PALACKÝ UNIVERSITY, OLOMOUC

Faculty of Science

Department of Physical Chemistry

M.S. HARESH AJANI

Computer-Aided Drug Design: Quantum Mechanical Investigation of Protein-Ligand and Protein-Ligand-Water Complexes

Dissertation Thesis

Supervisor

Prof. Ing. Pavel Hobza, DrSc., Dr. h. c., FRSC

Institute of Organic Chemistry and Biochemistry

of the Czech Academy of Sciences

Olomouc   **2018**

UNIVERZITA PALACKÉHO v OLOMOUCI

Přírodovědecká Fakulta

Katedra Fyzikální Chemie

M.S. HARESH AJANI

Počítačový návrh léků: Kvantově-mechanická studia komplexů mezi proteiny a ligandy a proteiny, ligandy a vodou

Disertační práce

Školitel

Prof. Ing. Pavel Hobza, DrSc., Dr. h. c., FRSC

Ústav organické chemie a biochemie, Akademie věd Česke Republiky, v.v.i.

Olomouc   **2018**

DECLARATION OF AUTHORSHIP

I, HARESH AJANI, hereby declare that the matter embodied in this dissertation "**Computer-Aided Drug Design: Quantum Mechanical Investigation of Protein-Ligand and Protein-Ligand-Water Complexes**" is the result of the work carried out by me. All the literature is properly cited, and it has not been submitted for the award of any degree or diploma or associateship of any university or institute.

Olomouc, April 2018   HARESH AJANI

                                                                                                    Prof. Ing. PAVEL HOBZA

Dedicated To

MY BELOVED DAUGHTER

(Zeel)

CONTENTS

TABLE OF CONTENTS

Attached Publications

   A. Publication 1 - Novel CDK2 Inhibitors

   B. Publication 2 - Discovery of Potent FLT3 Kinase Inhibitors

   C. Publication 3 - Novel Pseudo-Irreversible Inhibitors of AChE & BChE

   D. Publication 4 - SQM/COSMO Scoring Function - Pose Recognition

   E. Publication 5 - CDK2 Inhibitors: Role of Water

   F. Publication 6 - AKR1B10 Inhibitors: Role of Water

   G. Publication 7 – Malonate-Based Inhibitors of Serine racemase: Role of Water

   H. Publication 8 - Cucurbit[n]urils: Host-Guest Complexes- Role of Water

   I. Publication 9 – Cucurbit[n]urils: Host-Guest Complexes- Role of Water

   J. Publication 10 - σ-Hole and Halogen Bond

ACKNOWLEDGEMENT

"Research is to see what everybody has seen but think what nobody has thought." Working on a research project needs guidance, support, and encouragement. Getting such help, I feel, is comparable to our human body. The Almighty has given us a wonderful body in which all the organs and organ systems have their own function. These systems work synchronously to live a healthy life. Similarly, my project work was like a human body where many people helped me in some or the other way to its successful completion. At this juncture, I would like to express my deep gratitude to one and all.

The nervous system is the system that controls and integrates the functions of the human body. The brain is the seat from which many thoughts and ideas emerge. This seat of ideas for my work was my esteemed guide **Prof. Pavel Hobza**. His innovative ideas, thoughts, concepts, and notions made my work very interesting and at the same time enjoying. The brain detects and processes the sensory information from inside and outside the body. Similarly, **Prof. Hobza** detected this particular research problem and constantly helped me to process this work properly. With a deep sense of gratitude, I would like to thank him sincerely.

Nerve fibers are next important part of the nervous system. **Martin Lepšík** was like nerve fibers having the property of excitability and conductivity. At times he used to get excited but at the same time he conducted smiles on our faces with his sense of humor and coordinating various systems in the department. I thank him for making mine Ph.D. enjoyable.

Blood is a specialized bodily fluid which helps in transportation of gases and nutrients and maintains water and acid-base balance in the body. I would like to thank **Jan Řezáč** for supplying all the nutrients, i.e. extending his helping hand whenever needed. I would like to thank **Jindřich Fanfrlík** for maintaining the homeostasis of my project. Homeostasis is the mechanism regulating the integrity of the bodily fluids and is the key to longevity. With a smile on their faces, they both helped me in regulating the integrity of my work in an orderly fashion.

The cell is the basic structural and functional unit of the human body. I am extremely grateful to all the Prof. Hobza lab members. Many cells unite to form tissues, tissues form the organs, which in turn form the organ system and various organ systems form the entire human body. Similarly, there are many cells who have contributed to my development and I would like to thank all of them, in particular to **Adam**, **Susanta**, **Róbert**, and **Jiří.** The liver synthesizes, secrets, metabolizes and detoxifies various substances in the body. **Institute of Organic Chemistry and Biochemistry and Palacký University**, served as liver by helping me perform an important part of my work. I would like to thank them for helping me selflessly.

It is the two hands and two legs which ultimately perform any kind of work in the body. Without the limbs, the body will not be able to perform even the basic functions like eating or drinking. Such limbs in my work were my lab mates, especially **Rabindranath, Cemal**, and **Saltuk**. Without them, it would be impossible to reach up to this level and make the project a success. At this point, how can I forget my other labs-mates **Amrit, Debashree,** and **Vijay**, who supported the limbs to maintain the work properly. A special well-deserved appreciation to **Mrs. Helena Černá** and others who were like the bones – not only providing shape and support to the limbs but also allowing the easy and free movements with the help of the joints. I am extremely indebted to them for this.

Lastly, I would like to thank the heart of the whole work – my family, my friend, philosopher, and guide.

APRIL 2018,   **HARESH AJANI**

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AChE – Acetylcholinesterase

ADME/T – Absorption, Distribution, Metabolism, Excretion, and Toxicity

AKR1B10 – Aldo- Keto Reductase Family 1 Member B10

BChE – Butyrylcholinesterase

CADD - Computer-Aided Drug Design

CB – Cucurbituril

CCDC – Cambridge Crystallographic Data Centre

CDK2 – Cyclin-Dependent Kinase 2

COSMO – COnductor-like Screening MOdel

DFT – Density Functional Theory

FLT3 – FMS-Like Tyrosine Kinase 3

GBSA – Generalized Born, Surface Area

HTS – High Throughput Screening

IE – Interaction Energy

IFST – Inhomogeneous Fluid Solvation Theory

LBDD – Ligand-Based Drug Design

MD – Molecular Dynamics

MM – Molecular Mechanics

NMR – Nuclear Magnetic Resonance

PBSA – Poisson-Boltzmann, Surface Area

PDB – Protein Data Bank

PM6-D3H4X – Parametrized Model 6 – Dispersion, Hydrogen and Halogen Bonding Correction

PMF – Potential of Mean Force

PSO – Particle Swarm Optimization

QM – Quantum Mechanics

RMSD – Root Mean Square Deviation

SAR - Structure-Activity Relationship

SBDD – Structure-Based Drug Design

SMD – Solvent Model based on Density

SQM – Semiempirical Quantum Mechanics

VDW – van der Waals

ABSTRACT

Over the last few decades, computer-aided drug design has emerged as a most successful technique rendering the drug discovery process more efficient and less costly. In the structure-based drug design branch, three-dimensional information on the biomolecular targets is used by the docking and scoring methodologies to find and optimize new ligands.

In this dissertation, the following approaches are presented: ligand design, binding mode prediction, structure-activity relationship (SAR), molecular docking, receptor-ligand scoring and bridging water thermodynamics, followed by their application in protein-ligand complexes and host-guest systems.

Molecular docking, which has become a powerful and influential tool for studying molecular recognition, aims to predict the binding mode of small molecules toward their biological target. It has been used in several projects: we have predicted the binding modes of a novel and potent inhibitor of CDK2 and FLT3 kinases and covalent inhibitors of AChE, BChE. Further, semi-empirical quantum mechanics-based scoring functions (SQM/COSMO) were used in native pose recognition where the poses had been generated with several docking programs. The SQM/COSMO scoring was compared with classical scoring function. We observed that SQM/COSMO accurately predicted the native poses in two dozen of difficult and diverse protein-ligand systems. Lastly, we discuss the important role of explicit water molecules in protein kinases, hydrolases, serine racemase and host-guest systems. We determined their thermodynamical parameter ΔG which correlated very well with the experimental binding ΔG for protein-ligand complex and host-guest systems.

ABSTRAKT

Během posledních několika desetiletí se počítačový návrh léků (angl. "computer-aided drug design") ukázal jako užitečný přístup, díky němuž je proces objevování a vylepšování léků účinnější a méně nákladný. Návrh léků založených na struktuře terapeutického cíle (angl. structure-based drug design") se zabývá hledáním a optimalizací nových ligandů biomolekulárních cílů pomocí metod dokování a skórování.

V této disertaci jsou prezentovány následující přístupy: návrh ligandů, predikce vazebného módu, vztah struktura-aktivita (SAR), molekulární dokování, výpočty afinit mezi receptorem a ligandem a termodynamika molekul vod, následované jejich aplikací v komplexech protein-ligand a supramolekulárních systémů.

Molekulární dokování, které se stalo účinným nástrojem pro studium molekulárního rozpoznávání, má za cíl předpovědět vazebný způsob malých molekul k jejich biologickému cíli. Tato metoda byla použita v několika projektech: předpovídali jsme vazebné způsoby nových a účinných inhibitorů kináz CDK2 a FLT3 a kovalentních inhibitorů AChE, BChE. Dále byly použity skórovací funkce založené na semiempirické kvantové mechanice (SQM/COSMO) při rozpoznávání nativního vazebného způsobu, kde byly alternativní možnosti vazby vytvořeny několika dokovacími programy. Výsledky výpočtů SQM/COSMO byly porovnány s klasickými skórovacími funkcemi. Zjistili jsme, že SQM/COSMO přesně předpovídá nativní vazebné způsoby ligandů ve dvou desítkách obtížných a různorodých systémů protein-ligand. Nakonec jsme se zabývali významnou rolí explicitních molekul vody v proteinkinasách, hydrolasách, serinové racemase a supramolekulárních systémech. Stanovili jsme jejich termodynamický parametr ΔG, který velmi dobře koreloval s experimentální vazebnou energií ΔG pro komplexní systémy protein-ligand a supramolekulární systémy

INTRODUCTION

The discovery and development of new drugs are very important and exciting processes, helping to treat or cure many diseases. The projects of drug development are demanding in terms of time and cost – it can be over ten years and beyond billion US dollars. Over the past decades, the investments in new drug development have continuously increased. Lots of considerable efforts hampered by various ways, resulting in low efficiency and high failure rate in the drug discovery and development process. In an effort to overcome this situation, computer-aided drug design (CADD) was proven one of most effective methods for facilitating and expediting the process and therefore saving time, money and resources possible.

1.1 Computer Aided Drug Design and History Context

Thousands of years ago, only herbal remedies were in use,[1] later drugs of synthetic/semi-synthetic origin were discovered.2 In a prior time, compound development was not very efficient in terms of potency, safety, and optimization. In the trial and error process, rational strategies were developed to improve the potency of compounds.[3-6] First time in the 1980s, the utilization of computers was expanded for data handling and other more prominent roles in drug discovery.[7] Over the years, high-throughput screening technologies emerged, which expedite and facilitate the drug discovery and development process by enabling a great number of compounds to be screened in shorter time. The limitations of HTS techniques failed to produce the new hits and success rate changed into extremely low, as well as late-stage failure in ADME/T studies. Therefore, all these issues underline the need to develop alternative strategies that can help in promoting the success rates and reduce cost and time in whole drug discovery and development process. CADD tools are divided into the structure-based and ligand-based design methods. The former methods require the three-dimensional structure of protein target. Usually, the structure is experimentally determined using either X-ray crystallography or Nuclear Magnetic Resonance spectroscopy. Another option is to build a homology model with the help of the available experimental structure. All structure-based virtual screening tools attempt to predict the shape and electrostatics complementary of small molecules with binding sites of the protein. Nowadays, CADD is playing a larger and more important role in the field of pharmaceutical research, which facilitates and helps to improve the efficiency of the industry. (CADD)[8,9]

CADD tools have been applied in almost every stage, greatly changing the strategy and pipeline for drug discovery. (Figure 1.1) Traditionally, CADD application was limited to lead discovery and optimization, today the application extends in the direction of target identification and validation, and even forwards preclinical studies.

1.2 CADD Applications in Drug Discovery and Development
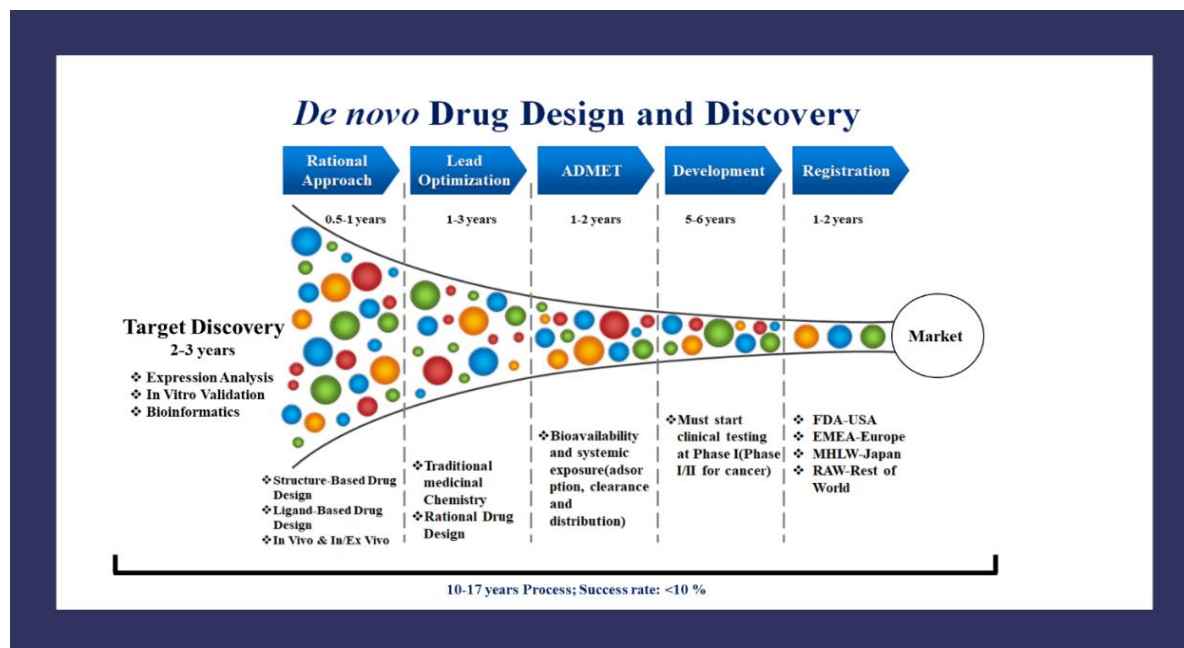
Figure 1.1: Drug Discovery & Development timeline.

There is a large list of successful application of CADD in development of novel and potent candidate in the de novo drug design.[10-18] (Figure 1.2) CADD is specialized discipline that utilizes computational methods to simulate drug-receptor interactions to determine whether a given molecule will bind to the target or not, and if it binds, what would be its affinity towards the target. For larger studies, traditional discovery and high throughput screening are more time to consume and costly then CADD.[19,20] This method has become the most widely used technique, which significantly decreases the number of potential medicinal compounds from a large library by predicting whether it will be active or inactive. Binding of ligands to receptors may occur via hydrophobic, electrostatics, halogen bonding, hydrogen-bonding, and other non-covalent interactions.[21-25] In addition, solvation phenomena play a major role in protein-ligand binding. LBDD and SBDD are two major techniques of computer-aided drug design. (Figure 1.3)
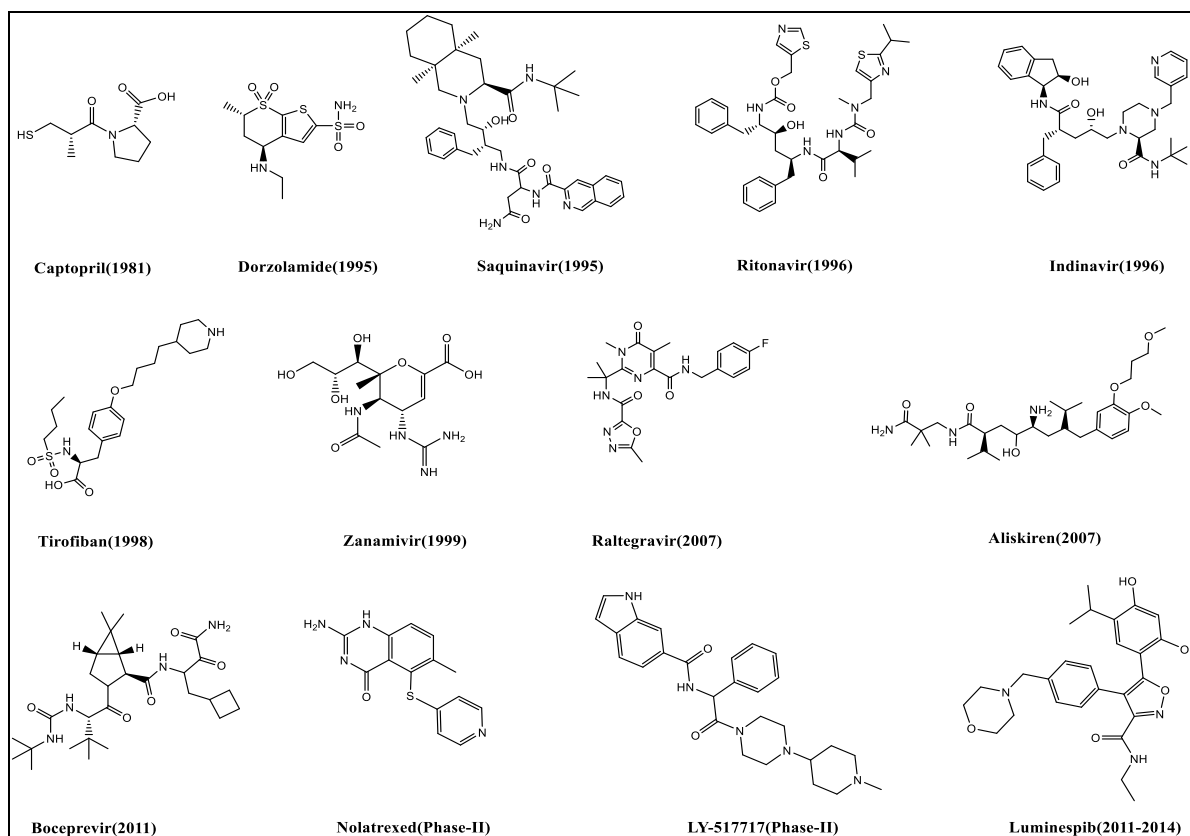
Figure 1.2: List of clinically approved drugs discovered by CADD.
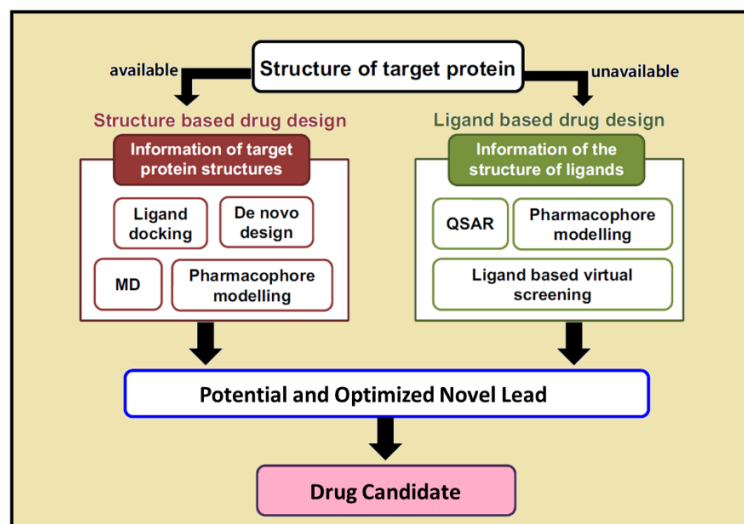
Figure 1.3: Flowchart of the Computer-aided drug design.

