2009

DEVELOPMENT OF HINT BASED COMPUTATIONAL TOOLS FOR DRUG DESIGN: APPLICATIONS IN THE DESIGN AND DEVELOPMENT OF NOVEL ANTI-CANCER AGENTS

Ashutosh Tripathi
Virginia Commonwealth University

DEVELOPMENT OF HINT BASED COMPUTATIONAL TOOLS FOR DRUG DESIGN: APPLICATIONS IN THE DESIGN AND DEVELOPMENT OF NOVEL ANTI-CANCER AGENTS

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

By

ASHUTOSH TRIPATHI
M.S. (Cheminformatics), University of Manchester, U.K., 2004
B.Pharm., M.J.P. Rohilkhand University, India, 2002

Director: GLEN E. KELLOGG, Ph.D.
ASSOCIATE PROFESSOR, DEPARTMENT OF MEDICINAL CHEMISTRY
INSTITUTE FOR STRUCTURAL BIOLOGY AND DRUG DISCOVERY

Virginia Commonwealth University
Richmond, Virginia

July 2009

I'm all yours.

Acknowledgement

I have worked with a great number of people whose contributions in different ways have influenced my life, my academic evolution, and overall development as a human being. It is a pleasure to convey my gratitude to all of them in my humble acknowledgment.

In the first place I would like to express my sincere gratitude to Dr. Glen E. Kellogg for his supervision, advice, and guidance in this research as well as giving me invaluable experiences through out the work. I consider myself extraordinarily fortunate in having Dr. Kellogg as my mentor. I can keep writing endlessly about how he has influenced my life and enrich my growth as a student, a researcher and a scientist I want to be. I am indebted to him more than he knows.

I gratefully acknowledge my committee members Dr. Martin K. Safo, Dr. John C. Hackett, Dr. David A. Gewirtz and Dr. J. Neel Scarsdale for sitting on my graduate committee. I am grateful in every possible way for their advice, supervision and support throughout my research. I also take this opportunity to thank my collaborators and coauthors in the research Dr. John T. Gupton, Dr. Susan L. Mooberry, Dr. Ray M. Lee and Dr. Daniele Simoni for generously sharing invaluable data with us. It was a great

honor to work with an exceptionally experienced team of scientists and I hope we continue our collaboration in the future.

I would also like to thank in particular Dr. Micaela Fornabaio, Dr. Philip D. Mosier, Dr. Alexander Bayden, Dr. Vishal Koparde and Dr. Kakali Sen for sharing their valuable research skills and experiences with me. I would forever be grateful for their scientific discussions, advice and precious time they granted to solve even my basic problem. I will always be indebted for their trust, patience and fruitful collaborations.

I am truly blessed with wonderful friends and I take this opportunity to heartily thank them for their love and friendship they bestowed on me. I would start in chronological order to express my gratitude. I would like to thank my B.Pharm friends Apurva, Arvind, Ajitabh, Gaurav, Sagar, Shashank, Ruchir, Shubha, Ritu, Rajiv, Keerti, Amita, Shruti…(well the list is practically endless) for being with me and supporting me in my tough times. Their continuous support and inspiration motivated me to go for higher studies. I would also like to thank my friend Majeed (from my masters' degree) for his valuable inputs on 'ideas' n' 'opportunities' in computation and chemistry. It is with immense pleasure I thank all my friends here who have made my stay in Richmond memorable. I would like to thank all my roommates and frequent visitors starting with Gaurav, Suhas, Kapil, Niraj, Preetpal, Jay, Sofia, Punam, and Koyal for wonderful

evenings, leisure trips and movie nights and of course amazing food we shared together. Furthermore, I would also like to thank my Biotech gang of Aurijit, Pooja, Anuja, Arjun, Chandravel, Tamara, Soumya, Sayali, Jigar, Max, Hardik and Garreth for keeping life interesting during work hours. Well the list is endless but how can I forget all my friends from 'VCU desis', Tiranga, Orkut and Facebook for scraping me and keeping me updated about the latest events. I could never have embarked and started all of this without the love and support of my friends.

I would also like to thank again Prajakta, Kakali and Vishal for meticulously proofreading my dissertation draft and giving me critical comments about it. I deeply appreciate them for their precious time, patience and suggestions.

I would also take this opportunity to thank my master's thesis advisor Dr. Vasudevan Ramesh for being a constant source of inspiration and encouragement and shaping up my academic career.

A very special thanks to my family especially my 'Mama', 'Mami' (Uncle and Aunt), Vaibhav and Sushmita for their inseparable support and prayers. They deserve a special mention for their selfless love and sacrifice they made for me. I wish my parents and my sister Nandu were with me to see this day.

Finally, I would like to thank Department of Medicinal Chemistry, School of Pharmacy and Institute for Structural Biology and Drug Discovery (ISBDD), VCU for providing the financial support and all the resources necessary for successful realization of this dissertation and a wonderful learning environment.

Table of Contents

List of Tables

Page

List of Figures

Page

ABSTRACT

DEVELOPMENT OF HINT BASED COMPUTATIONAL TOOLS FOR DRUG DESIGN: APPLICATIONS IN THE DESIGN AND DEVELOPMENT OF NOVEL ANTI-CANCER AGENTS

By

Ashutosh Tripathi Ph.D.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2009

Major Director: Glen E. Kellogg, Ph.D.
Associate Professor, Department of Medicinal Chemistry
Institute for Structural Biology & Drug Discovery

The overall aim of the research is to develop a computational platform based on HINT paradigm for manipulating, predicting and analyzing biomacromolecular-ligand structure. A second synergistic goal is to apply the above methodology to design novel and potent anti-cancer agents.

