Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

[Student's name typed] Date

Validation of Receptor-Based Drug Design and Applications in the Study of IKKs, Truncated Taxane and LRH-1

By

Haipeng Hu
Doctor of Philosophy
Chemistry

_____
Dennis C. Liotta, Ph.D.
Advisor

_____
James P. Snyder, Ph.D.
Advisor

_____
James T. Kindt, Ph.D.
Committee Member

_____
Vince Conticello, Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Validation of Receptor-Based Drug Design and Applications in the Study of IKKs, Truncated Taxane and LRH-1

By

Haipeng Hu
M.S. University of Science and Technology of China

Advisor:
Dennis C. Liotta, Ph.D.
James P. Snyder, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry
2011

Abstract

Validation of Receptor-Based Drug Design and Applications in the Study of IKKs, Truncated Taxane and LRH-1

By
Haipeng Hu

Since the first successful report of structure based drug design in the 1990's, it has been utilizing routinely in modern drug discovery. The performance of the design based on the fundamental assumption: **a)** the protein target does not undergo conformational change upon ligand binding or changes the same way independent on the ligand type; **b)** ligands which bind to the same protein share similar pharmacophore. In order to verify this assumption, a protein-ligand interaction survey with templates from a selected protein-ligand crystal structure database was performed. The result indicates that over 99% crystal structures obey the structure fundamental assumption which indicates although a few peculiar PDB pairs violate it (19 out of over 3000 cases), the structure base assumption is trustworthy in most cases. After verification of the fundamental assumption, it is utilized in several projects:

Liver receptor homolog 1 (LRH-1) is an orphan nuclear receptor (NR) which activates an array of genes responsible for development of endodermal organs. Based on the highly reserved structure properties of NRs, a series of LRH-1 antagonist candidates were developed from known NRs agonists and antagonists, yielding several compounds with mild inhibition to LRH-1 but poor solubility based on in vitro and in vivo assays. Further work is currently in process to improve their activities and ADME properties.

Two truncated taxane models were computationally generated by replacing the baccatin core with different fragments based on their structure similarity with PTX, and both of them lead to micro-molar level activities.

The inhibitor of Kappa B Kinases (IKKs) which bind with the rel homology domain of NFκB regulate the activation of NFκB by the phosphorylation-induced ubiquitination of the IκB proteins. Homology models were constructed and investigated in silico. The result provides a first-step to understand the mechanism of IKKs inhibition and offers a provisional guidance on the design and synthesis of novel IKKs inhibitors.

In the end, a quantum chemical calculation was performed to prove that the cuprate intermediate formed between nucleophilic attach of the dimethylcopper anion to the β-unsaturated carbon has a square-planar structure.

Validation of Receptor-Based Drug Design and Applications in the Study of IKKs, Truncated Taxane and LRH-1

By

Haipeng Hu
M.S. University of Science and Technology of China

Advisor:
Dennis C. Liotta, Ph.D.
James P. Snyder, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry
2011

Acknowledgements:

This thesis is the result of not only my efforts but also those of others who have supported me in different ways all along my Ph.D. program.

I would like to gratefully and sincerely thank Drs. James P Snyder and Dennis C Liotta for their guidance, understanding, patience, and most importantly, their friendship during my graduate studies at Emory University. Their mentorship was paramount in providing a well rounded experience consistent my long-term career goals. For everything you've done for me, Drs. Snyder and Liotta, I thank you.

I would also like to thank my Ph.D. committee members Drs. Conticello and Kindt for their valuable discussions and encouragement.

Additionally, I am very grateful for the friendship of all of the members of Liotta/Snyder research group, especially Drs. Ana Alcaraz, Ashutosh Jogalekar, Andrew Prussia, Matthew Geballe, Pahk Thepchatri, Tiao Xie, and Serdar Kurtkaya who preceded me and gave me so much mentoring. For those of whom behind me in the lab: Mi-Sun Kim, Jin Liu and Qi Shi, I have enjoyed working with all of you. A special thank for Dr. Pieter Burger for his encouragement and helps during the thesis writing process.

Finally, and most importantly, I would like to thank my wife Guannan. Her support, encouragement, quiet patience and unwavering love were undeniably the bedrock upon which the past six years of my life have been built. I thank my parents, Shudong and Baohua, for their faith in me and allowing me to be as ambitious as I wanted. It was under their watchful eye that I gained so much drive and an ability to tackle challenges head on.

Table of Contents

List of Figures

List of Tables

Chapter 1： Bird View of Structure Based Drug Design

1.1 Introduction

Structure-based drug design relies on knowledge of the three dimensional structure of the biological target obtained through various methods such as x-ray crystallography, NMR spectroscopy or homology modeling. With the 3D-structure of the biological target, candidate drugs that are predicted to bind with high affinity and selectivity to the target are designed using various automated computational procedures. In the early 1990s, the first successful stories utilizing structure based drug design in modern drug discovery were reported [1-3]. The X-ray crystal structure of HIV protease was determined in 1989, and this crystal structure was fed into a computer modeling program as a reference to determine the types of molecules that might block its function. These molecules can be retrieved from chemical libraries or can be designed on a computer screen and then synthesized in a laboratory. Since that time, the ability to rationally design drugs utilizing protein 3D-structures has become an achievable goal for many structural biologists [4, 5]. Such structure-based drug design strategies have the potential to reduce years and millions of dollars from the traditional trial-and-error drug development process as Dale Kempf, a chemist involved in the HIV protease inhibitor program at Abbott Laboratories, says "From the identification of HIV protease as a drug target in 1988 to early 1996, it took less than 8 years to have three drugs on the market."

Structure based drug design is a multi-step process. The first step includes the

selection of the target and the structure determination of the target molecules. Then, a series of compounds from compound database or from synthesis are positioned into a selected pocket on the target molecule and scored based on their steric and non-covalent interactions with the target molecule. The top score compounds are recommended to be investigated in the bioassay. Those compounds with micro molar or better activities can be used as lead compound in the consequent step. In the second step, the lead compound is optimized to increase its potency and selectivity. This step contains an iterative cycle including synthesis of the optimized lead, structure determination of the new target-lead complex, and further optimization of the lead compound. After several cycles of the drug design process, the optimized compounds usually show reliable potency and selectivity.

1.2 Target Selection

Target selection is primarily made by biological and biochemical basis. The ideal target macromolecule for structure-based drug design is the one which is closely related to human disease and can carry out a function after binding to a small molecule. The goal in developing drugs against the targets is often to modulate the function of human protein or inhibit the pathogenic organisms upon small compound binding. Under this circumstance, the target should be essential and unique. That is to say the target should be part of a crucial cycle in the cell, and its elimination should lead to the pathogen to death. No other pathway should be able to supplement the function of the target and overcome the malfunction of the target. Of course, the target should be able to be inhibited by

binding to a small molecule. Those macromolecules which contain a well defined ligand-binding pocket are often good drug targets, such as enzymes.

In order to utilize the structure based drug design, an accurate 3D structure of the target is essential. Currently, there are three major ways to obtain the 3D structure of a macromolecule: X-ray crystallography, NMR and homology modeling. Protein Data Bank [6] is the largest macromolecule structure databases, and it is the most important source to obtain 3D structure of the target. Until now, this database contains over 68,000 structures obtained from X-ray crystallography, NMR and electron microscopy. Typically, an X-ray crystal structure with 2.5Å or better resolution, 28% or lower R_{free} and R value is acceptable to be used in structure based drug design. For NMR structure, the one with small number of restraints per residue and less nuclear overhauster effect (NOE) violations is better for drug design.

For those targets without experimentally determined structure, a rational homology model is a good choice for drug design [7-8]. Typical homology modeling routinely needs three items of input, 1) the sequence of the target; 2) a 3D template which has the highest sequence identity or similarity with the target sequence. The template structure must be determined by experimental methods such as crystallography or NMR with reliable quality; 3) a sound sequence alignment between the target and template. The homology modeling will proceed to arrange the backbone of the target sequence according to that of the template, using the sequence alignment to decide where to position each residue. Therefore, the quality of the sequence alignment is of crucial

importance. Many drug design software packages provide homology modeling tools, such as Prime [9], Modeller [10] and Swiss-Model [11]. All of them construct 3D-model of the target based on a reasonable sequence alignment. In general, a good homology model 1) should have a reliable structure comparing with other reported crystal structures; 2) can explain the experimental data related to the target; 3) can be utilized in prediction of further experimental. The evaluation of the homology model should base not only on the statistic of reported crystal structures (the reliability of the structure), but also the reported experimental data related to the target (the reliability of the interactions prediction with small molecules). The programs PROCHECK [12] and WHAT IF [13] provide structure-based details for evaluating a homology model based on the statistic data obtained from reported crystal structures.

The aim of structure based drug design is to search for a small molecule that can alter target function upon binding. After the 3D structure of the target is determined, finding a proper ligand binding site on the target is the next goal. Ideally, target binding site is a pocket with a variety of potential hydrogen bond donors and acceptors, hydrophobic characteristics, and size of molecular surface. The ligand binding site can be the active site as in an enzyme, an assembly site with another macromolecule, or a communication site necessary in the mechanism of the molecule. Actually, many resolved crystal structures contain an initial small molecule which is co-crystallized with the target macromolecule. The binding site of the initial molecule can be invaluable for the determination of a good target binding site.

1.3 Drug Design

After 3D structure of the target molecule is obtained and the small molecule binding site is located, there are at least two paths to develop a good lead compound based on the structure of the target, the experimental path and computer-aided method. An example of the experimental methods is high-throughtput screening (HTS) [14-15]. The compound library used in the HTS can be a large compound database such as ACD, or small libraries made up from single compounds.

In the computer-aided path, there are three methods to develop a good lead compound based on the structure of the target: inspection, virtual screening, and de novo generation. In inspection, known molecules that bind into the site, such as substrates or cofactors in the cases of enzymes, or peptides in the case of protein-protein interactions, are modified to become inhibitors based on maximizing complementary interactions in the target binding site. In virtual screening, structures from a small molecules database are fit into the identified target binding site with proper docking programs. The energies of the resulting complexes are evaluated and those with high binding energies can be experimentally tested as possible lead compound. For de novo generation, the interactions of several chemical groups with the target are investigated, and these fragments can then be assembled into one compound with linking groups. The final compounds, created in silico from the linked fragments, are worth to be synthesized and bio-evaluated in the laboratory.

Once a small molecule potentially binding to the target molecule is identified, it

must be evaluated further before proceeding to the next stages. It is important to consider that the ranking assigned by the scoring function is not always trustable, since the model obtained from docking programs is just estimation. The scoring function is only verified by a subset of the reported structure database, and different scoring function may provide different ranking for the same molecules. In order to conquer this problem, several molecules with good score, not the top score compound only, during the docking procedure are introduced to the next step. They will be evaluated and optimized with computer graphic first and then recommended to synthetic chemists and biologists. The experimental feedback then will be utilized to verify previous optimization and lead to further optimization. The activity and selectivity of the lead is improved in this iterative cycle.

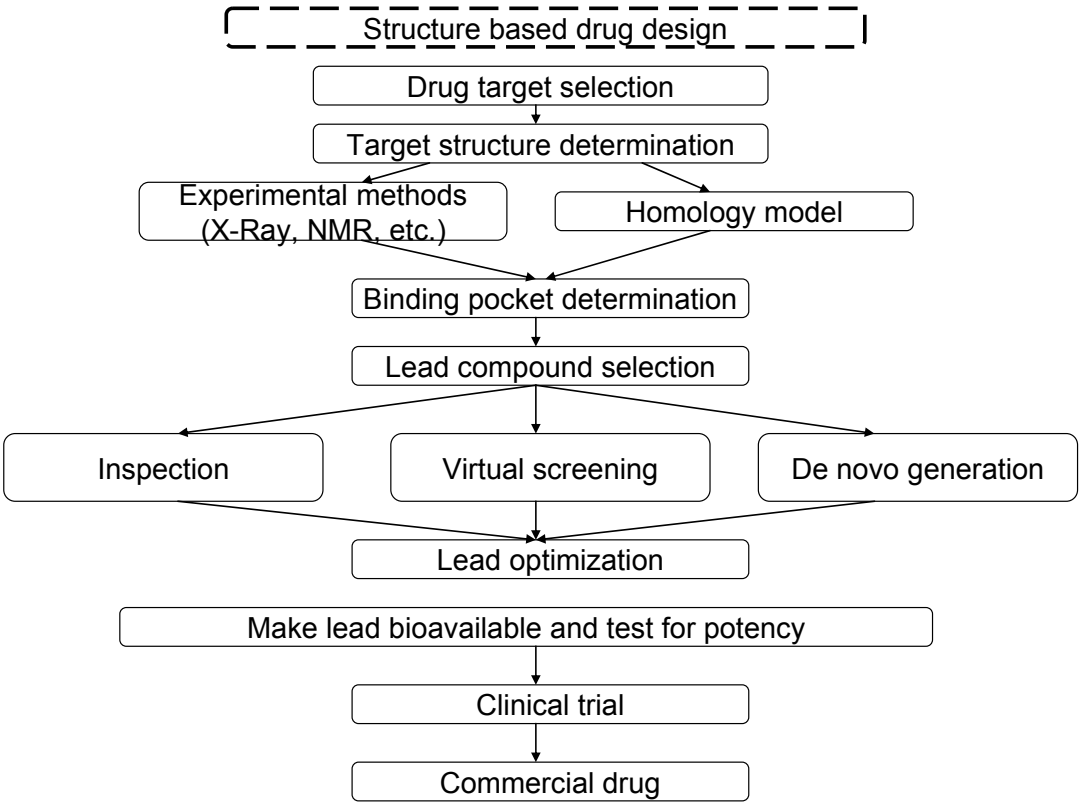


Figure 1.1 Process of structure based drug design.

1.4 Outline of Subsequent Chapters

Structure based drug design is an integral part of most industrial drug discovery program and it is the major subject of research for many academic laboratories. The fundamental assumption of structure based drug design is introduced and its rationality is investigated by a crystal structure database survey in Chapter 2. Chapter 3 and 4 describe the applications of the structure based drug design in two specific cases, the liver receptor homolog-1 (LRH-1) antagonist design and truncated taxane design for stabilizing microtubule. In Chapter 5, a case based on homology modeling is introduced, and in Chapter 6, a quantum calculation related project is described.

Chapter 2: Protein Ligand Interaction Survey

2.1 Interactions between Protein and Ligand

Most drugs have been discovered either by identifying the active ingredient from traditional remedies or by serendipitous discovery before Paul Ehrlich proposed his lock and key theory nearly a century ago. This theory has served as one of the cornerstones of rational drug development. After 1980's, more and more protein crystal structures were established and were utilized to analyze the 'fingerprint' of the drug target. Disrupting or enhancing the interaction between a key compound and its substrate based on the 3D crystal structure of the target become to the basis for numerous drugs targeting signal transduction processes [1-4]. Crystal structures of protein complexes with small compound provide insight to what kind of small molecules can bind as well as the mode. Under this circumstance, more and more attentions are paid to investigating the interactions between proteins and their key molecules, especially non-covalent interactions, so that the fingerprint of the target can be utilized in drug discovery.

Non-covalent interactions are considered the key thermodynamic factors that drive protein-ligand binding. They are bonding forces with strengths of 1-7Kcal/mol, and are much weaker than covalent interactions, which is over 50Kcal/mol. However, such weak interactions are the major driving force that leads to protein and ligand binding and

contributes most to the stability of the complex. Generally, the non-covalent bond interactions between protein and small molecules can be divided into five major categories:

Hydrophobic interactions exist widely in proteins and play an important role in both protein folding and protein ligand binding. Hydrophobic interactions are not the direct interaction between atoms, but due to the characters of non-polar groups dissolved in water. When ligand and protein bind together, the highly ordered solvent shell around the dissolved ligand and protein respectively are destroyed and part of the shell waters are extruded. The entropy of the system is increased due to the released water molecules. This increase in entropy is the major driving force of two non-polar groups binding with each other in aqueous solution.

Hydrogen bonds are formed when a hydrogen atom covalently bonded to an electronegative atom is shared with another electronegative atom. The atom to which the hydrogen atom is covalently bonded is known as the hydrogen bond donor. The atom with which the hydrogen is shared is the hydrogen bond acceptor. Protein consists of a large number of hydrogen bond donors and acceptors which help to form functional protein structure. The hydrogen bond donors and acceptors within the binding pocket help to form the frame of the protein ligand complex. However, hydrogen bond interactions may not contribute to binding energy as much as we expect. Because water is also a good hydrogen bond contributor, the hydrogen bond interaction in the complex can be considered as a substitute of the solvent molecules.

Polar interactions (salt bridge) are essentially columbic forces, and their strength depends on both the dielectric constant of medium and the distance between the two polar groups. The effect of long-range columbic forces in a solution with a low dielectric constant is generally weak. However, intermediate values of the dielectric constant, could lead to a strong electrostatic interactions and other polar interactions that may exist at membrane interfaces or in the binding clefts of proteins.

Van der Waals forces exist both in non-polar group and in polar groups of the protein ligand complex. Their potential energy obeys the Lennard-Jones formula. They have an attractive term which is due to favorable interactions among the induce instantaneous dipole moments that arise from fluctuations in the electron charge densities of neighboring non bonded atoms and a repulsive term which is caused by electron-electron repulsion. These two forces are balanced in atom's Van der waals (VDW) radii. The repulsive term increases rapidly with the decrease of distance between two atoms, and the attractive term becomes primary when the distance between two atoms are longer than their VDW radii,

Halogen bond interactions are attributed to the transfer of negative charge from an oxygen, nitrogen, or sulfur (a Lewis base) to a polarized halogen (a Lewis acid). They are referred to as halogen bonds by analogy to classical hydrogen bonds with which they share numerous properties. Extensive surveys of structures show that the interaction is primarily electrostatic, with contributions from polarization, dispersion, and charge transfer. The stabilizing potential of halogen bonds is estimated to range from half to

slightly greater than that of an average hydrogen bond.

2.2 Structure Base Assumption

A central theme in drug discovery is the search for lead candidates by investigating the interactions between drug candidates and the target protein. The most promising candidates will be optimized to improve their binding and ADMET properties. The rationale of this approach is that molecules which induce the same biological effect should share both 3D structure and non-covalent interactions with target protein. It also can be described as the structure base assumption of drug discovery: ***a)*** *the protein target does not undergo conformational change upon ligand binding or changes the same way independent on the ligand type;* ***b)*** *ligands which bind to the same protein share similar pharmacophore (display similar interactions and binding pose).* Initially the motivation for the structure base assumption in drug discovery was virtual screening, which first appeared at 1990's but has been around for a long time. Now, this assumption is embedded into the drug discovery paradigm and applied throughout the drug design process including pharmacophore model generation/HTS procedure, lead-like compounds modification and drug metabolism.

2.3 True or False?

The structure base assumption has been widely used in the drug discovery process for decades, yet seldom has the validity of this assumption been tested. In 2003, Sung, B., *et al* indicated in their research that three phosphodiesterase (PDE)-5 inhibitors, sildenafil, vardenafil and tadalafil, bind into the same PDE-5 inhibitor binding pocket with various scaffolds [5]. Sildenafil and vardenafil share similar structures with different nitrogen atoms position on the bicyclic ring and the hydrophobic group on the para- position of the distal piperazine. These differences don't alter the binding pose of these two compounds which share almost identical binding pose with each other. Both these compounds reserve the hydrogen bond interactions with the Glu on PDE-5, the phenyl ring and the bicyclic core locates in the same position (**Figure 2.1** a). Due to the similarity of structures and binding poses, they should present similar effect on PDE-5, and the experimental data verified this. However, Tadalafil, another PDE-5 inhibitor gives a different scaffold and binding orientation from sildenafil and vardenafil (**Figure 2.1** b). The bicyclic ring no longer exists in this compound. Instead a 4-membered ring structure occupies the corresponding position. Experimentally, tadalafil shows similar activity as sildenafil although it differs in structure and binding pose. That is to say, there are at least two different binding models for the PDE-5 inhibitor binding site with similar function. This fact violates the second term of the structure base assumption. Under these circumstances, several questions need to be reconsidered: does the structure-base fundamental

assumption always stand? If not, in which condition will it fail? In order to answer these questions and validate this assumption, a protein-ligand interaction survey was performed based on a selected protein-ligand crystal structure database.

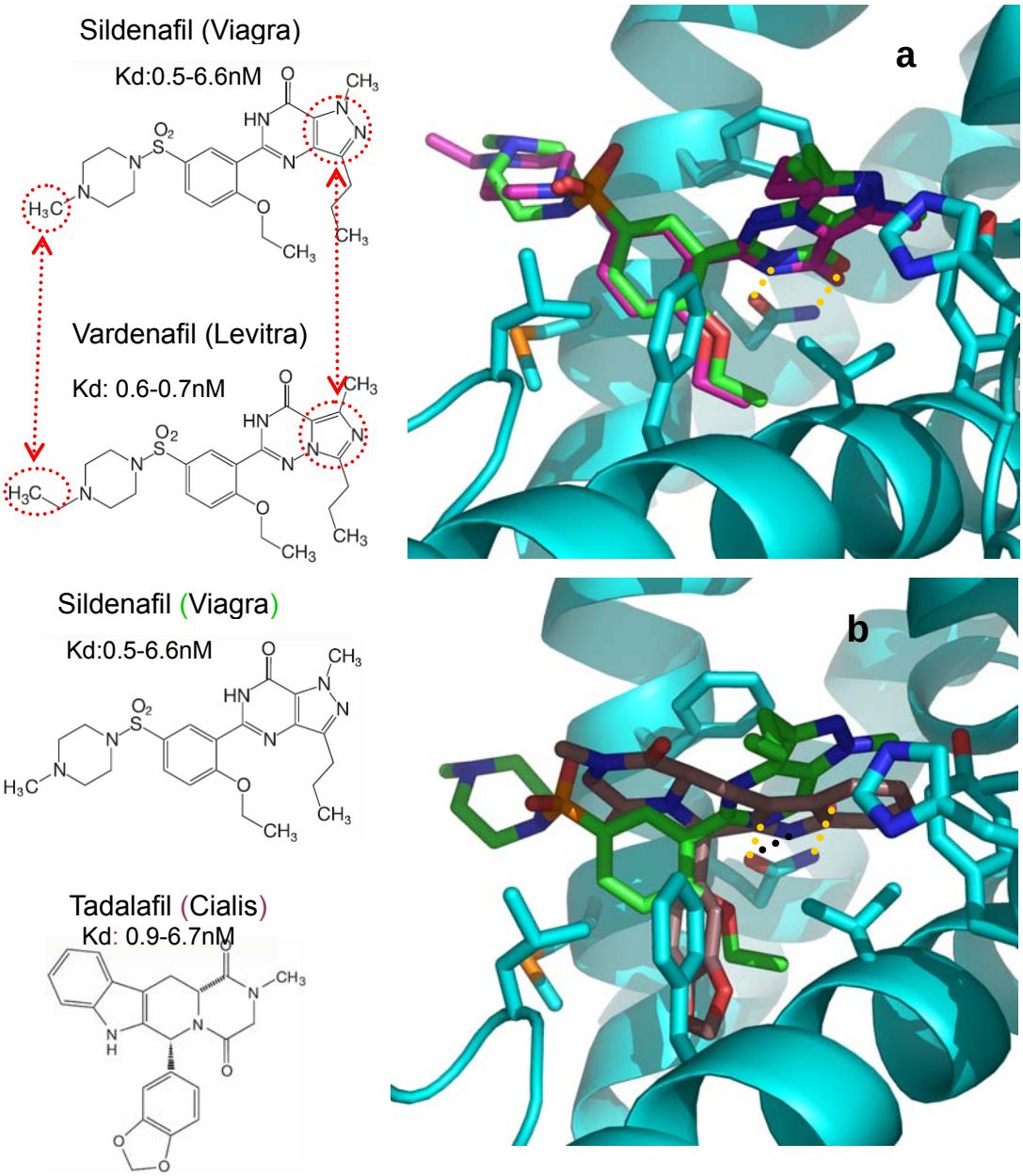


Figure 2.1 Structure and binding mode comparison among vardenafil, sidenafil and tadalafil.

Sildenafil is shown in green sticks, Vardenafil is shown in magenta sticks in a and Tadalafil is shown in pink sticks in b.

2.4 Experiment Detail

PDB-At-A-Glance [6] is a hypertext-based classification of the current release of the Protein Data Bank (PDB). It consists of a set of pre-defined bio-chemically meaningful search categories corresponding to keywords that represent the entire span of the database. Thus, PDB structures which contain the identifiers “complex”, “inhibitor” or “activator” in the “compound” or “header” fields are marked as complexes. In this survey, each crystal structures marked as complex in the database was utilized as a template to perform a similar binding pocket search in PDB. Those complexes with similar binding pocket as the template were investigated together by using several protein structure analysis tools. The major tool used to compare differences between complexes is Relibase+ [7], which is a web-accessible protein database containing all released crystal structures from PDB coupled with crystal structure analysis tools that are widely used to search and compare protein-ligand interactions across a range of complexes. Thus, it provides unique and easy access to crystal structures within PDB. The ‘similar binding site search tool’ assembled in Relibase+ was employed to examine similarities between the template structure and the complexes in PDB database. Complexes with similar binding sites can be superimposed and visualized against a user-selected template. The

RMSD of both the backbone and side chain as well as mutations between the template and query structure can also be obtained.

In the current study, the compared region is set as the residues within 10 Å around the ligand, and the criteria of the sequences similarity between the query and template protein to 95% covering genetic variation. With these setting, the Relibase+ application displays all the crystal structures with similar binding site region. In most but not all cases, the obtained complex structures contain similar ligand and share similar binding pose as the template structure. Those complexes with different protein-ligand binding properties in the same protein were identified for further investigation. After obtaining these proteins and their corresponding complexes crystal structures, they were subjected to analysis the protein-ligand interactions difference between them. Ligplot [8] was used to generate schematic 2D diagrams of protein-ligand interactions from the 3D coordinates. This program not only illustrates polar interactions between protein and ligand, but also displays non-polar protein moieties around the ligand, both of which affect binding affinity and selectivity. Masetro [9] and Pymol [10] provided the graphics to view 3D environments around ligands including hydrogen bonds as well as polar and nonpolar sub-sites. Cross dockings and scoring was derived using Glide [11] and Prime/MM-GBSA [12].

2.5 Classification

The PDB-At-A-Glance database search resulted in 19 pairs of complex structures when making use of over 3000 structures with title 'complex' as searching template. These cases represented complexes with two or more small compounds with different pharmacophores in the same binding pocket. For these 19 cases, the structure-based assumption is no longer held. In order to investigate the conditions under which the assumption is not true, a systematical analysis was performed. The case analysis indicated that there are four major reasons which result in the invalidation of the structure base assumption: (***i***) *pose difference*, ligands locate in the same pocket with different binding pose. They share a few or no common interaction features with protein; (***ii***) *protein movement*, due to the flexibility of the protein, its side-chain or backbone conformations are altered upon ligand binding; (***iii***) *solvent effect,* solvent molecules work as hydrogen bond bridges in protein ligand interactions; and (***iv***) *metal effect*, metal ion acts as important role in protein ligand interactions.

(*i*) Pose Difference:

a. Different binding poses, with shared common ligand-protein interaction features.

b. Different binding poses, with no shared ligand-protein interaction features.

(*ii*) Protein Movement

a. Similar binding pose, but different sizes of substitutes on the ligands induce protein movement due to steric conflict.

b. Difference binding pose of ligands induce protein movement due to steric conflict.

c. Difference binding pose of ligands induce protein movement due to H-bond interactions.

(*iii*) Solvent Effect

a. Water-mediated H-bonds with protein.

b. Water-mediated interaction with metal ions in the binding site.

(*iv*) Metal Effect

One interacts with the metal ion in the binding site, the other doesn't contain metal ion.

Table 2.1 Classification Table

Categories	Sub-categories	cases
Pose Difference	Share common interactions	(1CR6, 1EK2); (1ELA, 1INC); (1EJN, 1SQT); (1RWQ, 1X70); (1AZ8, 1MAX) (1B8Y, 1USN)
	No shared interaction	(5EST, 1E38); (1CVZ, 1BQI)
Protein movement	Similar pose with different size	(1SRF, 1VWJj); (1A4H, 2BRC); (1EAU, 1JIM)
	Different pose with Steric conflict	(1EJN, 1SQT); (1E1X, 2B53); (1DD6,1JJT); (1AZ8, 1MAX); (1B8Y, 1USN)
	Different pose with bond interactions	(1ELA, 1INC); (1A4H, 2BRC); (5EST, 1E38); (1AZ8, 1MAX)
Solvent effect	Water mediated H-bonds	(1IDB, 1IVQ);(2WEB, 2WEC);(1ELA, 1INC);(1E1X, 2B53); (1EZQ,1F0S); (5EST,1E38); (1EGH, 1IK4)
	Water-mediate with Metail ion	(1DMW, 4PAH); (1RWQ, 1X70)
Metal effect		(1C1U, 1KTT)

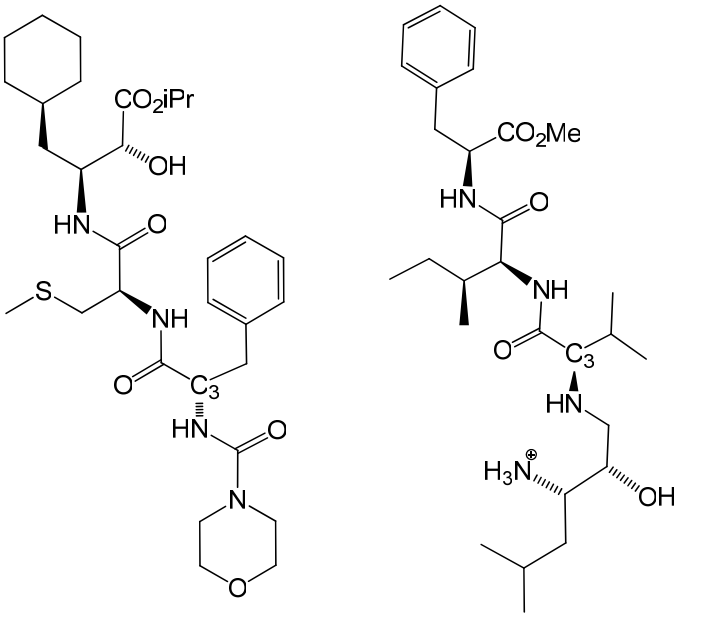
2.6 Case Discussions

2.6.1 Pose Difference: a. 1EPN & 5ER1 [13, 14]

Aspartic proteinases are an important enzyme family associated with several pathological conditions including hypertension, gastric ulcers, neoplastic disease and AIDS. Studies of inhibitor binding are therefore of great importance for design of novel blockers for potential therapeutic applications. Endothiapepsin is mainly composed of beta-sheets with the enzyme pocket located above the linkage between its C- and N-terminal domains at the back of the hinge loop. The structures of two inhibitors, XFCX-CH3 and LOL-CH2-VIFX, in complex with endothiapepsin have been solved by X-ray crystallography at 1.60 Å and 2.00 Å resolution (PDB codes: 1EPN and 5ER1), respectively. A comparison between the 2-D structures of the small compounds indicates both of them are short peptide like compounds with one 6-membered ring side chain on the top; the backbone of the next two residues are also similar to each other except the chirality of C3 labeled in **Figure 2.2**. Due to their structure similarity, it is expected that the pose difference is mainly caused by the chirality difference of C3. However, when superimposing these two complexes, the common interactions shared by them differ from the prediction based on their 2-D structure (see **Figure 2.3**). The cyclohexane moiety in XFCX-CH3 corresponding to the distal isopropyl in LOL-CH2-VIFX is found in a small

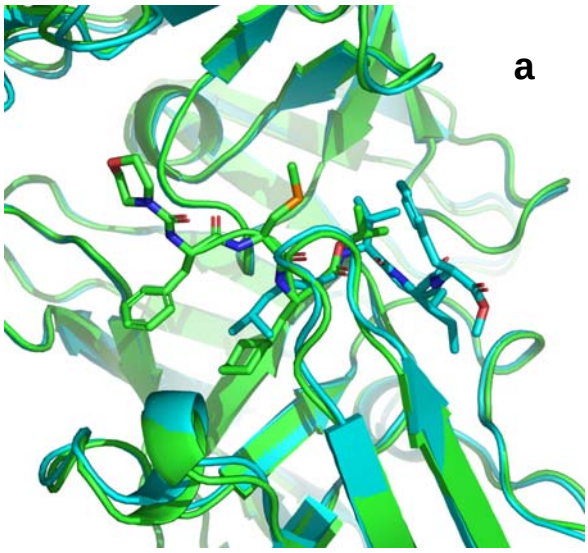
hydrophobic pocket surrounded by Phe111, Tyr75 and Leu120. A hydrogen bond between Asp215 and the ligand's hydroxyl oxygen is found in both complexes as is an isopropyl group surrounded by Phe189, Ile213 and Ile299.

Apart from these common interactions, the remaining fragments of the two ligands occupy different regions of the binding pocket. In 1EPN, the phenyl ring which parallels to the cyclohexane and morpholine ring on XFCX-CH3 are near the protein's C- terminal domain. Two amides adjacent to the methylthio group engage in H-bonds with Thr219, Asp77 and Gly76. A hydrogen bond is also formed between the amide oxygen and the backbone of Gly76. In LOL-CH2-VIFX, the two amide groups participate in hydrogen bond interactions with Gly76, Gly34 and Ser74 on the other side of the central region. For this compound, the reported binding affinity is 960 nM, but the binding affinity of XFCX-CH3 is unknown. Why do the structurally similar ligands demonstrate such different poses in the crystallographic complexes? Glide docking of LOL-CH2-VIFX into the binding pocket in a pose similar to that adopted by XFCX-CH3 results in a 15 Kcal/mol energy penalty. Its binding pose indicates neither Phe-Ile motif nor the two *i-Prs* would engage in the common interaction in the central region. Cross docking of XFCX-CH3 into the LOL-CH2-VIFX space indicates it would lose most of its hydrogen bonds with the receptor. In addition, due to the size of the pocket, its cyclohexane ring would be exposed to solvent area. This case is a good example to illustrate how ligands share some key interactions but adopt quite different poses.



XFCX-CH_3 (1EPN) LOL-CH_2-VIFX (5ER1)

Figure 2.2 2D structures of ligands in 1EPN(left) and 5ER1(right).



