2.0	8.0	2.8	117.5	123.6	4.3	6.3	92.7	81.3	3.4	86.0	28.0	22.1	504.0	432.0	415.0	5.0	7.8	179.8
F5485-0325	10.0	439.5	7.0	2.0	8.0	2.8	117.5	123.6	4.1	6.2	94.8	84.8	3.6	84.4	28.0	19.7	416.0	397.0	386.0	5.0	7.8	179.8
F5485-0325	1.0	439.5	7.0	2.0	8.0	2.8	117.5	123.6	4.2	6.3	90.5	80.6	3.1	85.4	29.1	22.2	479.0	427.0	414.0	4.0	7.8	179.8

Table A.2. continues...

Name	POSE_NUM	Mol_Weight	Num_H_Acceptors	Num_H_Donors	Num_RotBonds	ALogP	ALogP_MR	Mol_PolSurfArea	LigScore1_Dreiding	LigScore2_Dreiding	-PLP1	-PLP2	Jain	-PMF	-PMF4	-CDOCKER_EN	Ludi_1	Ludi_2	Ludi_3	Consensus	FitValue	Estimate
F5463-0082	1.0	416.9	4.0	3.0	7.0	1.4	108.9	128.4	5.2	6.8	109.8	101.3	6.3	90.0	16.8	46.1	600.0	493.0	488.0	9.0	7.7	231.8
F5463-0082	3.0	416.9	4.0	3.0	7.0	1.4	108.9	128.4	5.1	6.9	105.9	99.8	4.7	87.8	9.3	41.8	539.0	440.0	431.0	8.0	7.7	231.8
F5463-0082	5.0	416.9	4.0	3.0	7.0	1.4	108.9	128.4	4.1	6.1	110.6	99.2	4.6	118.2	26.2	41.4	510.0	428.0	482.0	7.0	7.7	231.8
F5463-0082	2.0	416.9	4.0	3.0	7.0	1.4	108.9	128.4	4.6	6.6	97.2	92.1	4.6	80.1	10.7	42.0	518.0	426.0	402.0	6.0	7.7	231.8
F5463-0082	4.0	416.9	4.0	3.0	7.0	1.4	108.9	128.4	4.2	6.2	92.9	86.7	3.3	75.7	10.2	41.4	473.0	405.0	383.0	6.0	7.7	231.8
F5463-0082	6.0	416.9	4.0	3.0	7.0	1.4	108.9	128.4	4.6	6.4	86.6	82.4	3.9	71.3	9.2	41.2	480.0	396.0	380.0	5.0	7.7	231.8
F5463-0082	7.0	416.9	4.0	3.0	7.0	1.4	108.9	128.4	4.5	6.4	88.5	83.6	3.2	77.7	11.1	41.1	467.0	399.0	377.0	5.0	7.7	231.8
F5463-0082	10.0	416.9	4.0	3.0	7.0	1.4	108.9	128.4	4.1	6.4	91.3	80.3	1.9	80.9	25.8	40.0	517.0	409.0	378.0	4.0	7.7	231.8
F5463-0082	8.0	416.9	4.0	3.0	7.0	1.4	108.9	128.4	4.3	6.2	84.9	76.8	2.0	79.3	12.5	40.9	496.0	417.0	397.0	3.0	7.7	231.8
F5463-0082	9.0	416.9	4.0	3.0	7.0	1.4	108.9	128.4	4.1	6.0	89.4	83.5	2.8	69.4	7.0	40.8	441.0	397.0	324.0	3.0	7.7	231.8
F5736-0414	3.0	397.4	8.0	2.0	9.0	3.0	106.4	135.1	4.6	6.4	98.1	87.9	2.7	98.3	40.3	40.4	583.0	479.0	462.0	10.0	7.6	270.9
F5736-0414	2.0	397.4	8.0	2.0	9.0	3.0	106.4	135.1	4.4	6.2	98.2	88.0	3.3	92.8	40.6	40.4	493.0	431.0	430.0	9.0	7.6	270.9
F5736-0414	5.0	397.4	8.0	2.0	9.0	3.0	106.4	135.1	5.3	6.4	86.9	68.4	3.8	99.9	48.1	39.8	558.0	476.0	473.0	9.0	7.6	270.9
F5736-0414	1.0	397.4	8.0	2.0	9.0	3.0	106.4	135.1	4.8	6.1	94.6	71.1	2.3	86.7	40.8	41.3	523.0	458.0	451.0	8.0	7.6	270.9
F5736-0414	4.0	397.4	8.0	2.0	9.0	3.0	106.4	135.1	4.2	6.0	95.9	84.1	3.0	99.1	41.8	39.8	458.0	396.0	447.0	8.0	7.6	270.9
F5736-0414	6.0	397.4	8.0	2.0	9.0	3.0	106.4	135.1	3.8	6.1	101.0	86.7	2.1	95.4	39.5	38.1	544.0	437.0	419.0	7.0	7.6	270.9
F5736-0414	8.0	397.4	8.0	2.0	9.0	3.0	106.4	135.1	4.2	6.3	105.9	91.4	2.2	96.9	41.0	37.6	473.0	408.0	409.0	7.0	7.6	270.9
F5736-0414	7.0	397.4	8.0	2.0	9.0	3.0	106.4	135.1	3.9	6.3	92.7	77.7	1.6	118.9	53.7	37.6	425.0	366.0	367.0	5.0	7.6	270.9
F5736-0414	9.0	397.4	8.0	2.0	9.0	3.0	106.4	135.1	3.8	6.0	85.2	71.2	1.4	76.6	38.8	37.5	444.0	349.0	343.0	2.0	7.6	270.9
F5736-0414	10.0	397.4	8.0	2.0	9.0	3.0	106.4	135.1	3.0	5.8	92.3	73.9	0.8	69.6	27.2	37.4	405.0	353.0	346.0	2.0	7.6	270.9