1.3 Docking and Prediction of Binding Modes

Molecular docking is one of the powerful influential tool and widely used computational techniques for structure-based drug design.[26-29] Docking is a method which predicts preferred orientation of one molecule to second when they bind to form a stable complex. In drug design

usually, the first molecule is the protein and second molecule is a small organic compound or a potential drug candidate. Knowledge of preferred orientation of ligand gives a relative prediction of ligand to protein binding mode, can be used to predict binding affinity, thus helps to differentiate high-affinity drug candidates over low-affinity compounds. There are lots of software available for docking to predict the binding mode and binding affinity. Here we have predicted the binding mode of Kinase (CDK2, FLT3 and AChE/BChE) inhibitors via induced fit docking.[30,31]

1.4 Scoring Functions: Concept and Application

The scoring function is a most important component of structure-based drug design for evaluating the binding efficacy of the ligands to the corresponding target proteins.[32] Ideally, the scoring function should be able to predict absolute binding affinity in the complex to ease identification of lead hits against any therapeutics target from any library. As molecular docking generates thousands of ligand conformations, scoring functions are used to rank and recognize these conformations and differentiate the accurate binding mode prediction from inaccurate predictions. However, it's still a huge challenge for existing scoring function to predict the binding affinities of diverse small molecules with high accuracy.[33,34]

In CADD and Computational chemistry, the binding free energies are often approximated by the scoring functions.[20,35-38] Three classes of scoring functions have been developed. Non-covalent interaction represents as the first class of scoring functions, have terms which are basically the sum of various energetics terms.[34,39,40] Second class of scoring functions have been developed based on experimental data, which have empirical parameters.[41,42] Another class comprises knowledge-based scoring functions.[37,43,44] These scoring functions are constructed based on available existing structures of protein-ligand complexes.

Almost every scoring functions contain binding free energy, which contains the interaction term, reflecting the ligand-protein interaction, solvation/desolvation term, and entropic term, which stand for flexibility changes of both protein and ligand. A popular type of scoring functions calculates interaction energy based on MM/PBSA or MM/GBSA. MM stands for molecular mechanics treatment of energies and deformation, PBSA and GBSA are the solvation models.[45,46]

During the past two decades, various scoring functions that exhibit different accuracies and computational efficiencies have been developed. Here, I briefly review existing scoring functions and compare with our scoring functions which have been developed in the group of Prof. Hobza for protein-ligand interactions in molecular docking. The energetic part of the scoring function is calculated at semi-empirical QM level, specifically PM6-D3H4X[47,48] and SCC-DFTB3-D3H4X.[49,50] Both these scoring functions employ energetics better than MM level theory.[51-54]

1.5 Waters

Water is a highly versatile component, which acts as both hydrogen bond donor or acceptor. Since last 10-15 years, the importance of waters molecules in drug design and host-guest systems has become of considerable interest.[55] Traditionally, it has been thought that there are

two major roles of water molecules in the systems. Firstly, to stabilize the system via creating the hydrogen bonding network with the inhibitor or small molecules. The second role is the ability of water molecules to be displaced upon the binding of ligands. The advantage of this displaced or release of water molecules into bulk carries an entropic gain, which may or may not is coupled with an enthalpic gain.

Gibbs free energy change (ΔG) has two thermodynamics contributions: the enthalpy changes ΔH and the entropy change ΔS (equation *1.1*).

$$\Delta G = \Delta H - T\Delta S \qquad\qquad \textit{1.1}$$

The enthalpic part of terms based on ligand binding or guest binding covers the changes in non-covalent binding patterns. The favorable non-covalent interaction between ligand and protein or host and guest are compensated by the unfavorable desolvation of ligand or guest. If noncovalent interactions of ligand-protein or host-guest are stronger enough than the noncovalent interaction between ligand or guest and solvent molecules, then the binding enthalpy is negative and favors binding.

On the contrary, the entropic part contains two major contributions:1). Loss of the conformational freedom of ligand or guest upon binding and 2). The releases of solvent molecules bound to the ligand or guest. Conformational entropy is unfavourable[55-57] for the former contribution. More tightly ligand or guest is bound in the active site, the more entropy losses by decreasing the flexibility.[58]

1.6 Non-Covalent Interactions

Non-covalent interactions are one or two orders magnitude weaker than the covalent bonding. This interaction differs from covalent bonding in that no electrons are shared between the participating atoms. Drugs produce their effect by interacting mostly non-covalently with their target protein in the systems. A fundamental understanding of event of ligand binding with the target protein requires deep insight, e.g. non-covalent interaction stabilizes the protein-ligand complexes as well as thermodynamics and dynamics of systems. There are different non-covalent interactions exist in several systems, such as induction, polarization, electrostatics interaction, charge transfer and dispersion. The non-covalent interactions are also seen to contribute vastly in the field of drug design, crystallography, and synthesis of new material.[59-64]

- Induction/Polarization: an attractive force that arises between dipole induced in a molecule by an electric field caused by a permanently charged or polar molecule.

- Electrostatics: an attractive or repulsive force that arises between two permanently charged molecules, two polar molecules and one permanently charged molecule and a polar molecule.

- Dispersion: an attractive force arises also two polar sites to correlated electron fluctuation in two molecules.

- Charge transfer: a fraction of electronic charge is transferred between the two or more molecules which are electron donor or acceptor.

SCOPE OF THE THESIS

Modern-day drug discovery is heavily dependent on computer-aided drug design. With this aspect, the objective of the research presented in this thesis is to provide insight into the application of the semi-empirical quantum mechanics-based method in the investigation of protein-ligand,[65-68] protein-ligand-water complexes,[69-71] and host guest systems[72,73] at the molecular level.

The first part of the thesis is based on structure-based drug design in the discovery, structural activity relationship (SAR), accurate prediction of binding mode of small drug-like molecules, novel core of CDK2 inhibitors, (imidazo[1,2-c]pyrimidin-5(6H)-one), FLT3,tyrosine kinase (2,6,9-trisubstituted purine derivatives), serine racemase (malonate-based inhibitors) and acetylcholinesterase and butyrylcholinesterase pseudo-irreversible (benzothiazole inhibitors) by a combination of biochemical approach, docking, and semi-empirical quantum mechanical scoring. The performance of currently used SQM based methods has been evaluated as a part of the research and the potential significance for drug discovery discussed.

The second part of the thesis is based on structure-based drug design method development followed by the application of SQM based scoring functions on diverse protein-ligand complexes with the implicit COSMO solvation model. This effort was encouraged by the limitations of classical molecular mechanics/empirical scoring functions in describing the non-classical P-L interaction in challenging systems. After comprehensive large-scale testing in near future, we propose SQM/COSMO as a useful computational tool in structure-based drug design which could serve as the reference method for further development of other scoring functions.

Beside the SQM methods, we analyzed and investigated the thermodynamic aspect of the important role of water molecules in structure-based drug design and host-guest systems. The first role is to stabilize the protein-ligand complex through creating hydrogen bonding network. This was studied in mammalian serine racemase and computational analyses reveal the profound effects of the thermodynamically-favorable active site water molecules on the molecular binding modes. Secondly, standard computational treatment failed to produce a correlation with the experimental binding free energies. The explicit treatment of active sites waters enhanced the description of CDK2 inhibitors binding affinity via a water hydrogen bond network. Lastly, in Aldo/keto reductase family protein target (AKR1B10) displacement of long-residence water molecules from hydration site play important role in ligand binding and an important source of binding free energy, which helps the shape complementarity and proper X-bonding interaction. Beside these protein targets, we have investigated hydration sites in the cavity of the host-guest system. (cucurbit[n]uril, n=5,6,7,8). We showed that high-energy waters need to be described explicitly to be able to obtain binding free energies which correlate with the experiments.

The performance of currently used semi-empirical quantum-based methods for non-covalent interaction in structure-based drug design and host-guest systems has been evaluated as part of the research and the potential significance of the results for drug discovery was shown.

STRUCTURE-BASED DRUG DESIGN: A DIRECT APPROACH

The direct drug design relies on available 3D structures of the biological target (protein) obtained through methods such as X-ray crystallography or NMR Spectroscopy.[74-76] A 3D protein structure should have indispensable to commence the paradigm of SBDD. To be usable for SBDD, the crystal structure needs to show the details of protein-ligand binding and have a reasonable resolution (< 2.5 Å).[77,78] In case, the 3D structure of a protein is not available, it can be modeled by means of homology modeling of nearest target related protein which the 3D structure available. Molecular Docking/Scoring and Free Energy Perturbation, and Pharmacophore Modeling have become most influential tools for SBDD to predict the affinities of small-molecule ligands to their protein targets.

3.1 Non-Covalent Molecular Docking

Molecular docking has great importance for computer-aided drug design since it should allow recognizing the binding pose and binding affinity of small molecules (ligands, L) to the protein (receptor(R)).(Eq. 3.1)

$$\mathbf{R + L \leftrightarrow RL} \qquad\qquad \mathbf{3.1}$$

This method works based on conformational search algorithm, a search algorithm is applied using systematic and stochastic search methods.[79,80] (See., 3.1.1 & 3.1.2) The sampling, the possible poses of the ligand into the binding pocket of the target protein are measured by the defined scoring methods.[28,81] Docking of small molecules are generally performed by one of the three ways:[29] (a) rigid docking, in which target and ligand are treated as rigid; (b) flexible ligand docking, in which target is held rigid and ligand treated as flexible; (c) flexible docking, in which target, and ligand both are considered as flexible. The sampling algorithm searches many possible poses of ligands provides many protein-ligand orientations to enable sufficient sampling of the binding modes. Search algorithm needs to have good speed and effectiveness. The scoring functions must be able to predict accurately thermodynamics of interaction. The complexity is increased in the order of rigid, flexible ligand and flexible docking.

3.3.1   Systematic (Local minimum) Search Method

In the conformational search methods, structural parameters of the small molecules, such as the degree of freedom i.e., rotational, translation and dihedral(torsional) are continuously modified during to search, thus explore the all the possible ligand degrees of freedom. This method works on incremental search and evaluations cycles, converges to local minimum

energy conformation to most likely binding mode.[82] This method has a major drawback in incremental search and can be overcome by performing a simultaneous search from a different point of the energy landscape.[83] (Figure 3.1A) With this approach, the method further classified as exhaustive, incremental search and conformational ensemble. Program Glide[84] and eHiTS[83] worked on the exhaustive search algorithm. Whereas, fragmentation approach followed by incremental search. i.e., Flex X,[85] DOCK,[86] and LUDI[35] Conformational ensemble methods, ligand flexibility[87] (conformation) pre-generated with docking program and ranked according to their binding energy scores. i.e., PhDOCK,[87] FLOG,[88] MS-DOCK.[89]

3.3.2 Stochastic (Global minimum) Search Method

The stochastic method works on the conformational search by randomly changing of all ligand degrees of freedom (translational, rotational and conformational) at each stage, generates a wide range of conformation in the energy landscape. (Figure 3.1B) This strategy avoids final confirmation at local energy minimum and increases the probability of finding the global minimum.[90] The main drawback of this method is tremendous computational cost. Monte Carlo method, (ICM,[91] ProDock[92]) Evolutionary Algorithms, (GOLD,[93] AutoDock[94]) Tabu Search method, (PSI-DOCK[95]) and Swarm Optimization(SO) (PSO@AutoDock,[96] PLANTS[97]) are based on a stochastic algorithm.
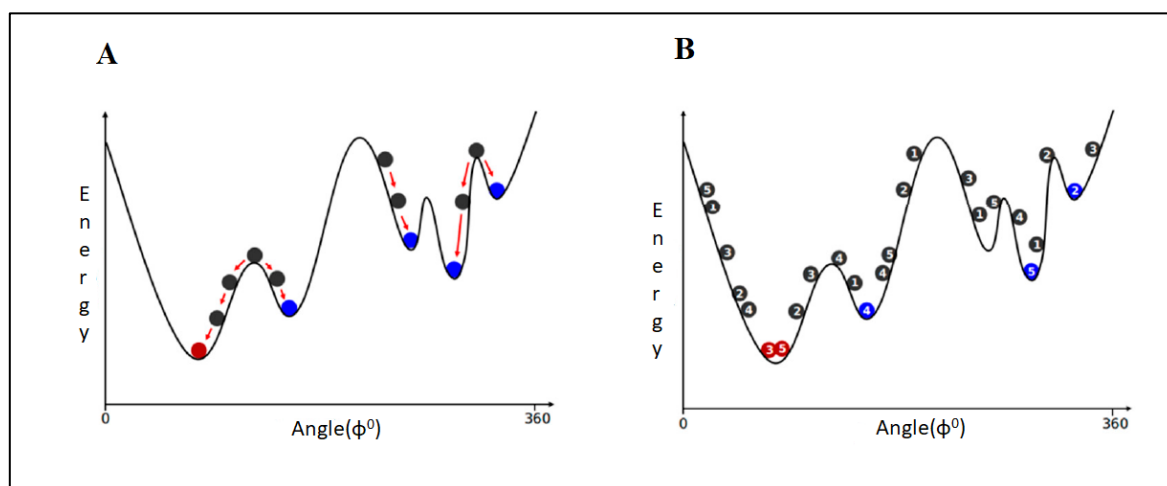
Figure 3.1: A). Systematic search algorithm changes all search parameter until local (blue spheres) or global (red spheres) energy minimum reached; **B).** Stochastic search explores the conformational space by randomly.

From past two decades, great variety of academic and commercial docking programs are available and most of them are dedicated to virtual screening. The most popular docking program include DOCK,[98] Autodock,[99] Autodock Vina,[100] FlexX,[85] GOLD,[101] and Glide,[84] among others. Following table 3.1 has some features of the existing docking programs.

Table 3.1 The features of existing docking programs.

Program	Feature	Website
Commercial Docking Software		
Glide[84]	Exhaustive search-based docking program.	http://www.schrodinger.com
GOLD[101]	GA-based docking program.	http://www.ccdc.cam.ac.uk
MOE Dock[102]	MOE Dock supplies a database of conformations or generate conformations on the fly, and then refines the poses using a force field-based method with MM/GBVI.	http://www.chemcomp.com
Surflex Dock[103]	Docking program based on a "protomol" that can be automatically generated and /or user-defined.	http://www.tripos.com
LigandFit[104]	Ligand conformation generated using Monte Carlo techniques are initially docked into an active site based on the shape and followed by CHARMm minimization.	http://www.accelrys.com
Academic Docking Software		
AutoDock[99]	LGA-based docking program.	http://autodock.scripps.edu
AutoDock Vina[100]	AutoDock Vina employs an iterated local search global optimizer.	http://vina.scripps.edu
rDock[105]	Combination of stochastic and deterministic search techniques to generate low energy ligand poses.	http://rdock.sourceforge.net
UCSF DOCK[98]	Anchor-and-grow based docking program.	http://dock.compbio.ucsf.edu
LeDock[106]	Based on Combination of simulated annealing and evolutionary optimization of the ligand pose and physics/knowledge hybrid scoring scheme.	http://lephar.com

3.2 Covalent Molecular Docking

In last decade, nearly 30% of marketed drugs has been based on the covalent attachment to their target, which traditionally been considered as conceptually distinct from conventional

non-covalent drugs.[107] Major benefit of the covalent interaction with the target protein has prolonged half-life, biological effects, and potential for improved selectivity.[108]

Although there has been much controversy over the role of covalent binding in the pathogenesis of idiosyncratic drug-related toxicity, the formation of reactive metabolites has been reported that, adds higher risk factor in drug discovery.[109] Despite this, there have been many successful and effective drugs that bind and function through the covalent mechanism. i.e. serine penicillin-binding protein which binds to B-lactams and B-lactone antibiotics, cysteine protease such as Cathepsin B, K, and S, which are covalently modified through vinyl sulfones, epoxides, and ketoamides, and hepatitis C virus protease, which covalently binds the ketoamide groups boceprevir and telaprevir.[108]
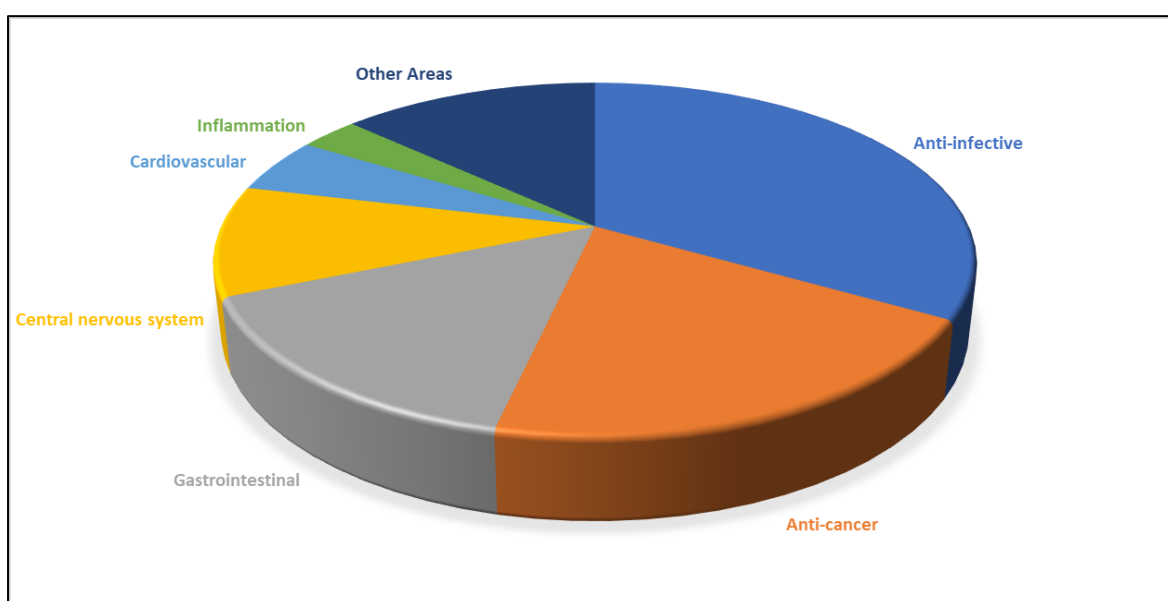
Figure 3.2: Approved covalent drugs by therapeutic indication (n= 39)

Several docking programs have been modified in focus of covalent docking. i.e. CovDock,[109] DOCKovalent,[110] AutoDock,[111] and GOLD.[101] All these programs are designed toward target-specific covalent inhibitors based on general mechanism shown in Eq. *3.2*

$$E + I \underset{K_i}{\rightleftharpoons} E \bullet I \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E - I$$

Intial non-covalent complex   Final covalent complex   *3.2*
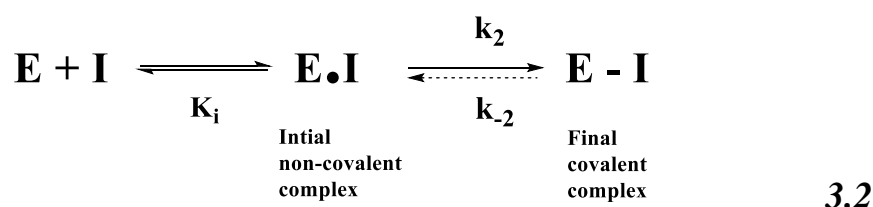
The entire process typically involves two notional steps: 1). The ligand binds non-covalently to the target protein, placing its moderately reactive electrophile close to a specific nucleophile

on the protein. Resulting complex, that undergoes specific bond formation from an initial non-covalent complex. 2). The covalent bond forms between the ligand and the target protein, later via the catalytic mechanism of the target protein, conversion of the unreactive ligand into highly reactive intermediate, which leads to covalent, irreversible inhibition of target protein. The ranking is based on a sampling of conformations with a scoring function to predict the best binding mode.