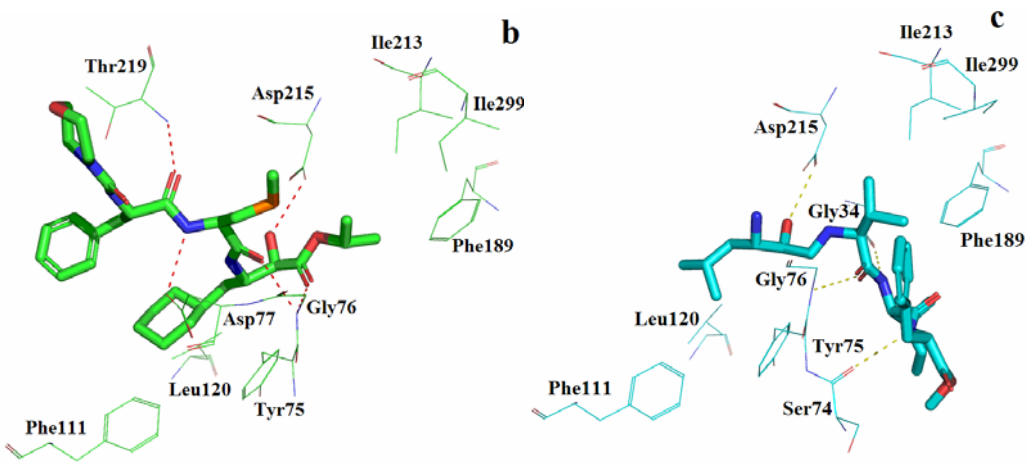
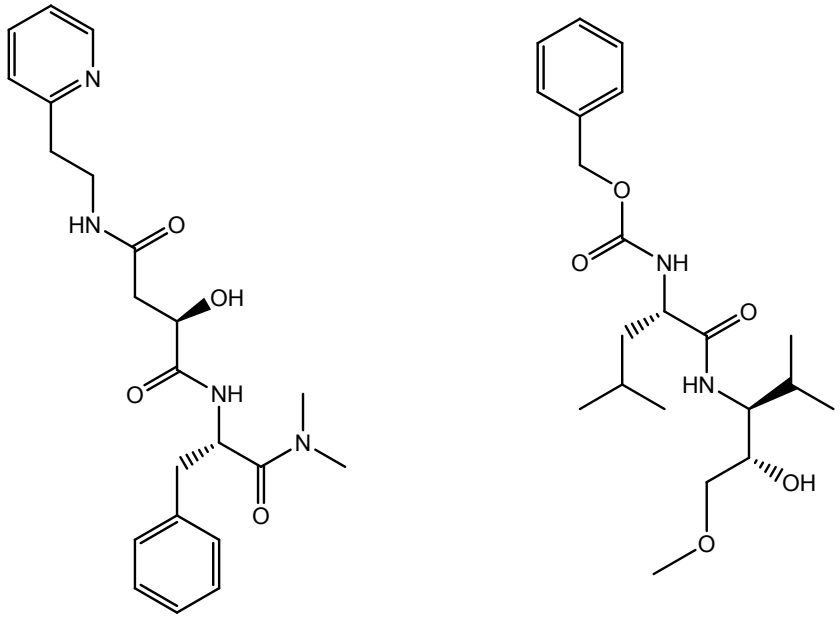


Figure 2.3 a) Superimposed crystal structures of 1EPN and 5ER1. Two ligands, XFCX-CH_3 (green stick) and LOL-CH_2-VIFX (cyan stick) locate in different region of the same binding pocket. b) Ligand XFCX-CH_3 in endothiapepsin ligand binding site. c) Ligand LOL-CH_2-VIFX in endothiapepsin ligand binding site.

2.6.2 Pose Difference: b, 1CVZ & 1BQI [15, 16]

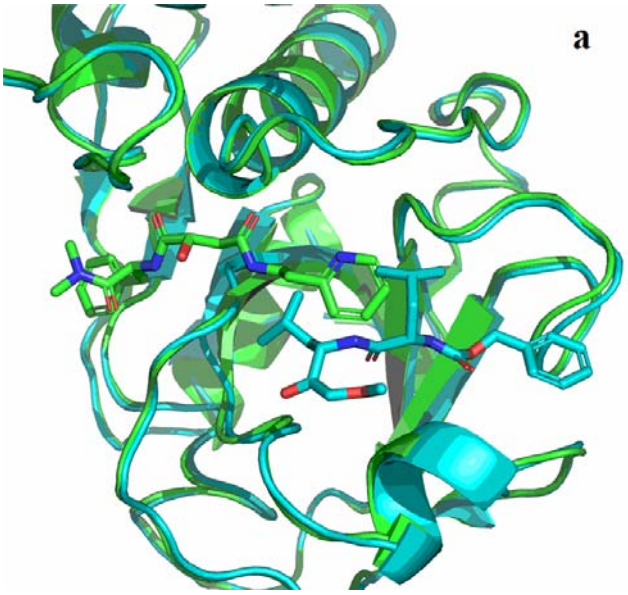
Papain has been used as a surrogate enzyme in a drug design effort to obtain potent and selective inhibitors of cathepsin K and L, new members of the papain super family of cysteine proteases that are selectively and highly expressed in osteoclasts and are implicated in bone resorption. The complex crystal structure of cathepsin L and compound C48 is reported as PDB entry 1CVZ at 1.70 Å resolution, and the complex of cathepsin K and compound SBA is reported as 1BQI at 2.50 Å resolution (ligands structure show in **Figure 2.4**). The primary structures of Cathepsin K and L are identical. The overall RMSD of these two crystal structures is less than 0.5 Å and the Cα

movement is less than 1.3 Å which indicates there is no significant difference between these two enzyme structures. The binding pocket of cathepsin locates between two domains, domain 1 from residue 1 to 110 and domain 2 from 111 to 212. In 1CVZ, the linkage between two aromatic rings of compound C48 locate on the bottom of a narrow cleft composed by His159, Cys25 and Gly23, and a covalent bond is formed between Cys25 and the C2 atom of C48 leading to 0.9 Å side-chain movement. The two amide groups on the linkage form direct hydrogen bonds with Gly66, Asp158, Gln19, Cys25 and one water-mediate hydrogen bond with Asp158. A π-π interaction is found between the distal pyridine ring and Trp177. The phenyl ring on the other end locates in the sub-site composed by Tyr67, Val157, Val133 and Ser205. For comparison, the two complexes structures were superimposed with each other shown in **Figure 2.5**. Due to steric conflict with the *i-pr* group on the linkage, compound SBA in 1BQI cannot locate in the same cleft as C48 without the covalent bond interaction. The linkage of compound SBA lies on the top of Trp177. Its phenyl ring interacts with the aromatic side chain of Trp181. This compound is further stabilized by a favorable electrostatic interaction with the carbonyl side chain of Gln142. Cross docking indicates if compound C48 located in the pocket where SBA is found, although it retains hydrogen bond interactions near C25, its dimethyl amide would explore into solvent area, and the phenyl group would be close to Gln142 which has electrostatic interaction with SBA. Based on the discussion above, although the two compound structures are very similar, they locate in different region of the binding site sharing no common interaction with papain.



C48 Ligand from 1CVZ SBA Ligand from 1BQI

Figure 2.4 2D structures of ligands in 1CVZ (left) and 1BQI (right).



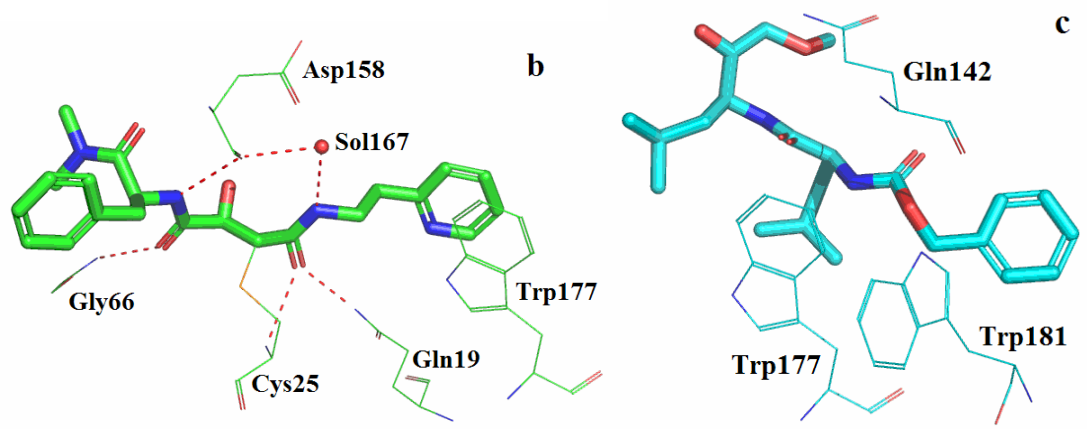


Figure 2.5. a) Superimposed crystal structures of 1CVZ and 1BQI. Two ligands, XFCX-CH_3 (green stick) and LOL-CH_2-VIFX (cyan stick) locate in different region of the same binding pocket. b) Ligand XFCX-CH_3 in cathepsin L ligand binding site. c) Ligand LOL-CH_2-VIFX cathepsin K ligand binding site.

2.6.3 Protein Movement: a, 1SRF & 1VWJ [17, 18]

Streptavidin is a tetrameric protein which binds tightly to small molecule biotin. The binding constant of this complex is extremely high and has made the streptavidin/biotin system the focus of a number of studies aimed at determining what particular intermolecular interactions give rise to such tight binding. Two reported X-ray crystal structures of streptavidin complexes with small drug-like compound MTB and short cyclic peptide ligand HPQ in the biotin binding site were reported at 2.00 Å and 1.45 Å resolutions as 1SRF and 1VWJ. In these two complexes, the loop composed by residue 43 to 55 is flexible and their conformation can be significantly changed upon ligand

binding.

In complex 1SRF, compound MTB is sandwiched by a series of beta-sheet and the loop from residue 43 to 55. The carboxyl group on MTB acts as hydrogen bond acceptor to interact with Asn23 and Ser27. The *tert*-butylphenyl is parallel to and stacks with the indole ring on trp79. The second phenyl ring on MTB is found in a small hydrophobic sub-site composed by Trp108, Thr90 and Trp92. In order to compare the interactions between these two ligands and streptavidin, complex 1VWJ was superimposed with 1SRF. Different from MTB, the ligand in 1VWJ is a cyclic peptide with sequence CHPQGPPK. The short motif HPQ occupies similar space as MTB in 1SRF: its imidazole ring of His3 and the backbone of Pro4 are found at the same position as the *tert*-butylphenyl on MTB. Backbone of His3 and Gln5 also form directly and solvent-bridged hydrogen bonds with Ser88, Tyr43 and Ser27. The side chain of Gln5 is superimposed with the second phenyl ring on MTB, and forms hydrogen bond with Thr90. The remaining residues of the ligand extend out of the binding pocket and alter the conformation of the flexible loop. **Figure 2.7** shows the loop conformational change. Because of size difference, the steric hindrance is quite remarkable in 1VWJ, the loop composed by residue 43 to 55 presents over 8 Å average side chain and Cα movements. In this region, the pyrrolidine rings of PRO7 and PRO8 on the cyclic peptide ligand hydrophobic interact with the side chain of Leu25 and Val47. There are also a few inner hydrogen bond and hydrophobic interactions between different parts of the cyclic peptide. Further docking study indicates the binding pocket of complex 1SRF is not big enough to

be occupied by the cyclic peptide ligand in 1VWJ. The significant loop conformational change is induced by ligand-receptor binding. This case represents an example that the conformation of the protein can be influenced by ligand size. Due to the movement of the protein, the shape of the binding pocket can be significantly changed.

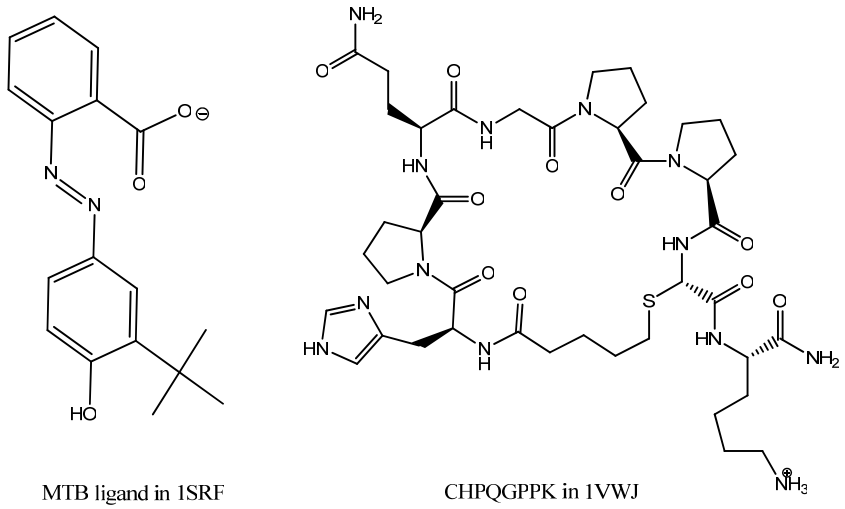
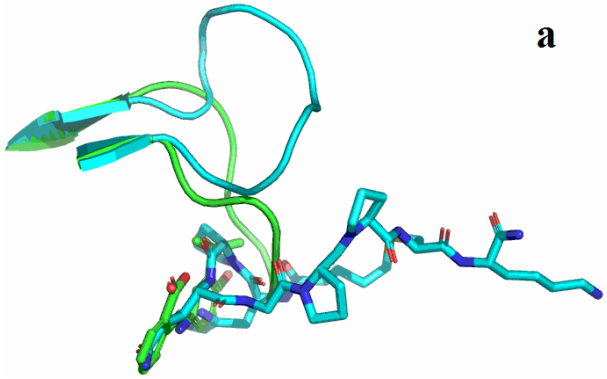


Figure 2.6 2D structures of MTB (left) and short peptide CHPQGPPK (right).



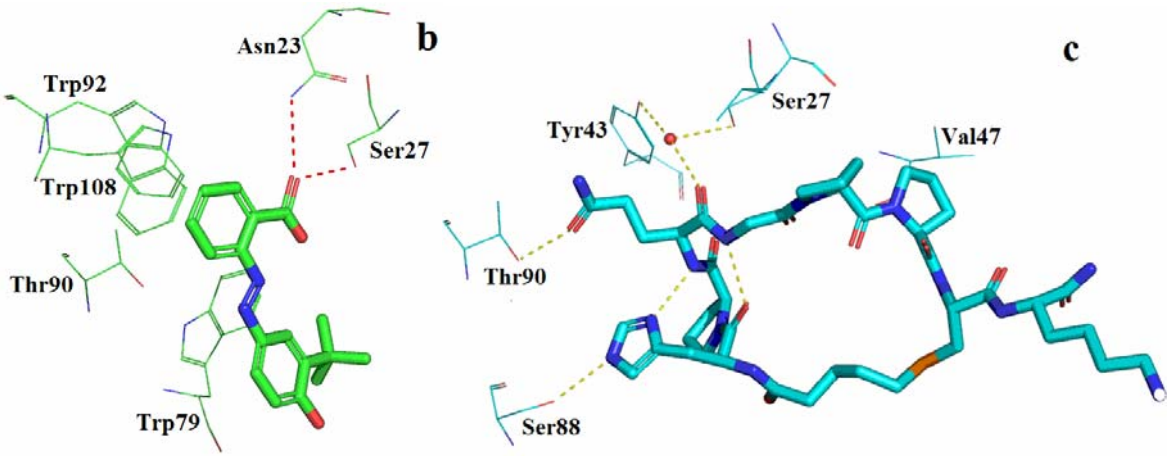


Figure 2.7, a) Superimposed crystal structures of 1SRF and 1VWJ. Two ligands, MTB (green stick) and CHPQGPPK (cyan stick) locate in the same binding pocket, but due to the size difference between these two ligands, a loop conformational change is observed. b) Ligand MTB in Streptavidin ligand binding site. c) Short peptide CHPQGPPK in Streptavidin ligand binding site.

2.6.4 Protein Movement: b, 1DD6 & 1JJT [19, 20]

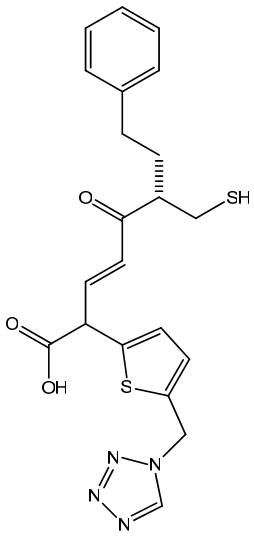
IMP-1 metallo-beta lactamase is a plasmid-borne zinc metalloenzyme that efficiently hydrolyze beta-lactam antibiotics, including carbapenems, rendering them ineffective. Because IMP-1 has been found in several clinically important carbapenem-resistant pathogens, there is a need for inhibitors of this enzyme that could protect broad spectrum antibiotics such as imipenem from hydrolysis and thus extend their utility. The structures of IMP-1 complexes with two drug-like compounds have been solved via X-ray crystallography at 2.0 Å and 1.8 Å resolutions and deposited into the Protein Date Bank as 1DD6 and 1JJT. In both structures, the flap composed by residue

Glu23 to Val30 presents a close conformation.

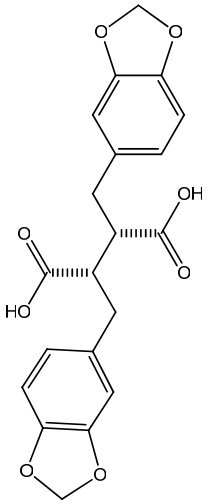
In the complex of IMP-1 and small compound MCI (PDB id: 1DD6), the carboxyl group on MCL interacts with Lys161, and the backbone amide nitrogen of Asn167 through H-bond interaction. Its amide oxygen involves a water mediated hydrogen bond with the side chain of Asn167. The MCL's thiolate bridges two Zn(II) ions in the active site displacing the bridging water in the native structure, and the phenylbutyry side chain binds in a hydrophobic pocket at the base of the flap. Trp28 on the flap edge-to-face interacts with the inhibitor's thiophene ring. In the complex of IMP-1 and BDS (1JJT), one carboxyl group mimicking the thiolate in MCI bridges the two Zn(II) ions in the active site, and the other carboxyl group occupies the same region as the corresponding carboxyl in MCI and forms hydrogen bond with Asn167. The two aromatic rings extend out from the tunnel like binding pocket. One of them conflicts with the side chain of Phe51 leading to 3 Å movement, and the other one edge-to-face interacts with Trp28 which requires a 100 degrees rotation of the Cα-Cβ bond of Phe51 comparing with 1DD6. The rest residues on the flap also present an average 2.5 Å movements. Further computational study is performed on these two complexes (see **Figure 2.9**).

The two ligands are cross docked into the corresponding crystal structures. Results indicate compound BDS can also locate into the MCL pocket, but the crystal structure pocket is energy preferred. For MCL, rotation of Phe51 side chain prevents the phenyl ring on MCL from entering the hydrophobic pocket. Instead, this phenyl ring is exposed into solvent. Hydrogen bonds between the carboxyl on MCL and IMP-1 also disappear in

the cross docking. Both these differences indicate the crystal structure is more reliable. In this case, the two drug-like compounds share similar interactions with the two Zn(II) ions. However, due to the different size of the distal ends, the aromatic rings on TRP28 and Phe51 present different conformations in two complexes. This difference also leads to a 2.5 Å movement of the flap composed by residue 23 to 30



MCL in 1DD6



BDS in 1JJT

Figure 2.8 2D structures of MCL (left) and BDS (right).

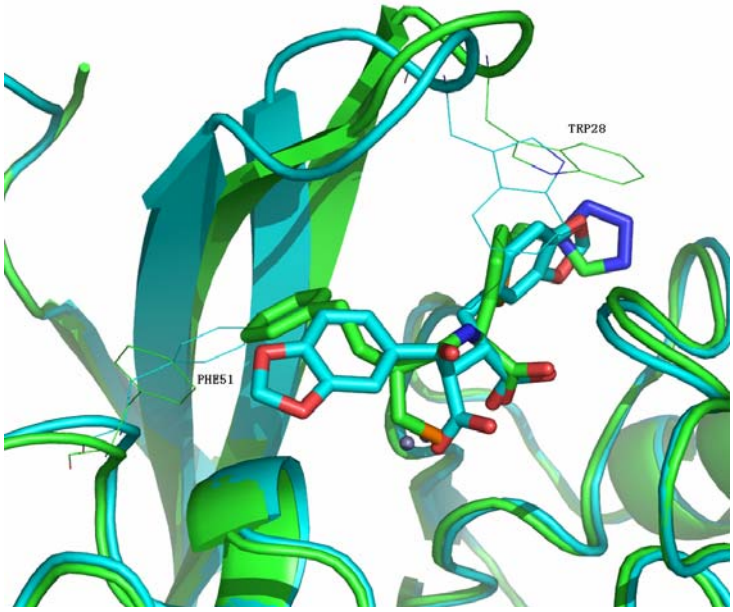


Figure 2.9 Superimposed crystal structures of 1DD6 and 1JJT. Two ligands, MCL (green stick) and BDS (cyan stick) locate in the same binding pocket, but due to the size difference between these two ligands. Due to the side chain movement of Trp28 and Phe51 (green and cyan line), two ligands present different binding models in the same binding pocket.

2.6.5 Solvent Effect: a, 1EGH & 1IK4 [21, 22]

Methylglyoxal synthase (MGS), which is only found in microorganisms, catalyzes the first reaction in the methylglyoxal bypass of the Embden-Myerhoff pathway (glycolysis). The enzyme eliminates phosphate from dihydroxyacetone phosphate (DHAP) to produce methylglyoxal. Phosphoglycolate (PGA), an intermediate analogue, and this has been used to understand the structural basis of how MGS and

triosephosphate isomerase (TIM) which produce chemically identical intermediates but different products. The structure of MGS, which is a homohexamer, has been solved by X-ray crystallography at 2.00 Å resolution in complex with PGA and deposited into Protein Date Bank as 1EGH. The phosphate group forms a cluster of hydrogen bonds with the residues 45-48 and 65-66 as well as Lys23 and Arg150 from the neighboring monomer. On the other end of PGA, in addition to the direct hydrogen bonds to His19 and Asp71, the carboxyl group forms three bridged hydrogen bonds through a water molecule to His19, Asp71 and His98.

For comparison, the structure of intermediate analogue inhibitor PGH has also been solved likewise in complex with MGS at 2.00 Å resolution and PDB-deposited as 1IK4. Alignment of the 1EGH and 1IK4 structures shows the binding site to be very rigid with 0.1 Å RMSD and no single atom shifts of over 0.3 Å (see **Figure 2.11**). Due to the high structural similarity, two ligands are superimposed exactly to each other. The phosphate group of PGH also forms exactly the same hydrogen bonding network with the active site as PGA. However, the replacement of OH by the bulkier NOH on the other end of PGH removed the water molecule associated with PGA. To make up for the loss of the bridged hydrogen bonds from the missing water molecule, the OH group extended further by the additional N atom reaches His19, Asp71 and His98 and forms three identical hydrogen bonds with them as the water molecule in 1EGH. Investigating other reported crystal structure of MGS, the H-bond bridge water molecule in 1EGH is not always in that position. That means the water molecule is not reserved for hydrogen bond interactions.

In fact, due to the trade of bridged hydrogen bonds by direct hydrogen bonds, the binding affinity of PGH has been improved to Ki = 39 nM from Ki=2 μM of PGA.

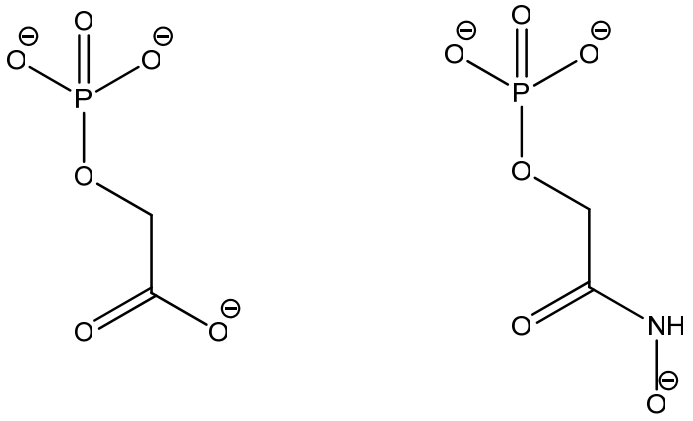
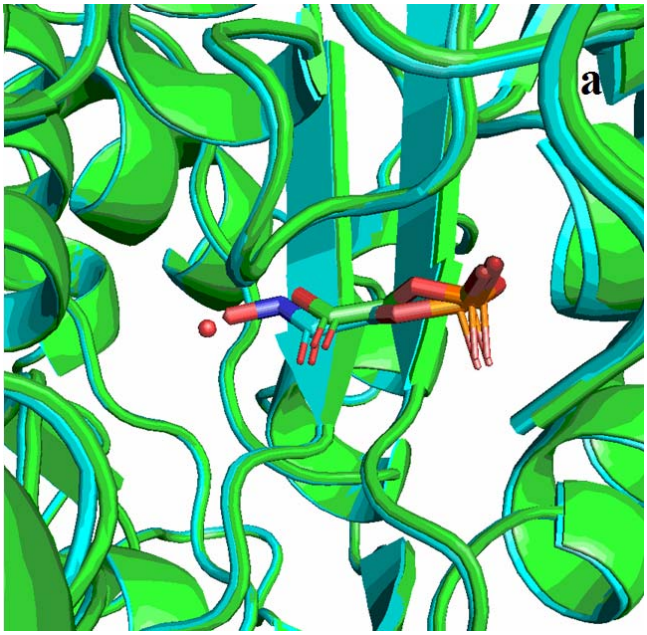


Figure 2.10 2D structures of MCL (left) and BDS (right).



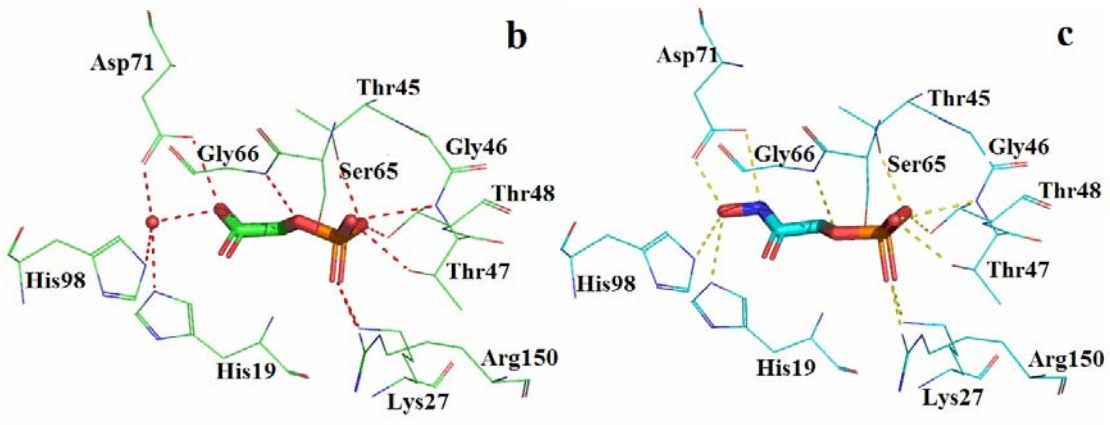


Figure 2.11 a) Superimposed crystal structures of 1EGH and 1IK4. Two ligands, PGA (green stick) and PGH (cyan stick) locate in the same binding pocket with quite similar binding pose except the extra water molecule in 1EGH which act as the bridge to connect the ligand with His19, His98 and Asp71. b) Ligand PGA in MGS ligand binding site. c) Ligand PGH in MGS ligand binding site.

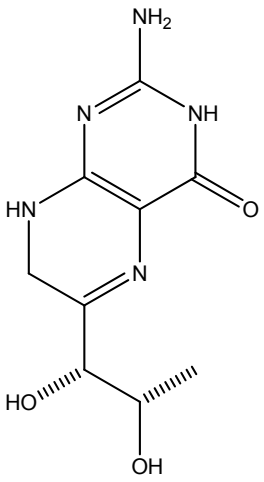
2.6.6 Solvent Effect: b, 1DMW & 4PAH [23, 24]

The mononuclear non-heme iron(II) enzyme phenylalanine hydroxylase catalyzes the hydroxylation of the essential amino acid l-phenylalanine to form l-tyrosin in the presence of the cofactor 6(R)-L-erythro-tetrahydrobiopterin. The crystal structures of phenylalanine hydroxylase complex with different cofactors are determined at 2.00 Å and deposited into PDB as 1DMW and 4 PAH.

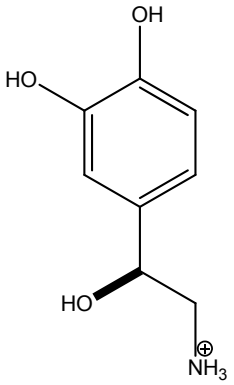
In 1DMW, the catalytic center retains its native state, the ferric ion is chelated by two His, one carboxyl group and three water molecules. There is no direct interaction between the cofactor and the ferric iron. The pterin analog HBI forms hydrogen bonds

with those chelating water molecules through the carbonyl oxygen on the pterin ring and this pterin ring π-stacks the phenyl side chain of Phe254. This cofactor also forms directly hydrogen bonds with Gly247, Leu249 and Ala322, and water mediate interactions with His 264 and Glu286. In 4PAH, the crystal structure indicates that a complex between ferric iron and the catechol was formed. Two of the three water molecules which help chelating the ferric iron in 1DMW are replaced by two hydroxyl groups on the catechol group. One of its hydroxyl also forms a hydrogen bond with Tyr325. A cross docking experiment is also performed in this case. Compound HBI can form direct interaction with Fe^{3+} ion through two hydroxyl group without the solvent molecules, but its polar group will locates in a hydrophobic environment surrounded by Val245, Pro281 and Thr266, which will loose most of its hydrogen bonds with receptor. For ligand LNR, the existence of the water molecules would prevent its direct interaction with metal ion which is the major polar interaction between LNR and receptor (details in **Figure 2.13**).

In this case, solvent molecules play as medium to connect ligand HBI with metal ion in 1DMW. It can be considered as an extension of the ligand, and functionalize as a fragment of the ligand. In this case, they act as part of the ligand and help chelating the metal ion in the binding site. However, these water molecules are not fixed in all crystal structures. In 4PAH, the two hydroxyls on catechol replace two of those three water molecules and directly interact with the ferric.

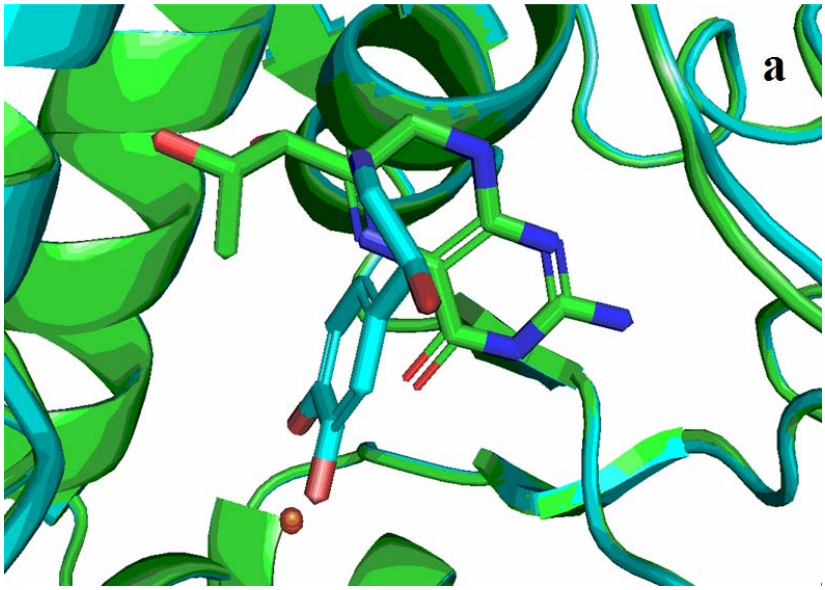


HBI in 1DMW



LNR in 4PAH

Figure 2.12 2D structures of HBI (left) and LNR (right).



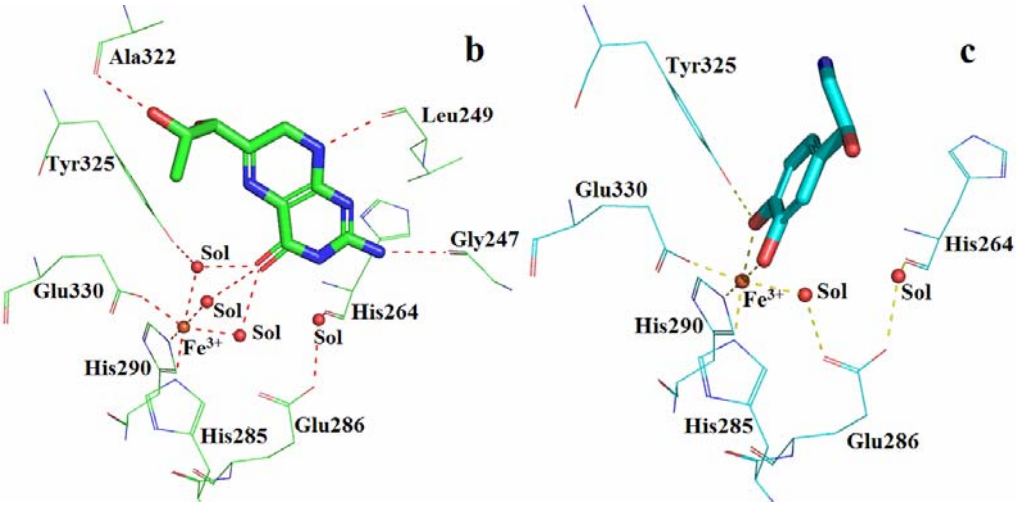


Figure 2.13 a) Superimposed crystal structures of 1DMW and 4PAH. Two ligands, HBI (green stick) and LNR (cyan stick) locate in the same binding pocket with different binding pose. The three solvent molecules in 1dmw act as the extension of the ligand and help chelating the ferric ion. The corresponding ferric ion is chelated by two hydroxyl oxygen from the ligand and one solvent molecule in 4pah. b) Ligand HBI in phenylalanine hydroxylase ligand binding site. c) Ligand PGH in phenylalanine hydroxylase ligand binding site.

2.6.7 Metal Effect: 1C1U vs. 1KTT [25, 26]

Myocardial infarction is the major cause of mortality in western societies, which is closely linked to excessive stimulation of the coagulation cascade. Thrombin is a coagulation protein (serine protease) that converts soluble fibrinogen into insoluble strands of fibrin. To regulate this process, hundreds of inhibitors have been designed to block the enzyme over the past decades, among which is the phase II compound

developed by Boehringer Ingelheim. The structure of the initial template for this phase II compound in complex with the enzyme has been solved via X-ray crystallography at 2.10 Å resolution and deposited into Protein Date Bank as 1KTT. Structure of this template molecule C02 and the 3-D complex are shown in **Figure 2.14.** The amidine group on C02 forms a bidentate salt bridge with the carboxyl group of Asp189 as well as hydrogen bonded to backbone oxygen of Gly219. The distal phenyl ring is bounded by a hydrophobic interaction in the distal (D) pocket and the N-methyl group nicely fits into the proximal (P) pocket of the thrombin active site, respectively. The binding affinity for C02 is pK = 5.8 (IC50 = 1.5 μM).

For comparison, the structure of another inhibitor BAI designed by Arris Pharmaceutical Corporation as shown in **Figure 2.14** has been solved in complex with the same enzyme at 1.75 Å resolution and deposited as 1C1U. Having 1KTT and 1C1U structures aligned (see **Figure 2.15**), the binding site is shown to be very rigid with a small 0.4 Å root-mean-square (RMS) deviation with no single atom shifts over 2 Å. BAI contains an amidine motif as well, which interacts via a salt bridge with Asp189 and hydrogen bonded to Gly219 in the same way as C02. However, the other end of the BAI points to the opposite direction of binding pocket. This is induced by the addition of Zn^{2+} ion in 1C1U binding site. The nitrogen atoms from two imidazole groups of BAI are positioned to couple the Zn^{2+} coordination, through which interact with His57 and Ser195 and fill the P-pocket. The additional metal-mediated interaction also explains the much higher binding affinity of BAI with pK = 8.3 (Ki = 6.0 nM).

This case is a perfect example to illustrate the potential of engineering metal ions in drug design process.