The crucial role of the microtubule in cell division has identified tubulin as an interesting target for the development of therapeutics for cancer. Pyrrole-containing molecules derived from nature have proven to be particularly useful as lead compounds for drug development. We have designed and developed a series of substituted pyrroles that inhibit growth and promote death of breast tumor cells at nM and μM concentrations in human breast tumor cell lines. In another project, stilbene analogs were designed and developed as microtubule depolymerizing agents that showed anti-leukemic activity. A molecular modeling study was carried out to accurately represent the complex structure and the binding mode of a new class of tubulin inhibitors that bind at the αβ-tubulin colchicine site. These studies coupled with HINT interaction analyses were able to describe the complex structure and the binding modes of inhibitors. Qualitative analyses of the results showed general agreement with the experimental *in vitro* biological activity for these derivatives. Consequently, we have been designing new analogs that can be synthesized and tested; we believe that these molecules will be highly selective against cancer cells with minimal toxicity to the host tissue.

Another goal of our research is to develop computational tools for drug design. The development and implementation of a novel cavity detection algorithm is also reported and discussed. The algorithm named VICE (Vectorial Identification of Cavity Extents) utilizes HINT toolkit functions to identify and delineate a binding pocket in a protein. The program is based on geometric criteria and applies simple integer grid maps to delineate binding

sites. The algorithm was extensively tested on a diverse set of proteins and detects binding pockets of different shapes and sizes.

The study also implemented the computational titration algorithm to understand the complexity of ligand binding and protonation state in the active site of HIV-1 protease. The Computational titration algorithm is a powerful tool for understanding ligand binding in a complex biochemical environment and allows generating hypothesis on the best model for binding.

CHAPTER 1

Computer-Aided Drug Design: Introduction and Applications in Anti-Cancer Drug Design

1.1 Introduction

1.1.1 Cancer Therapy

Cancer is a group of diseases in which there is an uncontrolled multiplication and spread of the body's own cells within body in abnormal forms. Cancer may affect almost any tissue of the body and may metastasize to other tissues within the body. If the spread is not controlled, it can result in death. According to the American Cancer Society's 'Global Cancer Facts & Figures 2007' and 'Cancer Facts & Figures 2008' cancer is the second leading cause of deaths after heart disease in developed countries and third leading cause of death in developing countries[1,2]. According to a report released by WHO, it is estimated that there will be 16 million new cases every year by 2020 and by 2050, the global burden is expected to grow to 27 million new cancer cases[3,4]. Despite considerable progress in its diagnosis and treatment cancer continues to be one of the major health and socio-economic problems.

Over the last couple of decades, research has revealed considerable information about the molecular biology, pathobiochemistry and the intricate pathways involved in cancer[5,6]. Multiple factors are involved in the initiation, promotion and progression of cancer which lead to changes in the host genome and aberrant expression of oncogenes or tumor suppressor genes[7-9]. The abnormal cell multiplication and spread can be attributed to both external and internal factors. External factors like tobacco, chemicals, some viruses and radiations can cause mutations. These mutations cause altered gene expression which may activate protooncogenes to oncogenes like erbB, ras, myc etc.[8,10,11]. This, in turn, may result in uncontrolled cell proliferation and dedifferentiation. In-addition, inherited mutations in metabolism and certain immune conditions can cause expression of oncogenes[11,12]. These inadvertent factors may act together or in sequence to initiate or promote carcinogenesis and develop primary tumor, which, subsequently develops into a full fledged malignant tumor.

Over the past decade, a number of chemotherapeutic drugs with different mechanism of actions and targeting various stages of metastatic cell growth have flooded the pharmaceutical market[13,14]. Based on their mechanism of action, these drugs can be classified into different classes as alkylating agents, antimetabolites, antibiotics, nucleoside analogues, antimitotic agents, etc[15]. However, most of these drugs are associated with severe toxicities and are not effective against all types of cancer. Thus, the search for new anticancer drugs and the development of more effective treatment strategies continues to be imperative[16].

Until recently, anti-cancer drug discovery had extensively focused on the critical cell growth stage, mitosis, and any alteration or miscontrol of which can lead to development of human tumors. After the clinical success of taxol, microtubules have received considerable attention as the potential targets in the drug discovery process[17-19]. Microtubules form the cytoskeleton of a cell and are critical in mitosis and cell division. Tubulin, a heterodimeric protein forms the fundamental structural unit of a microtubule and joins in a head-to-tail fashion to form a long, filamentous protofilament. These protofilaments join laterally to form a hollow tube shaped protein polymer, the microtubule, which forms a mesh like network in eukaryotic cells. Besides playing an important role in mitosis, microtubules are involved in diverse cellular processes such as locomotion and intracellular transport. In addition, they play an important role in the development and maintenance of cell shape, signaling and transport of cellular components such as vesicles, mitochondria, etc. Thus, microtubules become an important target for anticancer drugs[18]. Since chemically diverse groups of anti-mitotic drugs that target microtubules and induce mitotic arrest have been used with great success against cancer, microtubules are considered as one of the best identified cancer targets[17,18]. However, research in this area is held back due to lack of a high resolution crystal structure that will aid in efficient structure-based drug design. Although there are several drugs which bind to tubulin protein and disrupt microtubule dynamics[19], their clinical usefulness is limited by their unfavorable pharmacokinetic profiles and side effects. Furthermore, the limited availability and complexity in synthesis of pharmacologically active lead compounds and their chemical modification presents a major obstacle in improving the overall profile of a drug candidate. The complexity in development of anti-

cancer chemotherapeutics makes it difficult to treat the disease and as a result it is necessary to come up with a rational program to identify potent leads that will stimulate the development of more effective anti-cancer chemotherapeutics.