Table A.2. continues...

Name	POSE_NUM	Mol_Weight	Num_H_Acceptors	Num_H_Donors	Num_RotBonds	AlogP	AlogP_MR	Mol_PolSurfArea	LigScore1_Dreiding	LigScore2_Dreiding	-PLP1	-PLP2	Jain	-PMF	-PMF4	-CDOCKER_EN	Ludi_1	Ludi_2	Ludi_3	Consensus	FitValue	Estimate
F3407-2898	1.0	477.5	8.0	1.0	8.0	2.0	116.0	132.0	5.2	6.6	96.1	86.6	3.6	112.1	49.7	26.5	577.0	488.0	517.0	10.0	7.6	286.6
F3407-2898	2.0	477.5	8.0	1.0	8.0	2.0	116.0	132.0	5.1	6.5	97.0	87.4	3.2	113.2	48.3	26.4	564.0	492.0	525.0	10.0	7.6	286.6
F3407-2898	3.0	477.5	8.0	1.0	8.0	2.0	116.0	132.0	5.0	6.5	94.3	83.7	3.5	114.5	50.2	26.2	532.0	456.0	489.0	10.0	7.6	286.6
F3407-2898	7.0	477.5	8.0	1.0	8.0	2.0	116.0	132.0	5.1	6.8	103.6	91.2	3.3	119.0	56.8	25.5	592.0	486.0	509.0	10.0	7.6	286.6
F3407-2898	8.0	477.5	8.0	1.0	8.0	2.0	116.0	132.0	5.2	6.7	101.3	91.1	3.1	116.9	48.7	25.3	595.0	492.0	517.0	10.0	7.6	286.6
F3407-2898	9.0	477.5	8.0	1.0	8.0	2.0	116.0	132.0	4.7	6.3	105.2	95.9	3.2	96.3	45.4	25.3	581.0	491.0	482.0	10.0	7.6	286.6
F3407-2898	6.0	477.5	8.0	1.0	8.0	2.0	116.0	132.0	3.5	6.2	99.1	82.5	2.8	115.9	46.0	25.7	430.0	408.0	482.0	6.0	7.6	286.6
F3407-2898	4.0	477.5	8.0	1.0	8.0	2.0	116.0	132.0	3.5	6.1	93.7	73.8	0.4	91.0	41.3	25.8	466.0	406.0	390.0	3.0	7.6	286.6
F3407-2898	5.0	477.5	8.0	1.0	8.0	2.0	116.0	132.0	3.3	6.0	91.5	71.4	1.5	93.0	42.5	25.8	469.0	408.0	382.0	3.0	7.6	286.6
F3407-2898	10.0	477.5	8.0	1.0	8.0	2.0	116.0	132.0	3.1	5.9	90.2	71.5	0.4	95.5	44.0	25.2	454.0	379.0	384.0	2.0	7.6	286.6

CURRICULUM VITAE

Surname-Name : ECE Abdulilah

Place of Birth : Silopi

Date of Birth : 1981

Marital Status : Single

Education and Academic Career :

High School : Silopi High School, 1994-1997

BSc. : Hacettepe University, Department of Chemistry, Organic Chemistry Division, 1997-2002

MSc. : Hacettepe University, Department of Chemistry, Organic Chemistry Division, 2002-2005

PhD. : Hacettepe University, Department of Chemistry, Organic Chemistry Division, 2005-2011

Foreign Languages :

English (Advanced, KPDS 86)

German (Beginner)

Work Experience :

Research Assistant at Hacettepe University, Department of Chemistry, Organic Chemistry Division, 2002-2011.