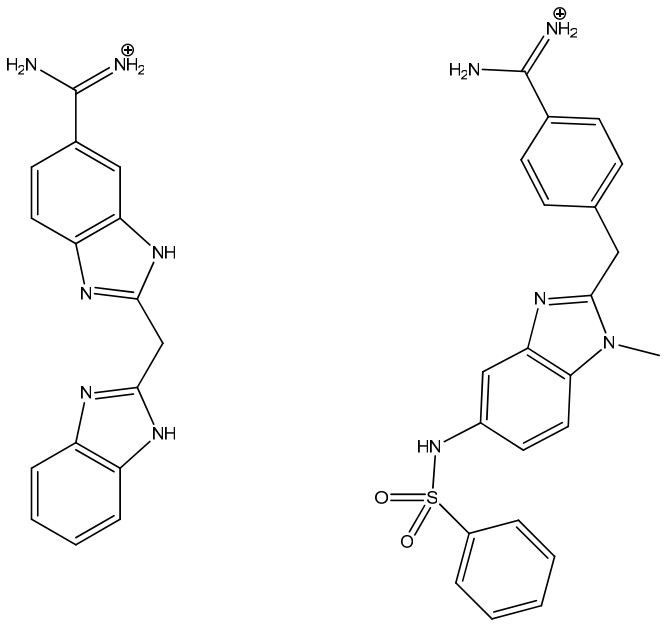
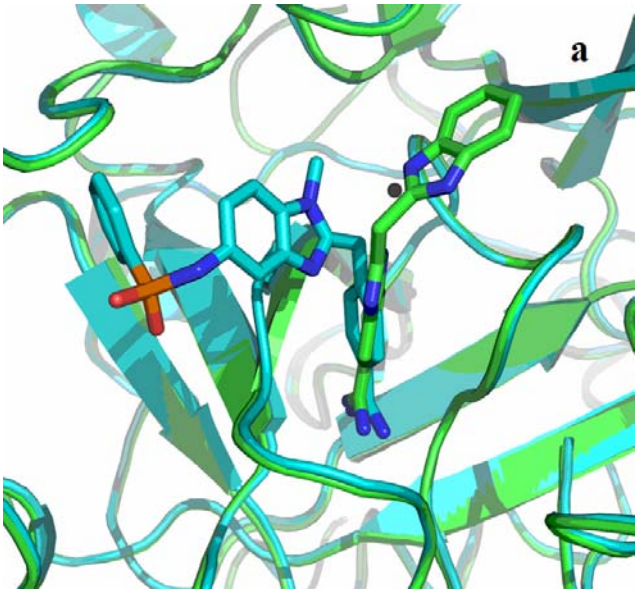


Figure 2.14 2D structures of HBI (left) and LNR (right)



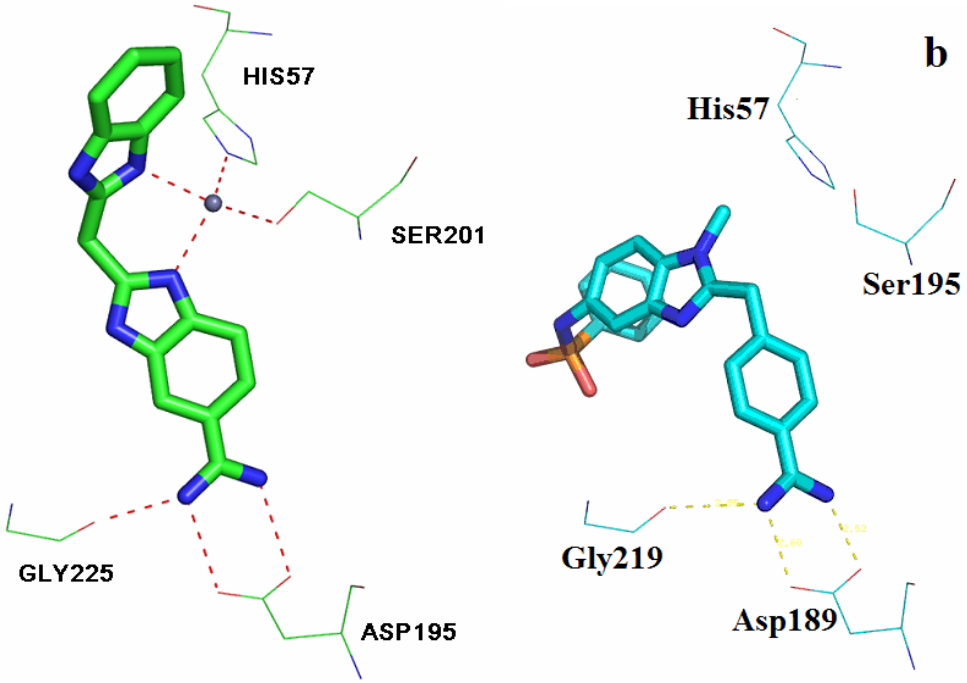


Figure 2.15 a) Superimposed crystal structures of 1C1U and 1KTT. Two ligands, C02 (green stick) and BAI (cyan stick) locate in the same binding pocket with different binding pose. b) Ligand HBI in phenylalanine hydroxylase ligand binding site. c) Ligand PGH in phenylalanine hydroxylase ligand binding site.

2.7 Conclusions

In this protein-ligand interaction survey, we didn't utilize all the published structures as searching templates, because that will lead to huge amount of duplicate results. Only the family and the subfamily members listed in the PDB-At-A-Glance database were utilized as searching template. These selections covered most of the known protein family and the result is biologically and statistically meaningful.

The survey result indicates 19 out of over 3000 cases violate the structure assumption. Although these peculiar cases violate the general assumption that similar ligands bind into a given protein in a similar way, the structure base assumption is trustworthy in over 99% cases. However, because of protein conformational changes, solvent and metal ion effects between proteins and inhibitors or size difference between compounds, small molecules with different pharmacophore properties can bind into the same protein binding pockets. Sometimes, this difference is remarkable. In such cases, adopting the assumption will lead to incorrect results and poorly active ligands designed.

Chapter 3: Computational Design of Liver Receptor Homolog 1 Antagonists Based on Helix-12 Conformational Reorganization

3.1 Nuclear Receptors

Nuclear receptors (NRs) are a classic example of receivers for small-molecule chemical messengers. They regulate the expression of responsive genes and thereby affect diverse biological processes including cell growth, development, and metabolism. The NRs are well-adapted for this type of function because they not only specifically bind the small molecule, but are capable of relaying or transducing a complex set of signals carried along by the properties of the ligand. The regulation of gene expression by nuclear receptors happens only when a ligand, a molecule that affects the receptor's behavior, is present. In more specific terms, ligand binding to a nuclear receptor results in a conformational change in the receptor, which in turn, activates the receptor resulting in up-regulation of gene expression [1-3].

A unifying feature of the NR superfamily is that each receptor consists of three major domains: an assembly of functional modules (AF-1) which varies among family members, a DNA binding domain (DBD) composed of two zinc finger motifs which allow for the specific recognition of short and imperfect inverted repeats of DNA or

direct repeats, and a ligand binding domain (LBD) typically about 250 amino acids in length which is composed of 12 helices and 1 hairpin turn (see Figure 3.1 left). The module most relevant to current drug-discovery approaches is the C-terminal LBD [4].

The activity of most NR superfamily members is controlled by small lipophilic ligands such as steroid hormones, retinoids, vitamin D and thyroid hormone [5]. In the ligand-dependent activation of NRs, a general and important structural feature illuminated by X-ray crystal structure analysis of LBDs is the ligand-induced re-folding of a loop to helix 12 (H12), an LBD substructure. In the inactivated-form of nuclear receptors, H12 locates in the co-activator binding site or extends to the solvent region (see **Figure 3.1** right). However, in the active-form, it forms a helix and covers the ligand-binding pocket. Based on these, an H12-folding inhibition hypothesis is introduced: the H12 folding is an essential condition but not sufficient condition for the activation of NRs. That is to say, H12 folding leads to either active or inactive form of NRs, but H12 unfolding only results in inactive NRs. The validity of the H12-folding inhibition hypothesis for molecular design of NRs antagonists is also supported by X-ray crystal structures of the LBD of holo-form NRs and known antagonists [6].

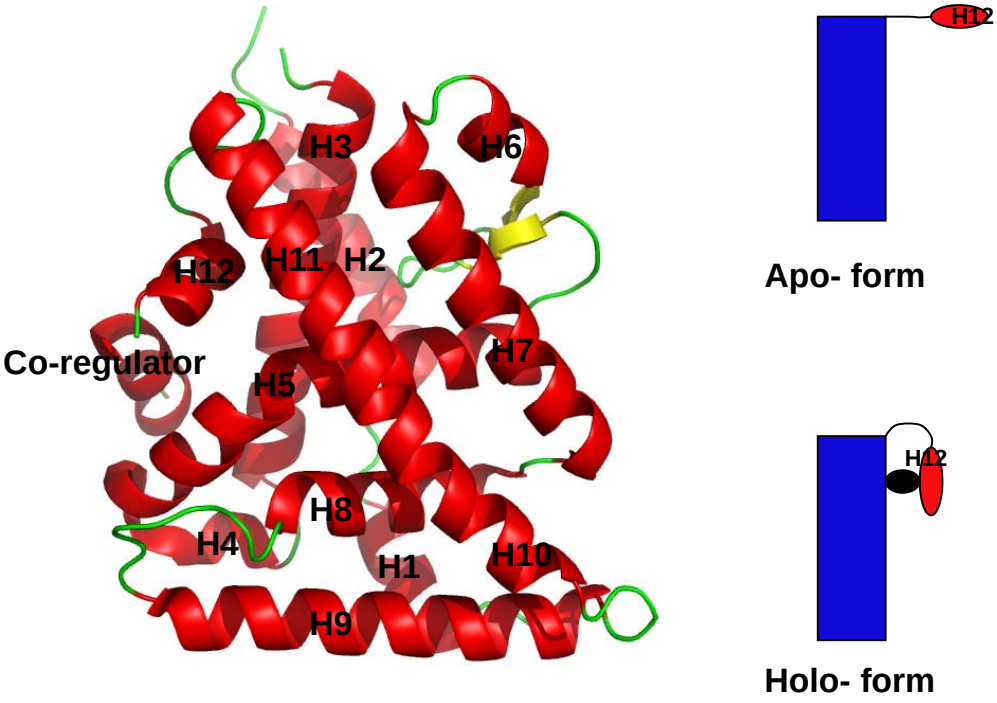


Figure 3.1 Structure of nuclear receptor's ligand binding domain. **Left**, the ligand binding domains of NRs are composed of 12 helices (red) and 1 hairpin turn (yellow) connected by several loose loops (green). The ligand binding pocket is at the top of this domain surrounded by helices 3, 5, 6, 7, 11, 12 and the hairpin. **Right**, in the inactive state, helix 12 takes an open conformation. With the agonist bound, helix 12 is folded and forms part of the ligand binding pocket.

3.2 Liver Receptor Homolog 1 (LRH-1)

Liver receptor homolog 1 (LRH-1) is a member of nuclear receptor subfamily V that also includes the steroidogenic factor 1 (SF1). LRH-1 activates an array of genes responsible for development of endodermal organs such as liver, intestine and pancreas. It

also plays a central role in lipid homeostasis by regulating genes involved in bile acid biosynthesis, reverse cholesterol transport and metabolism of lipoprotein complexes [7]. As for other NRs, LRH-1 is composed of several modular functional domains, including an N-terminal A/B domain, a characteristic zinc finger DNA binding domain, a hinge region and a C-terminal ligand binding domain. Different from other NRs, LRH-1 seems to be a constitutively active transcription factor in that it activates many reporters in the absence of any exogenous ligand. The crystal structure of the mouse LRH-1 LBD reveals a sandwich fold of four layers of helices instead of the three layers observed in other NRs. The extra fourth layer is constructed by H2 which is much longer than it is in other NRs. Mutation on R412M on LRH-1 Helix 2 displays moderately decreased activity relative to that observed for the analogous SF-1 R314M mutation. This suggests that additional interactions at the more extended H2-H3 interface in LRH-1 help to stabilize the receptor structure, so that its C-terminal activation helix H12 is packed in an active conformation even in the absence of a small ligand populating the ligand-binding pocket [8].

Based on the H12-folding inhibition hypothesis, LRH-1 antagonists were designed based on known antagonists of other nuclear receptors. Due to the structural differences between LRH-1 and other NRs, reported antagonists for NRs do not inhibit LRH-1. Nonetheless, these antagonists were modified to be potential LRH-1 inhibitors by substitution of additional moieties which were predicted to displace H12 from its active position. This design process leads to several candidates which were synthesized by Dr. Tony Barrett and assayed for LRH-1 inhibition by Dr. Simak Ali and his team at Imperial

College London, resulting in several novel lead compounds with mild activities.

3.3 Structural Comparisons

The X-ray crystal structure of LRH-1 reveals the ligand binding pocket in the middle of the tri-layer structure, formed by H3, H5, H6, H7, H11, H12 and the hairpin. The center of the pocket is mostly hydrophobic except for one polar region. This polar sub-site (P1) is located between the bottom of the hairpin and the distal end of helix 5 which contains four polar residues, Asp389, His390, Arg393 and Thr352, and an intra-hydrogen bond between the side chain of Asp389 and Arg393. A non-polar sub-site (N1) is adjacent to the P1 pocket and between H12 and H11. Beside these two small sub-sites, three tunnels (two similar size polar tunnels and one small non-polar tunnel) connect the main pocket with solvent region. Tunnel A (composed by H3, H11 and the loops between H6 and H7) is found in most NRs and is considered the major entrance of small molecules. The other polar tunnel (tunnel C) is adjacent to the P1 pocket, surrounded by H6, H7 and the hairpin. The smallest non-polar tunnel (tunnel B) is composed by H3, H6, and the top of the hairpin.

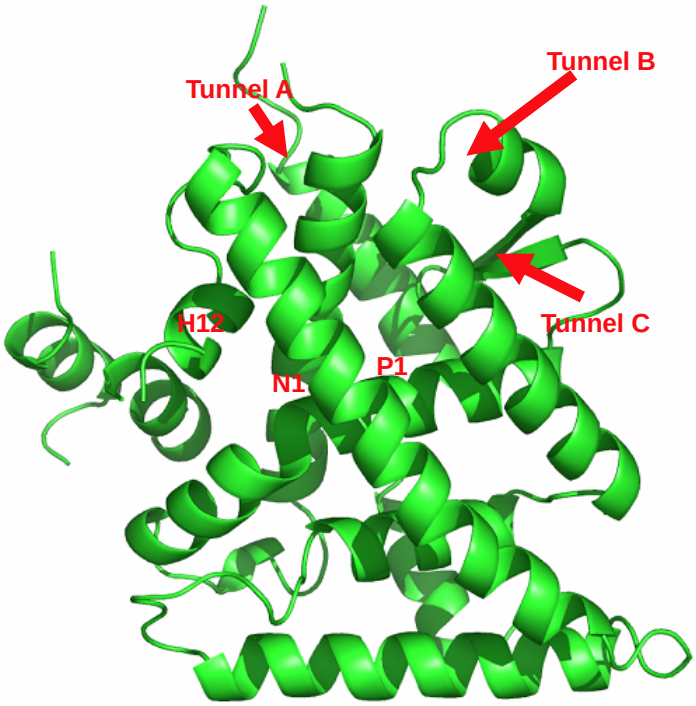


Figure 3.2 X-ray crystal structure of LRH-1, labeled by sub-sites and tunnels.

3.3.1 LRH-1 vs. Estrogen Receptor-α (ER-α)

The overall backbone RMSD of LRH-1 (PDB:1YOK) [9] with the estrogen receptor α unit (ER-α) (PDB:3ERT) [10] is 2.7 Å. The superimposed LRH-1 and ER-α structures show that the H2 helix in LRH-1 is replaced by a loose loop between H1 and H3 in ER-α (see Figure 3.3 left). This loop is close to the ligand binding pocket. It not only covers tunnel B but also alters the conformation of the hairpin, pushing the hairpin and H6 closer to H7 in ER-α compared with LRH-1. Another loop between H6 and H7 is extended to tunnel A and reduced in size in ER-α. Tunnel C is also eliminated in ER-α due to the conformational shifts of H6 and H7. The P1 sub-site residues are moved out of the pocket;

only Arg394 is left in the pocket forming a hydrogen bond with the hydroxyl group on the co-crystallized tamoxifen, an ER-α antagonist. The double bond center of Tamoxifen locates in the center of the pocket, with its dimethylamine head occupying in the space where H12 locates in the activated state (see Figure 3.7 bottom). Helix 12 is pushed away from its active conformation to the co-regulator pocket upon tamoxifen binding (cf. Figure 3.3). Thus, we predict that a compound that occupies the H12 region in LRH-1 with a tamoxifen-like binding pose (pose H) will function as an LRH-1 antagonist, preventing H12 holo-form conformation, agonist binding and subsequent LRH-1 activation (Figure 3.1).

3.3.2 LRH-1 vs. SF-1

Superimposed crystal structures of SF-1 (PDB:1YP0) [11] and LRH-1 (PDB:1YOK) [9] show that their backbones superimpose quite well except for the flipped loop between H2 and H3 and the shorter distance between H6 and H7 in SF-1 (see Figure 3.3 right). The overall backbone RMSD of these two crystal structures is 1.6 Å. In the SF-1 LBD, tunnel B is covered by the loop between H2 and H3, and tunnel C is blocked by side chains of the residues on H6 and H7. Only tunnel A is found in SF-1 LBD crystal structure. Docking investigation shows that small molecule agonists or antagonists of SF-1 can reside in tunnel C in LRH-1, leaving no structural elements near Helix 12. In other words, a blocker for Helix 12 folding in SF-1 is unlikely to prevent Helix 12 from

folding in LRH-1.

The comparison of the crystal structure of LRH-1 and other NRs indicates that LRH-1 not only exhibits the general properties of NRs LBD, but also several characteristics which are different from them. We can see from Table 3.1, that the extra tunnels B and C in LRH-1 provide 200-300 $Å^3$ additional volume to the ligand binding pocket. This extra space increases the capability of this pocket to accept large molecules. As a result, small molecule antagonists of other NRs may involves in the extra space of the pocket showing no conflict with H12.

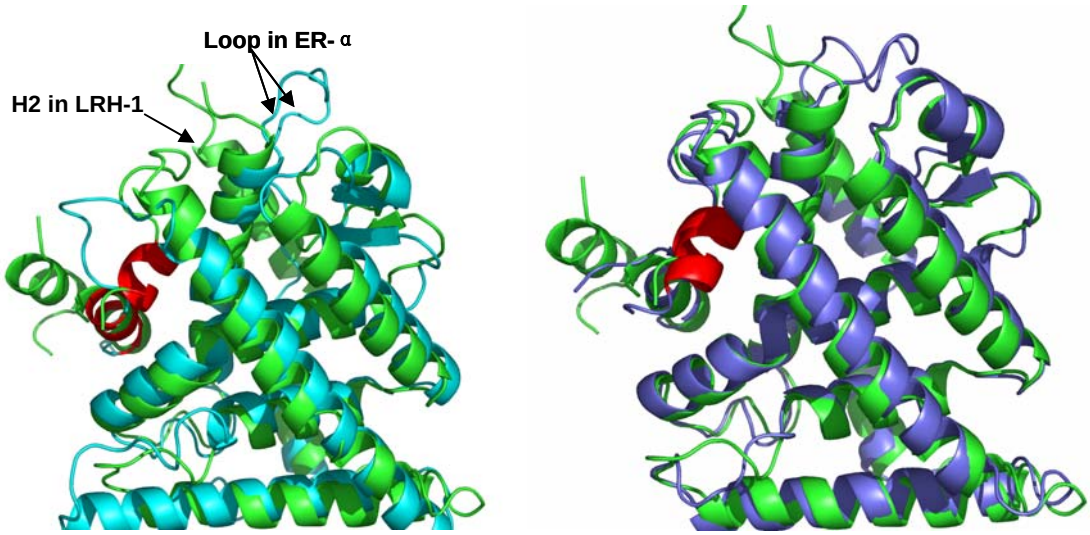


Figure 3.3 Superimposed crystal structures of LRH-1 (green) with ER-α (cyan on right) and SF-1 (blue on left). Helix 12 is colored red.

Table 3.1 Structural properties of LRH-1, ER-α and SF-1

	ER-α (3ERT)	SF-1 (1YP0)	LRH-1 (1YOK)
Volume Å ^{3}	1299	1680	1953
Tunnel A	Y	Y	Y
Tunnel B	N	N	Y
Tunnel C	N	N	Y
N1	N	Y	Y
P1	Y	Y	Y

3.4 Computational Methods

In this research, four published agonists, partial agonists and antagonists of various NRs were selected as lead compounds. Tamoxifen [12] is an antagonist of estrogen receptors that has been used for more than 30 years to treat breast cancer in women. Raloxifene [13], an oral selective estrogen receptor modulator (SERM), has estrogenic actions on bone and anti-estrogenic actions on the uterus and breast. SID-7969543 [14] is an NIH probe that shows antagonist properties against SF-1 but no effect on LRH-1. The last selected compound MFA-1, a steroid analog and agonist of the farnesoid X receptor (FXR), was obtained from PDB: 3BEJ [15].

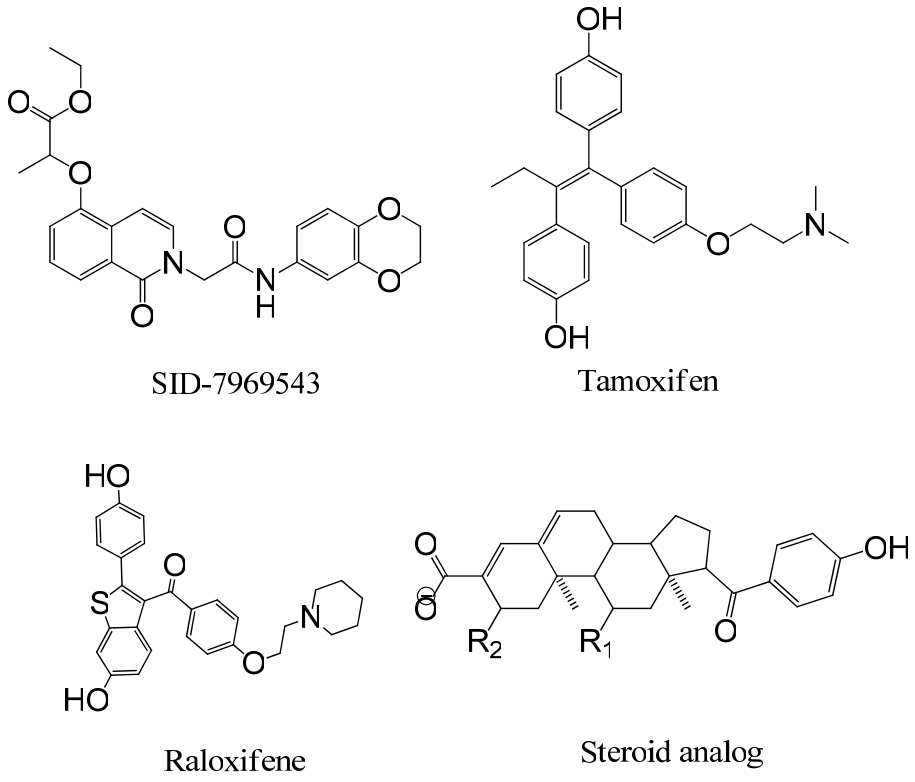


Figure 3.4 Four lead compounds selected for LRH-1 antagonist design.

The lead compounds were prepared by the Maestro GUI and Ligprep, [16] and the receptor structure was prepared by the 'protein preparation wizard' tools integrated in Maestro GUI. A docking exercise was performed to dock the leads into the prepared receptor structure with Glide [17]. In the published LRH-1 LBD crystal structures, H12 is folded into its active conformation. In order to mimic the inactive state of LRH-1, H12 and the adjacent loop were removed in the receptor preparation step. The docking results were scored by Prime/ MMGBSA [18] and the top-scoring poses of each analog were analyzed.

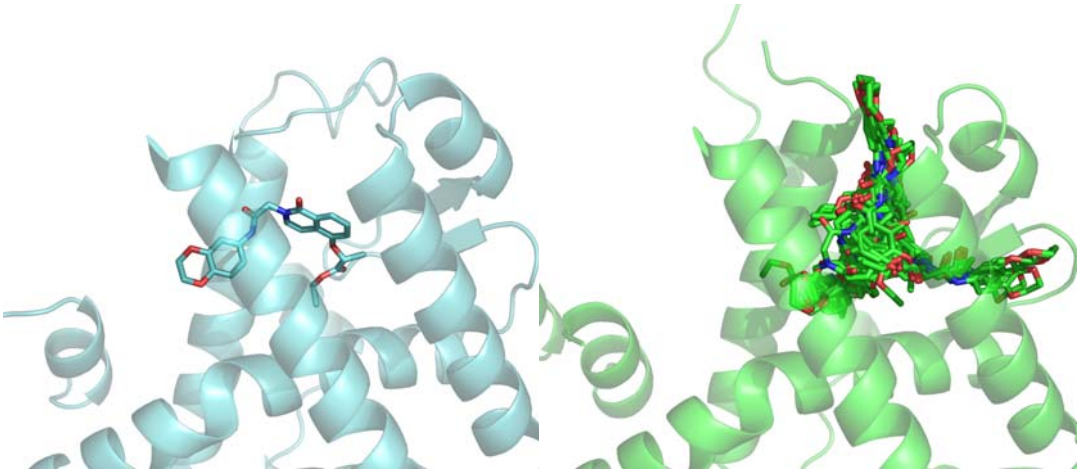
Analysis of the LRH-1 docking results indicates that the lead compounds always involve in the extra tunnel **C** or the enlarged tunnel **A** instead of the region close to H12. This docking result is not surprising based on the structural comparison between LRH-1 and other NRs. In order to obtain promising LRH-1 antagonists, modifications were performed on these lead compounds. The aim of these modifications is to design molecules which inhibit LRH-1 function upon blocking H12 from adopting its active conformation. The modified compounds should prefer binding near H12 in order to prevent H12 from folding. The easiest way to alter compound binding is to modify the moieties which interact with tunnel **C** to disfavor these interactions. Modifications on the central core moiety can also, in principle, induce a major change in the ligand binding mode. This change reorients the molecule's two distal ends among tunnels **A**, **C**, and the Helix-12 region. However, such changes are difficult to predict and control.

3.5 Analysis of Results

3.5.1 SID-7969543 Analogs

SID-7969543 was reported by Madoux *et al* as a probe for SF-1 [14]. It selectively inhibits SF-1 but has no effect on LRH-1. Docking results, shown in **Figure 3.5,** illustrate that in SF-1, SID-7969543 occupies the space utilized by the activated form of Helix 12,

but in LRH-1, the compound resides in tunnel **C** and shows no steric conflicts with Helix 12. This difference is consistent with the bioactivity data in **Figure 3.5** The LRH-1 docking result also indicates that the hydrogen atoms at R_4 to R_8 on the bicyclic moieties are close to the residues around them. Modifications at these positions are predicted to conflict with LRH-1 and alter the ligand's binding pose. All analogs were generated with Maestro 9.0 [19] and prepared by LigPrep2.3. Docking procedures were performed with Glide and rescored by Prime-MM/GBSA. The best pose identified for each analog was investigated.

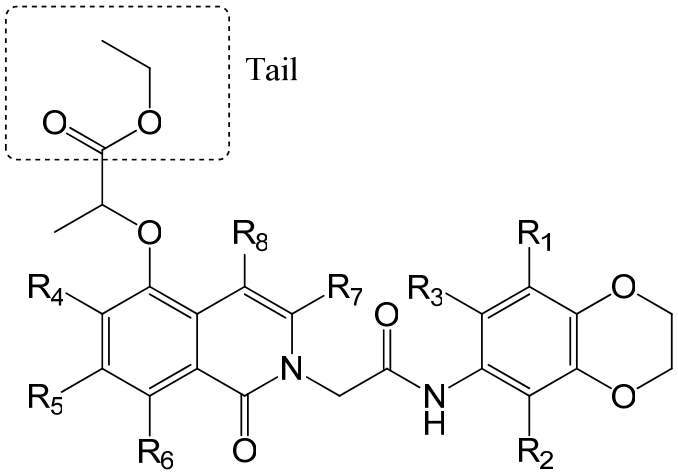


	Inhibition (cytotoxicity) %	IC_{50} (CC_{50}) nM
SFRE promoter assays with full-length proteins		
SF-1	136∓4	30 ∓15
LRH-1	0 ∓3	N.A.

Figure 3.5 SID-7969543 was prepared and docked into SF-1 (top left) and LRH-1 (top right). The top

docking pose of SID-7969543 in SF-1 has steric conflicts with H12. In LRH-1 the top 20 poses show no conflicts with H12. The bio assays results from [14] are listed in the bottom table.

Compound SID-7969543 can be divided into 3 moieties: the benzodioxane head, the bicyclic core and the ethyl tail. There are several positions that can be modified on each moiety. Modification at the R_1, R_2 and R_3 positions on the benzodioxane are predicted to prevent the head group from entering tunnel **C** because of steric conflicts with residues around tunnel **C**. Modifications at the R_4, R_5, R_6, R_7, and R_8 positions on the bicyclic core are predicted to alter the compound's binding pose through steric conflicts with the residues around the central pocket. Modifications on the ethyl tail is also believed to alter the analog binding pose by changing its hydrogen bond interactions with polar residues on the **P1** pocket or tunnel **A**. Individual and combined modifications to these regions were investigated.



Various non-polar and polar groups were attached at different positions of probe SID-7969543. Docking results indicate that the same moiety placed at different positions

results in different binding poses. For most of the SID-7969543 analogs, the bicyclic core locates in the center of the pocket, and the benzodioxane head prefers tunnel **A** or tunnel **C** resulting in no conflicts with H12 in LRH-1. However, there are a few exceptions. On the benzodioxane head group, a small hydrophobic substituent at R_1 or R_2 (analog 3-1 and 3-2 in **Figure 3.6** a and b) leads the benzodioxane head to overlapping with H12. However, modification at R_3 is different from R_2 and R_1. With small group modifications at R_3 such as methyl or hydroxyl, the benzodioxane head is always found in the small hydrophobic **N1** pocket, and no conflicts were observed between the compound and H12. When the moiety at R_3 is enlarged to an ethyl or hydroxyl methyl (analog 3-3 and 3-4 in **Figure 3.6** c and d), this substitution replaces the benzodioxane in the **N1** pocket. The benzodioxane head is simultaneously relocated to the H12 region and conflicts with the active form H12.

On the bicyclic core, a polar group such as CH_2OH at R_5 (analog 3-5) results in an antagonist-like pose in which the benzodioxane head overlaps with H12 (**Figure 3.6** e). A hydrogen bond between the hydroxyl and Arg393 is found which is absent in other analogs. Small hydrophobic groups at R_6 also lead to promising poses (analog 3-6 and 3-7 in **Figure 3.6** f and g). Modifications at R_4, R_7, and R_8 give no promising candidates. In these analogs, the head groups always reside within the **N1** pocket, and the ethyl tail lies into the P1 pocket or near tunnel **A**.

In the modifications above, the ethyl tail of the promising analogs always locates in a hydrophilic region such as the small polar cave P1 or near tunnel **A** (**Figure 3.6** a-g). If

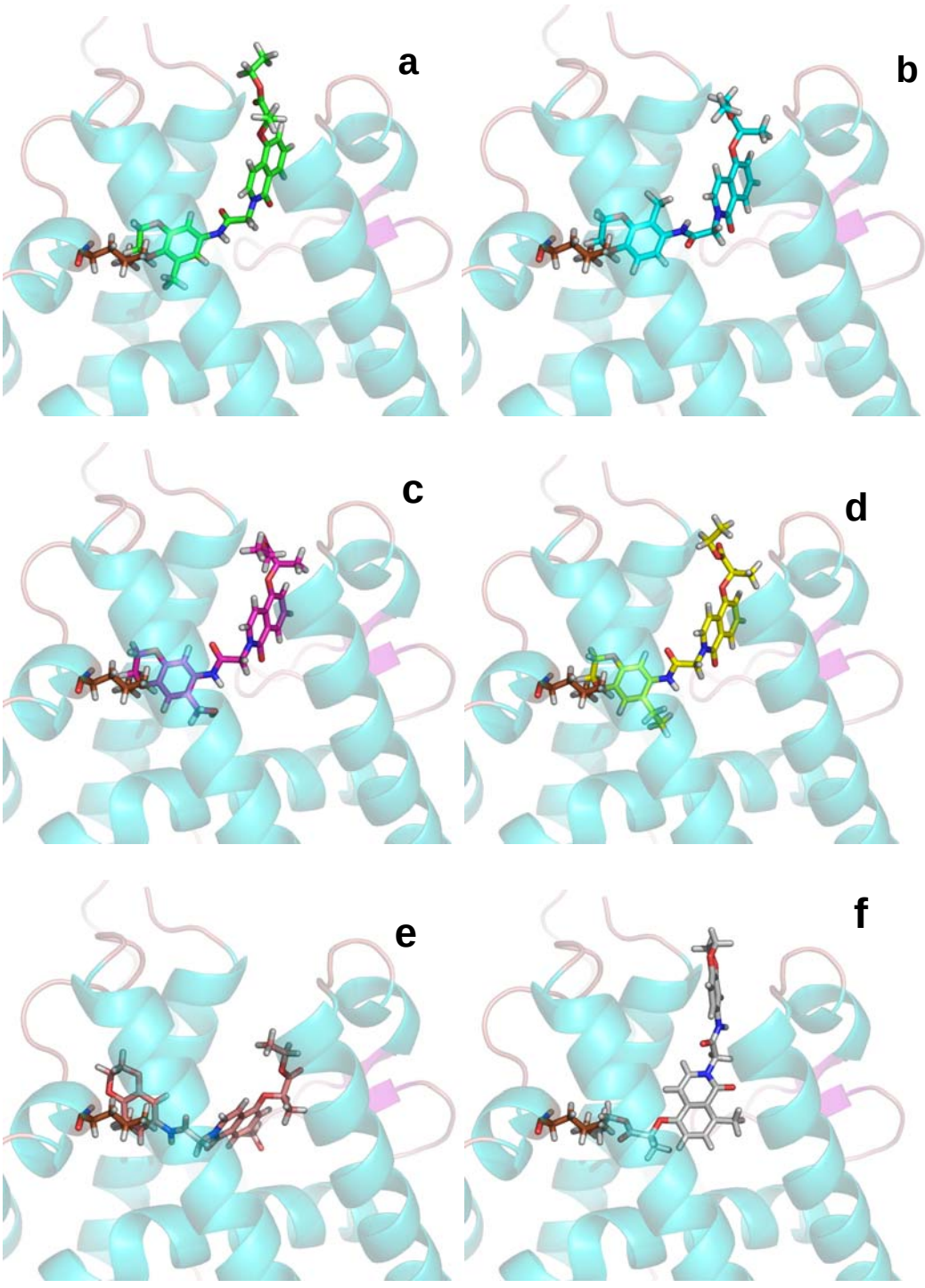
the ethyl tail is modified to a polar group, it should improve the activity of the compound by additional polar interactions. Thus, the ethyl tail was removed leaving the carboxyl group alone in order to form polar interactions with the polar cave **P1** or tunnel **A**. Docking study indicates that the modified tail forms hydrogen bonds with Arg393 and Val406 in the **P1** pocket. However, with these hydrogen bonds, the bicyclic core lies on the bottom of the pocket and parallels to H5, while the head group prefers tunnel **A** and shows no potential conflicts with H12. Thus, a single modification at this position would not likely lead to a LRH-1 antagonist. However, when it is combined with modifications on the benzodioxane head or the bicyclic core, such as a methyl group at R_1 or R_2 (analog 3-8 and 3-9 in **Figure 3.6** h and j, respectively), the head group is found in the H12 region. In these analogs, hydrogen bonds are always formed between the modified tail and Arg393 and His390, helping retain the position of the bicyclic core.