3.3 Scoring Functions

The scoring function is a most important component of structure-based drug design for evaluating the efficacy of ligands binding to their target proteins.[32,112,113] Ideally, scoring function should be able to predict absolute binding affinity of the complex to ease identification of lead hits against any therapeutics target from any library. As molecular docking generates thousands of ligand conformations, scoring functions are used to rank and recognize these conformations and differentiate the accurate binding mode prediction from inaccurate predictions. Till now it's a huge challenge for an available scoring function to predict the binding affinities of diverse small molecules with high accuracy.[37,114,115]

During the past two decades, various scoring functions that exhibit different accuracies and computational efficiencies have been developed. Here, I briefly review our scoring functions and existing scoring functions which have been developed for protein-ligand interactions in molecular docking. Figure 3.3 shows different scoring functions currently in use:
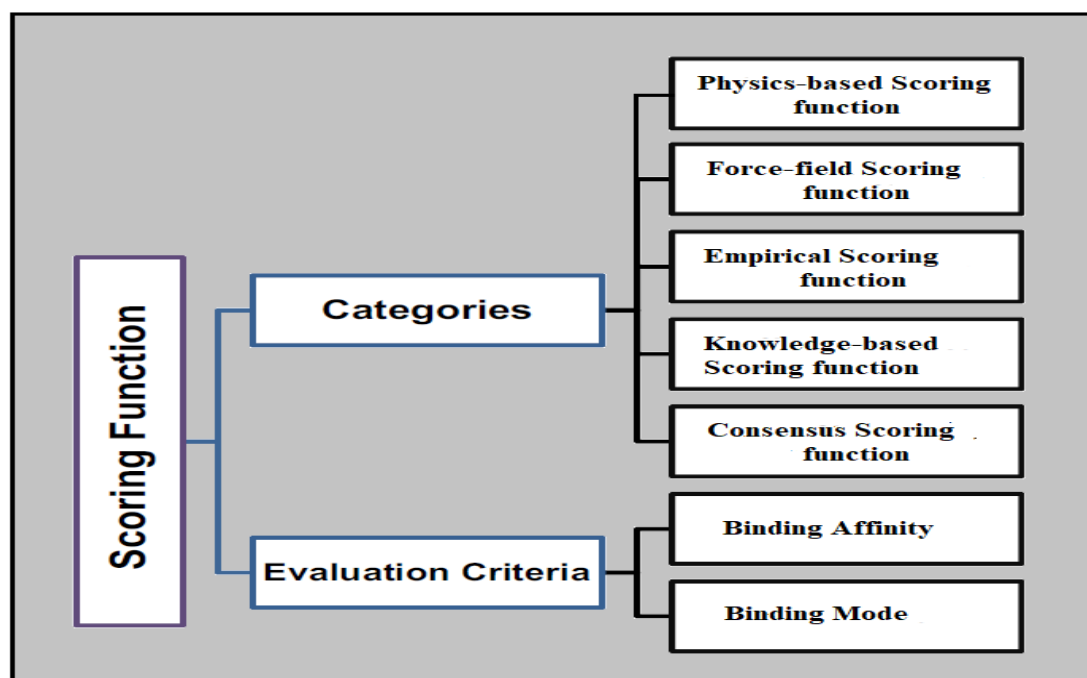

Figure 3.3: Different categories and evaluation criteria for Scoring function in protein-ligand interaction.

3.3.1 Physics Based Scoring Function

A few years ago, semi-empirical quantum mechanics methods were not widely used due to their computational cost, and some extended classical scoring function performs well. Because of computational cost, preference has been given to molecular mechanics methods, such as a combination of MM interaction energy with implicit solvation free energy term(GB, PB) to estimate affinities.[116] Additionally, wide coverage was available for organic chemical space of ligand parametrization via AMBER force field.[117,118] However, an explicit description of quantum mechanical effects in P-L interaction, such as charge transfer, polarization, covalent bond formation and σ-hole bonding were missing. These effects qualitatively well described by QM methods.[53,119,120] Presently SQM-based methods perform fast, due to powerful computing infrastructure and easy way to apply them for P-L complexes in numerous setups: efficient parallelization of SQM methods,[54,120-123] reliable linear scaling, and use of QM/MM,[120-127] DFT-D3 on truncated P-L complexes[128] or various fragmentations.[124,129,130] Specifically, AM1, RM1, PM6 or DF-TB SQM methods have been used[120-127,131] as such or with empirical correction for dispersion, hydrogen and halogen bonding.[47,132,133] The first SQM-based SF was introduced by Merz's group.[51,52] The combination of AM1 SQM method with an empirical dispersion(D), and the PB implicit solvent. The method was used for describing metalloprotein-ligand binding, but not sufficient to yield quantitative results.

$$Score = \Delta E_{int} + \Delta\Delta G_{solv} + \Delta G^{`w}_{conf} - T\Delta S \qquad\qquad 3.3$$

The above equation (3.3) represents the general physics-based SFs. The terms correspond to the gas-phase interaction energy (ΔE_{int}), the change of solvation free energy upon complex formation($\Delta\Delta G_{solv}$), the change of conformational "free" energy ($\Delta^{`w}_{conf}$) and the change of entropy upon ligand binding (-TΔS).

In our laboratory empirical corrections to SQM methods were developed, provide reliable, fast, and accurate description of the wide range of noncovalent interaction including dispersion, hydrogen and halogen bonding.[47,132,133] It is coupled with the COSMO implicit solvent model.[134] Correction was coupled with the PM6 SQM method[135] and resulting PM6-D3H4X method, does not require any system specific parameterization. It was used in various noncovalent complexes[131,136,137] such as aldose reductase,[138,139] carbonic anhydrase,[140] metalloprotein,[136,141,142] covalent inhibitors[143] and 17 diverse P-L complexes.[67]

3.3.2 Force-Field or Molecular Mechanics-Based Scoring Functions

The energy calculation is performed using classical molecular mechanics.[39] The scoring functions involve various physical features such as van der Waals (vdW) interaction, electrostatic interaction, and bond stretching/bending/torsional forces. (Eq. *3.4*)

$$E = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} \qquad\qquad 3.4$$

FF-based or MM-based SFs utilize parameters derived from both experimental and *ab initio* quantum calculation.[34] These scoring functions estimate the binding free energy of P-L complexes by the sum of van der Waals (VDW) interactions and electrostatic interactions. In prior time these methods received huge success, but most important problem associated with FF-based SFs is their inability to treat solvent molecules in protein-ligand complexes.[144]

3.3.3 Empirical Scoring Functions

Empirical scoring functions work based on the parameters obtained through experimental data.[41,42,145,146] The number of atoms that are in contact with each other within a ligand and receptor count by scoring functions or calculate changes in the solvent accessible surface area (ΔSASA) in the complex and the uncomplexed structure of the protein and ligand. These interaction terms may include favorable contacts (hydrophobic-hydrophobic), unfavorable contacts, favorable contributions to affinity (especially if shielded by solvent), no contribution if solvent exposed (number of hydrogen bonds), and unfavorable conformational entropy contribution (number of rotatable bonds immobilized in complex formation). Empirical functions have been used in several commercially available docking programs. i.e., Flex X, Surflex-Dock

3.3.4 Knowledge-Based Scoring Functions

Knowledge-based scoring functions work based on the available experimental knowledge about the receptor (target) - ligand binding or in the protein data bank (PDB).[37,147-149] Their energetic contribution to the binding is measured by the occurrence of individual contacts. A specific contact that occurs more frequently than an average or random distribution indicates attractive interaction, whereas less frequent occurrence indicates repulsive interaction, e.g., DRUGSCORE,[150] PMF score.[151]

3.3.5 Consensus Scoring Functions

Consensus scoring function is an approach which improves the probability of finding the correct solution via a combination of different scoring functions.[39] The best aspect of consensus scoring functions is their ability to score predicted binding poses using different scoring functions.[152,153]

Commonly used consensus scoring strategies include: (1) weighted combinations of scoring functions, (2) a voting strategy in which cut-off established for each scoring method is followed by decision-based on number of poses a molecule has, (3) a rank by number strategy ranks each compound by its average normalized score values, and (4) a rank by rank method sort compounds based on average rank determined by individual scoring functions.

3.4 Water Molecules in Protein Binding Sites and Host-Guest Systems

3.4.1   Water: Protein-Ligand Complexes

Hydration site is the structural feature of protein-ligand complexes and forms the hydrogen-bonding network in protein-ligand complexes.[154,155] Water mediating binding is more common in protein-ligand complexes. In 392 protein-ligand complexes,[156] its was found that 85% complexes have at least one or more water molecules making the interaction between protein and ligand. Previously, water molecules were ignored in docking studies and ligands were docked into desolvated binding sites. Now, there are many docking protocols where water molecules are included implicitly or explicitly.

In rational drug design, targeted displacement of water molecules into a bulk solvent, due to this favorable entropic gain occur, resulting in increase the translational and orientation degree of freedom of waters. If this targeted displacement of the water molecule is unsuccessful,[157,158] then the decrease of the binding affinity of ligands can be expected.[158,159] Previously, Water sites were predicted by classical molecular dynamics and Monte Carlo simulation with help of an explicit water model.[160] These methods have benefit including entropic calculation but, very time consuming to run, and spent more time on buried cavities due to permeate the water molecules in cavities. Later, lots of computational methods were discovered, which are fast, efficient and less time consuming, these methods are mostly based on grid-based Monte Carlo method and the Inhomogeneous fluid solvation theory(IFST), as developed by Lazaridis.[161,162] They use a short molecular simulation to identify water sites and then calculate the thermodynamics of water molecules in protein binding sites using IFST.[163-165] A key advantage of all these methods is the fact that computation is faster with a high degree of accuracy.

In the present study, we represented the importance of water molecules in binding cavities of the kinase, protease family, and hydrolase of protein targets.

3.4.2 Water: Host-Guest Systems

The condensation reaction between glycouril and formaldehyde form the macrocycle, named as cucurbituril,[166] which consist the 6 glycoluril units and 12 methylene bridges, was first reported by the Eberhard and Meyer, often abbreviated as CB[6]. The popularity of cucurbituril was increased during the 1980s and 1990s due to the crystal engineering and non-covalent interaction, which facilitate the non-covalent binding. Later in 2000 and 2002, Kim[167] and Day[168,169] modified the earlier reaction conditions and synthesize the variety of glycoluril-based cucurbituril macrocycles, named as cucurbit[n]uril, n = number of glycoluril. CB[5]-CB[8] have found a variety of uses given their ability to form binary and ternary host-guest complexes and therefore wide impact in scientific research areas.[170-172]

Cucurbit[n]urils consist of n glycouril molecules, alignments of the glycouril results in hydrophobic cavities with carbonyl-lined portals. The dipolar nature of these carbonyl-fringed portal of CB[n]s makes the portal highly attractive for cation binding through ion-dipole effects. While CB[n] portals are highly electronegative, the inner cavities of CB[n]s are hydrophobic and show a preference for the encapsulation of hydrophobic compounds. Hence,

alkali metals,[173] alkaline earth metals,[174-176] transition metals,[177,178] lanthanides and actinides[179] as well as ammonium and imidazolium ions[180] have been shown to bind the CB portals. Typically, due to aqueous environments, host-guest affinity drops compare to organic solvents, as water competes strongly for hydrogen bonds and solvent charged species. If high-energy water is present in cavity, supplies a driving force to form a host-guest complexation, and needs to review these high energy water release from the complexations. (Figure 3.4) For entropic reasons, all cavities contain the water molecules. Such water molecules reduced the number of hydrogen bonds compared to bulk water and are thus called high-energy waters. The binding of guest molecules in the CB[n]s cavities releases these high-energy water molecules, which lower the energy of the system, provide the entropic and enthalpic gain in favor of complex formation. Previously these hydrophobic effects have been studied by molecular dynamics simulations. In this thesis, we have reported the crucial role of the specific solvation in cucurbit[n]uril, which include the location of waters, occupancy, enthalpy and entropy of water in the cavity.
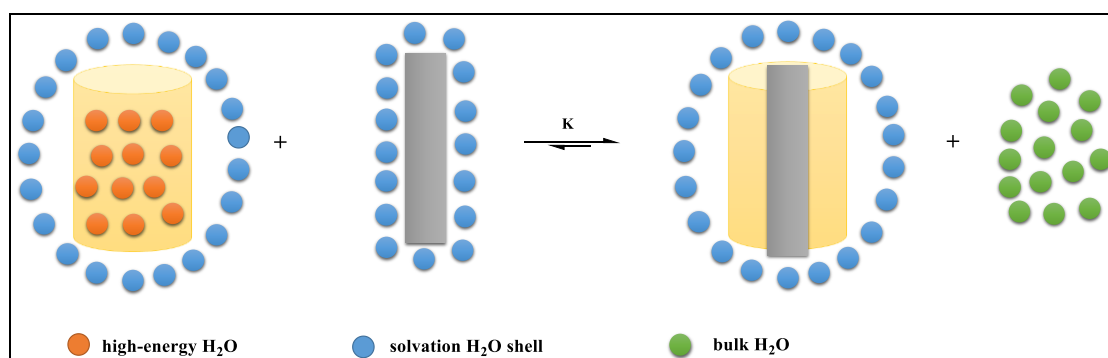
Figure 3.4: Schematic representation of the formation of CB[n] host-guest complexes driven by the release of high-energy water.

3.5 Non-Covalent Interactions

The non-covalent interactions are one or two orders of magnitude weaker than the covalent interaction. This interaction differs from covalent interaction in that no electron is shared between the participating atoms. Drugs produce their effect by interacting either covalently or non-covalently with their target protein in the systems. A fundamental understanding of ligand binding with the target protein requires deep insight. e.g. non-covalent interaction stabilizes the protein-ligand complexes as well as thermodynamics and dynamics of systems. There are fundamental non-covalent interactions in several systems, such as induction, polarization, electrostatics, charge transfer and dispersion.

The focus has been on the different classes of hydrogen bonds, halogen bonds, and σ-bond interactions. These interactions are briefly described in the following sections. There are, however, other types of interactions also important in protein-ligand complexes, like ionic bonds, salt bonds, and hydrophobic interactions.

3.5.1 Hydrogen bond interactions

The hydrogen bond X-H…Y is an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X-H in which X is more electronegative than H and electron donor molecule Y. The strength of the interaction is affected by the angle and distance between the donor and acceptor, which often reinforce in positive cooperative manner. This force contributes to the interaction energy which includes electrostatics, induction, and dispersion. In addition, hydrogen bond exhibits partial covalent character as result of charge transfer between the donor and acceptor. The work in the thesis is mainly related to the strong non-covalent interaction between the protein-ligand and ligand-water molecules which are thoroughly investigated with available QM/SQM based methods. e.g. kinases, hydrolase and host-guest systems. Examples of different types of hydrogen bond considered into protein-ligand complexes in this thesis are presented in Figure 3.5.
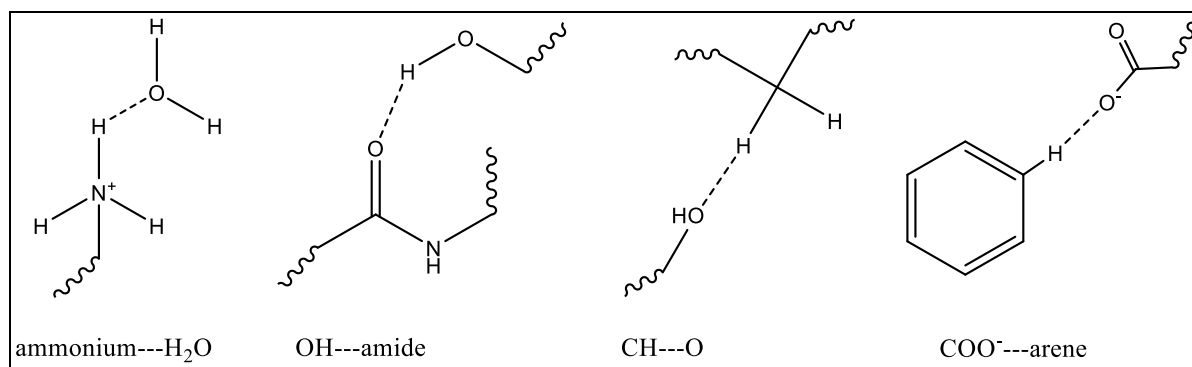
Figure 3.5: Examples of hydrogen bonds considered in Protein-ligands complexes.

3.5.2 Halogen Bonds interaction and σ-bond

A first halogen bond was proposed by Guthrie[181] in 1863 in a complex of ammonia and iodine. Later, Remsen[182] and Mulliken[183] proposed halogen bond with trimethylamine and bromine and dihalogens, respectively. During the1960s, Hassel (Nobel prize awarded) determined halogen bond by X-ray diffraction and first complex studied was between 1,4-dioxane and bromine. The halogen bond represented by (Fig 3.6),
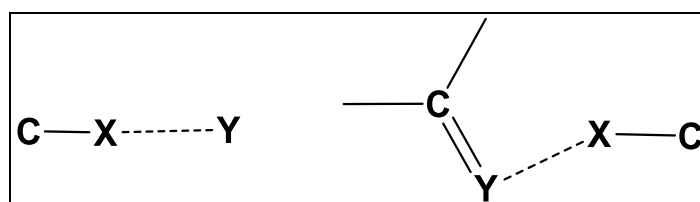
Figure 3.6: Representation of halogen bond.

where C-X is defined as the halogen-bond donor and can include dihalogen molecules. Y is the halogen bond acceptor or Lewis base and can include the lone pairs of electrons on an atom, an anion or a region of π-electron density. These include polar interaction, hydrophobic interaction and multipolar interaction.