1.1.2 Drug Discovery

Traditionally the strategy for discovering new drugs consisted of random screening of thousands of compounds derived from natural products to identify a lead[20]. Taking this lead structure, a program is developed for finding analogs exhibiting the desired biological properties. However, this entire process is highly laborious, expensive and conceptually inelegant. During last couple of decades pharmaceutical industry has experienced a paradigm shift in its approach to discover drugs. The traditional methods of drug discovery are now being supplemented by some more direct approaches which are derived from the understanding of the molecular processes involved in the underlying disease. Pure samples of protein targets are now being isolated and the three dimensional structure of both ligand and target site may be determined by X-ray crystallography or computational methods. It is now possible to learn how precisely topography of structures control the regulation of life processes. In order to further such progress, a rational approach to drug discovery has emerged in the pharmaceutical industry and has contributed to the rapid development of active candidates. New molecules are conceived either on the basis of the similarities they share with the known lead compound or their complementarities with 3D structures of known active sites[21-23].

Understanding the three-dimensional aspects of drug-receptor interactions and their specificity at the molecular level has become a focal point in the modern drug discovery. Recent advances in genomics, bioinformatics, high-throughput screening and combinatorial synthesis are providing valuable inputs for the small molecular leads. However, optimizing the chemical and biochemical properties of a lead warrants substantial resources, which, in turn, will assist in the rational selection of "druggable" chemical entities, i.e., the selection of drug candidates/chemical entities that are more likely to have favorable characteristics for the treatment of human disease, thus improving the efficiency of the drug discovery/development process[24].

The drug discovery process is an interplay between computational and experimental approaches. Computer-aided drug design contributes to the understanding of biomolecular processes in a qualitative and quantitative way. It not only presents means for analyzing the details of the molecular machinery involved in a system and understanding the way the biological system functions, but also, provides the tools for predicting potential possibilities of the prototype candidate molecules. The techniques currently available provide extensive insights into the precise molecular features that are responsible for the regulation of the biological processes. These structural and physicochemical characteristics are of primary importance in understanding the structure –activity relationships and hence, the rational design of drug. Computer-aided drug design has opened the way to the discovery of lead molecules by a rational approach, and its central role in rational drug design has become fully apparent[25].

1.1.3 Computer-Aided Molecular Design

Computer-aided molecular design is expected to contribute to the discovery of novel molecules conceived on the basis of precise three dimensional stereochemical and physicochemical considerations. Systematic investigation of 3D stereochemical and physicochemical features of a protein and its binding site are crucial for designing small molecules that modulate protein functions. The ability to accurately predict binding mode from computer simulations is an invaluable resource in understanding biochemical process and drug action. Several aspects of molecular recognition can be discerned from the computational modeling of protein-ligand complex. Predicting the affinity of a new putative inhibitor for a macromolecular receptor is therefore imperative for designing of more potent analogs. Though structural data available from x-ray crystallography and NMR has undeniably encouraged the efforts towards understanding the biological complexity of molecular recognition, fundamental uncertainties in binding site interactions and insufficient knowledge of well defined binding pockets are a major bottleneck in designing effective drugs with optimal activity.

One important aspect in modeling bio-molecular systems or molecular recognition events is to accurately model the energetics of the binding. Although computational tools have been used to corroborate experimental data, the interaction energies calculated without the pH, ionization and entropic contributions are not always expected to correlate well with the experimentally measured free energies of binding. Subtle variations in microenvironment of a protein due to changes in experimental conditions such as pH,

buffer, ionic strength and temperature can influence the bio-molecular interactions and complex formation[26]. The primary goal of modeling studies is to accurately simulate the microenvironment of an active site. Calculation of binding free energy involves evaluation of both enthalpic and entropic contributions and forms an integral part of structure-based drug design protocol[27,28]. A meaningful description of the system can be obtained from thermodynamic analysis of protein-ligand interaction. To date, several structure-based modeling methods have been developed that give vital insight into the free energy changes of the system. Some of these methods viz. Free Energy Perturbation (FEP), Linear Interaction Energy (LIE), and Molecular Mechanics Poisson-Boltzmann Surface area (MM-PBSA) calculations involve extensive sampling of conformational states. However, the computational cost and human interaction necessary to perform these calculations make them less applicable for screening of a very large number of compounds and more viable in a later phase of drug design. These models can contribute significantly to the understanding of structural and energetic basis of protein-ligand interactions and hence, the structure-based design of novel compounds. Virtual screening methods involving database searching and docking tools along with *de novo* drug design tools facilitate identification and optimization of lead candidate. This research proposes the development and application of such tools to identify new anti-cancer drugs. In the subsequent subsections, the role of different computational approaches and the research plan to design and develop novel anti-cancer agents will be discussed. In section 1.2 we start with a brief review of approaches and technologies used in structure-based drug design. In the subsequent section 1.3 we highlight the HINT model and describe the Toolkit design. In the next section 1.4 we present a research plan for the design and

development of computational tools and their application in designing novel anti-cancer agents.