In summary, computational study indicates that SID-7969543 analogs with a small hydrophobic group at R_1, R_2, R_3, or R_6 might function as LRH-1 antagonists. With modifications at these positions such as in analogs 3-1, 3-2, 3-3, 3-4, 3-6, and 3-7, the benzodioxane head can be directed to H12 in LRH-1 in a manner analogous to SID-7969543 in SF-1. The ethyl tail is no longer located on the bottom of the hydrophobic cavity. Instead, it is found at the ligand entrance of tunnel **A**. A $-CH_2OH$ attachment at R_5 (analog 3-5) is also a good candidate. The additional hydroxyl group engages in hydrogen bonds with Arg393, which helps the analog lie on the bottom of the pocket and present the head group near H12. Modification at the ethyl tail alone cannot

lead to promising LRH-1 antagonists. However, when combined with R_1 or R_2 modifications such as analogs 3-8 and 3-9, they also lead to LRH-1 antagonist candidates. In such analogs, the modified tail is found in the **P1** pocket, where it hydrogen bond interacts with Arg393 and His390. The analogs mentioned above are listed in **Table 3.2** and the structures complex with LRH-1 can be found in **Figure 3.6**.

Table 3.2 SID-7969543 analogs which show conflictions with H12.

#	R_1	R_2	R_3	R_5	R_6	tail
3-1	Me	H	H	H	H	H
3-2	H	Me	H	H	H	H
3-3	H	H	MeOH	H	H	H
3-4	H	H	Et	H	H	H
3-5	H	H	H	MeOH	H	H
3-6	H	H	H	H	Me	H
3-7	H	H	H	H	Cl	H
3-8	Me	H	H	H	H	Carboxyl
3-9	H	Me	H	H	H	Carboxyl


a
b
c
d
e
f

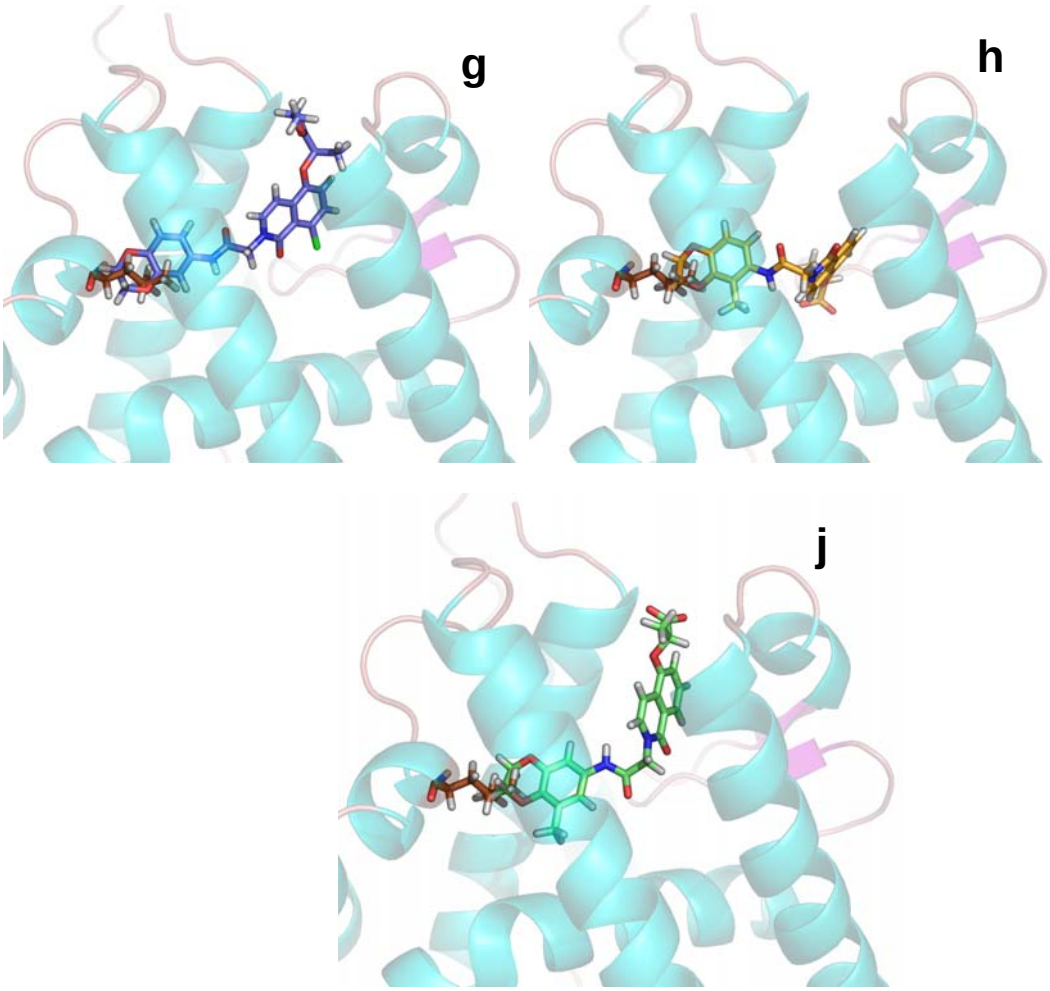


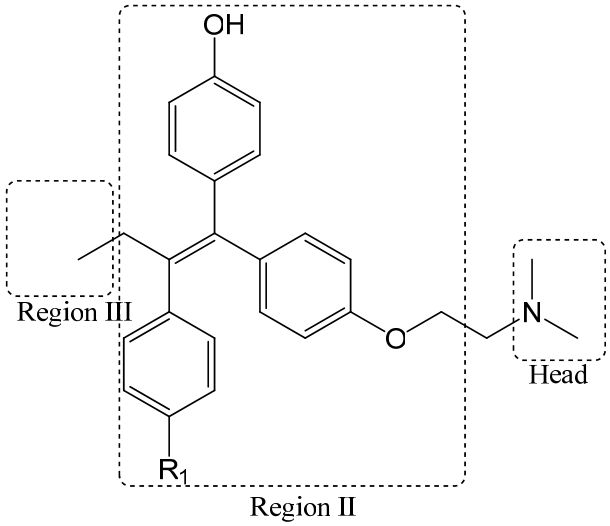
Figure 3.6 Best docking poses of SID-7969543 analogs in LRH-1. a) analog 3-1, b) analog 3-2, c) analog 3-3, d) analog 3-4, e) analog 3-5, f) analog 3-6, g) analog 3-7, h) analog 3-8, j) analog 3-9.

3.5.2 Tamoxifen Analogs

Tamoxifen is an antagonist of the estrogen receptor (ER) in breast tissue. It has been the standard endocrine (anti-estrogen) therapy for hormone-positive early breast cancer in post-menopausal women, although aromatase inhibitors have been proposed. Estrogen

binds to and activates the estrogen receptor of some breast cancer cells which require estrogen for growth. Tamoxifen is metabolized into compounds that also bind to the estrogen receptor but do not activate it. Because of this competitive antagonism, tamoxifen acts as a key broken off in the lock that prevents any other key from being inserted, preventing estrogen from binding to ER ligand binding domain [12].

Tamoxifen was prepared and docked into ER-α and LRH-1 with the same procedures as for the SID-7969543 analogs (**Figure 3.7**). In ER-α, the core structure of tamoxifen is composed of one double bond and three attached phenyl rings located in the center of the hydrophobic pocket. The dimethylamine head points to the H12 region and prevents H12 from folding into its active form. In LRH-1, the core structure of tamoxifen also sits in the center of the hydrophobic pocket, but the protonated dimethylamine head points into tunnel **C**. The cis- phenyl ring is in the non-polar pocket **N1**. With this pose, tamoxifen does not show any potential conflicts with H12.



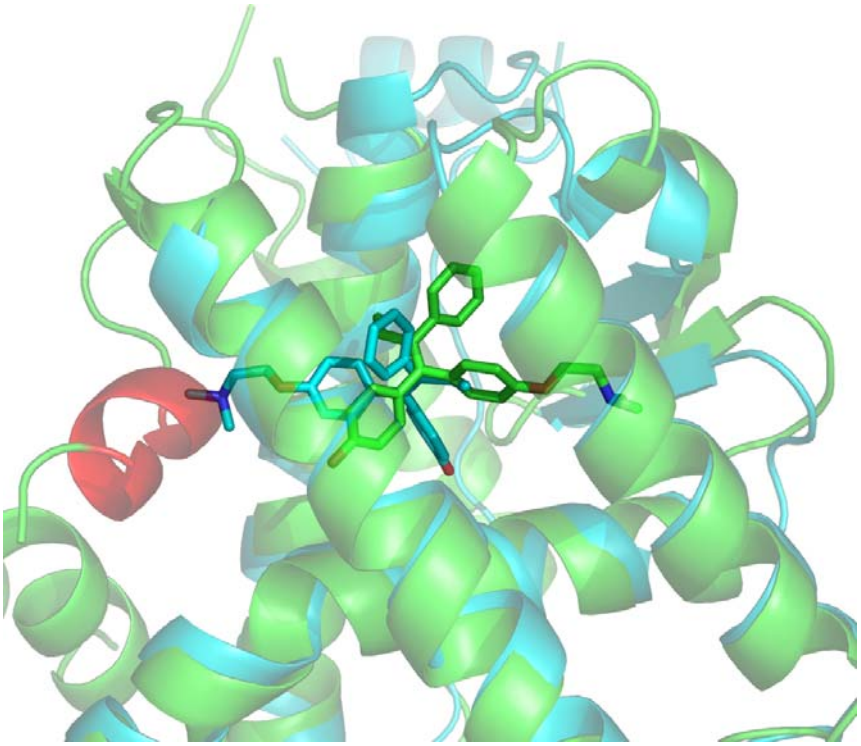
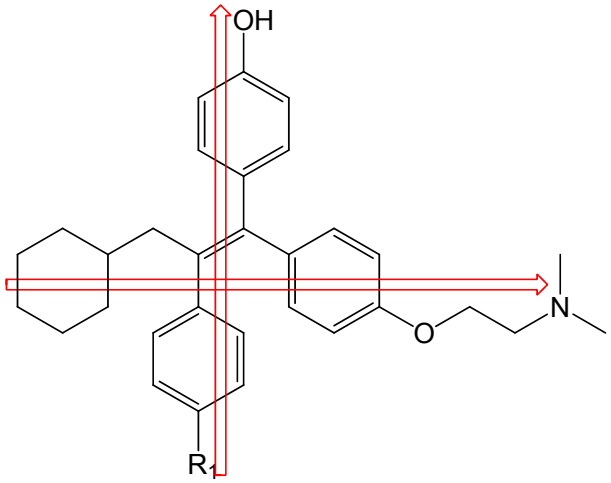


Figure 3.7 Tamoxifen 2D structure (top) and its binding pose in ER-α (cyan cartoon and stick on bottom) and LRH-1 (green cartoon and stick on bottom). The red helix is helix 12 of LRH-1. Due to the steric conflict of tamoxifen and H12 in ER-α, the H12 of ER-α is missing in the crystal structure.

Since the tamoxifen head is a small moiety, preferring to locate in tunnel **C** with no predicted conflicts with H12, the head group was enlarged to a 6-membered ring, either saturated (cyclohexane, piperidine) or unsaturated (phenyl or pyridine ring). Docking study indicates the modified head groups still prefer tunnel **C**. Modifications at R_1 position were also performed on tamoxifen. Neither polar nor non-polar groups induce an antagonist-like binding pose as tamoxifen does in ER-α. The head groups of all these analogs prefer tunnel C instead of the H12 region. A combination of region I and region

II modifications were also investigated, but such modifications did not change the binding poses of these analogs. Modification at region III leads to some promising results. When the ethyl tail is replaced by methylcyclohexane (analog 3-10 in **Figure 3.9** a), this analog binds with its benzohydroxyl group approaching the **P1** pocket and forms a hydrogen bond with Thr352. The dimethylamine is near tunnel **C**, and the cyclohexane points close to H12 showing a slight conflict with Ser532. The remaining phenyl ring points to the entrance of the pocket, tunnel A. With this modification, the new analog presents a cross shape in the ligand binding pocket of LRH-1, in which the four side chains (3 phenyl and one cyclohexane) point in four different directions: up to tunnel A, down to the bottom of the pocket, left to helix 12 region, and right to tunnel C. With this shape, the compound keeps its core structure in the center of the pocket and one of its side-chains near the space which will have steric conflicts with H12.



Region III modification provides an analog with potential to function as an LRH-1 antagonist. However, this cross-shaped pose doesn't prefer in MM/GBSA scoring. Further modifications were performed on this analog to favor binding in the desired pose.

With a modified region III, small group modifications at R1 lead to promising antagonist poses. When R_1 is replaced by Cl, OH, Et or i-Pr (analog 3-11 to 3-14 in **Figure 3.9** b), the top scoring poses bind in a manner similar to tamoxifen in ER-α. With these modifications, the dimethylamine head overlaps with H12, the benzohydroxyl is close to tunnel A, and the modified R1 group locates to the N1 pocket. When R_1 is replaced by methyl or methoxyl group, the antagonist pose is no longer the top-scoring pose. This series of analogs show that the double bond in the core structure always locates in the center of the hydrophobic pocket, while the four side chains (the cyclohexyl and three phenyl rings) point to four different directions. In the most promising analogs, binding occurred with the dimethylamine side chain pointing to and overlapping with the active form of H12.

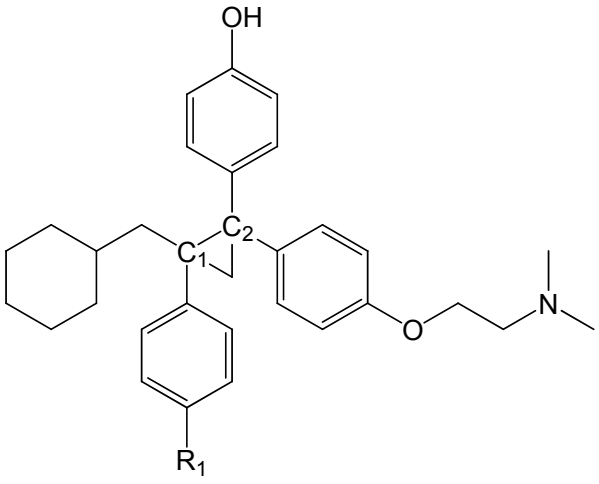


Figure 3.8 Scaffold of analogs 3-15 to 3-18

Advised by the synthetic group, replacing the double bond core with a cyclopropyl

group was also investigated. With this modification, there are four stereoisomers of each tamoxifen analog: R^1S^2, R^1R^2, S^1S^2 and S^1R^2. In order to keep the two short phenyl side chains in a trans-orientation, only R^1S^2 and S^1R^2 isomers were investigated. In this series, analog 3-15 with Cl at R_1 and cyclohexane at region III (the R^1S^2 isomer in **Figure 3.9** c, shown as green sticks) presents a pose with slight H12 steric conflicts. The cyclopropyl retains the same position as the double bond in the parent compound. But when R_1 is replaced by OMe in the R^1S^2 isomer (analog 3-16 in **Figure 3.9** c, shown as yellow sticks), it retains almost the same pose as the double bond form. But if R_1 is replaced by Et (analog 3-17 shown in **Figure 3.9** d), the head group in the top-scoring pose approaches tunnel A instead of the H12 region. However, in this pose the hydroxyl group has steric conflicts with H12 instead. This pose is similar to the top-scoring pose of the R^1S^2 analog with Cl substitution at position R_1. With a larger group such as i-Pr at R_1, the S^1R^2 isomer is predicted to bind more favorably than the R^1S^2 isomer, but with less H12 steric conflict.

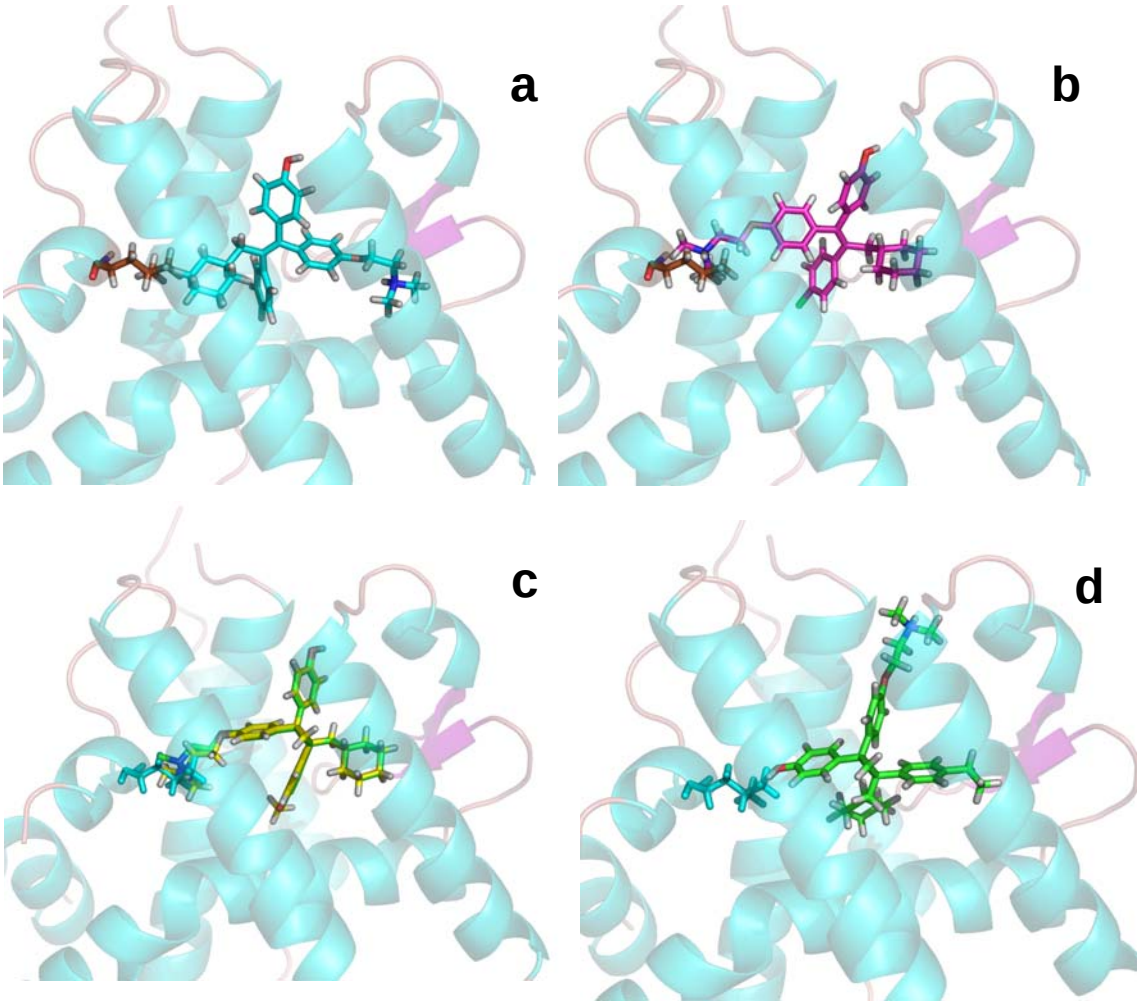


Figure 3.9 Tamoxifen analogs in LRH-1. a) analog 3-10, b) analog 3-11. analogs 3-12 to 3-14 bind with the same pose as 3-11. c) 3-15 (green stick) and 3-16 (yellow stick), d) 3-17

Furthermore, we tried to modify all three regions simultaneously: the head group was enlarged to a piperidine ring, region III was substituted by cyclohexane, and the R_1 group was replaced by different groups. Docking study indicates that when R_1 is Cl or OMe, none of them overlaps with H12 no matter where the nitrogen atom is positioned on the piperidine head. When R_1 is replaced by Et, the piperidine ring does have steric

conflicts with H12 if the piperidine nitrogen is at the *meta*- or joining position. With *i-Pr* at R_1 and the piperidine nitrogen at the *ortho*- or joining position, the analog also binds in an antagonist-like position but loses its rank as the most-favored. Replacing the double bond in these analogs with a cyclopropyl group alters their docking modes. All four analogs bind in antagonist-like poses with R^1S^2 stereochemistry. When R_1 is modified to OMe (analog 3-18), it binds with a perfect pose to sterically interfere with H12. The other analogs lack H12 conflicts with their head group, but a few of them present a pose with the head group buried within the **N1** pocket with their linkage portion showing slight H12 conflicts. These analogs are also predicted to function as LRH-1 antagonists, albeit with less activity than other candidates.

Table 3.3 Tamoxifen analogs which show conflict with H12.

	Head	R_1	Region III	Core /Stererochemistry
3-10			cyclohexyl	Double bond
3-11		Cl	cyclohexyl	Double bond
3-12		OH	cyclohexyl	Double bond
3-13		Et	cyclohexyl	Double bond
3-14		i-Pr	cyclohexyl	Double bond
3-15		Cl	cyclohexyl	Cyclopropene / RS
3-16		OMe	cyclohexyl	Cyclopropene / RS
3-17		Et	cyclohexyl	Cyclopropene /RS
3-18	Piperidine	OMe	cyclohexyl	Cyclopropene/RS

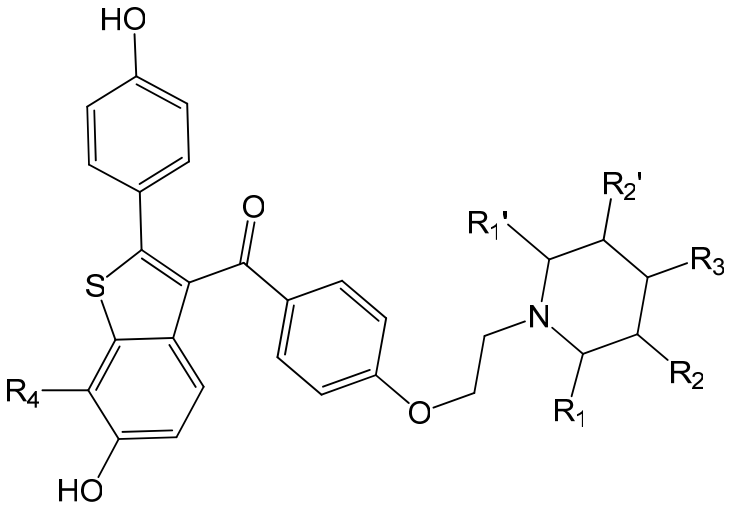
A structural feature important for potential LRH-1 antagonism was apparent from the computational analyses of tamoxifen analogs. As a T-shape compound, tamoxifen acts as antagonist in NRs with a small ligand binding pocket by preventing H12 from folding. However, in NRs with large ligand binding pockets like LHR-1, this kind of compound prefers different binding modes. In these large pockets, a cross-shaped compound binds more favorably and in a better position to disrupt H12. In this case, a methyl- cyclohexyl group added as the fourth side chain exhibits this cross-shape, with its fourth side chain extending close to H12 and potentially disrupting its agonist-related conformation.

3.5.3 Raloxifene Analogs

Raloxifene is an orally-active, selective estrogen receptor modulator (SERM) that has estrogenic actions on bone and anti-estrogenic actions on the uterus and breast. It is used in the prevention of osteoporosis in postmenopausal women. In 2006, the National Cancer Institute announced that raloxifene was as effective as tamoxifen in reducing the incidence of breast cancer in postmenopausal women at increased risk. A major adverse effect of tamoxifen is uterine cancer; raloxifene had fewer incidences of such cancers. Tamoxifen increases the risk of cataracts, but raloxifene does not. Both groups exhibited blood clots in veins and the lungs, but this side effect was more common with tamoxifen than raloxifene. On September 14, 2007, the U.S. Food and Drug Administration

announced approval of raloxifene for reducing the risk of invasive breast cancer in postmenopausal women with osteoporosis and in postmenopausal women at high risk for invasive breast cancer [17].

Raloxifene was prepared and docked in to LRH-1 as described previously. The docking results predict that the benzo-thiophene ring sits near tunnel A, the adjacent benzo-hydroxyl group points to helix 11 without steric conflicts with either helix 11 or 12, and the piperidine head locates between tunnel C and pocket **P1**. The head piperidine ring is close to Tyr431 and His390, suggesting that modifications at the *meta*- or *ortho*-positions would cause conflict with these two residues.



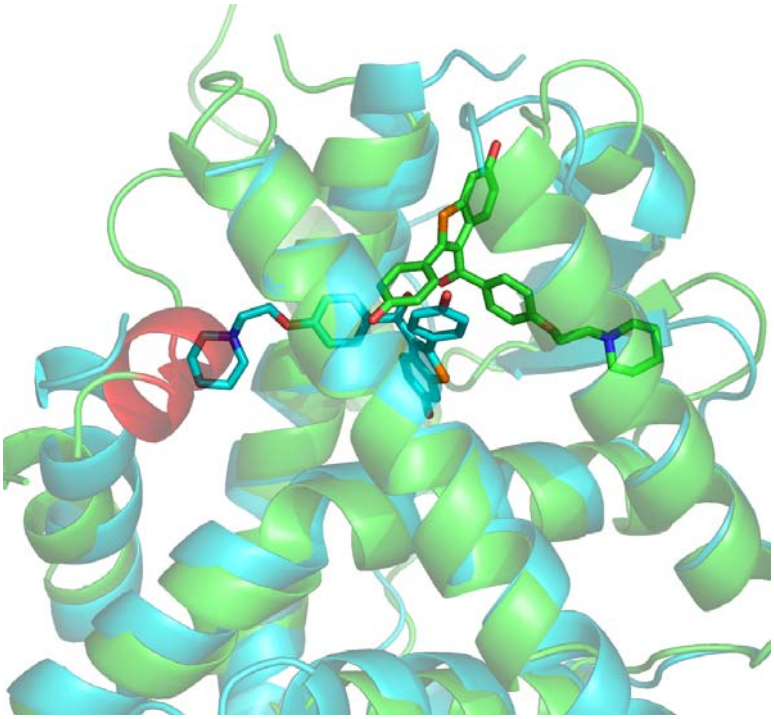


Figure 3.10 Raloxifene 2D structure and its binding pose in ER-α (cyan cartoon and stick) and LRH-1 (green cartoon and stick). Helix 12 of LRH-1 is colored red. Due to the steric conflicts of tamoxifen with H12 of ER-α, this portion of the protein is missing in the X-ray crystal structure.

These modifications were realized by introducing small hydrophobic and hydrophilic groups at the *ortho-*, *meta-*, or *para-* positions. Docking results indicate that single modifications at any of these positions directs the piperidine head to tunnel C, likely due to the positive charge on the piperidine nitrogen preferring a more polar region. One of the hydroxyl groups always locates to the **P1** pocket or near tunnel A to form a hydrogen bond with these polar residues. In these poses, no overlap between H12 and the raloxifene analogs is observed.

Since modification of any single position on the head group does not change the

binding pose of the analogs, a combination of modifications at multiple positions was performed on the head group. The resulting docking results show that only the analogs with a large hydrophobic group at the *ortho*-position and a small hydrophobic group at *meta*-position lead to an antagonist-like poses. For example, analogs 3-19 and 3-20 (**Figure 3.11** a), with OMe or Et at the *ortho*- position and Me at the *meta*- position, are two excellent candidates for disrupting H12. Substituting the methyl with a polar group causes the head group to relocate in tunnel C again. This result not only verifies the assumption that suitable modifications on the head group prevent it from locating in tunnel C, but also verifies that tunnel C is a polar region, causing polar modifications on the head group to prefer binding at this location.

The cross-shaped binding mode obtained from the tamoxifen analysis led us to attempt design of a cross-shaped raloxifene analog that can also act as an LRH-1 antagonist. Modification on the opposite side of the piperidine side chain was performed. Since the original side chain contains a phenyl ring, we added another phenyl ring at the opposite side of the piperidine side chain. The docking results indicate the attached phenyl ring experiences a slight steric conflict with the H12 residues. In this analog, the piperidine head is found in tunnel C. To increase interactions between the added phenyl ring and the H12 residues, we modified the phenyl ring further by adding a small group at the *para*- position (analog 3-21 and 3-22 in **Figure 3.11** b and c). Docking revealed that such a small group does have steric conflicts with H12. However, replacing the phenyl ring with a saturated ring causes the small groups on this ring to show less or no steric

conflicts with H12. With an unsaturated ring, the moiety is restricted to possible orientations directed at H12; with a saturated ring, more positions for the ring are possible due to bond rotations and ring flips. Hence, an aromatic ring with a small side chain at the para- position will lead raloxifene analogs to a promising LRH-1 antagonist.

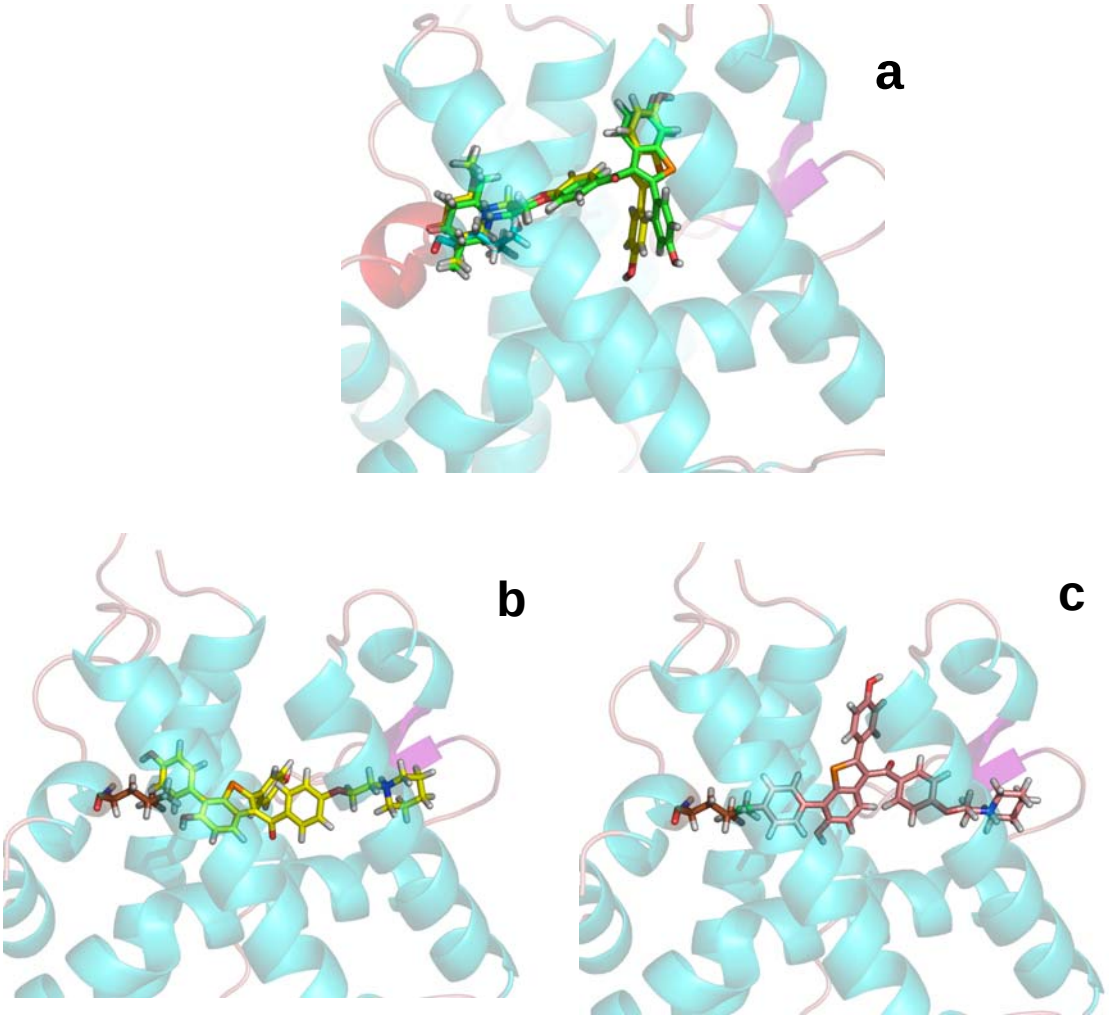


Figure 3.11 Raloxifene analogs which show steric conflicts with H12. a) analog 3-19 and 3-20, b) analog 3-21, c) analog 3-22.

Table 3.4 Raloxfene analogs which show steric conflicts with H12.

	R1	R2	R4
3-19	Et	Me	
3-20	Ome	Me	
3-21			PhOH
3-22			PhCl

3.5.4 Steroid Analogs

A steroid analog, MFA-1, co-crystallized with FXR was chosen as the starting design template (PDB code 3BEJ). This steroid was obtained by attaching a benzohydroxyl side chain on ring D and a charged carboxyl group on ring A. MFA-1 is a high–affinity ligand for FXR. Titration studies demonstrated that the compound has an EC_{50} of 16.9 nM in the coactivator recruitment assay which is nearly 500-fold more potent than the highest-affinity naturally occurring bile acid agonist, chenodeoxycholic acid.

When MFA-1 was prepared and docked into the LRH-1 ligand binding domain, its carboxyl group located at the bottom of the pocket, the two methyl groups resided between rings A and B, rings C and D pointed to helix 11, the benzohydroxyl positioned in tunnel A, and the hydroxyl oxygen formed hydrogen bonds with Gln419 and Ile415. With this pose, the hydrogen atoms at R_1 or R_2 point at H12. Moieties attached at these

two positions may have steric conflicts with H12.

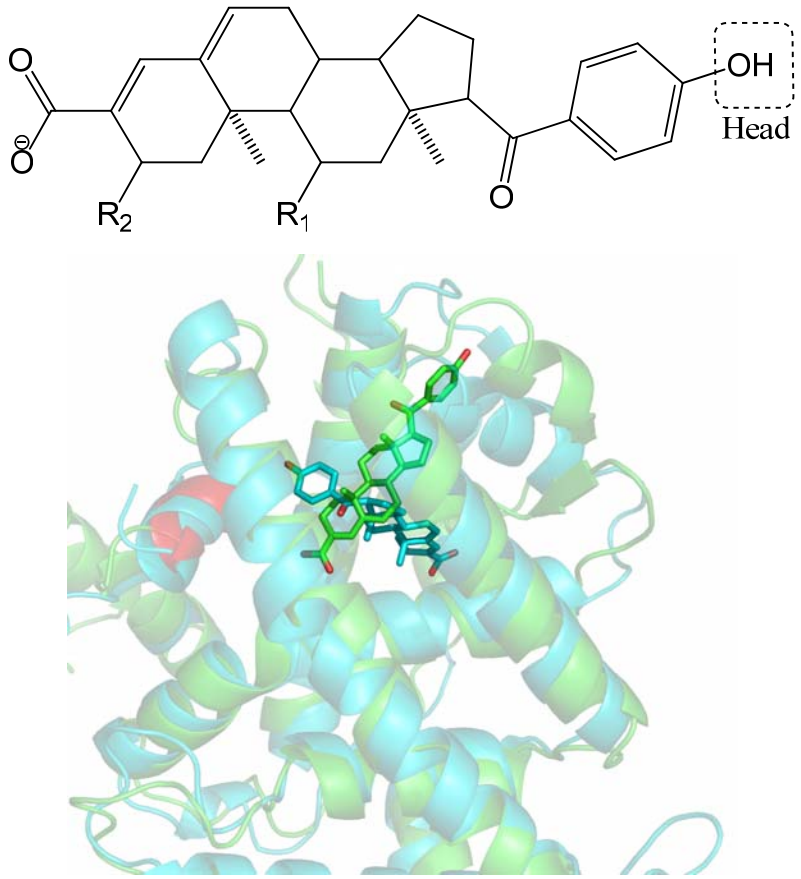


Figure 3.12 The steroid analog in the bile acid receptor (cyan cartoon and stick) and LRH-1 (green cartoon and stick). The red helix is helix 12 of LRH-1. Since the steroid analog acts as agonist in bile acid receptor, its H12 presents an active conformation in the crystal structure.