The X-bond "donor" is attributed to an anisotropic distribution of the charge density on the halogen atom, due to this polarization happen on the halogen (X) along the C-X σ-bond. According to the molecular orbital theory, the valence electrons of outer shell pz orbital participate in the formation of the covalent σ-bond, thus partially exposing the positive nuclear charge opposite the C-X σ-bond. This is referred as σ-hole, which leads to an attractive interaction with the linear arrangement. The strength increase as the size of halogen increase, means electrons are more polarizable: F << Cl < Br < I, where I am forming more stable halogen bonds. On the other hand, F is more electronegative and less polarizable and forms halogen bond only in certain conditions.

The attention of halogen bond in drug discovery increase in past few years. Approximately 20-25% of all drugs contain at least one fluorine atom and approximately 14.5% contain Cl, 1.5% Br and 1.2% I. Three out of ten bestselling drugs in 2011 contain F. These include some blockbusters, e.g. Prozac(depression), Celebrex(arthritis), Sustiva(anti-HIV), Januvia(diabetes) and Lipitor(dyslipidemia).

In our laboratory, we investigated the selectivity and binding affinity of CDK2 inhibitors, with a polyhalogenated derivative of AKR1B10 inhibitors, which explicitly includes the bromine atoms. If bromine is substituted by the other alkyl groups or hydrogen, then we have shown that binding affinity drastically reduce from potent to less potent in CDK2 inhibitors. Secondly, we characterized the nature of halogen bonding in 128 complexes using advance quantum mechanical calculations. The first subset of 38 complexes with small intermolecular distance and significant van der Waals distance have stabilization energies in the range of 7-32 kcal/mol, while the second subset with 90 complexes has stabilization energies smaller than 7 kcal/mol.

The dissertation is organized as follows, First, we discuss docking, binding mode prediction and covalent inhibitor design of kinase and hydrolase targets, respectively. (Attachments A+B+C). Next, we briefly describe the performance of SQM methods. Specifically, PM6 and DFTB3 based SQM scoring functions on protein-ligand diverse set. (Attachment D). Finally, we describe the important role of water on two different projects; one of them is based on protein-ligand complexes, like, kinase, hydrolase, and racemase, (Attachment E+F+G) while the second one we studied on the host-guest binding. (Attachment H+I)

4.1 Binding Mode Analysis and Design and Discovery of Kinase and Hydrolase Inhibitors

In Structure-Based Drug Design, docking method is considered a promising tool for designing, structure-activity relationship (SAR), and prediction of the binding mode of inhibitors and application on biological systems. Here, we did prediction of the binding mode of the novel and potent inhibitor of CDK2 and FLT3 kinases and, the structure-activity relationship of covalent inhibitors of AChE, BChE with its predecessor. For these reasons, we utilized docking as primary tool and data available in the literature.

4.1.1 Computational Method

Using structure-based approach, we have rationalized the observed structure-activity relationship and, binding mode analysis. Induced fit docking[31,184] and non-covalent docking were carried out in active Kinases structure (CDK2, FLT3) and covalent docking in hydrolase target (AChE, BChE) using Glide and understanding of binding mode in term of binding free energies was approached via semi-empirical quantum mechanics(SQM) scoring methods. (See the Methods in Publication A, B, and C.)

4.1.2 Result and Conclusion

Figure 4.1 summarizes a SAR and Binding mode analysis of novel inhibitor (imidazo[1,2-c]pyrimidin-5(6H)-one) which were subjected to biochemical assays to determine their activity against recombinant CDK2/cyclin E. The 2',3'-dihalogenated compound have (See methods, Publication A) exhibited better activity ($IC_{50} = 1.3\mu M$). Due to the small size of the inhibitors, we have observed several types of binding modes. The most active compound had standard type I inhibitor binding mode (bma1, and bm1b), featuring two hinge-region hydrogen bonds like their predecessors. In contrast, bm2 had reversed core with two hinge region hydrogen bonds, similar to purvalanol. The last binding mode bm3 had only one hinge region H-bond. Due to the lack of crystal structures, narrow potency range and, other factors correlation with experimental data was not quite possible. We used SQM approach to differentiate binding

mode via their calculated binding free energies. In conclusion, the smaller size of substituent was consistent with the micromolar potency. SQM rescoring identified the probable favorable binding mode which will guide future structure-based design and optimization to get potent compounds.
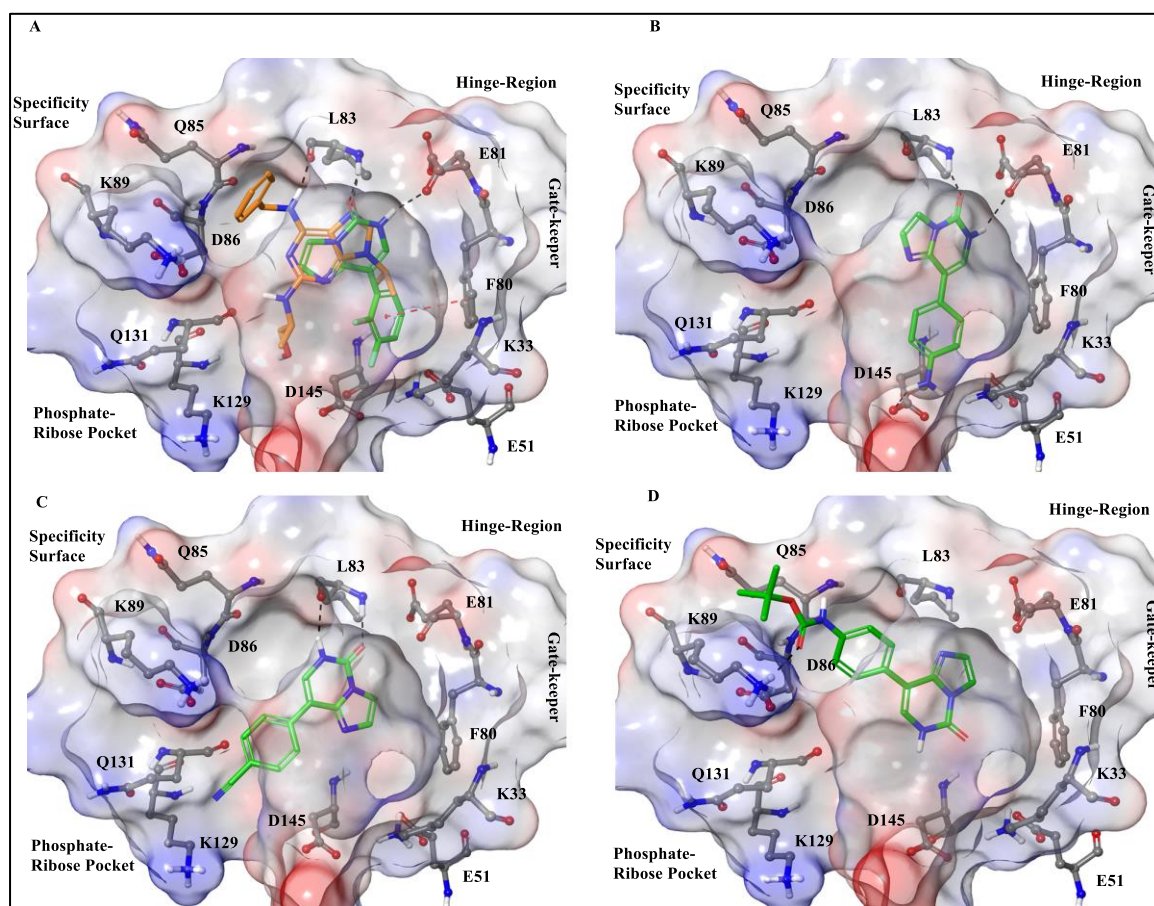
Figure 4.1: Binding modes of CDK2 inhibitors in active CDK2/cyclin E conformation. A). active Compound **3j**(bm1a) B). active compound **3b**(bm1b) C). least active compound **3n**(bm2) D). inactive compound **3s**(bm3)

Figure 4.2 summarizes the binding mode prediction of potent FLT3 kinase inhibitor for acute myeloid leukemia with FLT3 mutations. The most active compound **7d** is type-I inhibitor and bind to the ATP binding site, forming the hydrogen bond with hinge region and presenting the substituent in the hydrophobic region near the gatekeeper residue in active kinase conformation. Compound **6j** makes many non-hinge interactions resulting in loss of activity. In conclusion, docking of 7d to FKT3 suggest type -I inhibitor, binding mode and, explains the structural determinant to its potency. (See Detailed, Publication B)
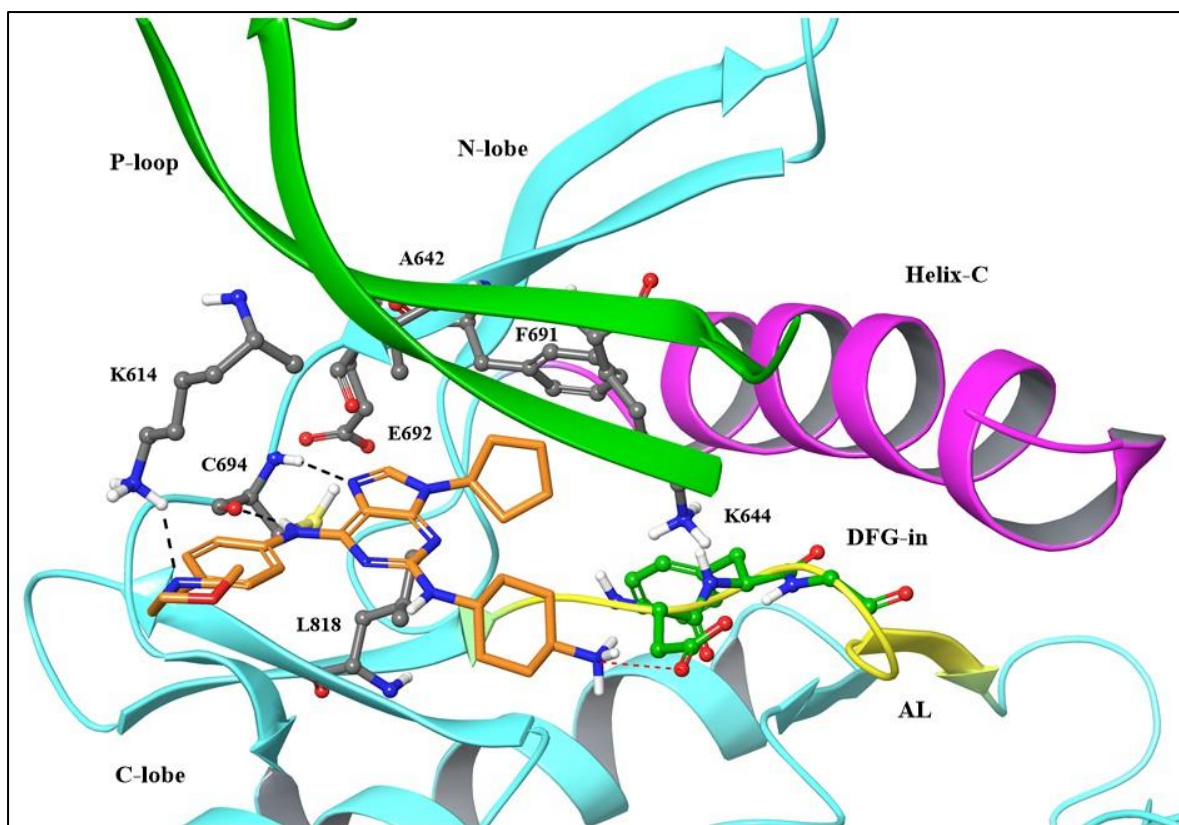
Figure 4.2: Docking binding pose of active compound 7d in active FLT3 Kinase conformation.

Last, sixteen novel derivatives of alkyl carbamates series compound were characterized, synthesized, as acetylcholinesterase(AChE) and butyrylcholinesterase(BChE) inhibitors. The result demonstrated they acted as pseudo-irreversible inhibitors. Compound (**3b,3d,3l, and 3n**) had the best AChE inhibitory activity and were tenfold more potent than standard drug rivastigmine. Covalent docking was performed. Binding mode (Figure 4.3) of all series you should describe the covalent score correlated with the experimental binding free energy. ($R^2 =$ 0.75). Thus, the weakest compounds in alkyl carbamates series were clearly distinguished from the strong binders. In conclusion, pseudo irreversible inhibition mechanism was confirmed by biochemical studies and computationally supported by state of the art covalent docking and scoring methodology. (See Detailed, Publication C)
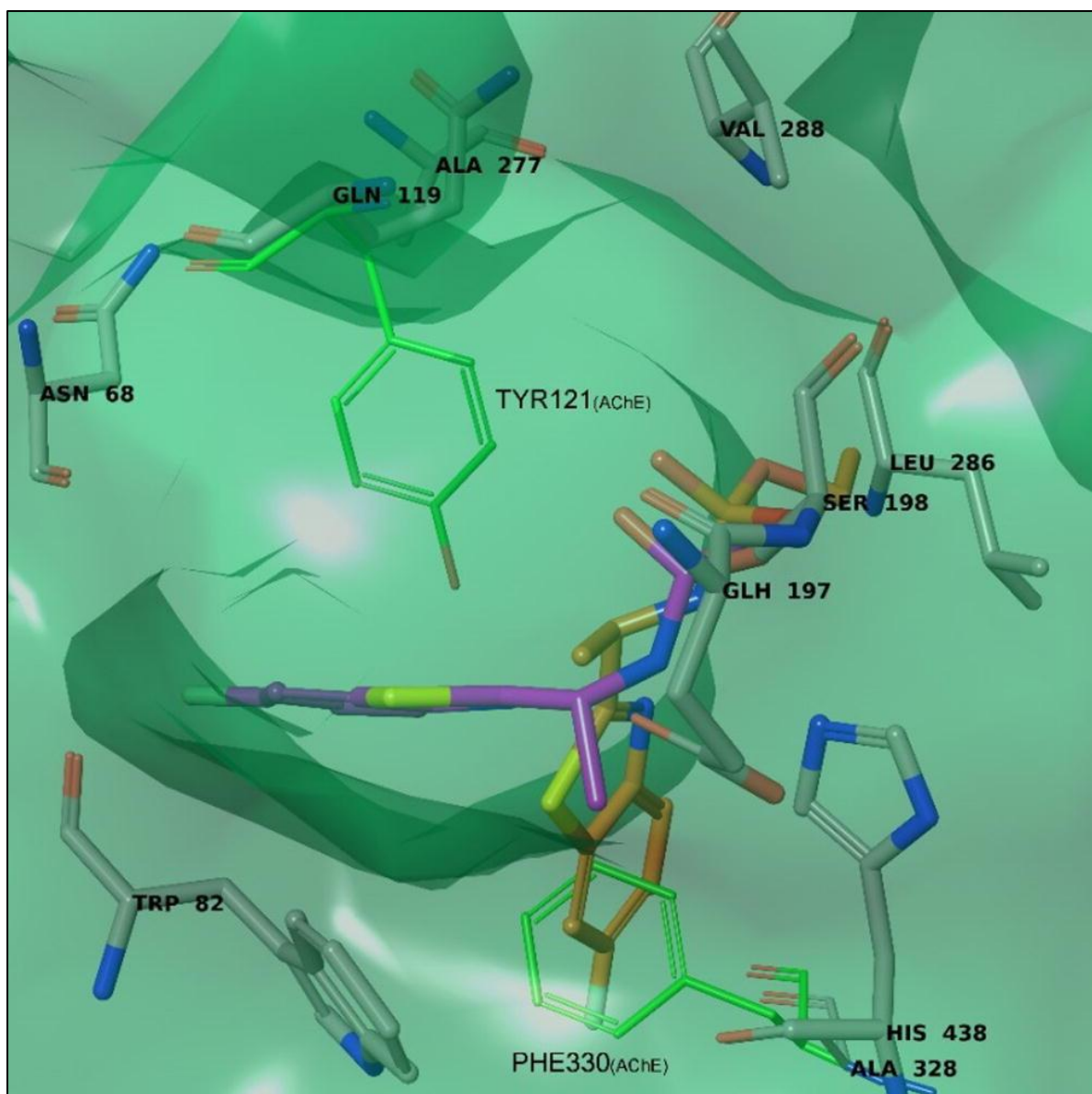
Figure 4.3: Comparison of covalently docked poses of compound 3b in TaAChE(purple) and BChE(Orange).

4.2 Comparing SQM/COSMO with Classical Scoring Functions

Docking is prime and well-established SBDD computational tool. The affinity of all the ligand geometries every generated by docking is approximated by their scores. Scoring power, ranking power, docking power, and screening power are most important parameters of any docking software. The remaining problem concerns accuracy in calculation of non-covalent interaction. Therefore, we developed SQM based Scoring function, which accurately treats non-covalent interactions. To aim this, we validate and extend the application of our SQM/COSMO SFs for native pose identification on a dataset containing diverse classes of P-L complexes.

4.2.1 Computational Method

QM based interaction energy calculations require structures of good quality. Therefore, we have applied strict criteria and selected based on them. (See Method, Publication D)

4.2.2 Result and Conclusion

We have extended our previous pilot studies, done on four difficult R-L systems by application of SQM/COSMO SFs to 17 pharmaceutically relevant and diverse complexes from five protein classes. The overall sampling power of all SFs is shown as enrichment plot (Figure 4.4A), where SQM based SFs, DFTB3-D3H4X level perform the best, followed by PM6-D3H4X, here as the worst performance was seen with GlideXP, AutoDock4, and AutoDock Vina. For detailed evaluation, we used strict False positive criterion. The SQM/COSMO SFs performed better than the classic SFs, the number of HFPs being up to 1 order of magnitude smaller compared to classic SFs. (See Results, Publication D, Figure 1B). Finally, in conclusion, the ability to recognize the native pose in cognate docking, SQM/COSMO SFs performed better than Classic SFs. Time requirement for SQM/COSMO SFs was higher than the classical SFs, but due to available supercomputer power can be evaluated in reasonable time.