1.2 Structure-Based Drug Design

The purpose of structure-based drug design is to identify or construct molecules that bind with high affinity to a structurally defined binding site of a target protein. It is the process of using the three-dimensional structural information from a macromolecular target or/and ligand-target complexes to design novel drugs that may modulate target protein for desired activity. Recent computational approaches facilitate extraction of all the relevant information from available structures to understand specific molecular recognition events and to elucidate the function of the target macromolecule[29].

Structure-based drug design uses an efficient and intelligent approach to design improved ligands for the target[24]. The first step in structure-based drug design is the elucidation of three-dimensional structure of the target macromolecule (protein or nucleic acid) by X-ray crystallography or NMR. In some cases, where direct structural information is unavailable, a homology model can be used as the starting point. However, there are instances where the function of a protein is unknown or the domain responsible for the activity is unidentified. Several methods are available for identification and characterization of the active site[30,31]. These programs can be sorted into different categories according to the approach they take to identify and delineate the active site (i) Evolutionary methods (structure/sequence alignments) (ii) probe/energy-based methods

(iii) geometric approaches. All these methods are likely to evolve with the availability of more structural and sequence data from structural genomics projects.

Recent advances in combinatorial chemistry, high throughput chemical synthesis and screening are providing valuable inputs for the identification of small molecular leads[21,22,32]. Once a lead compound has been identified, an iterative process of lead optimization begins that involves solving the three-dimensional structure of the lead compound bound to the target. Development of new computational drug design tools facilitates examination and characterization of the complex structure and the types of interactions that the bound ligand makes with the protein. The binding site specifies structural and physicochemical constraints that must be met by any putative ligand. Hence, it is imperative to analyze the binding site by mapping the characteristics that are essential for ligand recognition. Besides shape complementarity, which plays an important role in protein–ligand interactions, physicochemical complementarity is essential for the specificity of binding[33].

Many promising approaches towards the goal of automated ligand design and optimization have been reported[34-37]. In particular, there has been a surge in new methods in the past few years, greatly extending the approaches to ligand design. Lead optimization methods may involve improvements to existing lead compounds by introducing new functionality to the lead scaffold. A combinatorial library, thus generated, is screened against the target protein. Although, the approach is rational and has been successfully applied, the analogs generated are very similar to the lead

compound and this limits the discovery of diverse library of compounds. In another widely used and more successful approach, large databases of compounds are virtually screened[23]. It is a two pronged approach, in which, while maintaining pharmacophoric constraints on one hand, new lead compounds with different scaffolds can be identified, thereby, introducing diversity to the compound library. These virtual screening methods take each proposed ligand and attempt to position it in the active site of the receptor, or match it to a pharmacophore model[38]. The compounds are scored and ranked based on their steric and physicochemical complementarity with the target site and the best compounds are tested with biochemical assays[39]. Another promising approach to generate entire novel series of compounds is *de novo* lead generation programs[36]. In contrast to the whole molecule docking approach, in *de novo* methods, fragments of molecules, usually small functional groups are docked into the site, scored, and linked together using different scaffolds. The methodology allows the diverse set of fragments to exhaustively explore the binding site. The final *in silico* library of compounds can be scored, ranked and synthesized in the laboratory. However, the main drawback with this approach is the stability and synthetic feasibility of the compound suggested. There are many excellent drug design software methods available capable of either virtual screening or *de novo* generation[36,40]. However, the main advantage of virtual screening from a database is that the hit compounds can normally be purchased and tested easily. The success of all these approaches depends on how well the protein-ligand complex is characterized and scored. Various techniques have proven to be efficient tools for generating near-native conformations of complexes[41]. However, there are still some inherent limitations that need to be addressed. The robustness of any computational approach depends on how

accurately the experimental information is derived and parameterized to simulate a biological system. Due to the multiplicity of factors controlling the binding in any biological system, the phenomenon is particularly difficult to model computationally. Thus the most demanding and rewarding answer to this problem is hoped to lie in rational drug design.

1.2.1 Microenvironment & Active Site

Molecular interactions are regulated by subtle recognition and discrimination patterns where three dimensional features and microenvironment of the active site plays a vital role. The first step in any rational drug design protocol is to identify and elucidate the active site in a protein molecule. These active sites themselves might not be observed from an initial inspection. Protein surfaces are formed by numerous cavities and protrusions that are interlinked through small narrow channels and more than often interspersed with numerous holes/voids. The size and shape of protein cavities dictates the three-dimensional geometry of ligands and guides the important intermolecular interactions that mediate binding. The study of cavities may give an insight into the mechanism of such interactions and might help in the design of novel ligands, substrates or inhibitors. The determination of binding pockets is, therefore, an important step towards the rational design of novel ligands. An in-depth analysis and classification of pockets on the surface of protein structures might also improve our understanding of the processes involved in structure-based drug design.

A number of successful methods have been developed for predicting ligand binding sites[30]. These methods basically take different approaches depending on the kind of information available. Evolutionary methods apply a heuristic approach by finding characteristic sequence patterns based on structure/sequence alignments. The presence and evolutionary conservation of certain residue pairs, cofactors, metal ions and binding motifs can, therefore, give useful clues for deducing the biochemical function of an uncharacterized protein and location of the binding sites. These computational methods are useful for predicting the location of binding sites when direct experimental information is unavailable.

Methods that use structural information, when available, are particularly promising since they can potentially identify and characterize an active site based on the properties of the pocket. These methods apply energy/probe to generate functional maps of the binding site indicating the most favorable regions for placing ligand groups with similar properties to the probes. The atomic details can be further evaluated by mapping the physicochemical properties of the binding surface such as hydrophobicity or electrostatic potentials.