The docking results of compounds with R_1 modifications reveal that both lipophilic and hydrophilic groups alter the steroid's binding pose. In the new poses, the benzohydroxyl always lies on the bottom of the pocket and the carboxyl group is close to tunnel A. The group at R_1 always points to the opposite side of H12 and shows no steric

conflicts with any residues in the receptor. When superimposed with the original compound, the R1-modified analogs always have steric conflicts with Helix 3 and therefore cannot maintain the desired binding pose.

R_2 modifications do not cause the same problems as R_1 modifications. All the analogs with R_2 modifications retain poses similar to the original compound's pose. The groups applied on R_2 extend close to H12 and have steric conflicts with Leu532 in several cases. Based on these facts, modifications at R_2 lead to more promising LRH-1 antagonists than R_1 modifications. The docking procedure also revealed that the larger the group is at R_2, the more it overlaps with H12. Therefore, six-membered rings were added at R_2 and evaluated.

To avoid steric conflicts between the core structure and the new ring at R_2, a linkage atom was added between them. The results indicate when the linkage atom is a polar atom such as O (analogs **3-23** to **3-26**) and S (analogs **3-31** to **3-34)**, the new analogs bind the same way as the original compound, with the added six-membered ring causing steric conflicts with H12 (see **Figure 3.13** Top). When the linkage atom is carbon (analogs **3-27** to **3-30**), saturated six-membered rings present a pose similar to the original steroid analog's pose and show steric conflicts with H12. However, an aromatic ring at the same position locates in the small hydrophobic pocket N1 (see **Figure 3.13** Bottom). The rearrangement is caused by the polar interaction between the ligand's carboxyl group and the **P1** pocket. For the cyclohexyl analog, the ring does not fit in the small N1 group, and the hydroxyl group forms hydrogen bonds with tunnel A residues.

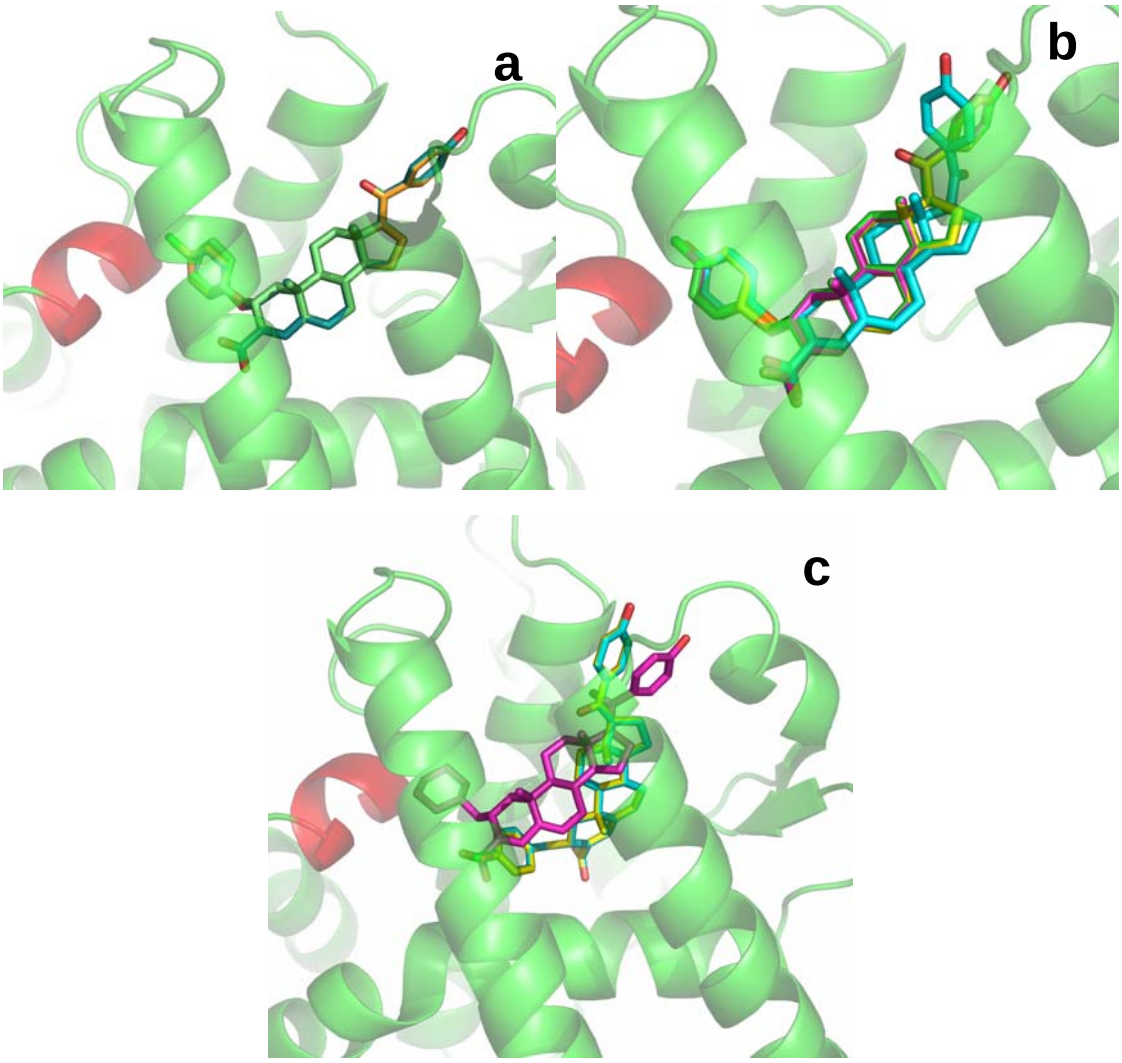


Figure 3.13 Steroid analogs with ring structures on R_2. The linkages are O (top left), S (top right), and C (bottom). In the C-series, only the saturated ring analog gives a promising pose (pink stick).

To determine the importance of the hydrogen bond between the D ring hydroxyl and the tunnel A residues, modifications on the hydroxyl group were performed. A methyl group was placed at the R_2 position in this analog series. Replacing the hydroxyl group by methyl, Cl or OMe changes the binding pose of the steroid analogs: the modified head no longer engages in hydrogen bond interactions with tunnel A but binds near H12,

although with little overlap with H12. An amino group at this position also alters the analog's conformation: the carboxyl group points near H12 and overlaps with H12 residues. The amino group binds at tunnel C instead of tunnel A, and the whole molecule lies parallel to helix 5. This result indicates that only a suitable polar group which forms a hydrogen bond with tunnel A residues can help stabilize an original-like docking pose. Small polar groups were applied to the head group, with an O-Ph group placed at R_2. The resulting docked poses showed that these analogs retain an original-like pose with CH_2OH and NH_2 on the head group.

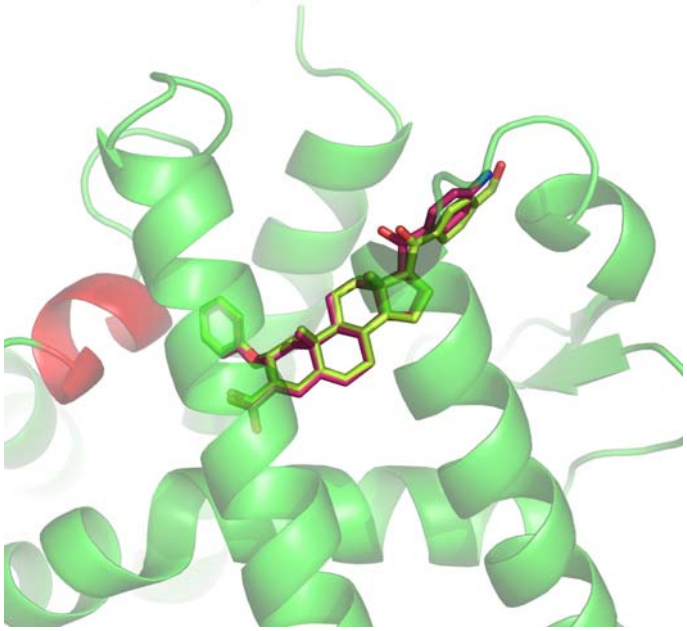


Figure 3.14 Steroid analogs with head group modifications. Both the amino (red stick) and the methyl-hydroxyl (green stick) head form hydrogen bonds with residues around tunnel A.

The docking results indicated that if the steroid analogs can maintain an antagonist-like binding pose with other modifications, a six-membered ring attached to

R_2 would overlap with H12. This docking study also indicated that a polar linkage between R_2 and the 6-membered ring, combined with a polar head group, causes the analog to retain an antagonist-like binding pose.

Table 3.5 Steroid analogs which show steric conflicts with H12.

	linkage	Ring	Head		linkage	Ring	Head
3-23	O	Ph		3-31	S	Ph	
3-24	O	PhOH		3-32	S	PhOH	
3-25	O	PHCl		3-33	S	PHCl	
3-26	O	Cyclohexane		3-34	S	Cyclohexane	
3-27	CH_2	Ph					
3-28	CH_2	PhOH		3-35	O	Ph	NH_2
3-29	CH_2	PHCl		3-36	O	Ph	CH_2OH
3-30	CH_2	Cyclohexane					

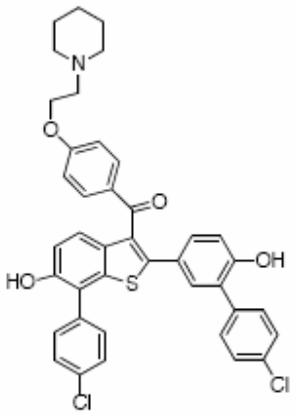
3.6 Biological Evaluation

Thus far, several compounds based on Raloxifene and Tamoxifen have been evaluated in a cell-based assay. The Raloxifene analogues tested so far either lack activity or actually activate LRH-1. Fortunately, several tamoxifen analogs have shown modest

inhibition, although solubility issues have plagued this compound series. Another frustrating aspect of this cell-based assay is that some of the compounds cause cell death when their concentrations are only slightly increased. An *in vitro* assay is currently being established that should give more focused data on the antagonism of LRH-1.

Table 3.6 Cell based activities of several LRH-1 antagonist candidates

Compound	**Class**	**MW**	**Solubility**	**Activity relative to Vehicle (DMSO)**
3-15	Tamoxifen analogue	490.0760	Difficult to dissolve	19% inhibition 1uM
3-16	Tamoxifen analogue	485.6570	Difficult to dissolve	8% inhibition 1uM
3-17	Tamoxifen analogue	483.6841	Difficult to dissolve	16% inhibition at 10uM
3-22	Raloxifene analogue	584.1243	soluble	20% activation at 1uM
JR 476(2)	Raloxifene analogue	694.6653	soluble	7% inhibition at 1uM



JR476-2

3.7 Conclusions

In this research, a structure based drug discovery method was utilized to investigate the structure of LRH-1 LBD. Based on structural comparisons of LRH-1's LBD with other NRs, a series of LRH-1 antagonist candidates were designed from lead compounds for other NRs. Several of them were tested in a cell-based assay and demonstrated modest antagonism, but with poor solubility and potential toxicity. These results indicate that the LRH-1 antagonist candidates need further modification to improve their activities and ADMET profiles.

From the comparison of LRH-1's X-ray crystal structure with other NRs, we found, as in other NRs, the ligand binding pocket of LRH-1 is mainly hydrophobic. Thus, a compound able to bind in this region should be mainly hydrophobic. The hydrophilic region located near the distal of H5 (**P1**) and around the entrance of tunnel A provides the necessary hydrogen bond donors or polar groups which may allow agonists and antagonists to enter the ligand binding pocket and maintain their binding pose. The LRH-1 ligand-binding pocket also differs from other NRs. Its volume is larger than most other NRs. The tunnels **B** and **C** provide 200-300 $Å^3$ additional volume to the ligand binding pocket. This extra space increases the capability of this pocket to accept large molecules. Small compounds are always capable of finding a suitable position in an agonist-active ligand-binding pocket of LRH-1. This is why the H12 blockers of other NRs increase rather than block the activity of LRH-1. The molecular weights for most of

the antagonist candidates we designed are above 500 Daltons. Such large molecules with hydrophobic fragments which fit for the LRH-1 binding pocket are always difficult to dissolve in water. However, we cannot say it is impossible for a small compound to have antagonist properties against LRH-1. There may be other mechanisms that influence LRH-1 activity apart from the H12 folding hypothesis. Thus far, however, our compound design based on H12 disruption has demonstrated the first-in-class antagonists for LRH-1.

Chapter 4 Truncated Taxane

4.1 Tubulin and Taxane Analogs

Tubulin belongs to a small family of globular proteins, and the most common members of this family are α-tubulin and β-tubulin. Unpolymerised tubulin exists as a tight α, β-tubulin heterodimer with two bound molecules of GTP, one on the dimer interface and the other in the end of the β-unit (**Figure 4.1**). The dimers assemble into microtubules. The latter make up the mitotic spindle and are in a particularly delicate state of balance between assembly and disassembly into their constituent subunits. Both the formation of the spindle and the movement of the two sets of chromosomes to opposite spindle poles depend carefully on coordinated extension and shrinkage of microtubules within the spindle. The central role of microtubules in the process of separating duplicated chromosomes during cell division makes them an important target for anticancer drugs. Paclitaxel (PTX) structure (shown in **Figure 4.2 a**) was discovered in a National Cancer Institute program at the Research Triangle Institute in 1967 when Monroe E. Wall and Mansukh C. Wani isolated it from the bark of the Pacific yew tree. PTX and its semi-synthetic analog docetaxel (DTX), both of which belong to taxanes, are clinically important antitumor agents [1]. Previous work reveals that PTX binds to a pocket on β-tubulin in a conformation described as a T-shape and thereby prevents

microtubule disassembly with high potency [2]. A few other microtubule stabilizing agents such as epothilone, eleutherobin, discodermolide and dictyostatin have been found and verified to bind to the taxane binding pocket.

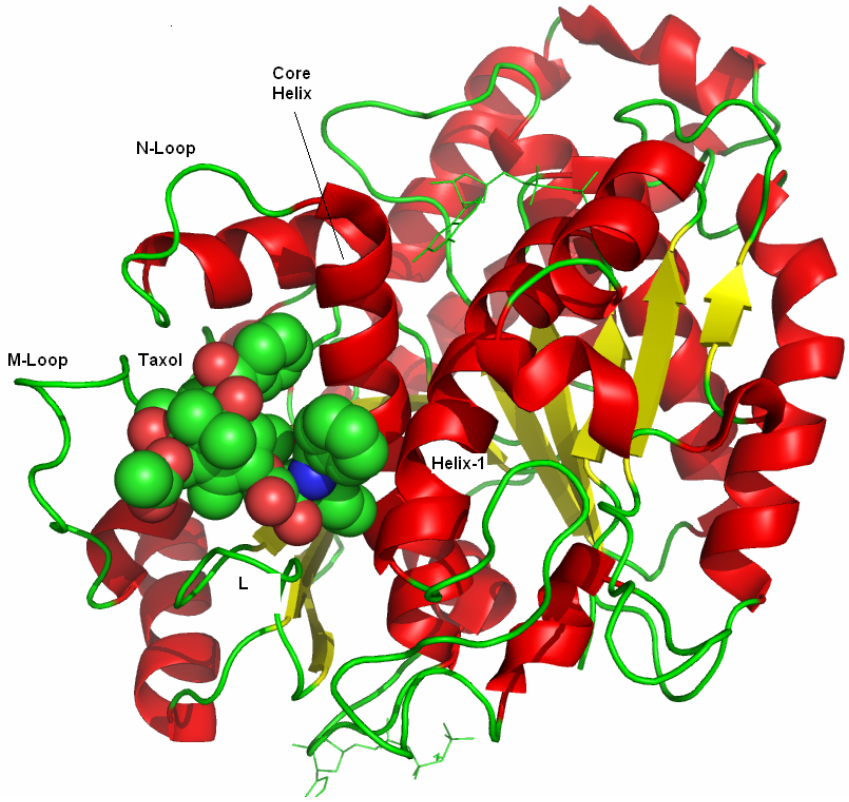


Figure 4.1 Atomic model of wild type β-tubulin (cartoon) complexes with PTX (ball).

Recently, a new model of β-tubulin optimized specifically for bridged taxane derivatives was constructed by Drs. Geballe and Alcaraz in our laboratory in collaboration with Drs. Kingston and Bane [3]. Among others, the bridged taxane derivative **282** (structure shown in **Figure 4.2 b**) was docked into the taxane site. A 10 ns MD simulation was performed to understand the changes in the protein upon **282** binding.

The distance between C-4 and C-13 in **282** is about 1 Å longer than PTX. With **282** in the taxane site, the new model predicts that the M-loop recedes from Helix-1. It is believed that this change in conformation leads to a 3-fold decrease in the critical concentration of GDP-tubulin relative to PTX and induces significant microtubule assembly in Mg^{2+} free buffer. They propose that the rigid bridged taxane operates within the β-tubulin taxane binding site by holding the M-loop in an open orientation with a conformation that strengthens the contact between adjacent microtubule protofilaments to an extent unachievable with the more flexible PTX ligand.

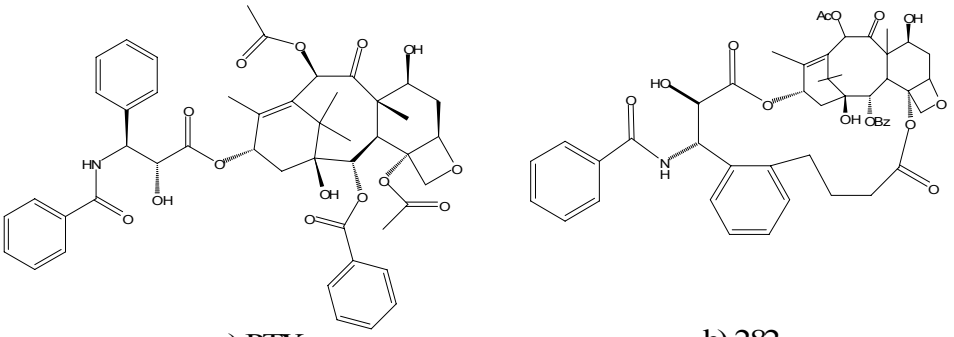
Previous research reveals PTX was effective against a variety of murine cancer models and human xenograft tumors, especially advanced human carcinomas refractory to conventional chemotherapy [4-6]. Now, PTX has emerged as one of the most active anticancer agents in clinic for the therapy of ovarian, breast and non-small cell lung cancer, AIDS-related Kaposi's sarcoma, bladder, prostate, esophageal, head and neck, cervical and endometrial cancers [5-11]. However, the development of PTX is tightly restricted, because it is a natural product, which accumulates in the bark of yew trees (*Taxus*). Yew trees grow extremely slow and the yields of PTX are too low to supply sufficient quantities for clinical use. Furthermore, the extraction of PTX from the bark of yew trees is not sustainable. Consequently, alternative strategies to guarantee its supply have extensively been investigated, including the production of taxanes in plant suspension cultures [12-14], induction of biosynthesis [15], or in fungal cultures [16], furthermore, total synthesis [17-19] and semi-synthesis [20,21]. It is a problem that most

these methods are not suited for large-scale production, because of the high production cost and low product yield. The synthesis of structurally simplified PTX analogs with similar or even improved activities is thus an attractive goal. Several attempts toward this goal have been reported, however, the analogs resulting from these studies were either inactive or significantly less active than the parent compound. An important consequence of the updated bridged analog binding model is the potential for designing other highly active analogs with simplified structure and reduced resistance profiles. Such molecules need to incorporate three characteristics: 1) molecular geometries that precisely span the space between Helix-1 and the M-loop; 2) molecular rigidity to prevent the M-loop from folding out of the inter-protofilament region; and 3) sufficient functionality to result in effective binding within the β-tubulin taxane cavity.

Two different ways were attempted to simplify the structure of the tubulin stabilizer: by replacing the baccatin core with a more easily synthesized multi-ring core; or searching for steroid analogs which fully mimic the binding properties of taxane.

Effect of Temperature and Taxane on the critical concentration of GDP-tubulin

Ligand	C_{cr}(μM)		
	12ºC	25ºC	37ºC
PTX	7.3±0.4	3.6±0.1	1.8±0.2
282	2.0±0.5	0.33±0.06	0.15±0.02



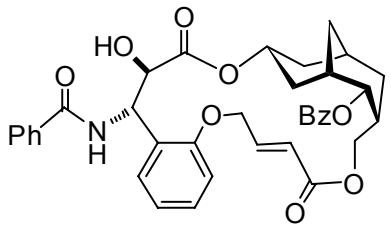
a) PTX b) 282

Figure 4.2 2D-structures of PTX and **282** and their activities.

4.2 Simplified PTX

4.2.1 Replacement of the Baccatin Core.

Gueritte-Voegelein, *et al* had previously designed and synthesized the simplified PTX analog (**4-1**) where a 2,2,1-bicyclononane moiety was used as a structurally simpler surrogate of the baccatin core [22].



4-1

The logic behind this design was that the northern hemisphere of the PTX structure is less crucial to its bioactivity based on its SAR, and a structure with the T-taxol pharmacophore could be achieved by installation of the PTX side chain to a simple

bridged bicyclic skeleton and subsequent constraint of the conformation with a suitable bridge. The resulting compound **4-1** demonstrated moderate activity in cytotoxicity assays. However, it was found to be highly insoluble in water, which made its biological evaluation very difficult, and it had only modest anti-proliferative and tubulin assembly inhibitory activity.

In our first approach to develop a new generation of water-soluble simplified PTX analogs, an aza-tricyclic moiety (structure see **Figure 4.3**) was designed to mimic the baccatin core of PTX. The basic tertiary nitrogen was included to increase water-solubility. Importantly, the aza-tricyclic moiety is functionalized with three hydroxyl groups where the key side chains can be attached. The final construct can be constrained to the T-taxol conformation by the selection of an appropriate length for the bridge linking the side chain and an appropriate conformation of the tri-cyclic core. The new tri-cyclic core bears three stereogenic centers which lead to eight possible stereo isomers.

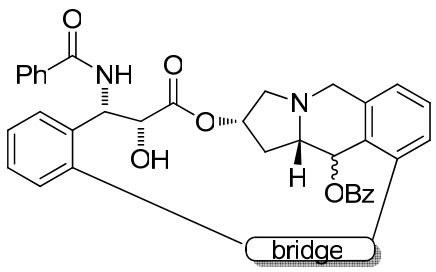
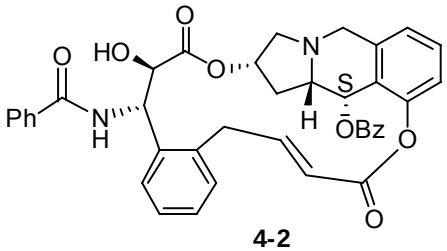
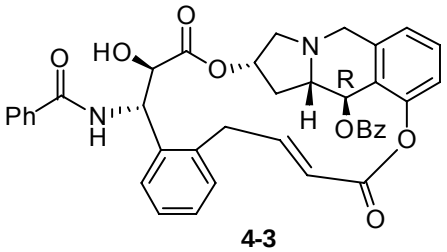


Figure 4.3 General structure of second generation T-taxol mimics

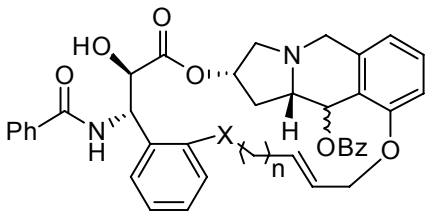
Molecular modeling studies were performed to determine which of the eight possible stereo isomers served as the best replacement of the baccatin structure. Thus, a conformational search was carried out for each of the eight stereo isomers with Macromodel 9.7 [23] using the OPLS-2005/GBSA/H_2O force field protocol. The corresponding eight global minima were fitted with appropriate side chains and various bridges. Each structure was optimized and docked into the PTX binding site of the updated conformation of tubulin. This structure was obtained from the equilibrated tubulin structure in the last 1 ns of the 10 ns molecular dynamics trajectory of the highly active bridged taxane **282** in tubulin [3]. Docking experiments in this context were meant to mimic the binding pose of **282**. Each docking exercise returned a variety of poses. Those that best mimicked the shape and extension of **282** were energy-rescored with MM-GBSA and compared with the original binding pose of **282** in the M-loop reorganized tubulin. **Figure 4.4** shows the two lowest energy structures (**4-2** and **4-3**) superimposed with **282** in a 3-D representation of the tubulin binding site. Compounds **4-2** and **4-3** are derived from two different stereo isomers of the aza-tricyclic moiety, with *S* and *R* configurations at C-10, respectively.



4-2



4-3



4-4 X = O or bond
n = 1-3

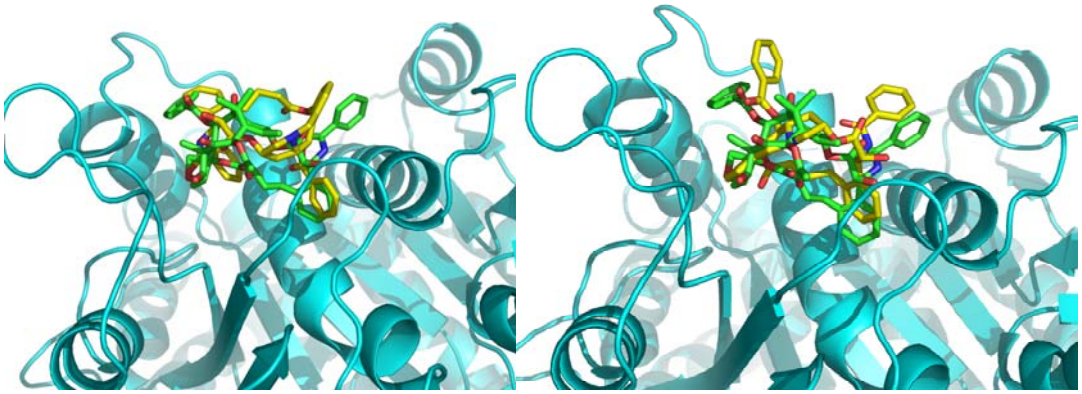
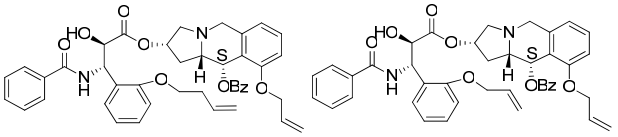


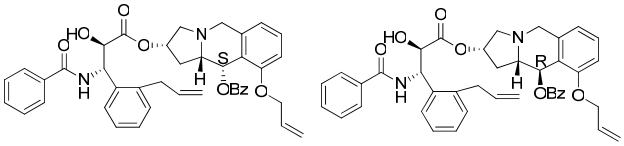
Figure 4.4 Left: low energy poses of compounds of **282** (in green) and **4-2** (in yellow) in the PTX tubulin binding site. Right: low energy poses of compounds **282** (in green) and **4-3** (in yellow) in the PTX tubulin binding site.

Encouraged by these model comparisons, we elected to prepare analogs arising from both stereo isomers and with open and bridged forms with different bridge lengths. For easier synthetic manipulation, the unsaturated ester tether was modified to that of **4-4**, incorporating ether rather than the ester linkage as in **4-2** and **4-3**. Finally, 7 aza-tricyclic

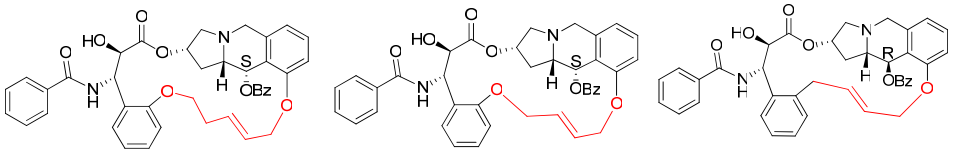
replacement taxane analogs were obtained including four open form analogs **4-5, 4-6, 4-7** and **4-8**, and three bridged form analogs **4-9, 4-10** and **4-11**. **4-5/4-9, 4-6/4-10** and **4-8/4-11** are the corresponding open and closed structures.



4-5 **4-6**



4-7 **4-8**



4-9 **4-10** **4-11**

4.2.2 Biological Evaluation

The open-chain analogs **4-5, 4-6, 4-7, 4-8** and bridged analogs **4-9, 4-10, 4-11** were evaluated for their anti- proliferative activities against the A2780 cell line. The results are summarized in **Table 4.1**. The simplified analogs showed moderate to good cytotoxicity. The bridged analogs **4-9, 4-10** and **4-11** exhibited slightly better activity than their open-chain counterparts **4-5**, **4-6** and **4-8** respectively. Compound **4-8** is over 4 times less active than the corresponding bridged analog **4-11**. However, all analogs are significantly

less active than the parent compound **282** in this bioassay.

Table 4.1 Anti- proliferative activities ($IC_{50,}$ μM)[a] of PTX and the simplified PTX.

Compound #	4-5	4-9	4-6	4-10	4-7	4-11	4-8	282
A2780[a]	5.7	4.4	8.3	6.2	4.8	5.1	23.8	0.015

[a] human ovarian cancer cell line.

4.2.3 Computational Evaluation

The structures of the seven taxane analogs were constructed in Maestro. The molecular frameworks are more flexible than those of **4-2** and **4-3** either due to the lack of a bridge (the open-form analogs **4-5 to 4-8**) or the absence of a planar carbonyl group on the bridge (the bridged form analogs **4-9 to 4-11**). To examine the potential binding poses of these taxane analogs, preliminary conformational analyses were performed with the OPLS-2005/ GBSA/ H_2O protocol within MacroModel 9.7. The highly flexible open form analogs **4-5, 4-6, 4-7** and **4-8** provide 4 to 10 times more conformers than the bridged form analogs **4-9, 4-10** and **4-11**. This flexibility leads to 900 to 4000 conformers of each open form compound which is difficult to handle in computation. The top 20 conformers of each analog were prepared and docked into the taxane site with GLIDE 5.5 [6] and rescored by PRIME/MM-GBSA 2.1 [7].

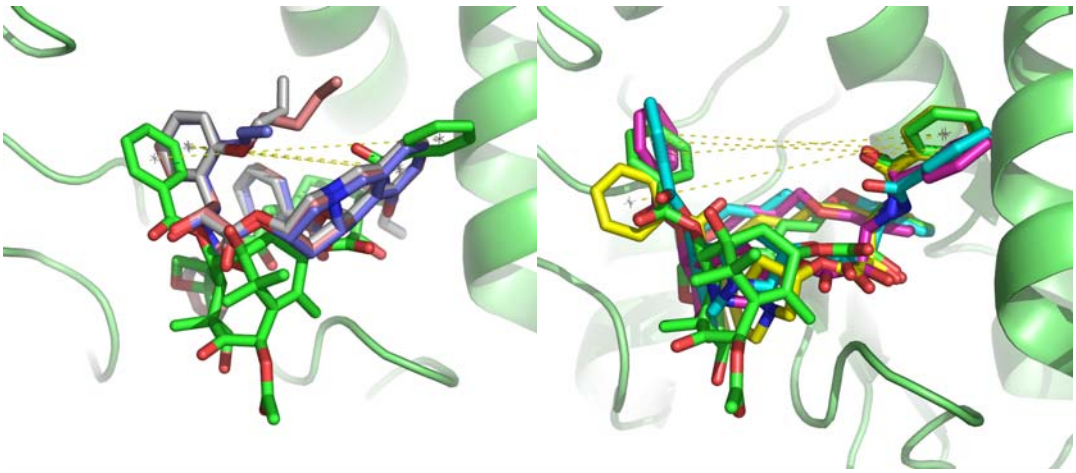
The docking results summarized in **Table 4.2** suggest the bridged analogs give similar binding energies to **282**, and a little bit lower than the open form analogs except

4-8. Investigation of the binding pose demonstrates that the bridged analogs bind into the taxane binding site similar to **282.** Namely, the two phenyl side chains on the bridge and the aza-tricyclic moiety are found in a similar position as the corresponding phenyl on **282**. The open form analogs give different binding poses. In the docking result of **4-5, 4-6** and **4-7**, the two phenyl side chains locate at the bottom of the pocket, while the phenyls on the open bridge and the aza-tricyclic moiety occupy the top region. Compound **4-8** presents a unique pose which is perpendicular to **282** (see **Figure 4.5**)**.**

The interatomic central distance between the two phenyl rings on the side chains of the bridge and the baccatin core **(S:** central distance between the two phenyl ring on the top of the pocket**)** is an important feature for compound **282** capable of stabilizing the tubulin dimmer. This is also the feature which we tried to maintain in the design of taxane analogs, along with the proper binding pose. Compounds **4-2** and **4-3** present similar binding pose as **282**. However, their **S** values are about 9.0 Å as compared with 11.8 Å for compound **282,** which is the largest **S** in the eight stereo isomers. This short **S** distance indicates compounds **4-2** and **4-3** should be less active than **282**. In the synthesized compounds, the conformers of the bridged and the open-form analogs present different **S** values. For the bridged form analogs**,** the S values are greater than 12 Å in their lowest energy conformations. However, in the low-energy open structures **4-5**，**4-6** and **4-7**, the two side chains are always close and parallel to each other resulting in **S <** 9Å (**Table 4.2**), and in the low-energy structure **4-9**, no phenyl ring structure appears in the corresponding region.

Table 4.2 MM-GBSA (kcal/mol) estimated relative binding energies and **S** values (Å) for open analogs **4-5 to 4-8** and bridged structures **4-9 to 4-11.**

analog	**4-5**	**4-6**	**4-7**	**4-8**	**4-9**	**4-10**	**4-11**	**282**
CS-comformer #	1598	896	1346	4105	466	236	148	200
S (Å)	9.0	8.9	9.0	-	12.3	12.5	12.6	11.8
$E_{binding}$(Kcal/mol)	-91.6	-92.3	-94.5	-98.1	-101.9	-99.2	-93.2	-98.9



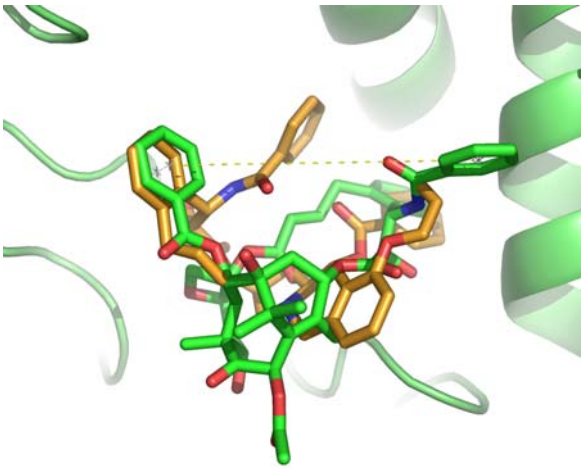


Figure 4.5 *Top left*: low energy poses of compounds **282** (in green) and **4-5**（in pink）**, 4-6** (in grey) and **4-7** (in blue) in **282** tubulin binding site. *Top right*: low energy poses of compounds **282** (in green) and **4-9** (in cyan)，**4-10** (in magenta)，**4-11** (in yellow) in **282** tubulin binding site. The **S** distance is marked by yellow dash line. *Bottom*: low energy poses of compounds **282** (in green) and **4-8** (in orange) in the PTX tubulin binding site.

4.2.4 Conclusions

Seven simplified taxane analogs containing an aza-tricyclic moiety have been synthesized. The bridged analogs **4-9**, **4-10** and **4-11**, designed to adopt the T-taxol conformation and the bridged geometry of **282**, show slightly better activities compared to their open-chain counterparts. The lack of anti-tubulin action relative to Taxol can, in part, be attributed to the complex conformational profile of the macro-cyclic rings relative to the parent molecule and to its bridged analogs. Analogs with modified structures that can better mimic the T-taxol conformation based on modeling studies will

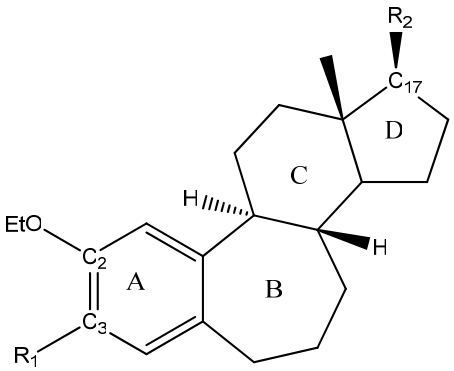
be synthesized in due course.