A).
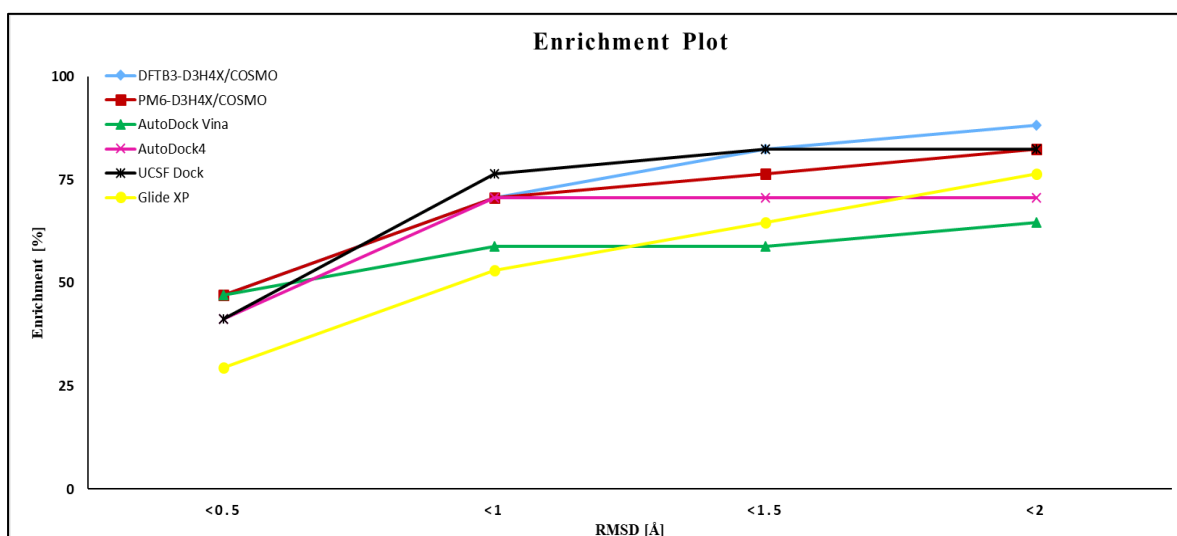
B).
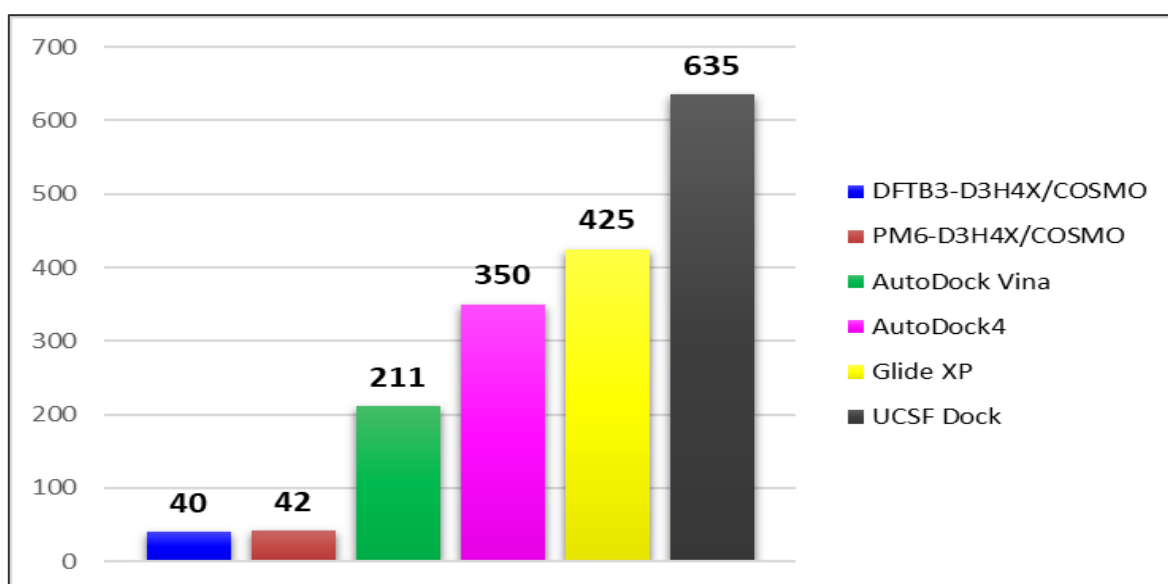
Figure 4.4: A). Enrichment plot for six scoring functions **B).** Number of HFP for Six scoring functions.

4.3 Active Site Explicit Water: Protein-Ligand Complexes

Prediction of structure and binding affinity are an essential tasks of structure-based drug design. The affinity of drugs for their target is described by thermodynamic properties. For this reason, calculation of thermodynamic properties has been the long quest of computational chemistry. As consequence, solvation patterns around drugs and their binding sites have become an important objective in molecular modeling. Hydrophobic surfaces have been estimated to cover ~75% of surfaces of drug sites, which indicates that hydrophobic effect may be an important energy contribution to ligand binding affinity.

4.3.1   Computational Method

WaterMap[185-187] is commercialized by Schrodinger Inc.[188] and based on the inhomogeneous solvation theory(IST).[161,162] The protein-ligand complexes (kinases, reductase, and hydrolase) based on X-ray structure was used for our computational protocol. (See detailed method Publication, E, F, and G) WaterMap tool analyses the results of a short (10ns) MD simulation in which the protein and ligand are typically held rigid and water molecules are able to enter and leave the binding pocket. A clustering algorithm is then applied to assign a population of sites from MD simulations. Free energy of each water site is calculated by IST, which estimates local average enthalpy and entropy contribution of individual waters binding to the site.

4.3.2   Result and Conclusion

Figure 4.5 describes the location of six water sites in CDK2 active site, which correspond to the location of crystallographic water molecules. These water molecules are present throughout the simulation time and forming the hydrogen bond chain, which interacts with inhibitor. Thermodynamically, the binding of all six water molecules in their protein sites is unfavorable

with respect to their ΔG in bulk solution (Table 4.1). This is mainly due to the entropic cost of trapping them in protein. These thermodynamic parameters of water were crucial for QM Scoring. In conclusion, active-site explicit waters and their thermodynamics helped to achieve good correlation with the experimental binding affinities and provided the information that the explicit solvent effect is much needed in the scoring procedure to obtain meaningful results. (See detailed Result, Publication E)
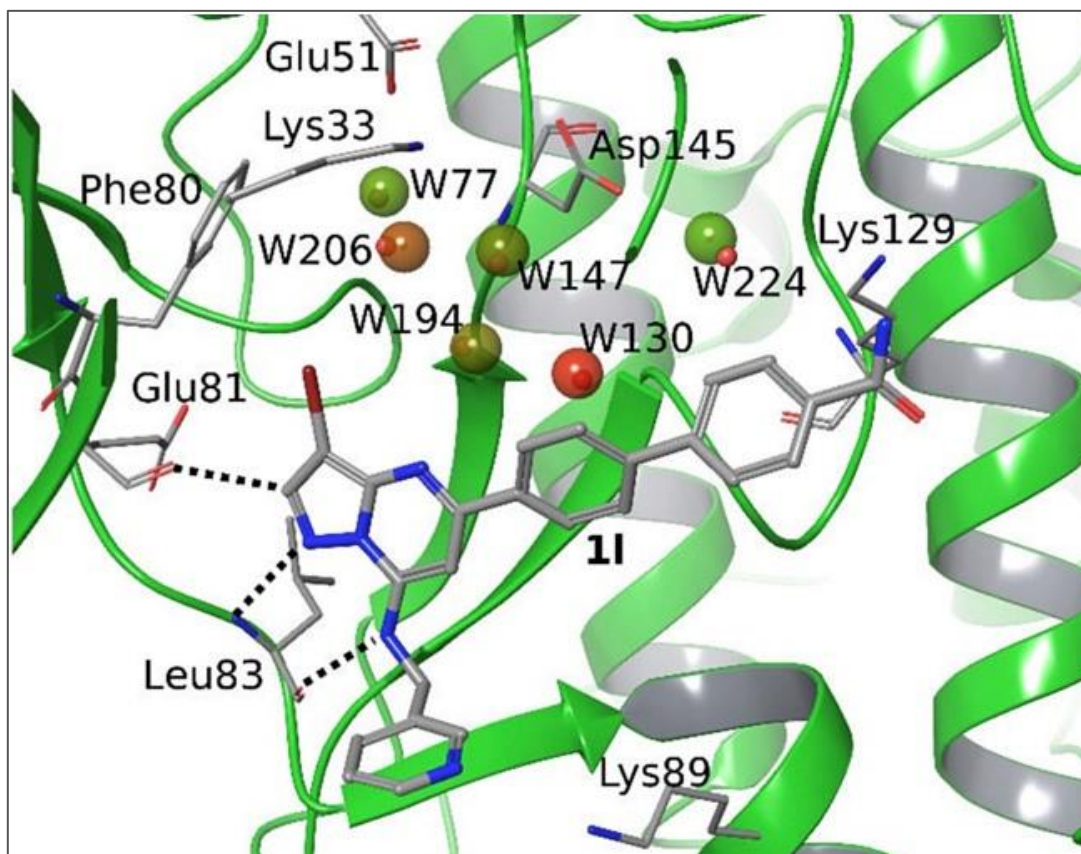
Figure 4.5: Structural water molecules in a crystallographic complex of CDK2/1l.

Table 4.1: Thermodynamic characteristics of water molecules in CDK2/1l calculated by WaterMap.(ΔG; free energy, ΔH; enthalpy, and -TΔS; entropy, all in kcal/mol)

Water	*ΔG*	*ΔH*	*-TΔS*
W77	1.2	-3.2	4.4
W206	4.4	0.5	3.9
W194	2.4	-1.6	4.0
W147	2.4	-2.0	4.5
W224	0.8	-4.4	5.1
W130	6.2	2.1	4.1

Further, displacement of long residence water molecule is usually energetically unfavorable in hydrophobic subpocket of AKR1B10 binding sites, due to their imperfect fitting and can lead the large enthalpic gain when displacing such water molecules. In the holoenzyme structure, WaterMap found buried water molecules, (Figure 4.6) which was very unfavorable in this position. Therefore, the bulkier substituent contains three ligands well fit in hydrophobic subpocket and makes strong interaction with native conformations. In conclusion, WaterMap was useful to identify the buried water molecules in AKR1B10 subpocket and their energy contribution to binding affinity. (See detailed Result, Publication F)
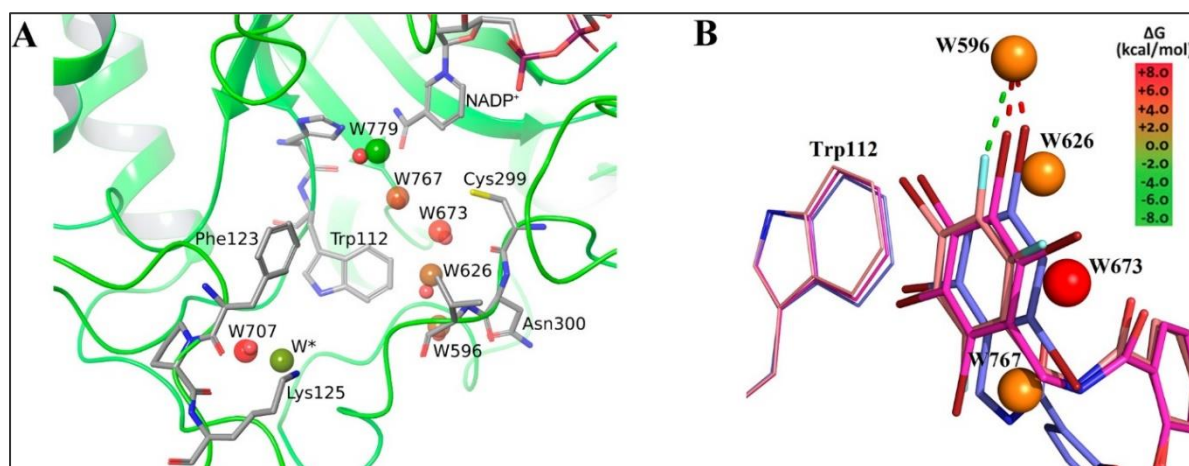
Figure 4.6: A). WaterMap Analysis of AKR1B10 apoenzyme. **B).** X-ray water positions and their energies overlaid with three inhibitors.

Finally, analysis of hydration site in Serine Racemase is described in Figure 4.7 The active site of **hSR** is highly polar, open to bulk solvent and filled with several structural water molecules. The X-ray structure contains a problematic pair of water molecules, which would have a repulsive contact due to their close distance. To overcome this problem, we have used WaterMap. In conclusion, positions/occupancies of structural water molecules are sometimes doubtful and thus they must be assessed by calculations of the hydration sites and their thermodynamics. (See detailed Result, Publication G)
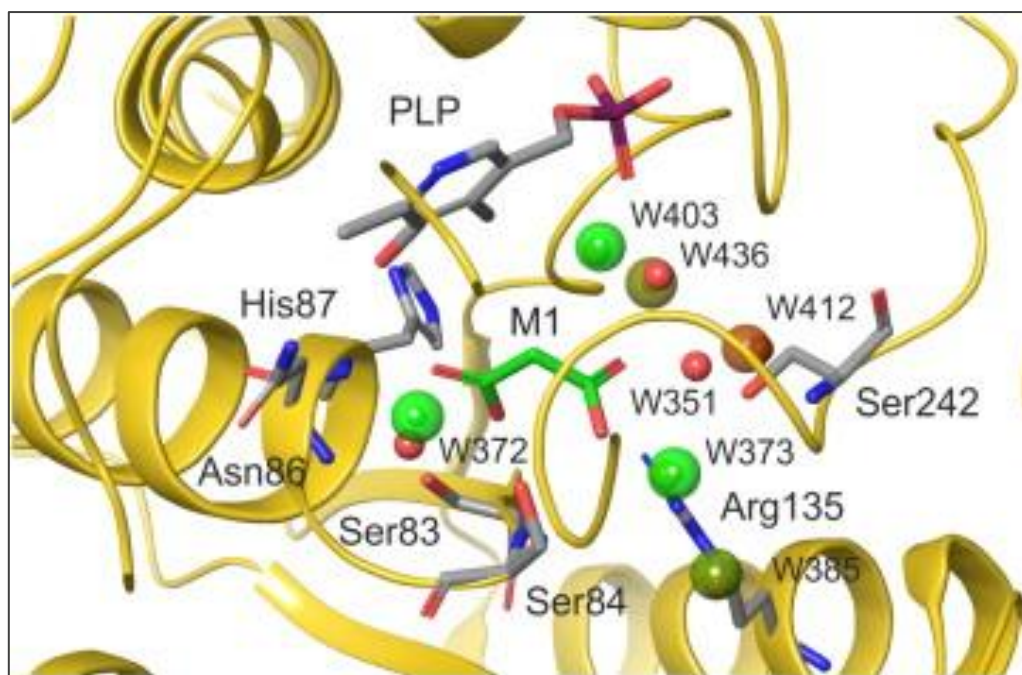
Figure 4.7: The hSR/malonate X-ray structure (PDB code: 3L6B) and active site waters. Brown; unfavorable energy, green; favorable energy, red; X-ray water positions and yellow ribbon hSR.

4.4 Explicit Water: Host-Guest Complexes

The supramolecular chemistry of host-guest complexes represents a logical step from noncovalently bound small molecules in vacuo, self-assembled triple helicate rigid cages in nonpolar solvent toward the protein-ligand complexes in water.[189] From last few decades, the chemistry of cucurbit[n]uril (CB[n], n = 5,6,7,8,10) has a promising application in material chemistry, molecular recognition, chemosensing and drug discovery.[190-194] The ability of CB[n] hosts to bind guest molecules within their cavities results from various effects nonspecific. Solvation plays a crucial role, called as nonspecific, because of damping the electrostatic effects between host and guest. The host is polar molecule and guest is charged molecule, therefore there has been significant electrostatic energy which is damped when passing from vacuum into water environment. For this reason, we have estimated the role of water molecules in Cucurbit[n]uril• host.(n=5,6,7,8) (See, Publication H, and I)

4.4.1   Computational Method

A four-isolated cucurbit[n]uril (n=5,6,7,8) and a training set of 11 complexes based on X-ray crystal structure was used for our computational protocol (Figure 4.8). We determined explicitly the high-energy water molecules inside the cavity of the host or without a host. We performed MD Water simulation (10ps long) of a CB[n] molecules dipped in 12 Å/side periodic cubic box filled with explicit water molecules. (see detailed method, Publication H).
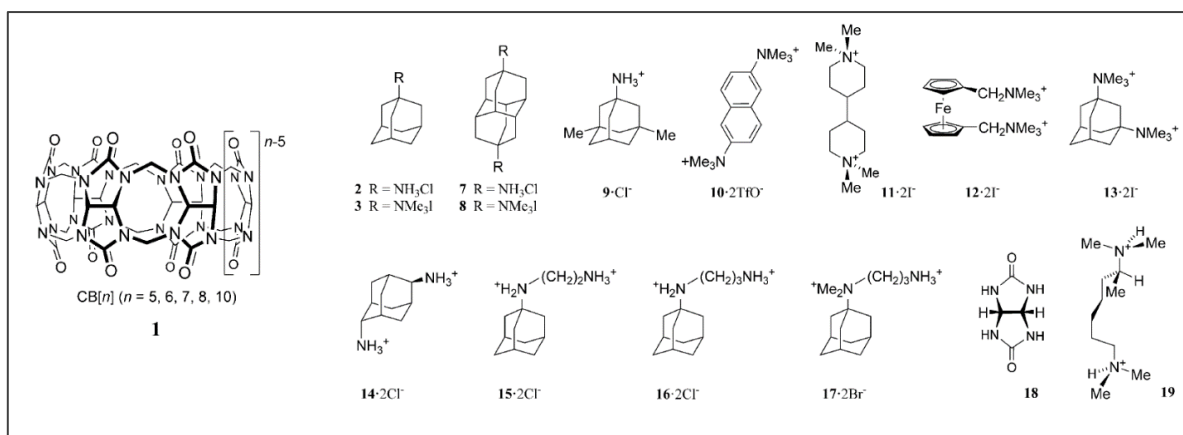
Figure 4.8: Illustration of host and guest molecules considered in our study: adamantane/diamantine, naphthalane, bipiperidine and ferrocene derivates.

4.4.2 Result and Discussion

In this section, the water thermodynamics of host and guest molecules will be discussed. Solvation energy of CB[n] hosts assessment by a WaterMap method for cucurbit[n]uril host-guest complexes will follow. Next, finding of high energy water molecules within CB[n] host molecules will be described.

4.4.2.1 Solvation Phenomena for CB[n] Host-Water-Guest Complexes.

The CB[n] host is polar and the studied guest molecules are charged. The significant part of the interaction energy originates from the electrostatic energy between them. However, it is strongly damped when passing from vacuo into water environment. It's clear that the environment can severely change the nature of binding and thus the accurate description of solvent effect represent a crucial point. Nau and co-workers,[195,196] pointed that the release of high energy water from the cavity of CB[n] macrocycles is a major determinant for guest binding in aqueous solution. Table 4.2 describes the solvation effect in detail. The free energy depends upon increasing diameter size of CB[n] hosts composed of a different number of glycouril units. (n= 5,6,7,8) It has been calculated as the energy difference between molecules in vacuo and in implicit solvent. The second column represents the theoretical difference of potential energy ($-\Delta E_{pot}$) of water molecules in a spherical cavity within the aqueous bulk and inside the host cavity, reported by Nau and co-worker(ref). The third column shows the WaterMap specific solvation free energies (ΔG_{pot}^{WM}) whereas the last column contains their enthalpy and entropy components. These results show that the most favorable solvation was found for CB[7] and CB[8]. (i.e. the smallest ΔG) It is noted that this finding is not in accord with results of Nau. However, if we take into account large error bars for all four hosts reported by Nau and co-worker, then it is apparent that reported potential energies for CB[7] and CB[8] are statically compatible. Therefore, our WaterMap generated results basically agree with their finding.

	$-\Delta E_{pot}{}^a$	$\Delta G_{pot}{}^{WM}$	$\Delta H/-T\Delta S_{pot}{}^{WM}$
CB[5]	41.6±28.8	12	7.6/4.4
CB[6]	51.1±29	21.4	12.3/9.1
CB[7]	102.4±31.3	5.3	-4.9/10.2
CB[8]	66.2±10.7	4.9	-10.5/14.9

Table 4.2: Calculated energies related to solvation of CB[n] hosts, all unit are kcal/mol. [a] data are taken from ref[197]

4.4.2.2 Finding of the high-energy of Cucrbit[n]uril (n=5,6,7,8) Host Molecules.

Table 4.3 describes the number of high energy water molecules residing within the host cavity in the absence of an encapsulated guest by WaterMap. The same number of water molecules as previously reported by MD simulation has been found in CB[n]uril and Packing Coefficient (PC) analysis.