The properties/microenvironment of the cavity are responsible for the substrate specificity. Theoretical calculations simulating the microenvironment of binding site are plagued by very drastic approximations. The phenomena of solvation/desolvation and the effects of varying pH and electrostatics are difficult to model computationally. Under physiological conditions a protein (enzyme/receptor) is present in a milieu of solvents,

even after crystallization a protein crystal may contain upto 70% of solvent[42], which includes the buffer solution as well. Exposure of protein crystal structure to these solvents and buffers during crystallization may affect the microenvironment of active site as these solvent molecules may distribute themselves within the different pH zones or solvent channels according to the nature of the residues lining these solvent pockets. The altered microenvironment may influence the ligand binding[43] due to change in the protonation states of the active site residues. The binding of ligand/small molecule to biological macromolecules is further complicated by peculiarities in the metal ion binding, preferred coordination numbers, relative affinity for specific ions, and preferred binding orientation and the lowest energy ionization state. It is, thus, not always possible to accurately resolve the biological microenvironment experimentally or to model it computationally. Thus calculation of optimum ionization/protonation state of complex formation is crucial for understanding binding process and step towards desigining more selective ligands.

1.2.2 Mapping the Binding Site

Structural and physicochemical characterization of an active site has become a major goal in drug discovery[44]. Understanding the chemistry behind molecular recognition is a central issue in drug design. With the advances in experimental techniques of X-ray crystallography and NMR, the elucidation of binding features on a protein structure has become more convenient. The computational mapping of a binding site to generate a pharmacophore model directly from a protein crystal structure can reveal key elements in protein-ligand binding[45]. Such knowledge is indispensable for rational drug design, since in majority of cases, receptor–drug interactions are specific in nature. Correctly mapping

the binding site is an important step in structure-based drug design and can be used as the starting point for finding new lead compounds or drug candidates[46].

The structure-based mapping of binding site or receptor-based pharmacophore generation provides an ensemble of steric and electronic features that ensures optimal interactions with a specific biological target structure, although a pharmacophore can be calculated from both ligands or/and protein structure[47]. A receptor-based pharmacophore defines essential features for molecular recognition and receptor-ligand interaction[48]. Structure-based mapping uses features complementary to a protein site to define the shape and physiochemical constraint of the target site[49]. The mapping of crucial features is challenging since the number of 'Hot Spots' and their strength is critical for hypothesis generation. This facilitates the docking process by defining a set of constraints that can be quantified in terms of how many and which pharmacophoric points can be matched by a ligand or a library of compounds.

A Receptor-based pharmacophore can be defined as an arrangement in three-dimensional space of several (typically three to six) features considered to be relevant for specific binding[50]. These features can be specific such as hydrogen bond donor or acceptor, or aromatic moiety. These physicochemical attributes can be mapped on a cavity surface using different approaches based on surface conservation or energetic contouring with interacting molecular probes[51]. However, the challenge is to pinpoint the specific electrostatic or hydrophobic interactions complementary to the residues lining the active site. To some extent, probe-based methods have found success in defining the polar and

hydrophobic regions within the cavity[52]. The knowledge-based approach of assigning interaction sites have also been implemented (LUDI) where statistical analysis of hydrogen bond patterns and geometries in crystal packing is taken into account by generating an ensemble of interaction sites distributed over the region of acceptable geometries[53]. Similarly diffused/non-directional interactions such as hydrophobic interactions due to aliphatic or aromatic moiety can be assigned after study of its distribution pattern. This approach has the advantage that it is purely geometrical and therefore avoids costly calculations of potential energy functions[53].

While geometric matching and pinpointing of specific interactions between ligand and receptor atoms provides a very intuitive picture, it is clear from basic physical principles and from measurements that the energetics involved in the thermodynamic association and dissociation processes determines the binding affinity of the ligand, which, after all, is one of the key properties of a molecule in the drug discovery process. Pharmacophores do not encode these energetic aspects, but do provide a rough sketch of the binding pocket. The potential binding pockets are considered as binding sites if a small molecule can bind in pocket such that it can form sufficient energetically favorable interactions with the protein. Numerous other techniques and approaches have been developed to calculate binding energies[28,54]. This has important implications in understanding of the nature of protein interactions, identifying the suitability of sites as drug targets and for identifying critical regions for docking and structure-based drug design.

1.2.3 Docking

Docking calculations are needed to predict how new hypothetical or existing compounds will bind to the protein. Given the three dimensional structure of a receptor, known ligands can be docked to examine how they fit so that a protein ligand complex can be characterized in detail and modifications that improve binding can be suggested. The docking methodology has found widespread applications in drug discovery since large number of small molecules can be virtually screened by docking compounds from a database into a receptor structure in an attempt to find novel compounds[39,55].

The docking approach can be thought of as a three step process. The first step involves characterization of binding site. This step identifies and delineates the binding site for docking. It is a critical step as it defines and sets the constraints to position the ligand in the defined binding region. The second step is to correctly position a set of ligands in the binding site based on the defined constraints. This step involves exploration of the configurational and conformational space for the interaction between target and the ligand. This step tries to correctly predict and identify the most favorable binding mode of the ligand into the target active site. The third step is an elaborate process of energetically assessing the docked position of ligand and scoring and ranking the proposed protein-ligand complex.