4.3 Truncated Taxanes

4.3.1 Steroid Analogs

Certain steroids have been reported as microtubule stabilizing agents that show competition with PTX and other taxane-site agents [26]. The 4-memberd ring moiety in the steroid analog is rigid and can be considered to mimic the baccatin core of the taxane derivatives. If a steroid analog binds into the taxane site on tubulin with an appropriate pose, analogous to **282**, in principle, it can span the space between Helix-1 and M-loop and prevent the M-loop from folding out of the inter-protofilament regions. In 2000, several steroid analogs shown in **Figure 4.6** were reported to have the ability to stabilize microtubules with mild activities [27]. **4-12** is a diacetate analog. In **4-13** and **4-14**, one of the acetate groups is replaced by a hydroxyl group (at C-3 in **4-13** and C-17 in **4-14**). This structure difference leads to a 50-fold difference in EC_{50} concentration (the compound concentrations when the enhancement of assembly rate yields to 50% of the rate of 100μM compound **4-12**). We hypothesized that modification at position C-17 would alter the binding pose of the steroid analogs and thereby influence the activities associated with microtubule polymerization. In an attempt to validate the hypothesis and to explain the unusual observations regarding the effects of steroids on tubulin polymerization, a

computational approach was undertaken.



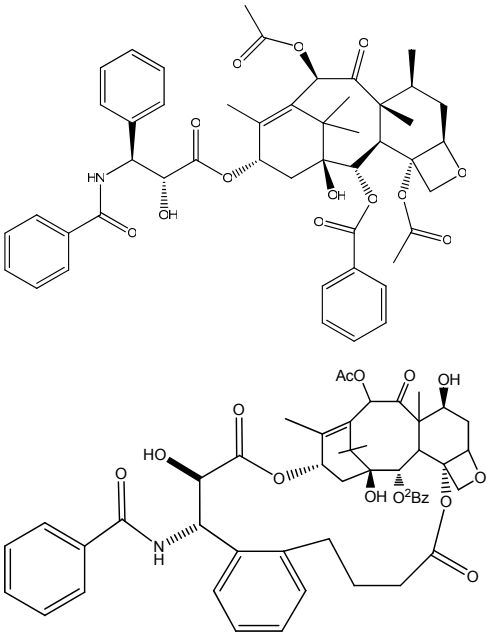
	R_1	R_2	Asembly
PTX			0.83
4-12	Ac	Ac	5.5
4-13	OH	Ac	8.2
4-14	Ac	OH	360

Figure 4.6 Reported steroid analogs which show effect on preventing microtubule disassemble.

4.3.2 Tubulin Binding Site

Comparison of the wild type tubulin-taxol binding site and the tubulin-**282** binding site shown in **Figure 4.7** reveals a shallower ligand binding pocket induced upon **282** binding. This is especially noticeable at the two phenyl arms region where the change is caused by side chain movement of Arg278. In this area, due to the receding of the M-loop and the movement of the His229 imidazole ring, the pocket displays a bowl-like shape. As a result, when a molecule binds to this pocket, there are more possible binding pose than the wild type binding pocket. Under this circumstance, **282** binds into the reshaped tubulin binding site with a pose perpendicular to Helix-1. The C-2 benzoyl is close to the M-loop, and interacts with the Arg276 side chain in what appears to be a cation-π

interaction. The C-13 side chain is close to Helix-1, and the C-12 hydroxyl group forms a hydrogen bond with the Asp26 side chain. In this pose, **282** occupies the entire space between Helix-1 and the M-loop. In order to mimic the **282** binding pose, the steroid analogs would be expected to orient the rigid 4-membered ring structure perpendicular to Helix-1 so that they also can occupy the space between the M-loop and Helix-1. In order to investigate whether the steroids fit into the reshaped binding pocket with the optimal pose (as described above) or not, all the steroid analogs were docked into the wild type PTX and the **282** sites with several docking methods and analyzed for the ligand orientations within the docking poses.



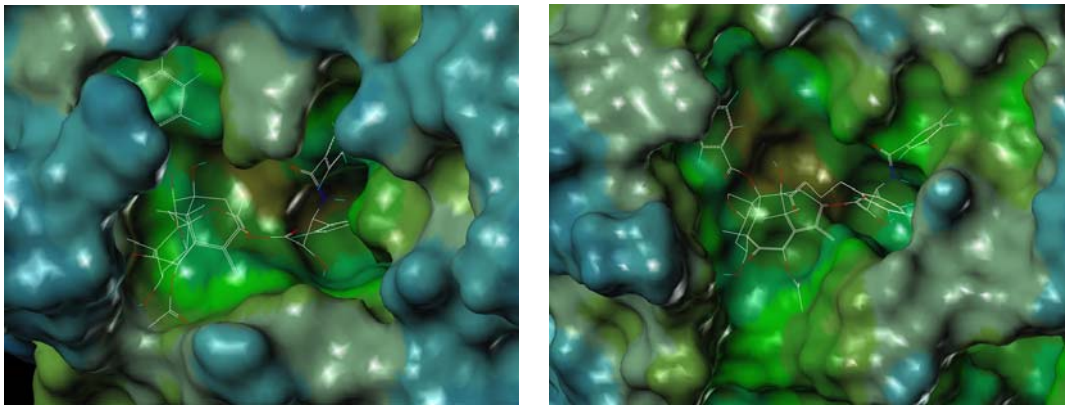


Figure 4.7 *Top*: 2D structures of PTX and **282** and superimposed wild tubulin-taxol (green cartoon and cyan stick from PDB: 1JFF) and reshaped tubulin-**282** (grey cartoon and yellow stick from MD simulation) binding pocket. The distance between C-4 and C-13 in 282 is 1 A longer than PTX. *Bottom left:* the surface of tubulin PTX binding site. *Bottom right:* the surface of tubulin **282** binding sites.

4.3.3 Ligands Preparation

Due to the lack of a crystal structure of any steroid bound to β-tubulin, a conformational search was necessary in the preparation of the steroid analogs **4-12**, **4-13** and **4-14**. The conformational search was performed by Macromodel 9.5 with the mixed torsional/ Low-model/ sampling method. Water properties were selected for the solvation continuum model. In order to obtain a complete conformational search, 10,000 structure generation and geometry optimization steps were performed on each analog. Hundreds of conformers were generated from the conformational search of each steroid analog. Comparing the conformers from the search results, most of them only differ in the rotations of the side chains at C-2, C-3, and C-17. It's unnecessary and too

time-consuming to investigate all these conformers. In the docking process, a quick conformational search is performed on the ligand's rotatable bonds. Therefore, most of the conformers obtained from the conformational search were redundant for the docking process. Redundant Conformer Elimination (RCE) was performed on the conformational search result of each analog through MacroModel 9.5. All the heavy atoms on the multi ring structure were selected as comparison atoms. That means only the conformers with different multi ring conformations would be recorded into the new conformer pool. Conformers only differing in side chain rotations are considered identical and eliminated except one. The RCE result showed only 4 conformers of each analog remained. The 4 conformers were prepared and utilized in the docking procedure.

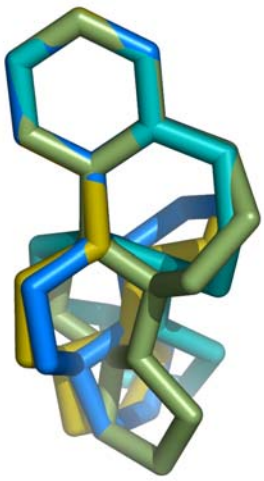


Figure 4.8 Redundant Conformer Elimination (RCE) result of the core region of the steroid analogs. For each analog, only 4 conformers of the steroid core were recorded.

4.3.4 Tubulin-Taxol Site Docking

Both the tubulin-taxol obtained from crystal structure (PDB:1JFF) and tubulin-**282** (obtained from MD simulation) complexes were prepared in Maestro using the ProteinPrep tool. The prepared steroid analogs were docked into the tubulin-taxol binding site with Glide5.5. The results indicate that all the compounds mimic the pose of PTX in the tubulin-taxol binding site with the absence of the C-13 side chain on PTX. All the poses obtained from the docking are partially parallel to Helix-1.

The lowest energy poses of **4-12** and **4-13** display similar conformations near the M-loop side. The C-3 and C-2 side chains sit between the N-loop and the M-loop similarly to the OBz group on PTX. The ethoxyl group at C-2 forms a hydrogen bond with Arg276 on the M-loop. Its carbonyl group on ring B forms a hydrogen bond with Gln280. Due to the ring flip, the length of this hydrogen bond in **4-12** is slightly longer than that in **4-13** (by 0.1Å). In both **4-12** and **4-13**, the C-17 acetate groups locate in a deep hydrophobic sub-site at the bottom of the tubulin-taxol binding site.

Compound **4-14** gives a different binding pose from **4-12** and **4-13**. The C-17 hydroxyl group on **4-14** takes the place of the C-3 acetate in **4-12**. Due to the short length of this side chain, it cannot reach the N-loop region. Instead, it bears close to the L-loop and hydrogen bond interacts with Thr274. On the Helix-1 side, the C-2 ethoxy group is predicted to locate in the same sub-site as the one occupied by the C-17 acetate group in **4-12**. Its C-3 acetate group sits between the L-loop and Helix-1. MMGBSA results indicate that the binding energies of **4-12** and **4-13** are about 10 KJ/Mol lower than that

of **4-14**.

More stringent docking methods, Glide XP docking and Induce Fit Docking (IFD), were also performed on this system. As discussed above, all three docking methods provided similar poses for **4-12** which are partially parallel to Helix-1. IFD gives a similar pose to Glide docking except the movement of the C-17 acetate to the bottom of the hydrophobic sub-site at the bottom of the binding pocket. For **4-13**, Glide XP docking results exhibited almost the same result as in Glide SP docking. IFD, on the other hand, provided an upside-down pose with the whole molecule more deeply imbedded into the tubulin-taxol binding site than in other results, placing it closer to the hydrophobic sub-site at the bottom of the pocket. This pose results in an unoccupied region between the N-loop and M-loop, and the hydrogen bonds between the acetate group at C-17 and Arg276 and Ser275 are absent due to the orientation of the steroid core pose.

For compound **4-14**, Glide XP docking does not provide a similar result to Glide SP. Instead, it delivers a pose similar to **4-12**. The C-3 acetate and C-2 ethoxy occupy the space between the M-loop and N-loop, and the C-17 hydroxyl group acts similarly to the C-17 acetate group in **4-12**. However, it is not placed as deeply into the hydrophobic sub-site as **4-12**. Compound **4-14**, when subjected to IFD, displays a pose similar to the one in Glide SP docking. Nevertheless, the C-3 hydroxyl group near the N-loop fails to provide interactions similar to the acetate groups in **4-12** or **4-13**. Instead, it participates in a hydrogen bond with Pro272. All the poses are presented in **Figure 4.9**, and all the corresponding MMGBSA energies are given in **Table 4.3**.

Comparing the docking results with the biological data, the C-17 side chain doesn't contribute as much as would be consistent with the IC_{50} differences; **4-12** and **4-13** are predicted to interact with tubulin with a lower energy than **4-14** (4~13 Kcal/mol) reflecting the 50- fold IC_{50} difference. The steroid analogs prefer Helix-1 parallel poses rather than a Helix-1 perpendicular pose. With such kind of poses, they cannot fully occupy the space between Helix-1 and M-loop as PTX does.

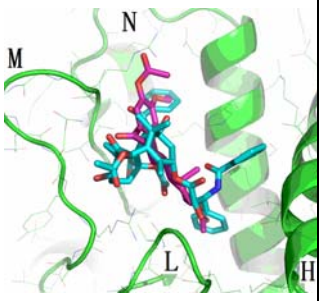
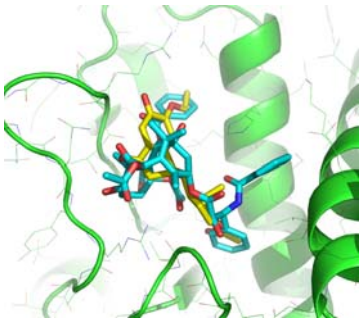
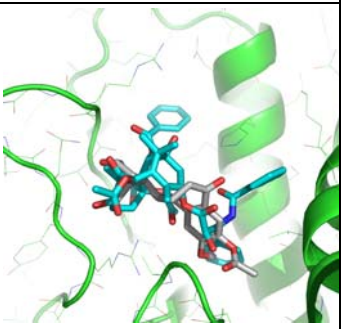
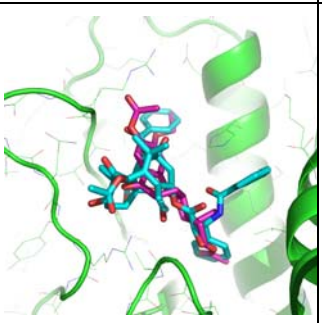
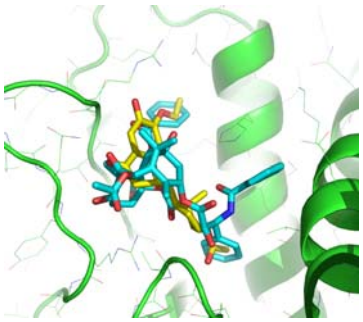
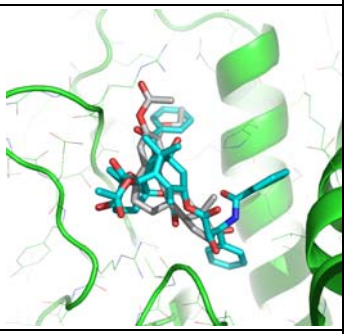
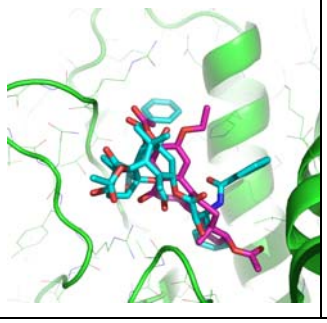
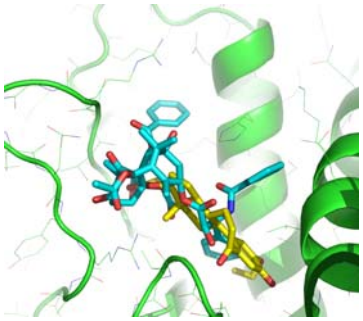
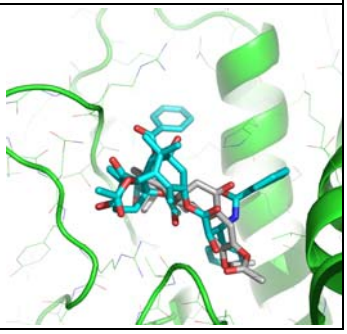
	4-12	4-13	4-14
Glide SP Docking			
Glide XP Docking			
Induce Fit Docking			

Figure 4.9 Steroids dock into tubulin-taxol binding site. The secondary structure of β-tubulin is shown in green cartoon, the PTX is shown in cyan as reference. Steroid analog **4-12** is shown as pink stick, **4-13** is shown as yellow stick and **4-14** is shown as white stick.

Table 4.3 MMGBSA energies of steroids in taxane binding site and 282 binding site.

kcal/mol	**4-12**	**4-13**	**4-14**
Tax_sp	-24.6	-27.8	-15.95
Tax_xp	-21.14	-27.55	-17.91
Tax_ifd	-29.22	-32.91	-39.65
282_sp	-19.7	-20.9	-13.9
282_xp	-18.8	-19.55	-11.00
282_ifd	N/A	-25.49	N/A

4.3.5 282 Site Docking Results

Compared with the wild type PTX tubulin binding site, the 282 tubulin binding site is wider and shallower. The deep hydrophobic sub-site at the bottom becomes shallow because of the side-chain movement of Arg218. The M-loop recedes from Helix-1, and the Arg282 side chain resides close to the M-loop back bone. The region previously occupied by the Arg282 side chain is empty now, and therefore the pocket becomes wider in the direction parallel to Helix-1. All this leads to a bowl-shaped pocket. **282** displays a pose perpendicular to Helix-1 in this pocket and spans the space between Helix-1 and the M-loop. Therefore, the optimal poses of the steroid analogs would also be perpendicular to Helix-1 spanning the space between Helix-1 and the M-loop with the side chains at C-2，C-3 and C-17 close to Helix-1 and the M-loop, respectively.

All the lowest energy poses of the three steroid analogs, when subjected to Glide SP docking, locate the C-17 side chain in the new sub-site which was occupied by the Arg282 side chain in the wild type tubulin-taxol site. The carboxylic oxygens at C-17 of **4-12** and **4-13** are close to the M-loop and form hydrogen bonds with the backbone of Met363 and Lys362. The C-17 hydroxyl on **4-14** also forms a hydrogen bond with Met363, but lacks the hydrogen bond with Lys362 due to its short length. In the tubulin-**282** ligand binding site, the imidazole ring of His227 separates the pocket into two sub-sites. The C-3 acetate on **4-12** sits in the sub-site on the M-loop side. That means **4-12** doesn't show an optimal pose because both the C-3 and the C-17 side chains are on the M-loop side, and the 4-memberedring core lies parallel to Helix-1. The C-3 side chain of **4-13** resides in the sub-site on the Helix-1 side, and its ethoxy group falls close to Helix-1. With this pose, **4-13** is perpendicular to Helix-1 similarly to **282** and spans more space between the M-loop and Helix-1 than **4-12**. The C-3 side chain of **4-14** resides between the two sectors.

The Glide XP docking results are quite similar to the Glide SP ones. **4-12** and **4-13** give similar binding poses and MM-GBSA energies which are about 8.0 kcal/mol lower than **4-14**. In Induce fit docking, **4-12** shows a different pose from the Glide result. Its C-3 and C-17 acetate groups are placed between the M- and N-loops. **4-13** keeps its Glide pose in the IFD regime, but the C-3 hydroxyl group relocates closer to Helix-1. Compound **4-14** shows a Helix-1-parallel pose and no optimal perpendicular pose was given in the IFD result. All the docking poses are shown in **Figure 4.10**.

The **282** site docking result leads to a similar conclusion as the wild tubulin-taxol site docking. The lowest energy poses of **4-12** and **4-13** always have similar energies and are 6- 9 kcal/mol lower than **4-14**. In the sub-site near His227, the steroids' side chains can be directed to either the M-loop side or Helix-1 side. The latter leads to a **282** like pose. With this pose, the steroids occupy most of the **282** site except the sub-site where the C-13 side chain on **282** is located. This difference may explain the weaker activities of the steroids as compared with **282**.

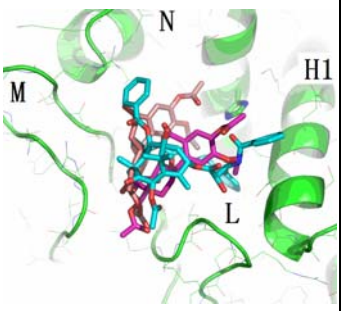
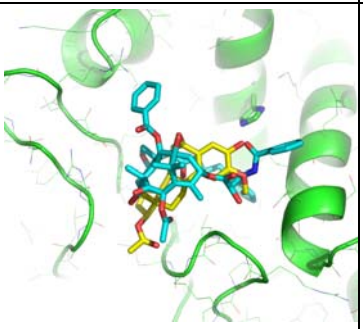
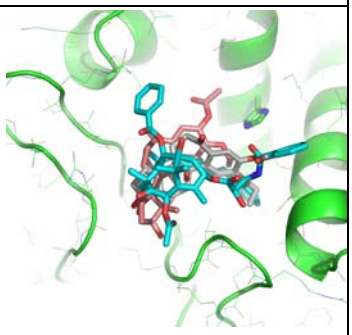
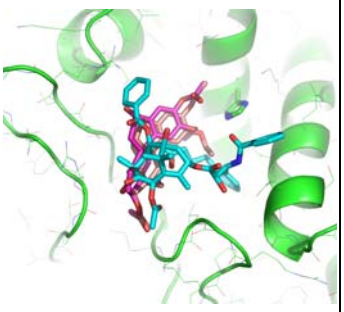
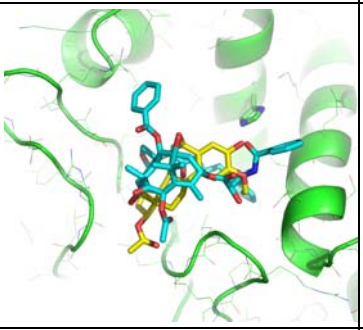
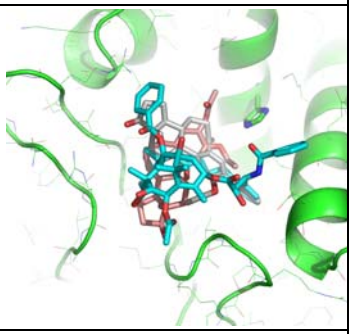
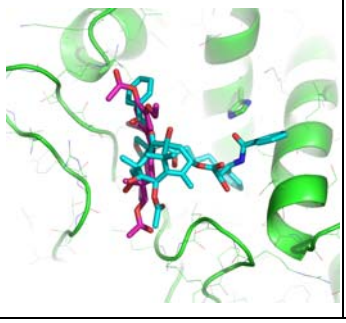
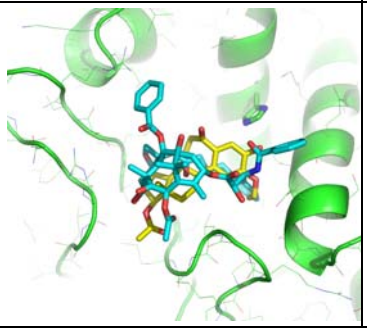
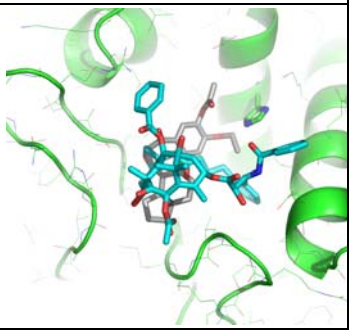
	4-12	4-13	4-14
Glide SP Docking			
Glide XP Docking			
Induce Fit Docking			

Figure 4.10 Steroids docked into the **282** site. The secondary structure of β-tubulin is shown in green; **282** in cyan. In the first row**,** the lowest energy poses of **4-12** is shown as pink stick and the H1 perpendicular poses is shown as magentas. **4-13** is shown as yellow stick in the second row. The lowest energy poses of **4-14** is shown as pink stick and the H1 perpendicular poses is shown as white stick. His227 is shown as green stick,

4.4 Conclusions

Two different ways were attempted to derive truncated taxanes that mimic the binding properties of **282**. However, neither results in compound with nano molar level activity. The aza-tricyclic replacements give **282**-like binding poses in the tubulin-**282** pocket, indicating that they should be candidates for tubulin stabilizing agents. However, the experimental data confirms they are much weaker than **282** or even PTX. This indicates there are other factors that we didn't include in our model which influence the activities. The steroid analogs cannot present a pose which fully mimics the pose of **282** or taxol. They bind parallel to Helix-1 leaving the space between the M-loop and Helix-1 partially empty. This unoccupied space in the binding pocket results in low activities. Both the steroid analogs and the aza-tricyclic replacement compounds have micro molar level activities. Further investigations are necessary to improve their activities.

Chapter 5: Homology modeling of IKKs and Analysis of Their Binding Properties through Molecular Modeling

5.1 NFκB Activation

Nuclear factor-κB (NFκB) proteins are a small family of closely related transcription factors. They play central roles in autoimmune, inflammatory and destructive mechanisms which drive the progression of disease such as rheumatoid arthritis. In this family, five different members have been identified in mammals: relA, relB, c-rel, NFκB1 and NFκB2. All family members share a rel homology domain (RHD) in common, which is required for DNA binding, dimerisation, nuclear localization and inhibitor binding. NFκB is rendered inactive in non-stimulated cells through binding of specific NFκB inhibitors, inhibitory κB (IκB) proteins, which bind to the rel homology domain of NFκB by means of a core domain. Activation of NFκB depends on phosphorylation-induced ubiquitination of the IκB proteins. This cytoplasmic modification in turn relies on a family of IκB kinases (IKKs) and the ubiquitin ligase complex. The IKKs complexes are composed of 3 subunits: IKK α unit, IKK β unit and a regulation subunit termed NFκB essential modulator (NMEO or IKK γ unit), which translates upstream signals into activation of the catalytic subunits [1].

Two different signaling pathways are involved in NFκB activation: canonical

pathway and non-canonical pathway [2, 5]. The canonical pathway is activated by various inflammatory stimuli, engagement of the T-cell receptor, exposure to bacterial products and some genotoxic stimuli. It is induced mainly by IKK β unit through phosphorylation of the canonical IκB proteins. This pathway is involved in development, immunity and cancer development [3]. The non-canonical pathway is activated by stimuli such as CD40, lnphotoxin-β receptors, the B-cell activating factor of TNF family, LPS and latent-membrane proteins or the Epstein-Barr virus. Activation occurs through IKK α unit and NF-κB inducing kinases and utilizes the P52 precursor protein P100. This pathway regulates survival of premature B-lymphocytes and development of peripheral lymphoid tissues [1]. Although the non-carnonical pathway might be dysregulated in autoimmune and inflammatory disorders, IKK α unit seems to regulate the duration of the NF-κB response in macrophages stimulated through TLRs, with prolonged expression of pro-inflammatory cytokines observed in IKKα-deficient cells [4]. This indicates that inhibitors selective for IKK β unit rather than α unit could be advantageous as anti-inflammatory agents.

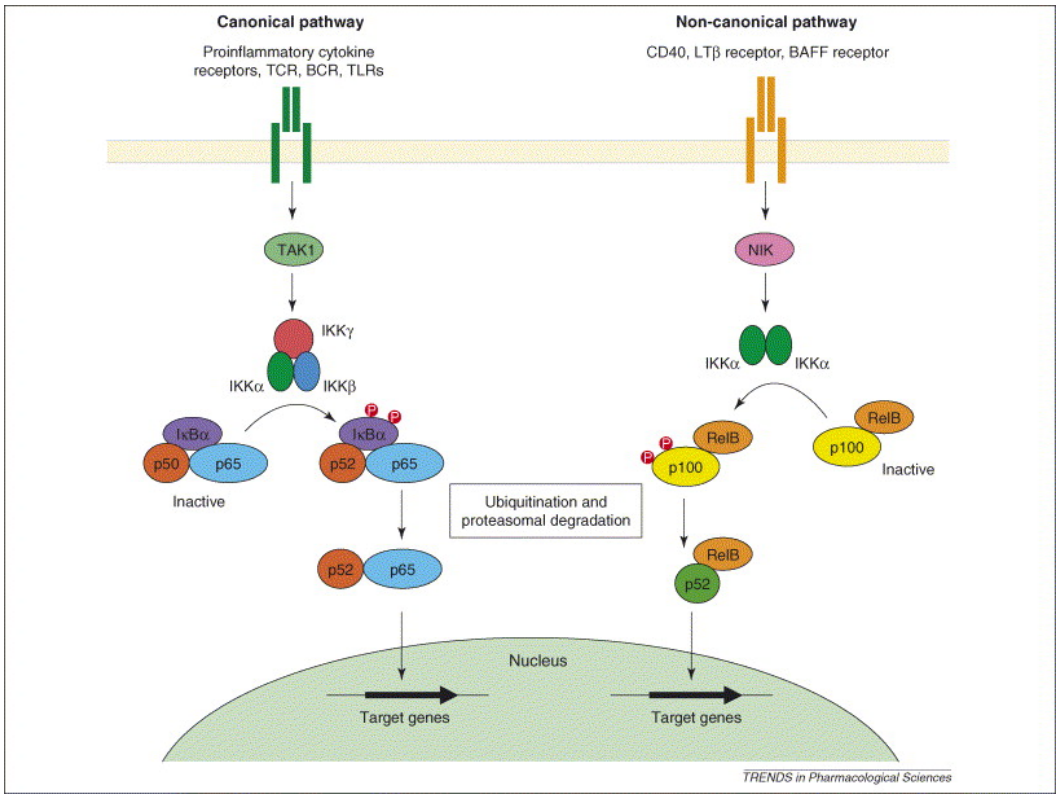


Figure 5.1 NFκB activation pathways dependent on IKKs. In the canonical pathway, the stimulation of receptors activates a multi-subunit IKKs complex containing IKK α unit, IKK β unit and NEMO, which catalyzes the phosphorylation of IκBα. This phosphorylation is essential for signaling the subsequent ubiquitination and proteolysis of IκBα, leaving NF-κB free to translocate to the nucleus and promote gene transcription. In the non-canonical pathway, activated IKK α unit phosphorylates p100 (NF-κB2) in a NEMO-independent manner, signaling a subsequent ubiquitination and partial proteolysis to form transcriptional active p52-RelB heterodimers [5].

With the recognition that inflammatory conditions are often associated with precede cancer, it was natural to suspect a link between NFκB and cancer, as was first suggested several years ago [6]. Since that time, experimental evidence revealing specific

mechanisms by which NF-κB influences cancer initiation, promotion and progression has been mounting as a dizzying pace. Hence IKK β unit, the canonical pathway regulator, is considered as an anticancer drugs target. Up to now, a few compounds have been reported to show inhibition to IKK β unit and selectivity between β unit and α unit. Of these compounds, BMS345541 [7] was identified as an ATP non-competitive inhibitor; TPCA1 [8], Bayer compound A [9], ML120B [10], NRDD-comp1 and NRDD-comp4 were identified as ATP-competitive inhibitors; NRDD-comp2, NRDD-comp5 and BMCL-5a [11, 12] also show inhibition on IKKβ. However, the inhibition mechanism is unknown. The structures and activities of these compounds are listed in **Figure 5.2** and **Table 5.1**. μ

Table 5.1 a) Activities of ATP competitors against IKKα and IKKβ.

Compound	TPCA1	Bayer A	ML120B	NRDD-comp1	NRDD comp4
IKKα IC_{50}	400nM	135nM	>100 μM	13 μM	>200 μM
IKK β IC_{50}	18nM	2nM	45nM	62nM	2.7-11.2 μM

b) Activities of compounds with unknown mechanism

Compound	BMCL-5a	NRDD-comp2	NRDD comp6
IKKα IC_{50}	450nM	5 μM	20 μM
IKK β IC_{50}	25nM	0.15 μM	0.6 μM

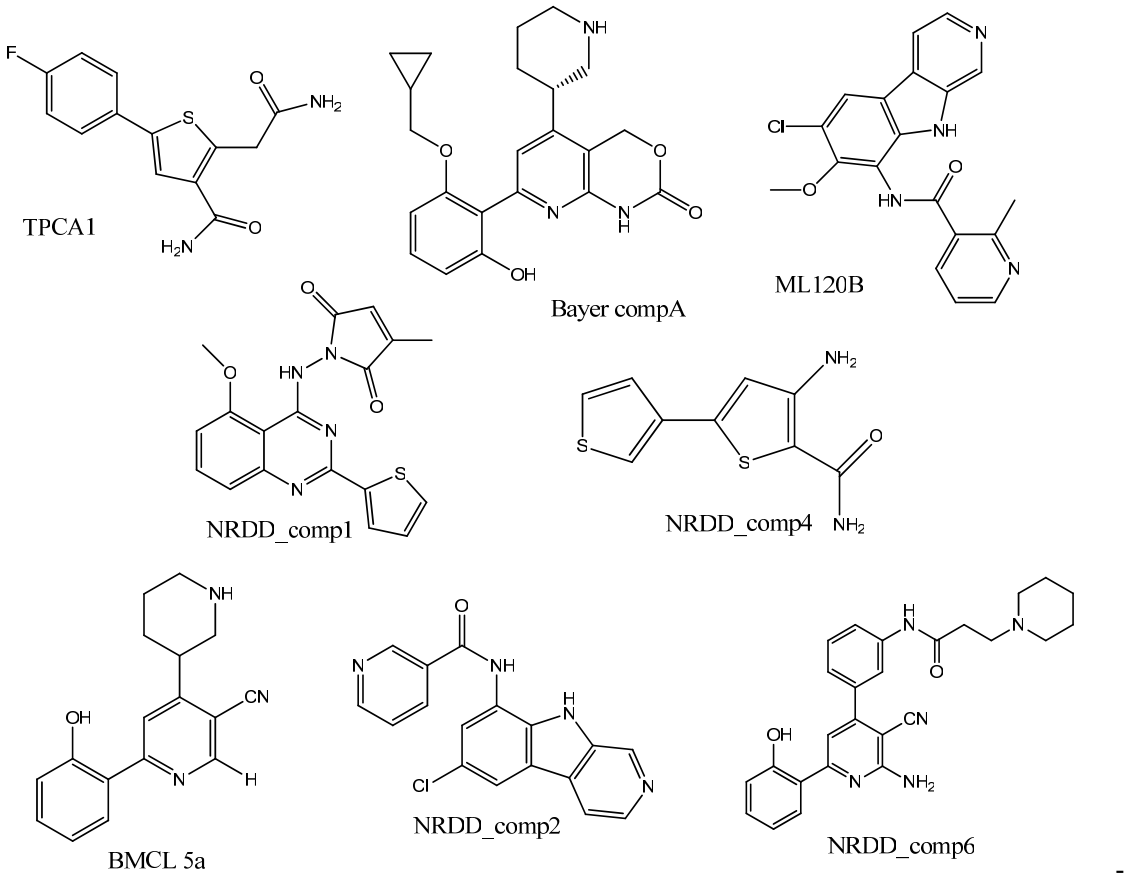


Figure.5.2 Structures of reported IKKs inhibitors [7-12].

5.2 Homology Models

IKK α and β units are proteins with 756 residues which are composed by N-terminal kinase domain, ubiquitin domain and C-terminal NEMO binding domain. Recently, several NEMO/IKKs association domain crystal structures were reported containing three long helices from NEMO and IKKs C-terminal NEMO binding domain, revealing the interactions between different units of IKKs. However, the structure detail of the N-terminal kinase domain is still unknown. A homology model of this functional kinase

domain is essential to investigate IKKs inhibitions. Previously, homolog models of IKK α and β units were generated by Los Alamos National Laboratory (LANL). They constructed a vast number of kinases homology models at the same time including IKKs with homology modeling program MODELLER from protein kinase A (PKA), insulin receptor kinase (IRK), and phosphorylase kinase (PHK) with no particular efforts which make the homology model off of the closest existing PDB structures [13]. In order to investigate the accuracy of these models, five selected ATP-competitive compounds listed in **Table 5.1** a) were prepared and docked into the ATP binding site of IKK α and β, respectively. The computational result reveals these five ATP-competitors always show similar binding pose in IKK α and β units and this cannot interpret their selectivity in different units. The structure similarity of IKKα and β units is mainly induced by their high sequence identity, especially the residues which surround the ATP binding pocket, of these two units. Under this precision, the homology models which constructed by the same templates and homology modeling tools cannot explain the activities difference from experiments. Thus, more precision homology models are required to analysis the difference between two IKK units and explain the selectivity of the IKK β inhibitors.