	$MD^{[a]}$	PC $analysis^{[a]}$	WaterMap		
			Water Sites	Avg. Occupancy(%)	N_{water} $molecules^{[b]}$
CB[5]	2[2.0]	2[b]	2	94	~2
CB[6]	4[3.3]	4[b]	6	56	~4
CB[7]	7[7.9]	8[b]	12	41	~8
CB[8]	10[13.1]	16[b]	19	41	~16

Table 4.3: Number of water molecules residing within a CB[n] hosts cavity as studied by molecular dynamics, packing coefficient, and WaterMap. [a] data are taken from the ref,[195,196] [b] Number of actual water molecules by analysis of water sites, the positions, and the occupancies.

In conclusion, we discussed the accuracy of previously reported solvation properties of CB[n] molecules by Nau & Co-workers and have been well produced by explicit(WaterMap). The presence of high energy water molecules was not required for modeling of the CB[n] host-guest complexes.

Summary and Outlook

The structure-based drug design is a complex and iterative process relying on pre-existing knowledge and driven by experimental data. Explanation of how and why these molecules interact with their target is essential, which not only give insights into the fundamental processes but also for drug discovery and other related fields. To obtain the comprehensive understating of non-covalent interaction between a ligand and a protein, many aspects need to be investigated and data from different sources are needed. The studies presented in this dissertation have used for computer-aided drug design a multidisciplinary approach: a framework for Binding mode analysis, Structure-activity relationship(SAR), a receptor-ligand scoring function and a docking, and Water thermodynamics.

In the thesis, first (*cf.* **A, B, C publications**) the binding mode prediction of protein targets, namely tyrosine-kinase CDK2, Fms-like tyrosine kinase 3(FLT3), and hydrolase (AChE, BChE) is made and exemplified by the prediction of inhibitors in binding sites by using a combination of molecular modelling and biochemistry. The discovered inhibitors have been used to probe the non-covalent interaction in kinases, hydrolase and have provided detailed insight into the binding of drug-like ligands in their respective targets. The discovered inhibitors may, however, also be used as starting point in the designing of new drugs for the treatment of anticancer and nerve-agent poisoning.

Next, (*cf.* **D publications**) we presented the SQM based scoring functions, a deterministic scoring algorithm, employing a complete non-covalent interaction and iterative pose prediction via different available docking programs. In order to validate the scoring functions, there are several caveats when trying to validate new scoring terms. The compositions of the dataset used to validate and evaluate the scoring function are one of such issues. This dataset of protein-ligand complexes, structures employed for the validation needs to be large and diverse enough and may not just contain data from certain classes of enzymes. Datasets like the Astex diverse set, PDB core set, DUD library are commonly used in in-silico procedures. We observed that in "PDB coreset" includes experimental three-dimensional information. In the future enhance validation of SQM based SFs even more. The optimization and time requirement for the SQM/COSMO SFs and selective integration of binding affinity. This option allows to calculate accurate scoring terms and reduce the false positive rates. Thus, whether the molecules actually fit into the target structure or not would be more reliably predicted. There are some promising approaches for our scoring function. One way would be a more detailed and accurate calculation of solvation/desolvation effects, although great care has been taken to ensure comprehensive performance on diverse datasets. This scoring function could be useful too in structure-based design.

In this dissertation, (*cf.* **E, F, G, H and I publications**) we next presented the hydration sites of protein targets and host-guest systems. Water molecules are ubiquitously found as the interfaces between protein and ligand, and it is often stated that the interfacial water must be considered as an integral part of binding sites. The work presented here that identify how the specific interactions and buried water molecules in binding sites affect the ligand binding,

which was perhaps the most challenging as it is important and may provide unexplored opportunities in drug discovery. To the date, these effects are not understood well enough to be generally applied in the design of new ligands. The enthalpy-entropy compensation observed for protein-ligand complexes and host-guest complexes constitutes an important discovery and can provide opportunities also in future studies. Indeed, we observe that the meaningful correlation between experimental and theoretical is achieved only with unified treatment of both the protein-ligand complexes, host-guest complexes and the interfacial bridging water. Our work thus demonstrates the impact of hydration dynamics on the protein-ligand complexes and host-guest systems.

To conclude, CADD is indeed a very useful tool for pharmaceutical companies and academic research group to search for potential drug candidates with reduced cost and time. Although, the contribution of this research has been presented in the context of structure-based drug design. However, there is still room for further improvement in CADD, such as more accurate scoring functions, the solvent effect in docking and of course in terms of increasing computational efficiency.

BIBLIOGRAPHY

1. Newman, D. J.; Cragg, G. M., Natural products as sources of new drugs over the last 25 years. *J Nat Prod* **2007,** *70* (3), 461-477.

2. Lourenco, A. M.; Ferreira, L. M.; Branco, P. S., Molecules of natural origin, semi-synthesis and synthesis with anti-inflammatory and anticancer utilities. *Curr Pharm Des* **2012,** *18* (26), 3979-4046.

3. Wikberg, J. E. S.; Spjuth, O.; Eklund, M.; Lapins M., Chemoinformatics Taking Biology into Account: Proteochemometrics. **2011**.

4. Reardon, S. Project ranks billions of drug interactions. *Nature* **2013,** *503* (7477), 449-50.

5. Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L., Principles of early drug discovery. *Br J Pharmacol* **2011,** *162* (6), 1239-1249.

6. Krasavin, M.; Karapetian, R.; Konstantinov, I.; Gezentsvey, Y.; Bukhryakov, K.; Godovykh, E.; Soldatkina, O.; Lavrovsky, Y.; Sosnov, A. V.; Gakh, A. A., Discovery and potency optimization of 2-amino-5-arylmethyl-1,3-thiazole derivatives as potential therapeutic agents for prostate cancer. *Arch Pharm (Weinheim)* **2009,** *342* (7), 420-427.

7. Kaul, P. N. Drug discovery: past, present and future. *Prog Drug Res* **1998,** *50*, 9-105.

8. Veselovsky, A. V.; Zharkova, M. S.; Poroikov, V. V.; Nicklaus, M. C., Computer-aided design and discovery of protein-protein interaction inhibitors as agents for anti-HIV therapy. *SAR QSAR Environ Res* **2014,** *25* (6), 457-471.

9. Song, C. M.; Lim, S. J.; Tong, J. C., Recent advances in computer-aided drug design. *Brief Bioinform* **2009,** *10* (5), 579-591.

10. Talele, T. T.; Khedkar, S. A.; Rigby, A. C., Successful applications of computer-aided drug discovery: moving drugs from concept to the clinic. *Curr Top Med Chem* **2010,** *10* (1), 127-141.

11. Vijayakrishnan, R., Structure-based drug design and modern medicine. *J Postgrad Med* **2009,** *55* (4), 301-304.

12. Van Drie, J. H., Computer-aided drug design: the next 20 years. *J Comput Aided Mol Des* **2007,** *21* (10-11), 591-601.

13. Hartman, G. D.; Egbertson, M. S.; Halczenko, W.; Laswell, W. L.; Duggan, M. E.; Smith, R. L.; Naylor, A. M.; Manno, P. D.; Lynch, R. J.; Zhang, G.; et al., Non-peptide

fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors. *J Med Chem* **1992,** *35* (24), 4640-4642.

14. Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A., Discovery of a novel binding trench in HIV integrase. *J Med Chem* **2004,** *47* (8), 1879-1881.

15. Kim, C. U.; Lew, W.; Williams, M. A.; Liu, H.; Zhang, L.; Swaminathan, S.; Bischofberger, N.; Chen, M. S.; Mendel, D. B.; Tai, C. Y.; Laver, W. G.; Stevens, R. C., Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J Am Chem Soc* **1997,** *119* (4), 681-690.

16. Cohen, N. C., Structure-based drug design and the discovery of aliskiren (Tekturna): perseverance and creativity to overcome a R&D pipeline challenge. *Chem Biol Drug Des* **2007,** *70* (6), 557-565.

17. Levin, J. I.; Chen, J. M.; Laakso, L. M.; Du, M.; Schmid, J.; Xu, W.; Cummons, T.; Xu, J.; Jin, G.; Barone, D.; Skotnicki, J. S., Acetylenic TACE inhibitors. Part 3: Thiomorpholine sulfonamide hydroxamates. *Bioorg Med Chem Lett* **2006,** *16* (6), 1605-1609.

18. Eccles, S. A.; Massey, A.; Raynaud, F. I.; Sharp, S. Y.; Box, G.; Valenti, M.; Patterson, L.; de Haven Brandon, A.; Gowan, S.; Boxall, F.; Aherne, W.; Rowlands, M.; Hayes, A.; Martins, V.; Urban, F.; Boxall, K.; Prodromou, C.; Pearl, L.; James, K.; Matthews, T. P.; Cheung, K. M.; Kalusa, A.; Jones, K.; McDonald, E.; Barril, X.; Brough, P. A.; Cansfield, J. E.; Dymock, B.; Drysdale, M. J.; Finch, H.; Howes, R.; Hubbard, R. E.; Surgenor, A.; Webb, P.; Wood, M.; Wright, L.; Workman, P., NVP-AUY922: a novel heat shock protein 90 inhibitor active against xenograft tumor growth, angiogenesis, and metastasis. *Cancer Res* **2008,** *68* (8), 2850-2860.

19. Jorgensen, W. L., Efficient drug lead discovery and optimization. *Acc Chem Res* **2009,** *42* (6), 724-733.

20. Stahl, M.; Rarey, M., Detailed analysis of scoring functions for virtual screening. *J Med Chem* **2001,** *44* (7), 1035-1042.

21. Muller-Dethlefs, K.; Hobza, P., Noncovalent interactions: a challenge for experiment and theory. *Chem Rev* **2000,** *100* (1), 143-168.

22. Riley, K. E.; Pitonak, M.; Jurecka, P.; Hobza, P., Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. *Chem Rev* **2010,** *110* (9), 5023-5063.

23. Morokuma, K., Molecular Orbital Studies of Hydrogen Bonds. III. C=O···H–O Hydrogen Bond in H2CO···H2O and H2CO···2H2O. *J Chem Phys* **1971,** *55*, 1236.

24. Nolte, R. T.; Wisely, G. B.; Westin, S.; Cobb, J. E.; Lambert, M. H.; Kurokawa, R.; Rosenfeld, M. G.; Willson, T. M.; Glass, C. K.; Milburn, M. V., Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. *Nature* **1998,** *395* (6698), 137-143.

25. Peter A. Kollman, a. L. C. A., Theory of the hydrogen bond. *Chem Rev* **1972,** *72* (3), 283-303.

26. Lengauer, T.; Rarey, M., Computational methods for biomolecular docking. *Curr Opin Struct Biol* **1996,** *6* (3), 402-406.

27. Gschwend, D. A.; Good, A. C.; Kuntz, I. D., Molecular docking towards drug discovery. *J Mol Recognit* **1996,** *9* (2), 175-186.

28. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* **2004,** *3* (11), 935-949.

29. Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P.; DesJarlais, R. L., Docking: successes and challenges. *Curr Pharm Des* **2005,** *11* (3), 323-333.

30. Sherman, W.; Beard, H. S.; Farid, R., Use of an induced fit receptor structure in virtual screening. *Chemical Biology & Drug Design* **2006,** *67* (1), 83-84.

31. Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R., Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem* **2006,** *49* (2), 534-553.

32. Wang, R. X.; Lu, Y. P.; Wang, S. M., Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem* **2003,** *46* (12), 2287-2303.

33. Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R., Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br J Pharmacol* **2008,** *153*.

34. Huang, S. Y.; Grinter, S. Z.; Zou, X., Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys* **2010,** *12* (40), 12899-12908.

35. Bohm, H. J., The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known 3-dimensional structure. *J Comput Aided Mol Des* **1994,** *8* (3), 243-256.

36. Muegge, I.; Martin, Y. C., A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J Med Chem* **1999,** *42* (5), 791-804.

37. Gohlke, H.; Hendlich, M.; Klebe, G., Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* **2000,** *295* (2), 337-356.

38. Paulsen, J. L.; Anderson, A. C., Scoring ensembles of docked protein: ligand interactions for virtual lead optimization. *J Chem Inf Model* **2009,** *49* (12), 2813-2819.

39. Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P., Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* **1999,** *42* (25), 5100-5109.

40. Rezac, J.; Hobza, P., Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chem Rev* **2016,** *116* (9), 5038-5071.

41. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P., Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* **1997,** *11* (5), 425-445.

42. Bohm, H. J., Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J Comput Aided Mol Des* **1998,** *12* (4), 309-323.

43. Muegge, I., A knowledge-based scoring function for protein-ligand interactions: Probing the reference state. *Perspectives in Drug Discovery and Design* **2000,** *20* (1), 99-114.

44. Huang, S. Y.; Zou, X., An iterative knowledge-based scoring function for protein-protein recognition. *Proteins* **2008,** *72* (2), 557-579.

45. Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A., Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J Am Chem Soc* **1998,** *120* (37), 9401-9409.

46. Feig, M.; Brooks, C. L., 3rd, Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr Opin Struct Biol* **2004,** *14* (2), 217-24.

47. Rezac, J.; Fanfrlik, J.; Salahub, D.; Hobza, P., Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes. *J Chem Theory Comput* **2009,** *5* (7), 1749-1760.

48. Korth, M.; Pitonak, M.; Rezac, J.; Hobza, P., A Transferable H-Bonding Correction for Semiempirical Quantum-Chemical Methods. *J Chem Theory Comput* **2010,** *6* (1), 344-52.

49. Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E., Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. *J Chem Phys* **2001,** *114* (12), 5149-5155.

50. Miriyala, V. M.; Rezac, J., Description of non-covalent interactions in SCC-DFTB methods. *J Comput Chem* **2017,** *38* (10), 688-697.

51. Raha, K.; Merz, K. M., A quantum mechanics-based scoring function: Study of zinc ion-mediated ligand binding. *J Am Chem Soc* **2004,** *126* (4), 1020-1021.

52. Raha, K.; Merz, K. M., Large-scale validation of a quantum mechanics based scoring function: Predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes. *J Med Chem* **2005,** *48* (14), 4558-4575.

53. Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; Wollacott, A. M.; Westerhoff, L. M.; Merz, K. M., Jr., The role of quantum mechanics in structure-based drug design. *Drug Discov Today* **2007,** *12* (17-18), 725-731.

54. Mikulskis, P.; Genheden, S.; Wichmann, K.; Ryde, U., A semi-empirical approach to ligand-binding affinities: dependence on the Hamiltonian and corrections. *J Comput Chem* **2012,** *33* (12), 1179-1189.

55. Zhou, H. X.; Gilson, M. K., Theory of free energy and entropy in noncovalent binding. *Chem Rev* **2009,** *109* (9), 4092-4107.

56. Karplus, M.; Kushick, J. N., Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules* **1981,** *14* (2), 325-332.

57. Tzeng, S. R.; Kalodimos, C. G., Protein activity regulation by conformational entropy. *Nature* **2012,** *488* (7410), 236-240.

58. Freire, E., Do enthalpy and entropy distinguish first in class from best in class? *Drug Discovery Today* **2008,** *13* (19-20), 869-874.

59. G C Pimentel, a.; McClellan, A. L., Hydrogen Bonding. *Annual Review of Phys Chem* **1971,** *22* (1), 347-385.

60. Burley, S. K.; Petsko, G. A., Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* **1985,** *229* (4708), 23-28.

61. Hunter, C. A.; Singh, J.; Thornton, J. M., Pi-pi interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. *J Mol Biol* **1991,** *218* (4), 837-846.

62. Bissantz, C.; Kuhn, B.; Stahl, M., A medicinal chemist's guide to molecular interactions. *J Med Chem* **2010,** *53* (14), 5061-5084.

63. Riley, K. E.; Hobza, P., On the importance and origin of aromatic interactions in chemistry and biodisciplines. *Acc Chem Res* **2013,** *46* (4), 927-936.

64. Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M., Principles and applications of halogen bonding in medicinal chemistry and chemical biology. *J Med Chem* **2013,** *56* (4), 1363-1388.

65. Ajani, H.; Jansa, J.; Köprülüoğlu, C.; Hobza, P.; Kryštof, V.; Lyčka, A.; Lepšík, M. Imidazo[1,2-c]pyrimidin-5(6H)-one as a Novel Core of Cyclin-Dependent Kinase 2 Inhibitors: Synthesis, Activity Measurement, Docking and Quantum Mechanical Scoring. *J Mol Recog*, DOI:10.1002/jmr.2720

66. Pejchal, V.; Stepankova, S.; Pejchalova, M.; Kralovec, K.; Havelek, R.; Ruzickova, Z.; Ajani, H.; Lo, R.; Lepsik, M., Synthesis, structural characterization, docking, lipophilicity and cytotoxicity of 1-[(1R)-1-(6-fluoro-1,3-benzothiazol-2-yl)ethyl]-3-alkyl carbamates, novel acetylcholinesterase and butyrylcholinesterase pseudo-irreversible inhibitors. *Bioorg Med Chem* **2016,** *24* (7), 1560-1572.

67. Ajani, H.; Pecina, A.; Eyrilmez, S. M.; Fanfrlik, J.; Haldar, S.; Rezac, J.; Hobza, P.; Lepsik, M., Superior Performance of the SQM/COSMO Scoring Functions in Native Pose Recognition of Diverse Protein-Ligand Complexes in Cognate Docking. *Acs Omega* **2017,** *2* (7), 4022-4029.

68. Gucký, T.; Řezníčková, E.; Jorda, R.; Muchová, T. R.; Klejová, Z.; Malínková, V.; Berka, K.; Bazgier, V.; Ajani, H.; Lepšík, M.; Divoký, V.; Kryštof, V. Discovery of N2-(4-Amino-cyclohexyl)-9-cyclopentyl-N6-(4-morpholin-4-ylmethyl-phenyl)-9H-purine-2,6-diamine as a Selective and Potent FLT3 Kinase Inhibitor for FLT3-ITD Positive Acute Myeloid Leukemia. *J Med Chem*, in revision

69. Hylsova, M.; Carbain, B.; Fanfrlik, J.; Musilova, L.; Haldar, S.; Kopruluoglu, C.; Ajani, H.; Brahmkshatriya, P. S.; Jorda, R.; Krystof, V.; Hobza, P.; Echalier, A.; Paruch, K.; Lepsik, M., Explicit treatment of active-site waters enhances quantum mechanical/implicit solvent scoring: Inhibition of CDK2 by new pyrazolo[1,5-a]pyrimidines. *Eur J Med Chem* **2017,** *126*, 1118-1128.

70. Cousido-Siah, A.; Ruiz, F. X.; Fanfrlik, J.; Gimenez-Dejoz, J.; Mitschler, A.; Kamlar, M.; Vesely, J.; Ajani, H.; Pares, X.; Farres, J.; Hobza, P.; Podjarny, A. D., IDD388 Polyhalogenated Derivatives as Probes for an Improved Structure-Based Selectivity of AKR1B10 Inhibitors. *Acs Chem Biol* **2016,** *11* (10), 2693-2705.