Although the first step of site characterization is crucial for docking, it is generally not considered to be a part of docking methodology as it has grown and developed as a separate field of study. This broadly leaves the docking approach as a two component

approach of 'search' and 'scoring'. Several different approaches have been developed for investigating the accessible conformational space of a ligand[56,57]. One of the earlier approaches involved was systematic search[58]. However, the search becomes more and more complex with the ligand flexibility as the number of degree of freedom of the ligand molecule increases. Such an approach was implemented in methods where ligand and binding pocket were considered to be rigid and ligand was fitted using shape complementarity via point complementarity or distance geometry approaches[59,60]. In such docking methods, the shape of both the receptor site and the ligand is investigated based on shape and pharmacophoric points. Orientations are generated through various alignment procedures in order to maximize the pharmacophoric constraints and shape complementarity. However, it is not feasible to exhaustively explore the available conformational space and a right balance between speed and accuracy has to be made so that as many binding modes as possible are explored. Fragment-based approaches that work by either incremental construction of ligand in the binding pocket or just by placing and joining the fragment circumvent the problem of combinatorial explosion of generated conformers in the previous approaches[61-63].

Stochastic methods involving random sampling of conformational space of ligand in the binding pocket have also been widely applied in many docking algorithms[55]. Algorithms using Monte Carlo sampling coupled with Metropolis criterion are applied to exhaustively search the conformational space[64]. Simulated annealing protocol combined with grid-based energy evaluation can be coupled with this approach to overcome high conformational energy barrier in conformational sampling[65]. Another such stochastic

approach that has been successfully implemented in docking algorithm is the genetic algorithm based sampling of conformational space[66-68]. In this approach, multi-conformers referred as chromosomes are evaluated, crossed and mutated and the best possible solution is kept based on the fitness function. The solution is represented by the best scored conformation of the total conformers after a set number of generations. GOLD (Genetic Optimization for Ligand Docking) is the most applied and well known for flexible molecular docking[69].

In contrast to these systematic and stochastic approaches, molecular dynamics-based and heuristic *tabu* searches have been implemented to explore the sample space[56,70,71]. However, molecular dynamics is computationally expensive which has restricted its use in docking. To circumvent this exhaustive sampling, *tabu* search approaches were adopted where a list of already explored conformations were maintained and only unexplored spaces were sampled[72]. This avoids reinvestigating the space already sampled by associating the sampled conformations with a degree of penalty. Apart from these deterministic approaches hybrid consensus approaches have also been implemented that combine features from other two approaches[73,74]. Although these approaches can exhaustively and successfully sample and generate all the possible conformations within the active site, the success of any docking program depends on how well it can reproduce the experimental binding mode of ligand within the cavity. The idea is to successfully predict the most energetically favorable pose out of thousands of sampled conformers.

1.2.4 Scoring

The success of whole molecule docking, *de novo* construction of molecules into a target site, or screening large virtual combinatorial libraries is ultimately dependent on the accuracy of the scoring function that ranks the compounds. Ligand orientations can be evaluated on the fly as the ligand or fragment is positioned within the cavity or all the generated poses can be scored in the end. The scoring methods that are used in high-throughput settings dealing with thousands of diverse compounds can be evaluated by how well the corresponding relative binding affinities can be predicted. This need has led to the development of a wide variety of methods, which can be subdivided in four major approaches: force field-based methods, semi-empirical approaches, empirical scoring methods, knowledge-based potentials and lastly consensus scoring functions that are combination of multiple scoring functions[55,75].

Force field-based scoring methods generally use a molecular mechanics force field, which contains terms for intramolecular forces like bond, angle and dihedral terms between the atoms that are bonded to each other plus energy terms for intermolecular forces that describes the forces between non-bonded atoms such as van der Waals and Coulombic terms. There are a number of widely applied and successful molecular mechanics-based scoring functions[76-79]. Due to their simplicity, they are widely applied and suitable for use in virtual screening. Though faster and simpler, they are ideally not meant for simulating biomolecular interactions as these methods were developed for calculating gas phase enthalpy of binding. However, this class of scoring approach is

associated with many drawbacks and does not account for hydrophobic interactions, solvation and entropic effects.

Empirical scoring methods offer an alternative approach to pure molecular mechanics-based force field methods for scoring[80]. The underlying idea is that the binding free energy of a non-covalent protein-ligand complex can be factorized into a sum of localized and chemically intuitive interactions. The terms accounting for different contributions such as hydrogen bonds, hydrophobic interactions, entropic effects are normalized by weighting factors derived from regression analysis of the data from training sets comprised of well characterized protein-ligand complexes. Based on the assumption of additivity, the binding affinity is estimated as a sum of interactions multiplied by weighting factors and solved by equation of the type (1).

$$\Delta G_{binding} \approx \sum \Delta G_i f_i (r_l, r_p) \qquad (1)$$

Where f_i is a simple geometrical function of the ligand (r_l) and receptor (r_p) coordinates[55]. However, accuracy of these methods depends upon the quality of the experimental binding data and crystallographic structural data of the training set.

Semi-empirical scoring functions, however, combine the above two approaches and incorporates empirical or empirically calibrated energetic terms for interactions that cannot be computed by pure molecular mechanics-based methods. Thus, implicit binding energy terms such as hydrogen bonding, solvent effects, hydrophobicity and entropic

terms can be included in the scoring functions. In contrast to force field-based scoring functions, semi-empirical scoring terms can more accurately estimate the binding energy by including entropic and solvation effects that are known to significantly affect biological interactions in aqueous medium[81-84].