Two homology modeling programs, Prime 2.0 [14] and Modeller 9.1 [15], were utilized to construct the 3D homology models of IKKs. The Amino acid sequences of human IKK α and β units were obtained from the databank at the National Center for Biotechnology Information, NCBI. Template selection was accomplished with Prime2.0 using BLAST searching protocol in the PDB database assembled in Schrodinger

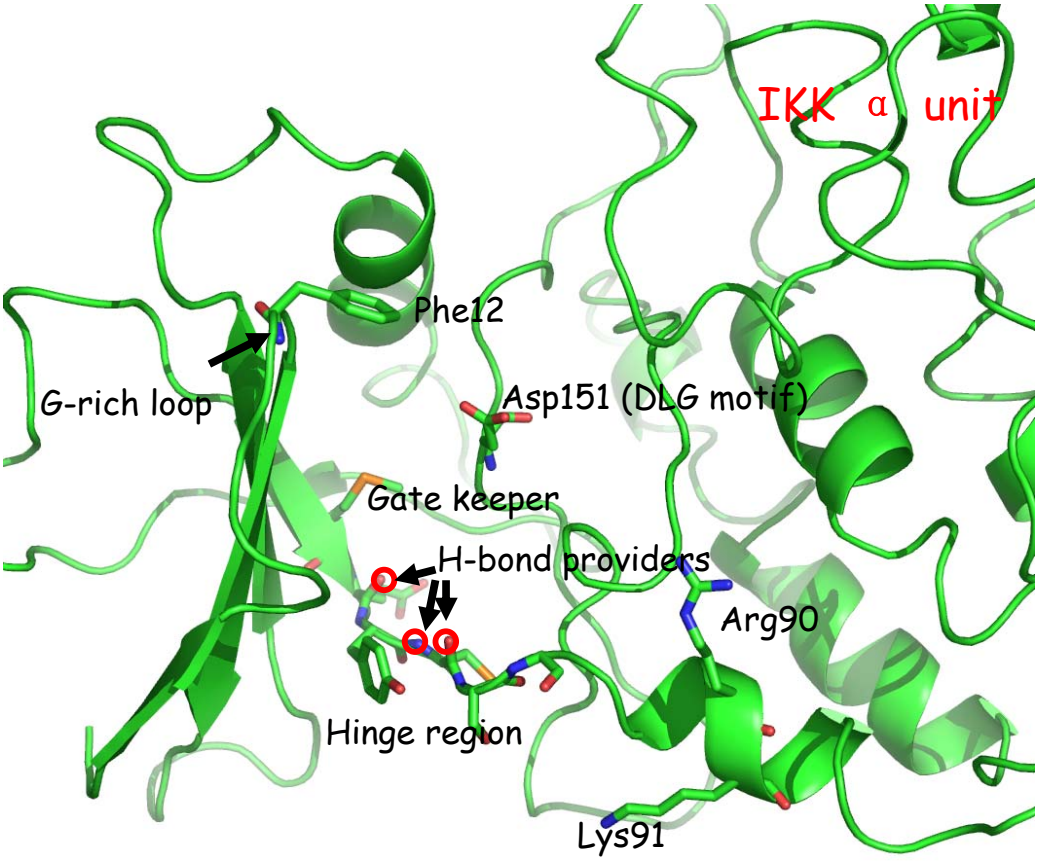
commercial package. The crystal structure of P-38γ MAP kinase chain A (PDB code: 1CM8) gives the highest kinase family overlap with IKK β unit, 33% sequence identity and 49% homology. However, this crystal structure doesn't contain the G-rich loop due to its flexibility in crystallization, Ribosomal protein S6 kinase (PDB: 2Z7R) which shows 32% identity with IKK β unit and presents similar sequence with P-38γ MAP kinase near the G-rich loop is well aligned with IKK β unit and P-38γ MAP kinase by multi-sequence alignment method ClustalW. Its G-rich loop was selected as template to generate the G-rich loop of IKK β unit.

The initial IKK β unit model was constructed by Modeller 9.1 with the co-crystallized AMP-PNP and two Mg^{2+} ions in the P-38 MAP kinase crystal structure. Homology model of IKK α unit was generated with the same method and templates. The initial models were refined by Prime 2.0 and Modeller 9.1. Until this step, we still have the same problem as the LANL models, the models of two units are quite similar with each other and they cannot locate the selectivity of the 5 ATP competitors. Considering the protein folding and ligand binding are dynamic processes, different conformational change can be obtained in these processes and these may lead to lower energy protein structure. Under this circumstance, 20ns MD simulations were performed on each refined IKK homology model in order to investigate their conformational changes. In the MD simulations, the co-crystallized AMP-PNP is retained in the starting structures. Both the MD simulations reached equilibrium in the end, and the last 1 ns structures were clustered with cutoff value 0.15 Å. The mean structures of the cluster were selected and

refined by Modeller and Prime until they obtained a reliable PROCHECK score.

As other kinases, the IKKs models we constructed retain the general properties of the kinase family. The ATP binding pocket is composed by three regions: the hydrophobic pocket **S1** which is above hinge region. Three hydrogen bond contributors from the backbone of hinge region residues, one donor and two acceptors, were pointed to the bottom of this hydrophobic pocket; a sugar pocket between the **S1** pocket and the solvent region which is composed by several polar residues; and the catalytic center near the DLG motif. Comparing the models of units α and β, the differences between them are in two regions: G-rich loop region and the distal of hinge region. In G-rich loop region, α unit gives a more open conformation, leaving a broader pocket both in the catalytic region and the **S1** hydrophobic pocket. In β unit, the Phe12 on the top of the G-rich loop π-π stack with Tyr155 which leads to a close conformation. With this conformation, the volume of the catalytic center is shrunk substantially. As a consequence, the β-strand adjacent to G-rich loop is about 2 Å closer to its DFG motif. That is to say, the hydrophobic sub-site **S1** of β unit is narrower that only planar and hydrophobic fragment is preferred to fit in, and this planar structure should parallel to the β-strand. On the hinge region residues, both models provide two hydrogen bond acceptors and one hydrogen bond donor which point into the ATP binding pocket. However, in the distal of the hinge region, a long loose loop exists in β unit model instead of a short helix in α unit. Two residues with long side chains, Arg91 and Lys92, on this loose loop not only blocks part of the hydrophobic entrances between the ATP site and solvent region, but also forms a

deep hydrophobic pocket **S2** with Leu90, Pro134, Glu135 and Ile137. Beside this, the **S1** pocket in β unit is deeper and narrower than the corresponding **S1** pocket in α unit. In α unit, due to the side chain of Ile150 and Tyr83 which is close to each other, the **S1** hydrophobic sub-site becomes shallow and broad. Its Arg90 and Lys91 bear close to Asp88 and Glu134 and form intra-hydrogen bonds with them leaving a broad entrance for small molecules. The side chain of Glu134 fills the space where the **S2** pocket locates in β unit.



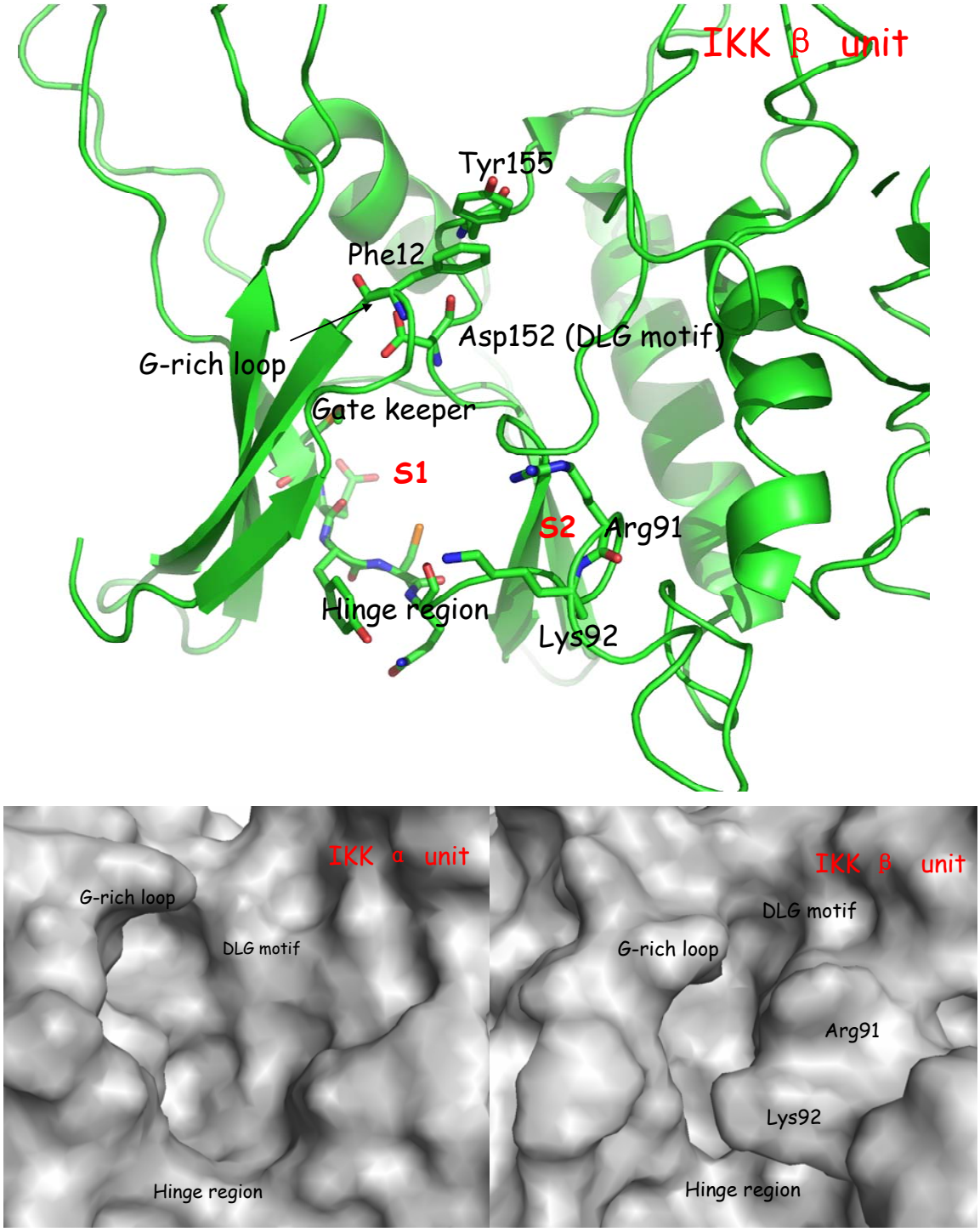


Figure5.3 Homology model of IKKs. Due to the different conformation of G-rich loop and Arg 91, Lys92, the hydrophobic pocket S1 becomes smaller and narrower in β unit, and an extra S2 hydrophobic pocket is obseved near the back bone of Arg91 and Lys92 in β unit.

After both models were constructed, a series of IKK β unit inhibitors were prepared and docked into the ATP binding sites of both models with Glide [16]. Five compounds from literature, TPCA1, ML120B, Bayer compound A, NRDD-comp1 and NRDD-comp4, were proved to be ATP competitive-inhibitors. The mechanism of another subset of the blockers is unknown. All the ATP competitors and the mechanistically opaque subset inhibitors were constructed and prepared with MacroModel [17] and subjected to a conformational search with the OPLS2005 force field, a continuum water solvent model, the mixed torsional/Low-mode sampling method, 5000 search steps and an energy cutoff window of 30 kJ/mol. The global minimum conformer of each compound was used as the starting structure for docking. Ligand-flexible docking was achieved with an extensive on-the-fly conformational search for each IKKs inhibitor, and the top 20 poses of each inhibitor were examined further. These poses were rescored by Prime MM/GBSA calculation [14].

5.3 Analysis on ATP Competitive Compounds

Previous research reveals most kinase inhibitors discovered to date are ATP competitive and present one to three hydrogen bonds to the amino acids on the hinge region of the target kinase, thereby mimicking the hydrogen bonds that are normally formed by the adenine ring of ATP. However, As Radha mentioned in his research, other binding pockets besides ATP binding site exist in different kinases subfamilies [18]. These pockets are also candidates which can block kinase function upon small ligand

binding. In the reported IKKs inhibitors, five of them are proved ATP competitive. The mechanism of the rest three is unclear.

All five ATP competitive compounds, Bayer A, ML120B, NRDD1, NRDD4 and TPCA1, selectively inhibits IKK β unit but has less effect on IKK α unit. Since they have been verified ATP competitive, the activities difference between different IKK units should be caused by the structure difference of the IKKs. Based on previous analysis, the structure differences of the two IKKs units are caused by the conformational change of G-rich loop and the post hinge region loop. These conformational differences lead to a small and deep **S1** pocket in β unit, and its ligand entrance is also blocked by Arg91 and Lys92 resulting an extra hydrophobic pocket **S2**. Both the IKKs units provide three hydrogen bond contributors from the backbone of hinge region, hydrogen bond interactions between ligand and these contributors are essential. In this case, four out of the five ATP competitors form hydrogen bonds with IKK β unit and none of them gives hydrogen bond interaction with these three H-bond contributors in α unit.

In β unit, compound Bayer A gives two hydrogen bonds with hinge region residues shown in **Figure 5.4** b. The amide oxygen and nitrogen atoms on the bicyclic ring form hydrogen bond interactions with the backbone of Cys85 and Gln86 which acts as an anchor to keep its bicyclic core deeply into the **S1** pocket. Due to the size of the pocket, this bicyclic ring is parallel to the β sheet adjacent to G-rich loop. Its phenyl side chain extends into the sugar pocket near Lys92 and its cyclopropane ring occupies exactly the same size as the **S2** pocket near the backbone of Lys92. Its piperidine ring resides into the

catalytic region and the protonated nitrogen points to the polar side chains of Asp152 and Asn136. When this compound was docked into α unit, shown in **Figure 5.4** a, its phenyl ring and the cyclopropene ring occupy **S1** pocket. However, the side chain of TYR83 and Ile150 occupy the space near the hinge region, bayer A cannot extend into the **S1** pocket as deep as is in β unit. Hence, no hydrogen bond interaction was observed with hinge region backbone. Instead, the hydroxyl on the phenyl ring hydrogen bond interacts with Lys30. Due to the open form of the G-rich loop, its bicyclic core presents in the sugar pocket and is perpendicular to the β sheet adjacent to G-rich loop. It spans the space between the β sheet and the DLG motif with the piperidine side chain. Hydrogen bond interaction is found between the piperidine ring and Leu7 on the G-rich loop. The catalytic center in α unit is unoccupied by Bayer A.

In this case, the complete occupation of **S1** pocket both in β unit and in α unit determines Bayer A is a good inhibitor for IKKs. However, the hydrogen bond interaction with hinge region is absent in α unit, and the **S2** pocket only exists and is occupied in β unit. These difference results in the activities difference in the experiment.

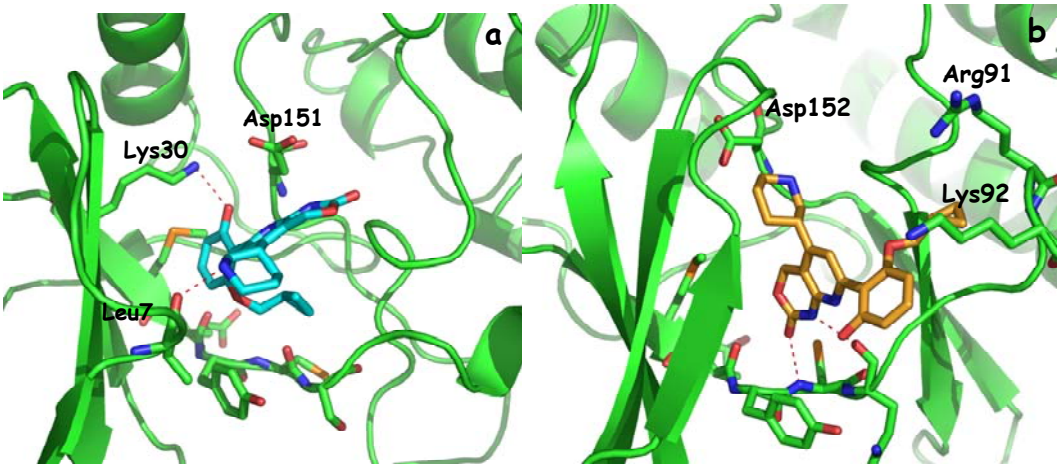


Figure 5.4 Compound Bayer A in IKK α unit (**a** as cyan stick) and β unit (**b** as orange stick). Hydrogen bonds with hinge region are only observed in β complex.

For compound ML120B, its 2-N- carbazole ring forms hydrogen bond interactions with Cys85 and Gln86 on the hinge region, and is deeply involved in **S1** hydrophobic pocket in β unit (see **Figure 5.5**). Its pyridine ring extends to **S2** pocket. However, the size of the pyridine is a little bit larger than **S2** pocket. It doesn't involve as deep as the cyclopropane in Bayer A. The amide bond between the carbazole ring and the pyridine ring forms a hydrogen bond with Glu135. When this compound was docked into IKK α unit, the tri-cyclic ring also locates into **S1** pocket. Due to the open conformation of the G-rich loop in α unit, its **S1** pocket is larger and broader. The tri-cyclic ring only takes the part near the gatekeeper. Its pyridine ring resides into the big pocket near the catalytic center which is composed by Phe12, Lys30, Arg41 and Trp44, and no hydrogen bond interaction is found with the protein.

Compare with different binding pose of ML120B in IKKs, the **S1** pocket is only

partially occupied by ML120B in α unit, and the hydrogen bond interactions with hinge region are only found in β unit. This difference makes ML120B a weaker inhibitor in IKK α unit. In addition, the H-bond interaction with Glu135 in β unit helps anchoring its binding pose.

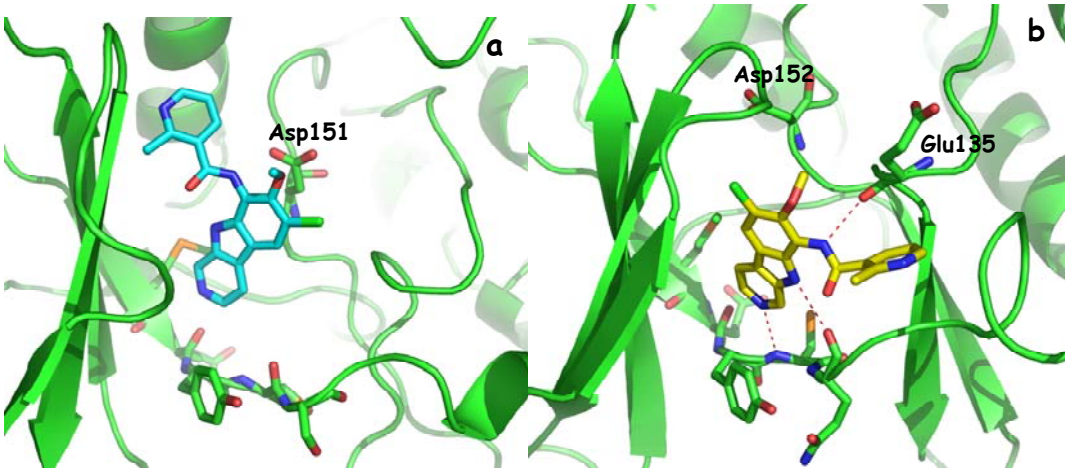


Figure 5.5 Compound ML120B in IKK α unit (**a** as cyan stick) and β unit (**b** as yellow stick). Hydrogen bonds with hinge region were observed in β complex only.

Compound NRDD1 is a little bit different from other ATP competitors. Due to the size of the **S1** pocket in IKK β unit, the three side chains attached on the quinazoline ring prevent it from entering **S1** pocket as Bayer A and ML120B. Instead, the quinazoline ring occupies the space above **S1**, and no hydrogen bond interaction is observed with the hinge region. The thiophene side chain on the bicyclic core extends into **S2** pocket and fills the space near the backbone of Lys91. The other 5-membered ring prefers the catalytic center. In IKK α unit, the pocket size is enlarged due to the open form of G-rich

loop. NRDD1 cannot fill all the pocket space with its size. Its quinazoline ring attaches in the inner surface of this enlarged pocket and is parallel to the side chain of Asp 152. Its thiophene ring extends into the pocket as the pyridine ring in ML120B. However, it only occupies the center of this pocket and is about 4Å to the nearest heavy atom around. The other 5-membered ring points to solvent region and one of its carbonyl oxygen forms hydrogen bonds with the side chain of Arg43. Details are shown in **Figure 5.6**.

In this case, the small molecule NRDD1 doesn't show any hydrogen bond interaction with hinge region either in β unit or in α unit. However, we found it spans the space on the top of the **S1** pocket and all **S2** pocket in IKK β unit which is believed to contribute to the activity. While, in α unit, it attaches in the inner side of the pocket and leaves over half of the pocket space empty. This difference can be used to explain the different effect between two IKK units in experiment.

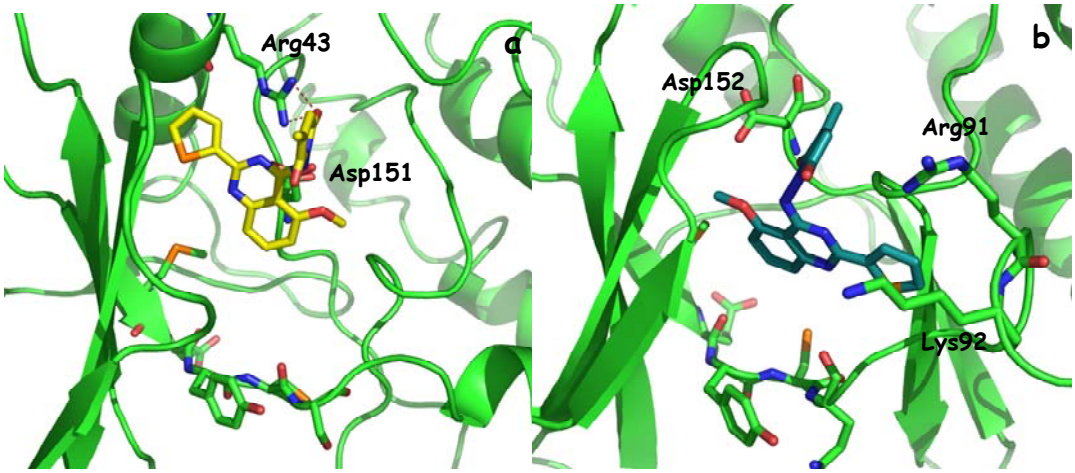


Figure 5.6 Compound NRDD1 in IKK α unit (**a** as yellow stick) and β unit (**b** as dark blue stick). None of them gives hydrogen bond interaction with hinge region. However, the occupation of **S2** and **S1** pocket can explain the activities difference.

NRDD4 and TPCA1 are two compounds sharing similar structure scaffold. Both of them have a two adjacent ring structures with two polar side chains attached, one amide and one amine in NRDD4 and two amides in TPCA1. Due to the high structure similarity, they should give similar binding pose in IKKs. Actually they do present similar binding pose in α unit. Both of them are parallel to the hinge region. They fill the **S1** pocket with their ring structures and the two polar side chains extend to the distal end of the hinge region. In NRDD4, its amide oxygen hydrogen bond interacts with the side chain of Tyr83, and the amine nitrogen forms hydrogen bond with Glu134. In TPCA1, the longer side chain involves hydrogen bond interactions with Gly86. Although their hydrogen bond interactions are close to or on the side chain of hinge region, they do not form hydrogen bond with the three contributors on the back bone of hinge region.

In IKK β unit, these two compounds give different binding pose. For NRDD4, the two polar side chains occupy the **S1** pocket, and they form hydrogen bonds with the hinge region residues GLN86, Cys85 and Glu83. One of the two thiophene rings sits above the S1 pocket and the other one extends near the catalytic center. For TPCA1, it presents an up-side-down pose to NRDD4, the para-F atom on the phenyl ring is deeply inserted into the **S1** pocket and forms a weak hydrogen bond with Cys85. The two polar side chains extend to the catalytic center and hydrogen bond interacts with Phe12 and Asn136.This pose difference is caused by the size of the polar side chains. The two side chains on TPCA1 are larger and they no longer fit into the **S1** pocket in IKK β unit. Instead, TPCA1 fills the **S1** pocket with the phenyl ring which is more hydrophobic

preferred. The hydrogen bond interactions between its polar side chains and the residues near the catalytic center help lock the two distal ends of this molecule. This also can be used to explain the higher activity of TPCA1.

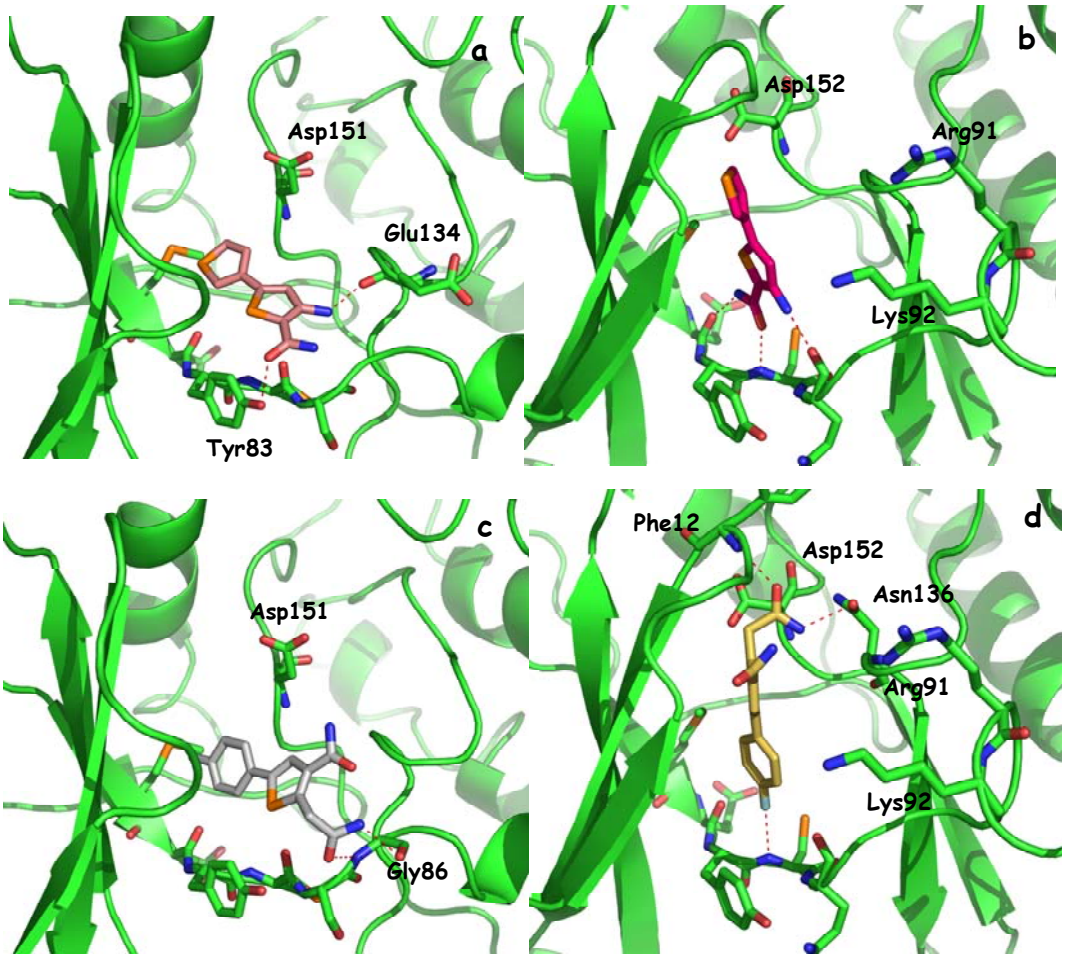


Figure 5.7 Compound NRDD4 in IKK α unit (**a** as pink stick) and β unit (**b** as magenta stick). TPCA1 in IKK α unit (**c** as white stick) and β unit (**d** as yellow stick).Two β unit complexes present hydrogen bonds on the back bone of hinge region.

Based on the discussion above, several conclusions for IKKs were obtained: a) the β

unit of IKK contains a smaller and deeper hydrophobic pocket **S1** due to the close conformation of G-rich loop. On the contrary, the **S1** pocket of α unit is wider and shallower caused not only by the conformational change of G-rich loop, but also by the side chain movement of Tyr83 and Ile150. b) Hydrogen bond interactions between ligand and the three H-bond providers are of importance. They act as an anchor so that the small ligand is able to deeply involve in **S1** pocket. c) The occupation of **S1** pocket contributes major activity of the ligand. The occupation of the **S2** pocket by ligands also improves their activity, although it is not essential. d) The **S1** pocket occupation by a polar group contributes much less on the activity than a non-polar group.

5.4 Mechanism Determination of IKKs Inhibitors

For BMCL-5a, NRDD2 and NRDD6, experiment has not yet shown whether they are ATP-competitive or not, although they exhibit good selectivity for IKK β unit over α unit. Based on the conclusion above, we can predict they are most likely ATP competitive if the computational results well match the experimental data, otherwise, there may be other possible binding pocket exist or other inhibition mechanism for these compounds. These three IKKs inhibitors were prepared and docked into the IKK models with the same protocol to speculate on the possibility for ATP competition. BMCL-5a incorporates a stereogenic carbon on the piperidine ring and therefore two enantiomers exist. Both these two enantiomers were prepared for the docking procedure.

The BMCL-5a docking result indicates only R isomer gives piperidine with

equatorial conformation which is energy preferred. Herein, only the docking result of this isomer is considered. In the docking complexes, the phenyl ring extends into **S1** pocket and the hydroxyl oxygen forms hydrogen bond with the hinge region residue Cys85 in both IKK α and β units. In α unit, the piperidine ring also locates in **S1** pocket with a hydrogen bond interaction with Cys84 which leads the compound to a hinge region parallel pose. The linkage pyridine ring sits above the **S1** pocket and the ethinyl points to the solvent region. However, due to the shrinkage of **S1** pocket in β unit, there is not enough space for both the phenyl and the piperidine group. This piperidine group moves out from this region and extends into the catalytic center. Hydrogen bond is observed in this region between the piperidine nitrogen and Glu135.

In this case, BMCL-5a forms hydrogen bond with hinge region, and **S1** pocket is occupied by the adjacent phenyl ring in both α unit and β unit. This leads to the inhibition of IKKs. While, the pyridine ring is deeply involved in a small region above the gatekeeper residue in β unit which may contribute more to the activity than a corresponding group above **S1** pocket. With this explanation, it is concluded that BMCL-5a is most likely an ATP competitor.

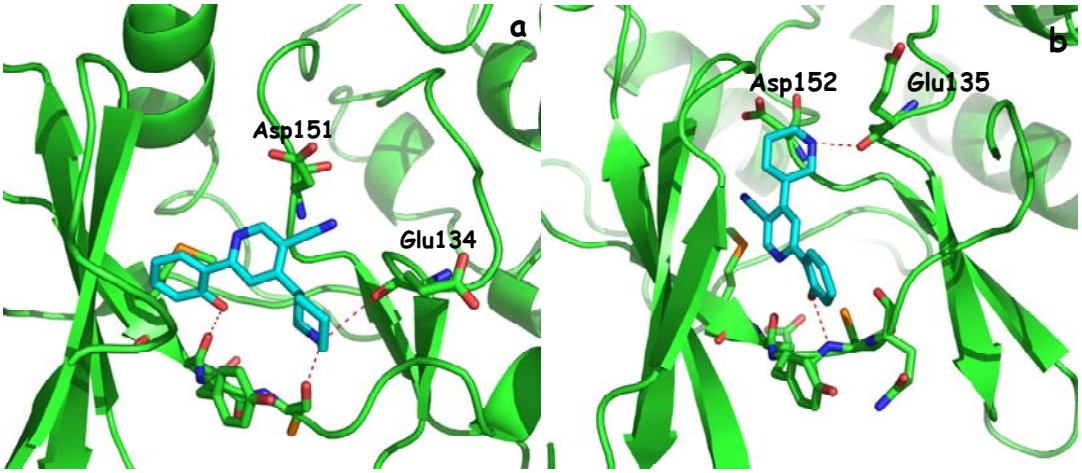


Figure 5.8 Compound BMCL-5a in IKK α unit (left) and β unit (right) as cyan stick.

NRDD2 is structurally similar to ML120B which also contains a 2-N- carbazole core with a pyridine side chain. Without the methoxyl side chain, the carbazole core attaches to the inner surface away from the G-rich loop and the pyridine ring extends into **S1** pocket in α unit. With this pose, NRDD2 doesn't show hydrogen bond with those three H-bond contributors. Instead, a hydrogen bond is observed between the carbazole ring and Gly86 in the end of hinge region which is outside of the ATP binding pocket. In β unit, nitrogen on the carbazole ring interacts with Cys85 as a hydrogen bond acceptor which anchors this core structure deeply in **S1** pocket. The pyridine ring resides into S2 pocket and the amide linkage hydrogen bond interacts with Glu135. With this pose, NRDD2 is more likely an inhibitor of IKK β unit. This gives the same conclusion as the experiment. Hence, NRDD2 is predicted ATP competitive.

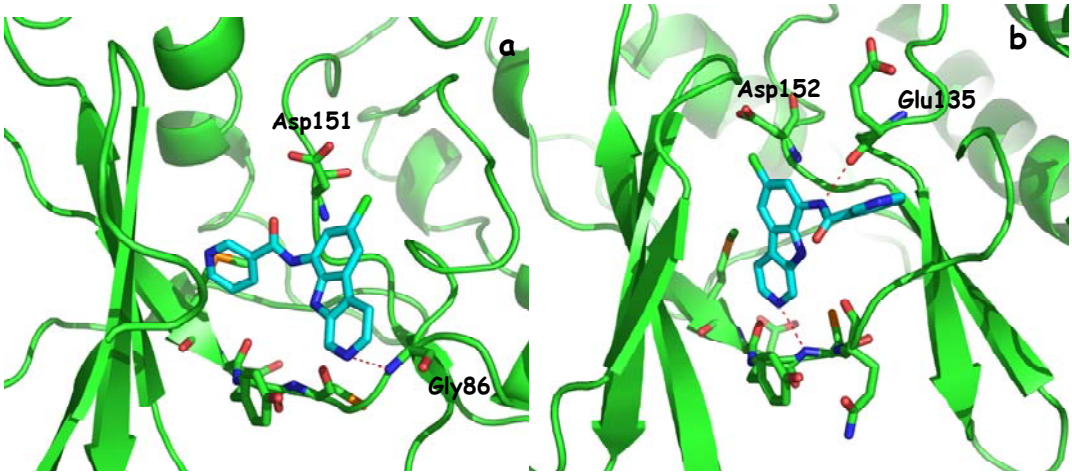


Figure 5.9 Compound NRDD2 in IKK α unit (left) and β unit (right) as cyan stick.

For compound NRDD6, it shares a similar scaffold as BMCL-5a with a long side chain on the phenyl ring. With this side chain, NRDD6 has to present a different binding pose from BMCL-5a. In β unit, the amine group attached on the pyridine ring forms hydrogen bond with hinge region residue Gln86, Although, this pyridine ring doesn't involve deeply in **S1** pocket due to the conflict of the ethinyl group and the gatekeeper residue, its ethinyl group presents in the region above the gate keeper residue. The phenyl-hydroxyl group fills **S2** pocket and its oxygen forms an intra-hydrogen bond with the pyridine ring. The added side chain extends into catalytic center and its protonated piperidine ring hydrogen bond interacts with the DLG-motif. In α unit, the phenyl-hydroxyl forms a hydrogen bond with Gln86 and its phenyl group deeply involved in **S1** pocket. The pyridine ring and the attached amine group extend into the tunnel to solvent region. The added side chain points to the catalytic center but no hydrogen bond is observed in this region. With this pose, it is found that this compound

only fills the G-rich loop side of the ATP pocket in IKK α unit. There is enough space for several water molecules between this compound and the DFG-motif. This may give negative contribution to the activity. However, based on the hydrogen bonds with hinge region, the occupation of **S1** pocket, it's difficult to predict which unit this compound should have to give higher activity. So, for this compound, its mechanism of inhibition is still unknown.