71. Vorlova, B.; Nachtigallova, D.; Jiraskova-Vanickova, J.; Ajani, H.; Jansa, P.; Rezac, J.; Fanfrlik, J.; Otyepka, M.; Hobza, P.; Konvalinka, J.; Lepsik, M., Malonate-based inhibitors of mammalian serine racemase: kinetic characterization and structure-based computational study. *Eur J Med Chem* **2015,** *89*, 189-197.

72. Hostas, J.; Sigwalt, D.; Sekutor, M.; Ajani, H.; Dubecky, M.; Rezac, J.; Zavalij, P. Y.; Cao, L.; Wohlschlager, C.; Mlinaric-Majerski, K.; Isaacs, L.; Glaser, R.; Hobza, P., A Nexus between Theory and Experiment: Non-Empirical Quantum Mechanical Computational Methodology Applied to Cucurbit[n]urilGuest Binding Interactions. *Chemistry* **2016,** *22* (48), 17226-17238.

73. Sigwalt, D.; Sekutor, M.; Cao, L.; Zavalij, P. Y.; Hostas, J.; Ajani, H.; Hobza, P.; Mlinaric-Majerski, K.; Glaser, R.; Isaacs, L., Unraveling the Structure-Affinity Relationship between Cucurbit[n]urils (n=7, 8) and Cationic Diamondoids. *J Am Chem Soc* **2017,** *139* (8), 3249-3258.

74. Congreve, M.; Murray, C. W.; Blundell, T. L., Structural biology and drug discovery. *Drug Discov Today* **2005,** *10* (13), 895-907.

75. Lyne, P. D., Structure-based virtual screening: an overview. *Drug Discov Today* **2002,** *7* (20), 1047-1055.

76. Philip Prathipati, A. D., Anil K. Saxena, Computer-Aided Drug Design: Integration of Structure-Based and Ligand-Based Approaches in Drug Design. *Current Computer-Aided Drug Design* **2007,** *3* (2), 133-148.

77. Kirchmair, J.; Wolber, G.; Laggner, C.; Langer, T., Comparative performance assessment of the conformational model generators omega and catalyst: A large-scale survey on the retrieval of protein-bound ligand conformations. *J Chem Inf Model* **2006,** *46* (4), 1848-1861.

78. Warren, G. L.; Do, T. D.; Kelley, B. P.; Nicholls, A.; Warren, S. D., Essential considerations for using protein-ligand structures in drug discovery. *Drug Discovery Today* **2012,** *17* (23-24), 1270-1281.

79. Yuriev, E.; Agostino, M.; Ramsland, P. A., Challenges and advances in computational docking: 2009 in review. *J Mol Recognit* **2011,** *24* (2), 149-164.

80. Agrafiotis, D. K.; Gibbs, A. C.; Zhu, F.; Izrailev, S.; Martin, E., Conformational sampling of bioactive molecules: a comparative study. *J Chem Inf Model* **2007,** *47* (3), 1067-1086.

81. Bissantz, C.; Folkers, G.; Rognan, D., Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* **2000,** *43* (25), 4759-4767.

82. Sousa, S. F.; Fernandes, P. A.; Ramos, M. J., Protein-ligand docking: current status and future challenges. *Proteins* **2006,** *65* (1), 15-26.

83. Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S. B.; Johnson, A. P., eHiTS: a new fast, exhaustive flexible ligand docking system. *J Mol Graph Model* **2007,** *26* (1), 198-212.

84. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004,** *47* (7), 1739-1749.

85. Kramer, B.; Rarey, M.; Lengauer, T., Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* **1999,** *37* (2), 228-241.

86. Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D., DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* **2001,** *15* (5), 411-428.

87. Joseph-McCarthy, D.; Alvarez, J. C., Automated generation of MCSS-derived pharmacophoric DOCK site points for searching multiconformation databases. *Proteins* **2003,** *51* (2), 189-202.

88. Miller, M. D.; Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P., FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *J Comput Aided Mol Des* **1994,** *8* (2), 153-174.

89. Sauton, N.; Lagorce, D.; Villoutreix, B. O.; Miteva, M. A., MS-DOCK: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. *BMC Bioinformatics* **2008,** *9*, 184.

90. Gorelik, B.; Goldblum, A., High-quality binding modes in docking ligands to proteins. *Proteins* **2008,** *71* (3), 1373-1386.

91. Abagyan, R.; Totrov, M.; Kuznetsov, D., Icm - a New Method for Protein Modeling and Design - Applications to Docking and Structure Prediction from the Distorted Native Conformation. *J Comput Chem* **1994,** *15* (5), 488-506.

92. Maurer, M. C.; Trosset, J. Y.; Lester, C. C.; DiBella, E. E.; Scheraga, H. A., New general approach for determining the solution structure of a ligand bound weakly to a receptor: structure of a fibrinogen Aalpha-like peptide bound to thrombin (S195A) obtained using NOE distance constraints and an ECEPP/3 flexible docking program. *Proteins* **1999,** *34* (1), 29-48.

93. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R., Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **1997,** *267* (3), 727-748.

94. Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J., Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J Comput Aided Mol Des* **1996,** *10* (4), 293-304.

95. Pei, J.; Wang, Q.; Liu, Z.; Li, Q.; Yang, K.; Lai, L., PSI-DOCK: towards highly efficient and accurate flexible ligand docking. *Proteins* **2006,** *62* (4), 934-946.

96. Namasivayam, V.; Gunther, R., pso@autodock: a fast flexible molecular docking program based on Swarm intelligence. *Chem Biol Drug Des* **2007,** *70* (6), 475-484.

97. Korb, O.; Stutzle, T.; Exner, T. E., Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J Chem Inf Model* **2009,** *49* (1), 84-96.

98. Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C., DOCK 6: Impact of New Features and Current Docking Performance. *J Comput Chem* **2015,** *36* (15), 1132-1156.

99. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem* **2009,** *30* (16), 2785-2791.

100. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **2010,** *31* (2), 455-461.

101. Jones, G.; Willett, P.; Glen, R. C., Molecular Recognition of Receptor-Sites Using a Genetic Algorithm with a Description of Desolvation. *J Mol Biol* **1995,** *245* (1), 43-53.

102. Corbeil, C. R.; Williams, C. I.; Labute, P., Variability in docking success rates due to dataset preparation. *J Comput Aided Mol Des* **2012,** *26* (6), 775-786.

103. Jain, A. N., Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem* **2003,** *46* (4), 499-511.

104. Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M., LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J Mol Graph Model* **2003,** *21* (4), 289-307.

105. Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A. B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D., rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput Biol* **2014,** *10* (4), e1003571.

106. Zhao, H. T.; Caflisch, A., Discovery of ZAP70 inhibitors by high-throughput docking into a conformation of its kinase domain generated by molecular dynamics. *Bioorg Med Chem Lett* **2013,** *23* (20), 5721-5726.

107. Kumalo, H. M.; Bhakat, S.; Soliman, M. E., Theory and applications of covalent docking in drug discovery: merits and pitfalls. *Molecules* **2015,** *20* (2), 1984-2000.

108. Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A., The resurgence of covalent drugs. *Nat Rev Drug Discov* **2011,** *10* (4), 307-317.

109. Ouyang, X.; Zhou, S.; Su, C. T.; Ge, Z.; Li, R.; Kwoh, C. K., CovalentDock: automated covalent docking with parameterized covalent linkage energy estimation and molecular geometry constraints. *J Comput Chem* **2013,** *34* (4), 326-336.

110. London, N.; Miller, R. M.; Krishnan, S.; Uchida, K.; Irwin, J. J.; Eidam, O.; Gibold, L.; Cimermancic, P.; Bonnet, R.; Shoichet, B. K.; Taunton, J., Covalent docking of large libraries for the discovery of chemical probes. *Nat Chem Biol* **2014,** *10* (12), 1066-1072.

111. Bianco, G.; Forli, S.; Goodsell, D. S.; Olson, A. J., Covalent docking using autodock: Two-point attractor and flexible side chain methods. *Protein Sci* **2016,** *25* (1), 295-301.

112. Leach, A. R.; Shoichet, B. K.; Peishoff, C. E., Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *J Med Chem* **2006,** *49* (20), 5851-5855.

113. Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S., A critical assessment of docking programs and scoring functions. *J Med Chem* **2006,** *49* (20), 5912-5931.

114. Jones, S.; Thornton, J. M., Analysis of protein-protein interaction sites using surface patches. *J Mol Biol* **1997,** *272* (1), 121-132.

115. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D., Improved protein-ligand docking using GOLD. *Proteins* **2003,** *52* (4), 609-623.

116. Gohlke, H.; Klebe, G., Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chem Int Ed Engl* **2002,** *41* (15), 2644-2676.

117. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J Comput Chem* **2004,** *25* (9), 1157-1174.

118. Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A., *{Amber 14}*. 2014.

119. Xu, M.; Lill, M. A., Induced fit docking, and the use of QM/MM methods in docking. *Drug Discov Today Technol* **2013,** *10* (3), 411-418.

120. Mucs, D.; Bryce, R. A., The application of quantum mechanics in structure-based drug design. *Expert Opin Drug Discov* **2013,** *8* (3), 263-276.

121. Hayik, S. A.; Dunbrack, R., Jr.; Merz, K. M., Jr., A Mixed QM/MM Scoring Function to Predict Protein-Ligand Binding Affinity. *J Chem Theory Comput* **2010,** *6* (10), 3079-3091.

122. Hennemann, M.; Clark, T., EMPIRE: a highly parallel semiempirical molecular orbital program: 1: self-consistent field calculations. *J Mol Model* **2014,** *20* (7), 2331.

123. Muddana, H. S.; Gilson, M. K., Calculation of Host-Guest Binding Affinities Using a Quantum-Mechanical Energy Model. *J Chem Theory Comput* **2012,** *8* (6), 2023-2033.

124. Soderhjelm, P.; Kongsted, J.; Ryde, U., Ligand Affinities Estimated by Quantum Chemical Calculations. *J Chem Theory Comput* **2010,** *6* (5), 1726-1737.

125. Wichapong, K.; Rohe, A.; Platzer, C.; Slynko, I.; Erdmann, F.; Schmidt, M.; Sippl, W., Application of Docking and QM/MM-GBSA Rescoring to Screen for Novel Myt1 Kinase Inhibitors. *J Chem Inf Model* **2014,** *54* (3), 881-893.

126. Chaskar, P.; Zoete, V.; Rohrig, U. F., Toward On-The-Fly Quantum Mechanical/Molecular Mechanical (QM/MM) Docking: Development and Benchmark of a Scoring Function. *J Chem Inf Model* **2014,** *54* (11), 3137-3152.

127. Burger, S. K.; Thompson, D. C.; Ayers, P. W., Quantum Mechanics/Molecular Mechanics Strategies for Docking Pose Refinement: Distinguishing between Binders and Decoys in Cytochrome c Peroxidase. *J Chem Inf Model* **2011,** *51* (1), 93-101.

128. Antony, J.; Grimme, S.; Liakos, D. G.; Neese, F., Protein-Ligand Interaction Energies with Dispersion Corrected Density Functional Theory and High-Level Wave Function Based Methods. *J Phys Chem A* **2011,** *115* (41), 11210-11220.

129. Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V., Fragmentation methods: a route to accurate calculations on large systems. *Chem Rev* **2012,** *112* (1), 632-672.

130. Antony, J.; Grimme, S., Fully ab initio protein-ligand interaction energies with dispersion corrected density functional theory. *J Comput Chem* **2012,** *33* (21), 1730-1739.

131. Lepsik, M.; Rezac, J.; Kolar, M.; Pecina, A.; Hobza, P.; Fanfrlik, J., The Semiempirical Quantum Mechanical Scoring Function for In Silico Drug Design. *Chempluschem* **2013,** *78* (9), 921-931.

132. Rezac, J.; Hobza, P., A halogen-bonding correction for the semiempirical PM6 method. *Chem Phys Let* **2011,** *506* (4-6), 286-289.

133. Rezac, J.; Hobza, P., Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J Chem Theory Comput* **2012,** *8* (1), 141-151.

134. Klamt, A.; Schuurmann, G., COSMO - A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J Chem Soc-Perkin Trans2* **1993,** (5), 799-805.

135. Stewart, J. J., Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J Mol Model* **2007,** *13* (12), 1173-213.

136. Fanfrlik, J.; Bronowska, A. K.; Rezac, J.; Prenosil, O.; Konvalinka, J.; Hobza, P., A Reliable Docking/Scoring Scheme Based on the Semiempirical Quantum Mechanical PM6-DH2 Method Accurately Covering Dispersion and H-Bonding: HIV-1 Protease with 22 Ligands. *J Phys Chem B* **2010,** *114* (39), 12666-12678.

137. Dobes, P.; Rezac, J.; Fanfrlik, J.; Otyepka, M.; Hobza, P., Semiempirical Quantum Mechanical Method PM6-DH2X Describes the Geometry and Energetics of CK2-Inhibitor Complexes Involving Halogen Bonds Well, While the Empirical Potential Fails. *J Phys Chem B* **2011,** *115* (26), 8581-8589.

138. Fanfrlik, J.; Kolar, M.; Kamlar, M.; Hurny, D.; Ruiz, F. X.; Cousido-Siah, A.; Mitschler, A.; Rezac, J.; Munusamy, E.; Lepsik, M.; Matejicek, P.; Vesely, J.; Podjarny, A.; Hobza, P., Modulation of Aldose Reductase Inhibition by Halogen Bond Tuning. *Acs Chem Biol* **2013,** *8* (11), 2484-2492.

139. Fanfrlik, J.; Ruiz, F. X.; Kadlcikova, A.; Rezac, J.; Cousido-Siah, A.; Mitschler, A.; Haldar, S.; Lepsik, M.; Kolar, M. H.; Majer, P.; Podjarny, A. D.; Hobza, P., The Effect of Halogen-to-Hydrogen Bond Substitution on Human Aldose Reductase Inhibition. *Acs Chem Biol* **2015,** *10* (7), 1637-1642.

140. Pecina, A.; Lepsik, M.; Rezac, J.; Brynda, J.; Mader, P.; Rezacova, P.; Hobza, P.; Fanfrlik, J., QM/MM calculations reveal the different nature of the interaction of two carborane-based sulfamide inhibitors of human carbonic anhydrase II. *J Phys Chem B* **2013,** *117* (50), 16096-16104.

141. Pecina, A.; Meier, R.; Fanfrlik, J.; Lepsik, M.; Rezac, J.; Hobza, P.; Baldauf, C., The SQM/COSMO filter: reliable native pose identification based on the quantum-mechanical description of protein-ligand interactions and implicit COSMO solvation. *Chem Commun* **2016,** *52* (16), 3312-3315.

142. Pecina, A.; Haldar, S.; Fanfrlik, J.; Meier, R.; Rezac, J.; Lepsik, M.; Hobza, P., SQM/COSMO Scoring Function at the DFTB3-D3H4 Level: Unique Identification of Native Protein-Ligand Poses. *J Chem Inf Model* **2017,** *57* (2), 127-132.

143. Fanfrlik, J.; Brahmkshatriya, P. S.; Rezac, J.; Jilkova, A.; Horn, M.; Mares, M.; Hobza, P.; Lepsik, M., Quantum Mechanics-Based Scoring Rationalizes the Irreversible Inactivation of Parasitic Schistosoma mansoni Cysteine Peptidase by Vinyl Sulfone Inhibitors. *J Phys Chem B* **2013,** *117* (48), 14973-14982.

144. Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R., Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* **2002,** *47* (4), 409-43.

145. Wang, R. X.; Lai, L. H.; Wang, S. M., Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des* **2002,** *16* (1), 11-26.

146. Gohlke, H.; Klebe, G., Statistical potentials and scoring functions applied to protein-ligand binding. *Curr Opin Struct Biol* **2001,** *11* (2), 231-235.

147. Huang, S. Y.; Zou, X., An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials. *J Comput Chem* **2006,** *27* (15), 1866-1875.

148. Spitzmuller, A.; Velec, H. F. G.; Klebe, G., MiniMuDS: A New Optimizer using Knowledge-Based Potentials Improves Scoring of Docking Solutions. *J Chem Inf Model* **2011,** *51* (6), 1423-1430.

149. Ishchenko, A. V.; Shakhnovich, E. I., SMall molecule growth 2001 (SMoG2001): An improved knowledge-based scoring function for protein-ligand interactions. *J Med Chem* **2002,** *45* (13), 2770-2780.

150. Velec, H. F.; Gohlke, H.; Klebe, G., DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J Med Chem* **2005,** *48* (20), 6296-303.

151. Muegge, I., PMF scoring revisited. *J Med Chem* **2006,** *49* (20), 5895-902.

152. Liu, S.; Fu, R.; Zhou, L. H.; Chen, S. P., Application of consensus scoring and principal component analysis for virtual screening against beta-secretase (BACE-1). *PLoS One* **2012,** *7* (6), e38086.

153. Perola, E.; Walters, W. P.; Charifson, P. S., A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **2004,** *56* (2), 235-249.

154. Sleigh, S. H.; Seavers, P. R.; Wilkinson, A. J.; Ladbury, J. E.; Tame, J. R., Crystallographic and calorimetric analysis of peptide binding to OppA protein. *J Mol Biol* **1999,** *291* (2), 393-415.

155. Roe, S. M.; Prodromou, C.; O'Brien, R.; Ladbury, J. E.; Piper, P. W.; Pearl, L. H., Structural basis for inhibition of the Hsp90 molecular chaperone by the antitumor antibiotics radicicol and geldanamycin. *J Med Chem* **1999,** *42* (2), 260-266.

156. Lu, Y.; Wang, R.; Yang, C. Y.; Wang, S., Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J Chem Inf Model* **2007,** *47* (2), 668-675.

157. Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G. J., Involvement of water in carbohydrate-protein binding. *J Am Chem Soc* **2001,** *123* (49), 12238-12247.

158. Kadirvelraj, R.; Foley, B. L.; Dyekjaer, J. D.; Woods, R. J., Involvement of water in carbohydrate-protein binding: concanavalin A revisited. *J Am Chem Soc* **2008,** *130* (50), 16933-16942.

159. Mikol, V.; Papageorgiou, C.; Borer, X., The role of water molecules in the structure-based design of (5-hydroxynorvaline)-2-cyclosporin: synthesis, biological activity, and crystallographic analysis with cyclophilin A. *J Med Chem* **1995,** *38* (17), 3361-3367.

160. Henchman, R. H.; McCammon, J. A., Extracting hydration sites around proteins from explicit water simulations. *J Comput Chem* **2002,** *23* (9), 861-869.

161. Lazaridis, T., Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *J Phys Chem B* **1998,** *102* (18), 3531-3541.

162. Lazaridis, T., Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids. *J Phys Chem B* **1998,** *102* (18), 3542-3550.

163. Li, Z.; Lazaridis, T., Thermodynamic contributions of the ordered water molecule in HIV-1 protease. *J Am Chem Soc* **2003,** *125* (22), 6636-6637.

164. Li, Z.; Lazaridis, T., Thermodynamics of buried water clusters at a protein-ligand binding interface. *J Phys Chem B* **2006,** *110* (3), 1464-1475.