Knowledge-based scoring functions[85] are rule-based scoring functions where rules are derived from the analysis of structural data of known and well characterized receptor-ligand interactions. Exponential growth and availability of structural data on protein-ligand complexes has allowed deriving and formulating set of rules based on the frequency of interactions. Scoring functions of this type try to capture the knowledge about protein-ligand binding that is implicitly stored in the protein data bank by means of statistical analysis of structural data. The potentials are obtained by statistical analysis of atom-pairing frequencies observed in crystal structures of protein-ligand complexes[86]. However, the accuracy of knowledge-based scoring function depends on the quality of experimental data, as it incorporates structural knowledge without referring to inconsistency in experimental and structural data.

Though several approaches have been implemented in deriving a robust scoring function, none of the scoring functions are ideal, as a variety of approximations are made to make a compromise between speed and accuracy. Taking into consideration the limitations of a single scoring function, the concept of consensus scoring evolved based on the premise that combination of different scoring functions would probably lead to a better performance by overcoming inherent weaknesses in individual functions[87]. A consensus

between a set of scoring functions can be reached either by averaging the rank assigned by each scoring function or averaging the score value calculated by different functions. Although a number of scoring functions as well as their combinations have been implemented, ideally, the best scoring function should be able to discriminate between native and non-native binding modes and be able to calculate the actual free energy of binding.

1.2.5 Free Energy Prediction

Predicting the structure and binding affinity of ligand-receptor complex is a challenging process and forms an integral part of structure-based drug design protocol. Ligand binding is governed by kinetic and thermodynamic principles. The quantity of interest in determining binding constants is the free energy difference between the complexed and uncomplexed state.

$$\Delta G_{binding} = \Delta G_{complex} - \Delta G_{free} \qquad (2)$$

The free energy is measured with the Gibbs free energy (G) of the system. The standard free energy change of the ligand-receptor binding process can be expressed as ΔG°. The most common measurement for ΔG is through the equilibrium constant for the complex.

$$\Delta G_{binding} = -RT \ln K_{eq} = RT \ln K_d \qquad (3)$$

Where R is the gas constant, T is the absolute temperature in Kelvin. Since the equilibrium constant (K_{eq}) or dissociation constant (K_d) is a direct measure of ligand affinity; calculation of ΔG, the free energy of binding is of great interest in drug design. Computationally, the goal is to estimate binding free energy, $\Delta G_{binding}$, which can be directly related to the experimentally measured association constant (K_a). A useful way to consider the binding free energy is in terms of the changes in enthalpy and entropy on formation of complex, as expressed by the following equation:

$$\Delta G_{binding} = \Delta H - T\Delta S \qquad (4)$$

Enthalpic (ΔH) and entropic (ΔS) factors that contribute to ligand binding include the hydrophobic effect, van der Waal and dispersion interactions, hydrogen bonding, other electrostatic interactions and solvation effects.

During the complex formation, both receptor binding site and the ligand become partially desolvated and there is structural reorganization of solvent molecules between the receptor and ligand. In a majority of cases the solvent molecules are displaced between receptor and ligand thereby replacing existing hydrogen bonds between solvent and receptor or ligand. Also there is a change in entropy on binding as it is accompanied by change in translational and rotational degrees of freedom for the ligand, receptor and solvent molecules. In order to calculate the binding energy accurately, the scoring function should include all the enthalpic and entropic terms. So far, there are a wide

variety of different techniques available for predicting the free energy of protein-ligand binding, but most scoring functions do not include all of these terms[27,28,54,88].

Most of the functions are based on the assumption that different contributions to free energy of binding are additive and can be calculated separately. The free energy of binding can be partitioned and calculated as an additive contribution from different parts. The "master equation" can be written as:

$$\Delta G_{binding} = \Delta G_{solvent} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{motion} \quad (5)$$

Where, $\Delta G_{solvent}$ is the free energy contributions from the solvation or hydration free energy, ΔG_{conf} is due to conformational changes in the protein and ligand, ΔG_{int} is free energy due to specific protein-ligand complex as a result of their proximity and ΔG_{motion} is free energy as a function of motion of receptor and ligand[89].

Ideally, all the energy contributions should be ensemble averages as the complex, the free protein, and the ligand are dynamic entities and these free energies can be calculated by integrating over all possible configurations of the protein, ligand and solvent system. Free energy perturbation methods allow direct calculations of certain differences in free energies[90]. The basic idea of free energy perturbation is derived from statistical mechanics as it relates the free energy of a system and ensemble average of an energy function that describes the system. However, free energy perturbation still requires relatively large amounts of computer time, provides very limited conformational

searching and remains technically difficult. Consequently, most approaches use approximate methods[54,89].

The force-fields used for calculations of free energy and intermolecular interactions assume that steric and electrostatic forces are sufficient to account for the observed biological interactions. However, these alone are never sufficient in accurate prediction of biomolecular interactions as they do not always include solvation/desolvation effects. The entropic contributions to binding are much less well-defined and often poorly quantitated or even ignored in most of the cases. Most approaches sum up these interactions separately as distinct enthalpic and entropic contributions, whereas, in reality, the ligand-protein recognition is a concerted event and thermodynamic quantities cannot be just simply summed.