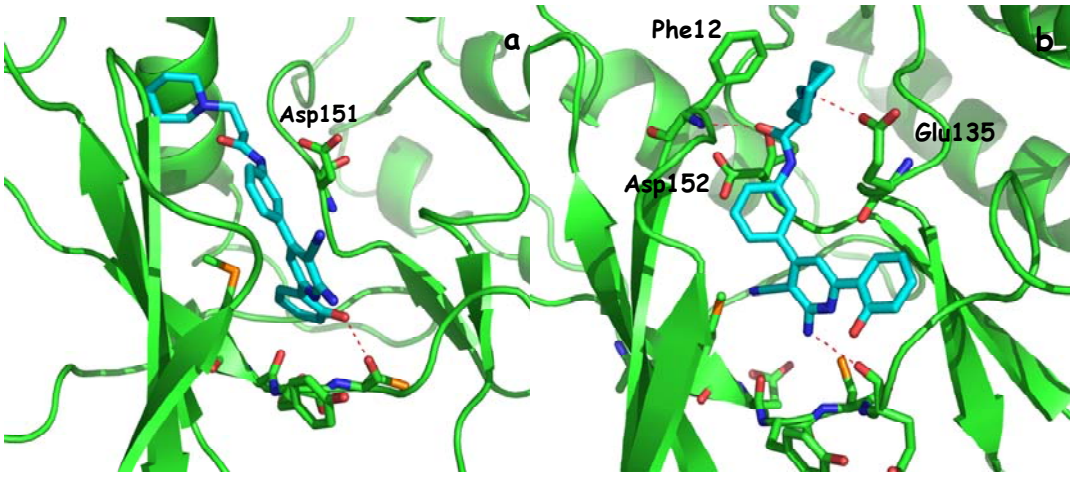


Figure 5.10 Compound NRDD6 in IKK α unit (left) and β unit (right) as cyan stick.

5.5 Conclusions

In this work, homology models of both IKK α unit and β unit were generated and their binding properties were analyzed by means of molecular modeling. The computational results indicate that hydrogen bond interactions between small ligands and the three H-bond providers on hinge region residues are very important for ATP competitive ligands. Such interactions help stabilizing the occupation of the hydrophobic pocket **S1** which contribute to the major activities. Occupation of the **S2** pocket in β unit

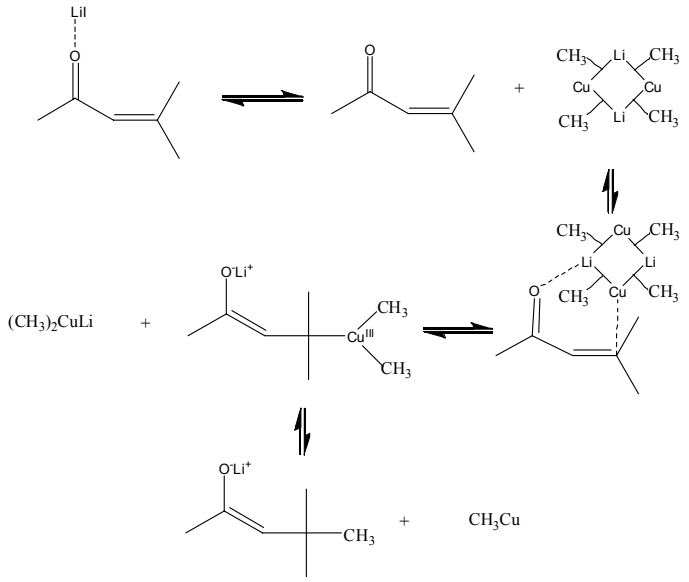
is also believed to provide extra activities. The corresponding **S1** pocket in α unit is broad and shallow due to the open form of the G-rich loop and movement of several residues' side chain. The reported IKKs inhibitors cannot fill and involve deeply in this pocket. There is always extra space for several solvent molecules which may alter the binding pose and binding energy in the ATP binding pocket of α unit. Such solvent effect is not considered in this study. Another drawback of this study is the small number of the reported IKKs inhibitors. Only five IKKs competitors and three mechanism unknown compounds were utilized to describe the selectivity of IKKs. Although more experimental data is necessary to verify the conclusion of this study, it provides a first-step to understand the mechanism of IKKs inhibition and offers a provisional guidance on the design and synthesis of novel IKKs inhibitors.

Chapter 6: Prediction of Structure of Cuprate Intermediate

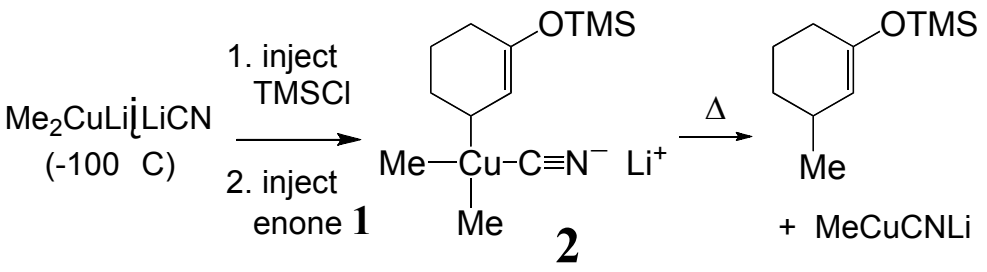
6.1 Introduction

Conjugate addition of organocopper reagents to α-enones is both a classic and a contemporary reaction in organic synthesis. Discovered in 1941 and examined further in the 1950s and 1960s, it has been the subject of many studies since [1]. Although the conjugate addition of lithium diorganocuprates to α, β- unsaturated ketone is widely used in organic synthesis, description of the detailed mechanism of this reaction has been impeded by the lack of kinetic data and by the scarcity of information regarding the structure and composition of these reagents in solution. In 1981, Krauss *et al* [2] utilized stopped-flow ultraviolet spectroscopy to investigate the reaction of several α, β-unsaturated ketones with excess lithium dimethylcuprate in the presence and absence of lithium in diethyl ether solution at 25.0˚C by measuring the rate of disappearence of a spectroscopically observable intermediate. They proposed that when the dimethylcopper anion acts as a nucleophile to attack an unsaturated carbon, a Cu^{III} intermediate is formed (***Scheme I***).

Scheme I



In 2002, Bertz and Ogle applied low-temperature rapid-injection NMR (RI-NMR) to observe cuprate π-complexes from 2-cyclohexenone 1 [3].



They injected TMSCl into a solution containing a solution of the cyanocuprate $Me_2CuLi{\cdot}LiCN$ at –100˚C, and then the cyclohexenone. At this low temperature, NMR spectra were obtained. An intermediate with four carbons around the copper atom in the

same plane (square planar) was proposed as the structure 2 based on analysis of NMR chemical shifts and coupling constants. Although the NMR analysis provides a starting point for the structure of the intermediate, the exact nature of the geometry around copper, the intermediate conformation, and the lithium association cannot be deduced from Bertz and Ogle's analysis. In this research, quantum mechanical calculations were performed to predict the NMR properties of the possible structures and their compatibility with the experimental data were evaluated, leading to a specific structure as the probable intermediate.

6.2 Experiment and Discussion

The present investigation was initiated by recognizing that intermediate 2 embodies considerable structural freedom. The cyano group can be either cis or trans to the ring methine carbon, while the cyclohexene ring can exist in two nonequivalent half-chair forms positioning the copper moiety in either a pseudoequatorial or a pseudoaxial position. Additionally, the Cu−methine bond can rotate to direct the Cu−CN bond syn or anti to the methine hydrogen. Given these degrees of freedom, there are six possible geometries for the complex (see **Figure 6.1**). The distal $OSi(CH_3)_3$ group was replaced by OCH_3 to make the calculations tractable, and they were optimized by a density functional theory (DFT) protocol employing an effective core potential for the copper atom (i.e., B3LYP/6-31G*/LANL2DZ [4]).

The optimizations were repeated for the same molecular geometries at the same level of theory with and without a lithium cation coordinated to the cyano nitrogen. The relative energies of all the lithiated and unlithiated optimized structures were listed in **Figure 6.1**. The result suggests that the energy difference in each series is less than 2.3 kcal/mol. In both series, the lowest energy forms show the square-planar copper unit as equatorial and the cyanide group cis to the methine carbon. The major difference is the orientation of the Cu−CN bond: the neutral lithiated form directing CN anti to the pseudoaxial hydrogen, the anion directing it syn, as shown in **Figure 6.2**.

The lowest energy structures were selected for chemical shift and coupling constant calculations. To ensure that the OMe models do not compromise the results, the geometry of lithiated 3a with $OSi(CH_3)_3$ instead of OCH_3 was optimized with B3LYP/ 6-31G*/LANL2DZ and likewise subjected to NMR parameter prediction.

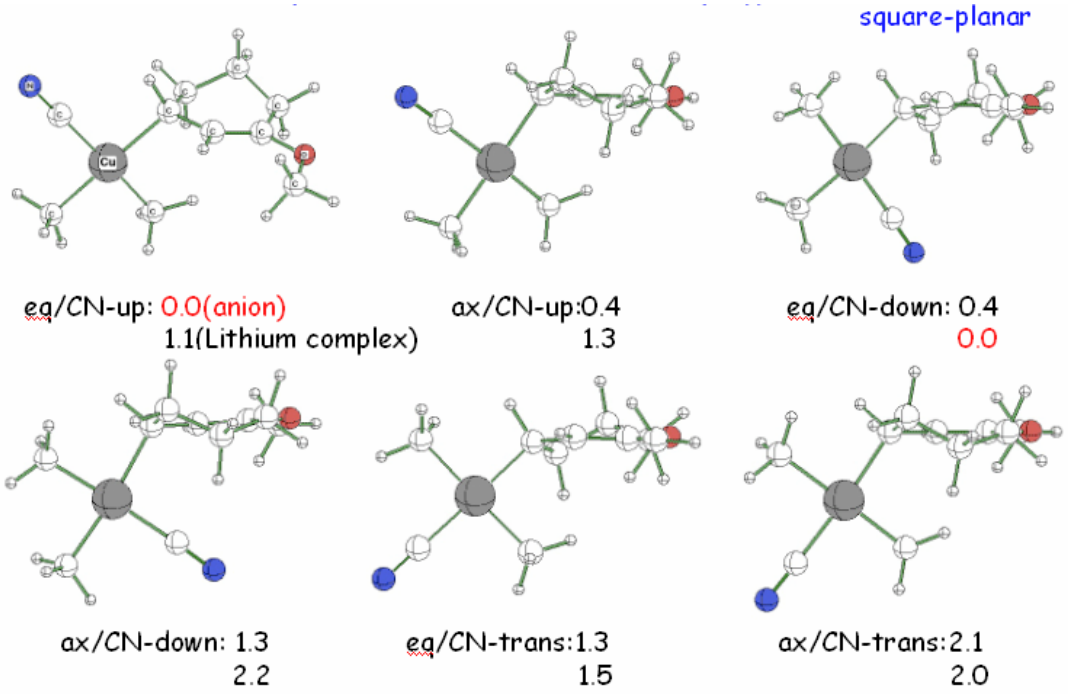


Figure 6.1 Conformation of the six isomers optimized with B3LYP/6-31G*/LANL2DZ [5]. The blue balls indicate negative charges or N coordinated to a lithium cation. The numbers show the energy of the optimized conformations. First line is the energy of ainon, the second line is the energy of the lithiated complex.

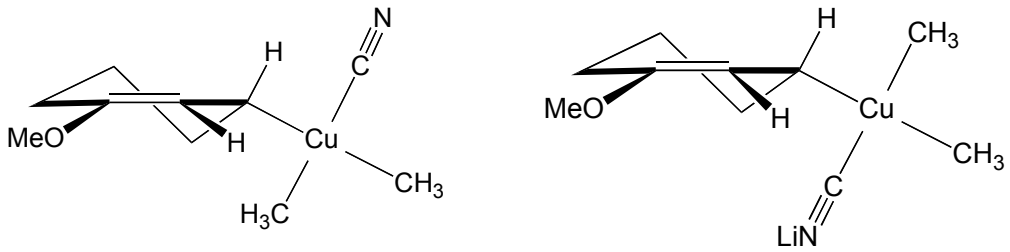


Figure 6.2 The lowest energy conformation within the two series: the unlithiated intermediate with equatorial/CN *syn* to the adjacent C-H bond; and the lithiated intermediate with equatorial/ CN directed *anti* to the C-H bond.

Chemical shifts were calculated for the optimized geometries using the GIAO method and a more elaborate DFT model (B3LYP/6-311+G*/Cu: pCVDZ/SDD//B3LYP/6-31G*/LANL2DZ). The results are shown in **Figure 6.3**. The ^{1}H and ^{13}C shift predictions vs. TMS are compared with experimental data for both the lithiated and unlithiated square-planar intermediates to show that the empirically assigned proton and carbon chemical shifts are reproduced with reasonable fidelity. The ^{1}H chemical shift result reveals the calculation data is in quantitative agreement with the experiment. The individual atomic differences between the calculation data and experimental data are less than 0.7 ppm. The differences between calculated and experimental ^{13}C chemical shift are larger than the difference of ^{1}H chemical shift, the largest one is around 20 ppm, but still in acceptable agreement. The calculated ^{13}C shifts for the anion seem to match experiment slightly better, particularly at the cyano carbon, suggesting either a solvent-separated ion pair or rapid cation−anion exchange.

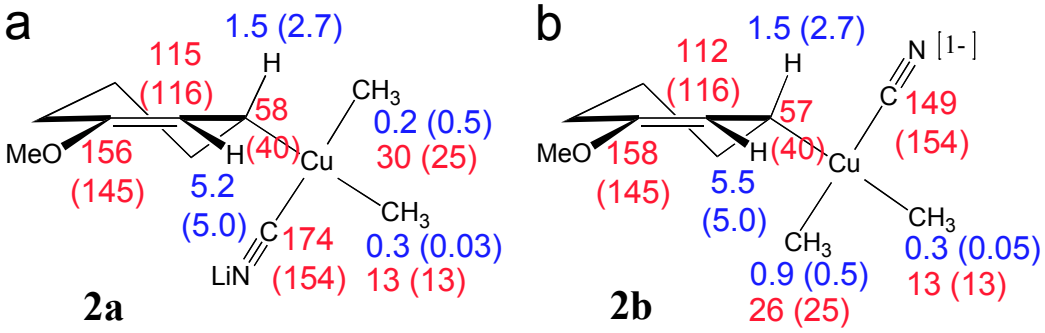


Figure 6.3 Calculated ^{1}H (blue) and ^{13}C (red) chemical shifts for lithiated **2a** and lithiated **2b** relative to TMS and compared with experimental values (parentheses); B3LYP/6-311+G* (pCVDZ) [7]/SDD

method.

Two bond ^{13}C-^{13}C coupling constants were calculated with two recently described methods. The coupled carbons were treated with the aug-cc-pcvdz-sd [5] and aug-cc-pvtz [6] basis sets in the B3LYP framework, and the 2-bond coupling (^{2}J) values were calculated as the sum of the four Ramsey terms [7] (FC (tight s-type function), SD, DSO and PSO). All other atoms were accommodated with the B3LYP/6-311+G*/SDD model. The results shown in **Table 6.1** closely agree with the experiment. Carbons trans across copper are characterized by large values falling between 35-60Hz, whereas cis couplings are well below 10 Hz. This observation coincides nicely with a 2.0 Kcal/mol calculated relative energy for the trans- vs. the cis-CN isomer of lithiated structure, both results ruling out the trans relationship. Furthermore, the trans methyl-cyano coupling is predicted to be smaller than the trans methyl-ring C coupling, while the relative magnitudes of the corresponding cis values are reversed; both in agreement with observation. The two cis couplings that are not observed are predicted to be smallest (<0.3 Hz). The ^{13}C two-bond coupling constants for the lithiated vs. unlithiated intermediate suggest the low temperature complex to be a tight ion-pair instead of a solvent-separated species. The approximation using the OMe group to replace TMSO group was validated by computing only slight differences in the ^{2}J couplings. Comparing the NMR computational data of OMe vs. TMSO in **Table 6.1**, the less than 0.6 Hz differences indicate the substitution is a suitable structural model.

Table 6.1 Experimental and calculated two bond C—C coupling constants[a,b] ^2J around square-planar copper for lithium salt **2a** and anion **2b**

		OCH_3				$OSi(CH_3)_3$	
coupling[c]	Exp.[a]	Li^+ salt **2a**		Anion **2b**		Li^+ salt	
		DZ[c]	TZ[c]	DZ[c]	TZ[c]	DZ[c]	TZ[c]
t-C--CN	35.4	36.4	34.5	44.0	41.8	36.9	35.1
t-C--C_r	38.1	46.1	45.9	52.2	51.7	45.7	45.4
c-C--C	2.9	–3.4	–5.6	–4.4	–6.4	–3.2	–5.4
c-C-CN	5.4	–6.1	–8.1	–6.4	–8.5	–5.9	–8.1
c-C_r--CN	ND[d]	0.01	–0.7	–1.7	–2.0	0.0	–0.6
c-C_r--C	ND[d]	2.5	1.5	1.9	0.9	2.4	1.3

[a]Experimental values.[3] [b]The ^2J values sum all four Ramsey terms. [c]DZ = pCVDZ; TZ = pVTZ; *t* = *trans*, *c* = *cis*, r = *ring*. [d]Not determined owing to low S/N at the unlabeled ring C.

Finally, to investigate whether the proposed square planar geometry was truly the energetically favored geometry, a tetrahedral cuprate was constructed and evaluated with the B3LYP/6-31G*/ LANL2DZ basis set. Optimization causes this structure to fall to the square planar geometry. To examine the potential NMR properties of this unfavorable tetrahedral structure, an analog of the lithiated cuprate with copper-carbon bond angles frozen to 109.5° was evaluated using B3LYP/6-31G*/ LANL2DZ. The average chemical

shift difference increases to 16 ppm compared with 9 ppm for planar structure. The coupling constants also differ from experiment by 20-200 Hz. These results along with the chemical shift and coupling constant comparisons described above strongly support the hypothesis that the copper intermediate does indeed have a square-planar structure as proposed by Bertz and Ogle.

6.3 Conclusions

The calculation results are in excellent accord with experimental data. It means the square-planar structure of the low temperature copper intermediate is confirmed. The work also shows that quantum chemical calculations can be used to verify the experimental data and to assist in the interpretation of structure for highly unstable species.

References

Chapter 1

1. Roberts N., Martin J., Kinchington D., Broadhurst A., Craig J., Duncan I., Galpin S., Handa B., Kay J., krohn.A., et al, *Science*, 1990, **248**, 358-361.
2. Erickson J., Neidhart D., VanDrie J., Kempf D., Wang X., Norbeck D., Plattner, J., Rittenhouse J., Turon M., Wideburg Erickson J., Neidhart N., et al. *Science*, 1990, **249**, 527–533.
3. Dorsey B.D., Levin R.B., McDaniel S.L., Vacca J.P., Guare J.P., Darke P.L., Zugay, J.A., Emini E.A., Schleif W.A., Quintero J.C., et al, *J. Med. Chem.,* 1994, **37**, 3443–3451.
4. Anderson A.C., *Chem. & Bio*. 2003, **10,** 787-797.
5. Henry C.M., *Sci. & Tech Pharm*. 2001, **79**, 69-74.
6. www.rcsb.org.
7. Enyedy I., Lee S., Kuo A., Dickson R., Lin C., Wang S., *J. Med. Chem.*, 2001, **44**, 1349–1355.
8. Enyedy I., Ling Y., Nacro K., Tomita Y., Wu X., Cao Y., Guo R., Li B., Zhu X., Huang Y., et al., *J. Med. Chem.,* 2001, **44**, 4313–4324.
9. Prime, Version 2.2, Schrodinger, LLC, New York, NY, 2010.

   A. Sali & T.L. Blundell., *J. Mol. Biol.* 1993, **234**, 779-815.
10. Peitsch M., Guex N., Large-scale comparative pro- tein modeling. In Proteome

Research: New Frontiers in Functional Genomics, M. Wilkins, K. Williams, R. Appel, and D. Hoch- strasser, eds. (New York: Springer), 1997, 177–186.

11. Laskowski R., Rullman J., MacArthur M., Kaptein R., Thornton J., *J. Biomol. NMR,* 1996, **8**, 477–486.
12. Vriend, G.J., *Mol. Graph.*, 1990, 8, 52–56.
13. Mishra K.P., Ganju L., Sairam M., Banerjee P.K., Sawhney R.C., *BioMedicine & Pharmacotherapy*, 2008, **62**, 94-98.
14. Agrestia J.J., Antipovc E., Abatea A.R., Ahna K., Rowata A.C., Barete J.C., Marquezf M., Klibanovc A.M., Griffiths A.D., Weitz D.A., *Proc, Nat. Acad. of Sci.*, 2010, **107**, 4004–4009.

Chapter 2

1. Angelo V., *Journal of Computational Chemistry* 1988, **9**, 269-280.
2. Nobuo T., Akiko I., Yoichi I., *Journal of Computer-Aided Molecular Design* 1987, **1**, 197-210.
3. Grootenhuis P.D.J., Van Geerestein V.J., Haasnoot C.A.G., Karplus M., *Bulletin des Societes Chimiques Belges* 1992, **101**, 661-662.
4. Greer. J., Erickson. J.W., Baldwin. J.J., Varney. M.D. *J.Med.Chem.,* 1994, **37**, 1035-1054.
5. Sung, B., *et al. Nature*, 2003, **425**, 98-102.
6. http://cmm.cit.nih.gov/modeling/pdb_at_a_glance.html.

7. Databases for Protein-Ligand Complexes M. Hendlich, *Acta Crystallographica*, 1998, **D54**, 1178-1182.

8. Wallace A.C., Laskowski R.A., Thornton J.M., *Prot. Eng.*, 1995, **8**, 127-134.

9. Maestro, version 9.0, Schrödinger, LLC, New York, NY, 2009.

10. DeLano, W.L. The PyMOL Molecular Graphics System (2002) on World Wide Web http://www.pymol.org.

11. Glide, version 5.5, Schrödinger, LLC, New York, NY, 2009.

12. Prime, version 2.2, Schrödinger, LLC, New York, NY, 2010.

13. Bailey, D., Cooper, J.B., *Protein Sci.*, 1994, **3,** 2129-2143.

14. Cooper J.B., Foundling S.I., Blundell T.L., Arrowsmith R.J., Harris C.J., Champness J.N., *Topics in Medicinal Chemistry* , 1988, **65**, 308-313.

15. Tsuge H., Nishimura T., Tada Y., Asao T., Turk D., Turk V., Katunuma N. *Biochem. Biophys. Res. Commun*., 1999, **266,** 411-416.

16. LaLonde J.M., Zhao B., Smith W.W., Janson C.A., DesJarlais R.L., Tomaszek T.A., Carr T.J., Thompson S.K., Oh H.J., Yamashita D.S., Veber D.F., Abdel-Meguid S.S., *J.Med.Chem.,* 1998, **41,** 4567-4576.

17. Weber P.C., Pantoliano M.W., Simons D.M., Salemme F.R., *J. Am. Chem. Soc.*, 1994, **116,** 2717.

18. Katz B.A., Cass R.T., *J.Biol.Chem.* 1997, **272,** 13220-13228.

19. Concha N.O., Janson C.A., Rowling P., Pearson S., Cheever C.A., Clarke B.P., Lewis C., Galleni M., Frere J.M., Payne D.J., Bateson J.H., Abdel-Meguid S.S.,

Biochemistry, 2000, **39,** 4288-4298.

20. Toney J.H., Hammond G.G., Fitzgerald P.M., Sharma N., Balkovec J.M., Rouen G.P., Olson S.H., Hammond M.L., Greenlee M.L., Gao Y.D., **,***J.Biol.Chem.* 2001, **276,** 31913-31918.
21. Saadat D., Harrison D.H., *Biochemistry*, 2000, **39,** 2950-2960.
22. Marks G.T., Harris T.K., Massiah M.A., Mildvan A.S., Harrison D.H., *Biochemistry*, 2001, **40,** 6805-6818.
23. Erlandsen H., Bjorgo E., Flatmark T., Stevens R.C., *Biochemistry,* 2000, **39,** 2208-2217.
24. Erlandsen H., Flatmark T., Stevens R.C., Hough E., *Biochemistry*, 1998, **37,** 15638-15646.
25. Katz B.A., Clark J.M., Finer-Moore J.S., Jenkins T.E., Johnson C.R., Ross M.J., Luong C., Moore W.R., Stroud R.M., *Nature*, 1998, **391,** 608-612.
26. Hauel N.H., Nar H., Priepke H., Ries U., Stassen J.M., Wienen W., *J.Med.Chem.,* 2002, **45,** 1757-1766.

Chapter 3

1. Evans R.M., *Science,* 1988, **240**, 889–895.
2. Olefsky J.M., *J. Biol. Chem.,* 2001, **276**, 36863–36864.
3. Mangelsdorf D.J., Thummel C., Beato M., Herrlich P., Schutz G., Umesono K., Blumberg B., Kastner P., Mark M., Chambon P., Evans R.M., *Cell,* 1995, **83**,

835-839.

4. Weatherman R.V., Fletterick R.J., Scanlan T.S., *Annu Rev Biochem*, 1999, **68**, 559-581.
5. Whitby R.J., Dixon S., Maloney P.R., Delerive P., Goodwin B.J., Parks D.J., Willson T.M. *J. Med. Chem.* 2006, **49**, 6652-6655.
6. Yuichi Hashimoto, Hiroyuki Miyachi, *Bioorg. & Med. Chem*. 2005, **13**, 5080-5093.
7. Fayard E., Auwerx J. & Schoonjans K, *Trends Cell Biol.* 2004, **14**, 250-260.
8. Irina N. Krylova, Elena P. Sablin et al, *Cell*, 2005, **120**, 343-355.
9. Shiau A.K., Barstad D., Loria P.M., Cheng L., Kushner P.J., Agard D.A., Greene G.L. *Cell*, 1998, **95,** 927-937.
10. Krylova I.N., Sablin E.P., Moore J., Xu R.X., Waitt G.M., MacKay J.A., Juzumiene D., Bynum J.M., Madauss K., Montana V., Lebedeva L., Suzawa M., Williams J.D., Williams S.P., Guy R.K., Thornton J.W., Fletterick R.J., Willson T.M., Ingraham H.A. *Cell*, 2005, **120:** 343-355.
11. Li Y., Choi M., Cavey G., Daugherty J., Suino K., Kovach A., Bingham N., Kliewer, S., Xu H., *Mol. Cell*, 2005, **17,** 491-502.
12. Jordan V.C., *Br J Pharmacol* (2006) **147** (Suppl 1): S269–76; "Tamoxifen Information: reducing the incidence of breast cancer in women at high risk", http://www.fda.gov/cder/news/tamoxifen/. July 3, 2007.
13. Vogel Victor; Joseph Constantino, Lawrence Wickerman *et al.*. *The Journal of the American Medical Association,* 2006, **295**, 2727–2741; Barrett-Connor E., Mosca L.,

Collins P., et al.. *New England Journal of Medicine*, 2006, **355**,125–137.

14. Roth J., Madoux F., Hodder P.. Roush W.R., *Bioorganic &Medicinal Chemistry Letters,* 2008, **18**. 2628-2632; Madoux F., Li X., Chase P., Zastrow G., Cameron M.D., Conkright J.J., Griffin P.R., Thacher S., Hodder P., *Mol. Pharmacol*, 2008, **79**, 1776-1784.
15. Soisson S.M., Parthasarathy G., Adams A.D., Sahoo S., Sitlani A., Sparrow C., Cui J., Becker J.W., *PNAS* , 2008, **105:** 5337-5342.
16. LigPrep, version 2.4, Schrödinger, LLC, New York, NY, 2010.
17. Glide, version 5.6, Schrödinger, LLC, New York, NY, 2010.
18. Prime, version 2.2, Schrödinger, LLC, New York, NY, 2010.
19. Maestro, version 9.0, Schrödinger, LLC, New York, NY, 2009.

Chapter 4

1. Crown J. & O'Leary M., *Lancet,* 2000, **355**, 1176-1178.
2. Ganesh T., Guza R.C., Bane S., Ravindra R., Shanker N., lakdawala A.S., Snyder J.P., Kingston D.G. I. *PNAS*, 2004, **101**, 10006-10011.
3. Shanker N., Kingston D.G.I., Ganesh T., Yang C. Alcaraz A.A., Geballe M.T., Banerjee A., McGee D., Snyder J.P., Bane S., to be published.
4. Kingston D.G.I., *Chem. Commun.*,2001, **10**, 867-80.
5. McGuire W.P., Rowinsky E.K., Rosenshein N.B., Grumbine F.C., Ettinger D.S., Armstrong D.K., Donehower R.C., *Ann. Int. Med.* 1989, **111**, 273-9.

6. Holmes F.A., Walters R.S., Theriault R.L., Forman A.D., Newton L.K., Raber M.N., Buzdar A.U., Frye D.K., Hortobagyi G.N., *J. Nat. Cancer Inst.* 1991, **83,** 1797-805.

7. Rowinski E.K., Donehower R.C., *Pharmacol. Ther.* 1991, **52**, 35-84.

8. Thigpen T., Blessing J., Ball H., Hummel S., Barret R., *Proc. Am. Soc. Clin. Oncol.* 1990, **9**, 156.

9. Einzig A.I., Wiernik P.H., Sasloff J., Runowicz C.D., Goldberg G., *J. Clin. Oncol.* 1992, **10**, 1748-1753.

10. Seidman A., Reichman B., Crown J., Begg C., Heelan R., Hakes T., Surbon A., Gilewski T., Lebwohl D., Currie V., Huddis C.,Klecker R., Collins J., Toomasi F., Berkery R., Quinlivan S., Kelsen D., Norton L., *Proc. Am. Soc. Clin. Oncol.* 1992, **11**, 59.

11. Murphy W.K., Fossella F.V., Winn R.J., Michael C.N., Susan C.R., *J. Natl. Biotech. Bioeng.* 2005, **90**, 491-500.

12. Michael C.N., Susan C.R., *Biotech. Bioeng.* 2005, **90**, 491-500.

13. Yuan Y.J., Li C., Hu Z.D., Wu J.C., *Enzyme Microb. Tech.* 2002, **30**, 774-778.

14. Huang Y.F., Lan W.Z., Chen C., Yu L.J., *Process Biochem.* 2005, **40**, 2793-2797.

15. Hajnos M.L., Zobel A.M., lowniak K.G., *Phytomedicine* 2001, **8**, 139-143.

16. Xu F., Tao W.Y., Cheng L., Guo L.J., *Biochem. Eng. J.* 2006, ***31***, 67-73.

17. Nicolaou K.C., *J. Am. Chem. Soc.* 1995, **117**, 624-629.

18. Nicolaou K.C., *Nature* 1994, **367**, 630-634.

19. Nicolaou K.C., *J. Am. Chem. Soc.* 1997, **119**, 7960-7973.

20. Mandai T., Kuroda A., Okumoto H., Nakanishi K., Mikuni K., Hara K.J., Hara K.Z., *Tetrahedron Lett.* 2000, **41,** 243-246.

21. Hu Z., Ou L.L., Zhang G.L., Yu Y.P., *Chinese Chem. Lett.* 2008, **19**, 130-132.

22. Ganesh T., Norris A., Sharma S., Bane S., Alcaraz A. A., Snyder J. P., Kingston D.G. I., *Bioorg. Med. Chem.* 2006, **14**, 3447-3545.

23. MacroModel, version 9.6, Schrödinger, LLC, New York, NY, 2009.

24. Glidel, version 5.5, Schrödinger, LLC, New York, NY, 2009.

25. Prime, version 2.1, Schrödinger, LLC, New York, NY, 2009.

26. Buey R.M., Barasoain I., Jackson E., Meyer A., Giannakakou P., Paterson I., Mooberry S., Andreu J.M., Diaz J.F., *Chemistry & Biology*, 2005, **12**, 1269-1279.

27. Wang Z., Yang D., Mohanakrishnan A.K., Fanwick P.E., Nampoothiri P., Hamle E., Cushman M., *J. Med. Chem.* 2000, **43**, 2419-2429; Verdier-Pinard P., Wang Z., Mohanakrishnan A.K., Cushman M., Hamel E., *Molecular Pharmacology,* 2000, **57**, 568-575.

Chapter 5

1. Schomer-Miller B. et al, *J. Biol. Chem,* **281**, 15268-15276.

2. Zand E., Rothwarf D.M., Delhase M., Hayakawa M., Karin M., *Cell*, 1997, **91,** 243-252.

3. Ghosh S., Karin M., *Cell,* 2002, **109**, S81-S96.

4. Dejardin E., *Biochem. Pharmacol,* 2006, **72**, 1161-1179; Lawrence T., et al, *Nature*,

2005, **434**, 1138-1143.

5. Strnad J., Burke J.R., *Trends Pharmacol Sci.* 2007, **28**, 142-148.
6. Karin M., Cao Y., Greten F.R., Li Z.W., *Nat Rev. Cancer* 2002, **2**, 301-310.
7. Burke J.R., et al, *J Boil. Chem.* 2003, **278**, 1450-1456.
8. Podolin P.L., *J. Pharmacol. Exp. Ther.* 2005, **312**, 373-381.
9. Ziegelbauer K., *Br. J. Pharmacol,* 2005, **145**, 178-192.
10. Wen D., et al, *J. Pharmacol. Exp. Ther,* 2005, **317**, 989-1001.
11. Karin M., Yamamoto Y., Wang Q.M., *Nature Review| Drug Discovery,* 2004, **3,** 17-26.
12. Murata T., Shimada M., Kadono H., Sakakibara S., Yoshino T., Masuda T., Shimazaki M., Shintani T., Fuchikami K., Bacon K.B., Ziegelbauer K.B., Lowinger T.B., *Bioorganic & Medicinal Chemistry Letters*, 2004, **14**, 4013-4017.
13. http://cellsignaling.lanl.gov/structure/kinase/kinase.html.
14. Prime, Version 2.0, Schrodinger, LLC, New York, NY, 2008.
15. Modeller, Version 9V1.
16. Glide, version 5.0, Schrodinger, LLC, New York, NY, 2008.
17. MacroModel, Version 9.6, Schrodinger, LLC, New York, NY, 2008.
18. Akella R., Moon, T.M., Goldsmith E.J., *Biochimica et biophysica acta* 2008. **1784**, 48-55.

Chapter 6

1. Gilman H., Straley J.M., *Recueil. Des. Travaux. Chimiques Des. Pays Bas* 1936, **55,** 821; Karash M.S., Tawney P.O., *J. Am. Chem. Soc.* 1941 , **63** , 2308; Gliman H., Jones R.G., Woods L.A., *J. Org. Chem*. 1952, **17,** 1630.
2. Krauss S.R., Smith S.G., *J. Am. Chem. Soc.* 1981, **103**, 141-148.
3. Bertz S.H., Carlin C.M., Deadwyler D.A., Murphy M.D., Ogle C.A., Seagle P.H., *J. Am.Chem.Soc.* 2002, **124**, 13650-13651; Bertz S.H., Cope S., Murphy M., Ogle C.A., Taylor B.J., Accompanying paper.
4. Frisch M.J., et al. *Gaussian 03*, revision A.1.; Gaussian, Inc.: Wallingford C.T., 2003.
5. Peralta J.E., Scuseria G.E., Cheeseman J.R., Frisch M.J., *Chem. Phys. Lett.* 2003, ***375,*** 452-458; Peralta J.E., Barone V., Scuseria G.E., Contreras R.H., *J. Am. Chem. Soc.* 2004, **126**, 7428-7429.
6. Deng W., Cheeseman J.R., Frisch M.J., *J. Chem. Theory. Comp*. 2006, **2l**, 1028-1037.
7. Ramsey N.F., *Phys. Rev.* 1953, **91**, 303-307.