165. Li, Z.; Lazaridis, T., The effect of water displacement on binding thermodynamics: concanavalin A. *J Phys Chem B* **2005,** *109* (1), 662-670.

166. W. A. Freeman, W. L. M., and N. Y. Shih, Cucurbituril. *J Am Chem Soc* **1981,** *103* (24), 7367-7368.

167. Jaheon Kim, In-Sun Jung, Soo-Young Kim, Eunsung Lee, Jin-Koo Kang, Shigeru Sakamoto, Kentaro Yamaguchi, Kimoon Kim. New Cucurbituril Homologues: Syntheses, Isolation, Characterization, and X-ray Crystal Structures of Cucurbit[n]uril (n = 5, 7, and 8). *J Am Chem Soc* **2000,** *122* (3), 540-541.

168. Day, A. I.; Blanch, R. J.; Arnold, A. P.; Lorenzo, S.; Lewis, G. R.; Dance, I., A cucurbituril-based gyroscane: a new supramolecular form. *Angew Chem Int Ed Engl* **2002,** *41* (2), 275-277.

169. Day, A.; Arnold, A. P.; Blanch, R. J.; Snushall, B., Controlling factors in the synthesis of cucurbituril and its homologues. *J Org Chem* **2001,** *66* (24), 8094-8100.

170. Aboul-Enein, A. A. E. H. Y., Supramolecular Analytical Application of Cucurbit[n]urils Using Fluorescence Spectroscopy. *Critical Reviews in Anal Chem* **2014,** *45* (1), 52-61.

171. Pemberton, B. C.; Singh, R. K.; Johnson, A. C.; Jockusch, S.; Da Silva, J. P.; Ugrinov, A.; Turro, N. J.; Srivastava, D. K.; Sivaguru, J., Supramolecular photocatalysis: insights into cucurbit[8]uril catalyzed photodimerization of 6-methylcoumarin. *Chem Commun* **2011,** *47* (22), 6323-6325.

172. Ying-WeiYang, H., Gold nanoparticles functionalized with supramolecular macrocycles. *Chinese Chem Lett* **2013,** *24* (7), 545-552.

173. You-Moon Jeon, J. K., Dongmok Whang, and, and Kimoon Kim, Molecular Container Assembly Capable of Controlling Binding and Release of Its Guest Molecules: Reversible Encapsulation of Organic Molecules in Sodium Ion Complexed Cucurbituril. *J Am Chem Soc* **1996,** *118* (40), 9790-9791.

174. Jing-Xin Liu, L.-S. L., Rong-Bin Huang, and Lan-Sun Zheng, Molecular Capsules Based on Cucurbit[5]uril Encapsulating "Naked" Anion Chlorine. *Cryst Growth Des* **2006,** *6* (11), 2611-2614.

175. Fedin, O. D. a. V. P., Solid-State Supramolecular Assemblies of Tryptophan and Tryptamine with Cucurbit[6]Uril. *Cryst Growth Des* **2012,** *12* (2), 550-555.

176. Ayhan, M. M.; Karoui, H.; Hardy, M.; Rockenbauer, A.; Charles, L.; Rosas, R.; Udachin, K.; Tordo, P.; Bardelang, D.; Ouari, O., Comprehensive Synthesis of Monohydroxy-Cucurbit[n]urils (n = 5, 6, 7, 8): High Purity and High Conversions. *J Am Chem Soc* **2015,** *137* (32), 10238-10245.

177. Schollmeyer, H. J. B. E. C. K. J. E., Determination of complex stabilities with nearly insoluble host molecules: cucurbit[5]uril, decamethylcucurbit[5]uril and cucurbit[6]uril as ligands for the complexation of some multicharged cations in aqueous solution. *Analytica Chimica Acta* **1994,** *437* (1), 157-163.

178. Gerasko, O. A.; Mainicheva, E. A.; Naumova, M. I.; Neumaier, M.; Kappes, M. M.; Lebedkin, S.; Fenske, D.; Fedin, V. P., Sandwich-type tetranuclear lanthanide complexes with cucurbit[6]uril: from molecular compounds to coordination polymers. *Inorg Chem* **2008,** *47* (19), 8869-8880.

179. Liu, J. X.; Long, L. S.; Huang, R. B.; Zheng, L. S., Interesting anion-inclusion behavior of cucurbit[5]uril and its lanthanide-capped molecular capsule. *Inorg Chem* **2007,** *46* (24), 10168-10173.

180. Lu Zhou, C. Z., Meng Wang, and Lu Li, Solubility of Hydroxyl Cucurbit[6]uril in Different Binary Solvents. *J. Chem. Eng. Data* **2014,** *59* (9), 2879-2884.

181. Guthrie, F., XXVIII. On the iodide of iodammonium. *J. Chem. Soc* **1863**.

182. Norris, I. R. J. F., *J Am Chem Soc* **1896,** *18*, 90-95.

183. Mulliken, R. S., Structures of Complexes Formed by Halogen Molecules with Aromatic and with Oxygenated Solvents. *J Am Chem Soc* **1950,** *72* (1), 600-608.

184. Zhong, H.; Tran, L. M.; Stang, J. L., Induced-fit docking studies of the active and inactive states of protein tyrosine kinases. *J Mol Graph Model* **2009,** *28* (4), 336-346.

185. Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A., Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J Am Chem Soc* **2008,** *130* (9), 2817-2831.

186. Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A., Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc Natl Acad Sci* **2007,** *104* (3), 808-813.

187. Schrödinger Release 2017-1. *WaterMap, Schrödinger, LLC, New York, NY, 2017* **2017**.

188. Small-Molecule Drug Discovery Suite 2017-3, Glide, Schrödinger, LLC, New York, NY,. **2017**.

189. Lagona, J.; Mukhopadhyay, P.; Chakrabarti, S.; Isaacs, L., The cucurbit[n]uril family. *Angew Chem Int Ed Engl* **2005,** *44* (31), 4844-4870.

190. Wang, R.; Yuan, L.; Macartney, D. H., Inhibition of C(2)-H/D exchange of a bis(imidazolium) dication upon complexation with cucubit[7]uril. *Chem Commun* **2006,** (27), 2908-2910.

191. W. L. Mock, a. N. Y. S., Host-guest binding capacity of cucurbituril. *J. Org. Chem* **1983,** *48* (20), 3618-3619.

192. Kim, C.; Agasti, S. S.; Zhu, Z.; Isaacs, L.; Rotello, V. M., Recognition-mediated activation of therapeutic gold nanoparticles inside living cells. *Nat Chem* **2010,** *2* (11), 962-966.

193. Reczek, J. J.; Kennedy, A. A.; Halbert, B. T.; Urbach, A. R., Multivalent recognition of peptides by modular self-assembled receptors. *J Am Chem Soc* **2009,** *131* (6), 2408-2415.

194. Li, W.; Bockus, A. T.; Vinciguerra, B.; Isaacs, L.; Urbach, A. R., Predictive recognition of native proteins by cucurbit[7]uril in a complex mixture. *Chem Commun* **2016,** *52* (55), 8537-8540.

195. Biedermann, F.; Uzunova, V. D.; Scherman, O. A.; Nau, W. M.; De Simone, A., Release of high-energy water as an essential driving force for the high-affinity binding of cucurbit[n]urils. *J Am Chem Soc* **2012,** *134* (37), 15318-15323.

196. Assaf, W. M. N. M. F. K. I., Deep Inside Cucurbiturils: Physical Properties and Volumes of their Inner Cavity Determine the Hydrophobic Driving Force for Host-Guest Complexation. *Israel J Chem* **2011,** *51* (5), 559-577.

197. Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K., The SAMPL4 host-guest blind prediction challenge: an overview. *J Comput Aided Mol Des* **2014,** *28* (4), 305-317.

ÚOCHB ᴬᵛ꜀ᵣ
IOCB PRAGUE

Ústav organické chemie a biochemie
Akademie věd České republiky, v. v. i.
Institute of Organic Chemistry and Biochemistry
of the Czech Academy of Sciences

DECLARATION OF THE SHARED AUTHORSHIP

Prohlášení

Prohlášení spoluautorů upřesňující podíl Mgr. Haresh Ajani na publikacích přiložených k disertaci:

- **Ajani, H.;** Jansa, J.; Köprülüoğlu, C.; Hobza, P.; Kryštof, V.; Lyčka, A.; Lepšík, M. Imidazo[1,2-c]pyrimidin-5(6H)-one as a Novel Core of Cyclin-Dependent Kinase 2 Inhibitors: Synthesis, Activity Measurement, Docking and Quantum Mechanical Scoring. *Journal of Molecular Recognition, (2018),* in press, DOI:10.1002/jmr.2720.

- Gucký, T.; Řezníčková, E.;Jorda, R.; Muchová, T. R.; Klejová, Z.; Malínková, V.; Berka, K.; Bazgier, V.; **Ajani, H.;** Lepšík, M.; Divoký, V.; Kryštof, V. Discovery of N2-(4-Amino-cyclohexyl)-9-cyclopentyl-N6-(4-morpholin-4-ylmethyl-phenyl)-9H-purine-2,6-diamine as a Selective and Potent FLT3 Kinase Inhibitor for FLT3-ITD Positive Acute Myeloid Leukemia. *The Journal of Medicinal Chemistry, (2018), accepted.*

- **Ajani, H.;** Pecina, A.; Eyrilmez, S. M.; Fanfrlík, J.; Haldar, S.; Řezáč, J.; Hobza, P.; Lepšík, M. Superior Performance of the SQM/COSMO Scoring Functions in Native Pose Recognition of Diverse Protein-Ligand Complexes in Cognate Docking. *ACS Omega, (2017),* 2, 4022-4029.

- Sigwalt, D.; Šekutor, M.; Cao, L.; Zavalij, P. Y.; Hostaš, J.; **Ajani, H.;** Hobza, P.; Majerski, K. M.; Glaser, R.; Isaacs, L. Unraveling the Structure–Affinity Relationship between Cucurbit[n]urils (n = 7, 8) and Cationic Diamondoids. *Journal of American Chemical Society*, *(2017),* 139, 3249-3258.

- Hylsova, M.; Carbain, B.; Fanfrlik, J.; Musilova, L.; Haldar, S.; Kopruluoglu, C.; **Ajani, H.;** Brahmkshatriya, P. S.; Jorda, R.; Krystof, V.; Hobza, P.; Echalier, A.; Paruch, K.; Lepsik, M. Explicit treatment of active-site waters enhances quantum mechanical/implicit solvent scoring: Inhibition of CDK2 by new pyrazolo[1,5-a]pyrimidines. *European Journal of Medicinal Chemistry, (2017),* 126, 1118-1128.

- Hostaš, J.; Sigwalt, D.; Šekutor, M.; **Ajani, H.;** Dubecký, M.; Řezáč, J.; Zavalij, P. Y.; Cao, L.; Wohlschlager, C.; Mlinarić-Majerski, K.; Isaacs, L.; Glaser, R.; Hobza, P. A Nexus between Theory and Experiment: Non-Empirical Quantum Mechanical Computational Methodology Applied to Cucurbit[n]uril·Guest Binding Interactions. *Chemistry- A European Journal, (2016),* 22, 17226.

- Cousido-Siah, A.; Ruiz, F. X.; Fanfrlik, J.; Gimenez-Dejoz, J.; Mitschler, A.; Kamlar, M.; Vesely, J.; **Ajani, H.;** Pares, X.; Farres, J.; Hobza, P.; Podjarny, A. D. IDD388 Polyhalogenated Derivatives as Probes for an Improved Structure-Based Selectivity of AKR1B10 Inhibitors. *ACS Chemical Biology, (2016),* 11, 2693-2705.

- Pechal, V.; Stepankova, S.; Pejchalova, M.; Kralovec, K.; Havelek, R.; Ruzickova, Z.; **Ajani, H.;** Lo, R.; Lepsik, M. Synthesis, structural characterization, docking, lipophilicity and cytotoxicity of 1-[(1R)-1-(6-fluoro-1,3-benzothiazol-2-yl)ethyl]-3-alkyl carbamates, novel acetylcholinesterase and butyrylcholinesterase pseudo-irreversible inhibitors. *Bioorganic & Medicinal Chemistry, (2016),* 24, 1560-1572.

- Vorlova, B.; Nachtigallova, D.; Jiraskova-Vanickova, J.; **Ajani, H.;** Jansa, P.; Rezac, J.; Fanfrlik, J.; Otyepka, M.; Hobza, P.; Konvalinka, J.; Lepsik, M. Malonate-based inhibitors of mammalian serine racemase: kinetic characterization and structure-based computational study. *European Journal of Medicinal Chemistry, (2015),* 89, 189-97.

- Kolář, M. H.; Deepa, P.; **Ajani, H.;** Pecina, A.; Hobza, P. The Characteristics of a σ-hole and the Nature of a Halogen Bond. *Topics in Current Chemistry, (2015),* 359, 1–26.

Mgr. Haresh Ajani je prvním autorem na většině publikací přiložených k disertaci, což jednoznačně vymezuje jeho podíl. Ve všech případech je tento podíl dominantní a to ve všech fázích přípravy publikace, od zadání tématu až k jejímu sepsání.

V Praze, 21. únor 2018

Prof. Ing. Pavel Hobza, Dr.Sc., dr. h. c., FRSC

LIST OF PUBLICATIONS

INCLUDED IN THE THESIS

1   **Ajani, H.;** Jansa, J.; Köprülüoğlu, C.; Hobza, P.; Kryštof, V.; Lyčka, A.; Lepšík, M. Imidazo[1,2-c]pyrimidin-5(6H)-one as a Novel Core of Cyclin-Dependent Kinase 2 Inhibitors: Synthesis, Activity Measurement, Docking and Quantum Mechanical Scoring. *Journal of Molecular Recognition*, in press, **DOI:10.1002/jmr.2720**

2   Gucký, T.; Řezníčková, E.;Jorda, R.; Muchová, T. R.; Klejová, Z.; Malínková, V.; Berka, K.; Bazgier, V.; **Ajani, H.;** Lepšík, M.; Divoký, V.; Kryštof, V. Discovery of N2-(4-Amino-cyclohexyl)-9-cyclopentyl-N6-(4-morpholin-4-ylmethyl-phenyl)-9H-purine-2,6-diamine as a Selective and Potent FLT3 Kinase Inhibitor for FLT3-ITD Positive Acute Myeloid Leukemia. *Journal of Medicinal Chemistry*, *accepted*.

3   **Ajani, H.;** Pecina, A.; Eyrilmez, S. M.; Fanfrlík, J.; Haldar, S.; Řezáč, J.; Hobza, P.; Lepšík, M. Superior Performance of the SQM/COSMO Scoring Functions in Native Pose Recognition of Diverse Protein-Ligand Complexes in Cognate Docking. *ACS Omega. 2017*, 2, 4022-4029

4   Sigwalt, D.; Šekutor, M.; Cao, L.; Zavalij, P. Y.; Hostaš, J.; **Ajani, H.;** Hobza, P.; Majerski, K. M.; Glaser, R.; Isaacs, L. Unraveling the Structure–Affinity Relationship between Cucurbit[n]urils (n = 7, 8) and Cationic Diamondoids. *Journal of American Chemical Society 2017,* 139, 3249-3258.

5   Hylsova, M.; Carbain, B.; Fanfrlik, J.; Musilova, L.; Haldar, S.; Kopruluoglu, C.; **Ajani, H.;** Brahmkshatriya, P. S.; Jorda, R.; Krystof, V.; Hobza, P.; Echalier, A.; Paruch, K.; Lepsik, M. Explicit treatment of active-site waters enhances quantum mechanical/implicit solvent scoring: Inhibition of CDK2 by new pyrazolo[1,5-a]pyrimidines. *European Journal of Medicinal Chemistry. 2017,* 126, 1118-1128.

6   Pejchal, V.; Stepankova, S.; Pejchalova, M.; Kralovec, K.; Havelek, R.; Ruzickova, Z.; Ajani, H.; Lo, R.; Lepsik, M. Synthesis, structural characterization, docking, lipophilicity and cytotoxicity of 1-[(1R)-1-(6-fluoro-1,3-benzothiazol-2-yl)ethyl]-3-alkyl carbamates, novel acetylcholinesterase and butyrylcholinesterase pseudo-irreversible inhibitors. *Bioorganic & Medicinal Chemistry. 2016,* 24, 1560-1572.

7 Hostaš, J.; Sigwalt, D.; Šekutor, M.; **Ajani, H.;** Dubecký, M.; Řezáč, J.; Zavalij, P. Y.; Cao, L.; Wohlschlager, C.; Mlinarić-Majerski, K.; Isaacs, L.; Glaser, R.; Hobza, P. A Nexus between Theory and Experiment: Non-Empirical Quantum Mechanical Computational Methodology Applied to Cucurbit[n]uril·Guest Binding Interactions. *Chemistry- A European Journal. 2016,* 22, 17226.

8 Cousido-Siah, A.; Ruiz, F. X.; Fanfrlik, J.; Gimenez-Dejoz, J.; Mitschler, A.; Kamlar, M.; Vesely, J.; **Ajani, H.;** Pares, X.; Farres, J.; Hobza, P.; Podjarny, A. D. IDD388 Polyhalogenated Derivatives as Probes for an Improved Structure-Based Selectivity of AKR1B10 Inhibitors. *ACS Chemical Biology. 2016,* 11, 2693-2705.

9 Vorlova, B.; Nachtigallova, D.; Jiraskova-Vanickova, J.; **Ajani, H.;** Jansa, P.; Rezac, J.; Fanfrlik, J.; Otyepka, M.; Hobza, P.; Konvalinka, J.; Lepsik, M. Malonate-based inhibitors of mammalian serine racemase: kinetic characterization and structure-based computational study. *European Journal of Medicinal Chemistry. 2015,* 89, 189-97.

NOT INCLUDED IN THE THESIS

10 Kolář, M. H.; Deepa, P.; Ajani, H.; Pecina, A.; Hobza, P. The Characteristics of a σ-hole and the Nature of a Halogen Bond. Topics in Current Chemistry. 2015, 359, 1–26

PRESENTATION OF THE RESULTS

POSTERS

- "Superior Performance of the SQM/COSMO Scoring Functions in Native Pose Recognition of Diverse Protein-Ligand Complexes in Cognate Docking." (2017) QM/MM Methods and Applications probing complex systems in biology and materials, University of Manchester, Manchester, UK.

- "Semiempirical quantum mechanical scoring function accurately recognizes ligand native poses on a diverse set of protein-ligand complexes." (2016) 21st European Symposium on Quantitative Structure-Activity Relationship. Aptuit Conference Center, Verona, Italy.

- "Active-site Water, Docking and Quantum Mechanical Scoring: Explaining Malonate Inhibitor Binding to Mammalian Serine Racemase." (2015) New Frontiers in Computer-Aided Drug Design." Gordon Research Conference. West Dover, VT, USA.

- "Drug Discovery Summit." (2014) 11th Swiss Course on Medicinal Chemistry. Leysin, Switzerland.

- "Schrödinger-14th Annual European User Meeting." (2014) Frankfurt, Germany.

- "International Scientific and Practical Conference of IOCB." (2014) Harrachov, Czech Republic.

ATTACHED PUBLICATIONS