1.3 HINT Model

This work focuses on an alternative force field-based on the experimental information from log $P_{o/w}$ (partitioning coefficient for water/1-octanol). Since log $P_{o/w}$ is an experimentally derived thermodynamic quantity, it directly correlates with the free energy of interaction and encodes all non-covalent interactions in the biological environment as well as solvent effects and entropy[91,92]. The HINT model describes specific atom–atom interactions between two molecules, using the equation:

$$H_{TOT} = \sum\sum b_{ij} = \sum\sum (a_i S_i \, a_j S_j R_{ij} T_{ij} + r_{ij}) \qquad (6)$$

where *a* is the hydrophobic atom constant (derived from log $P_{o/w}$), *S* is the solvent accessible surface area, *T* is a function that differentiates polar – polar interactions (acid – acid, acid – base or base – base), and *R* and *r* are functions of the distance between atoms *i* and *j* . The interaction score b_{ij} describes the specific atom–atom interaction between atoms *i* and *j*, and H_{TOT} represents the total HINT interaction score. The HINT model is based on the assumption that each b_{ij} approximates a partial δg value for a specific atom–atom interaction, thus the total HINT score, which is the cumulative sum of all the atom–atom interactions, correlates with $\Delta G_{interaction}$, i.e., $\Delta G = \sum(\delta G)$. The HINT paradigm has been extensively applied for predicting the free energy of binding for protein –ligand and other systems[91,93-95].

1.3.1 HINT Hypothesis

Hydropathic interactions comprising of polar and hydrophobic interactions form the very foundation of intra and intermolecular interactions. Although polar interactions can be accurately quantified using quantum chemistry methods, hydrophobic interactions are hard to quantify. The thermodynamic factors which give rise to the hydrophobic effect are complex and still incompletely understood. The hydrophobic effect is considered to be the major driving force in protein folding resulting in the burial of the hydrophobic residues in the core of the protein. It is exemplified by the fact that oil and water do not mix and oil droplets coalesce together to form a separate layer. The Hydrophobic effect, which is considered to be mainly entropy driven, is also responsible for protein-ligand interaction and bio-macromolecular association. However, commonly used molecular mechanics-based forcefields do not include terms for quantifying entropic energy. There

are, however, computationally intensive free energy perturbation methods that are able to accurately calculate free energy of binding[90,96].

The HINT forcefield is based on empirically-derived Log $P_{o/w}$ values that intuitively estimate free energy of binding, ΔG[96]. Log $P_{o/w}$ is a thermodynamic quantity that encodes both enthalpic and entropic term. These include hydrogen-bonding, acid-base interactions, coulombic attractions as well as hydrophobic interactions. All of these are related to solvent partitioning phenomena and the free energy of transfer of a compound from one reference state, such as an organic solution, into water, involves free energy change comprised of enthalpic and entropic contributions. The two phases, *i.e.*, octanol and water, correspond to hydrophobic and polar residues and also are analogous to the hydrophobic microenvironment inside the binding cavity and solvent system. Since the dissolution of any compound in a mixed solvent system involves the same fundamental processes as biomolecular interactions within or between proteins and ligands, HINT derives the information from bulk molecular solvent partitioning to discrete interactions between the atoms. The program calculates empirical atom-based hydropathic parameters (hydropathic atom constant) that encode all significant intermolecular and intramolecular non-covalent interactions implicated in drug binding or protein folding. The hydropathic atom constant forms the core of HINT paradigm.

The hydrophobic atom constant, *a*, is the key parameter in the HINT model. The atom constant, *a*, is calculated using an adaptation of the hydrophobic fragment constant approach of Hansch and Leo[97]. However, contrary to the Hansch and Leo approach, that

calculates the total solvent partition constant for a molecule by summing up the fragment constants, HINT backtracks from this point and distributes and assign atom constant to each atom of a fragment or molecule. Facilitating this are two databases for small molecules and bio-macromolecules, with re-parameterized force-field atom types, modified factors for bond, ring, and branched chain fragment data. The molecular database is used for structure/connection based calculation of Log P. The database contains a data tree of partition information and a bare hydrophobic atom constant for each atom type. These bare hydrophobic atom constant values can be modified by a variety of factors specific to the molecular environment. A series of simple rules based on atom connectivity and proximity to other atoms identify and apply these modification factors to atom constants. In case of bio-macromolecules, dictionary-based calculations are carried out for log P depending upon the acidic, basic or neutral solvent condition where the modification rules are implicitly applied. Hydrogen atoms can be treated explicitly to model interactions more accurately. These modification factors characterize actual biophysical phenomena related to the molecular structure and properties. These hydrophobic atom constants along with modification factors encode the thermodynamic information and reveal the potential type and strength of interaction that the atom may encounter. Coulombic, hydrogen bonding, dispersion as well as hydrophobic effects may be extracted from the hydrophobic atom constant by examination of the sign and magnitude of the constant.

1.3.2 HINT Calculation

The basic principle and application of HINT methodology is calculation of experimental, information-rich hydropathic atom constants and calculation of atom-atom interactions in a bio-molecular system. The partition constant LogP is a thermodynamic parameter, which due to its unprocessed and unbiased experimental nature, contains interaction information specific to the biological environment as well as solvent effects and entropy[98,99]. HINT was created specifically to calculate all non-covalent interactions. In practice, the conceptually simple HINT model scores each atom–atom interaction within or between biological molecules according to equation (6). As mentioned b_{ij} is the interaction score between atoms i and j, a is the hydrophobic atom constant, S is the solvent accessible surface area (H_2O probe), T_{ij} is a logic function describing acid-acid, acid-base and base-base interaction, and R_{ij} and r_{ij} are functions of the distance between atoms i and j (i.e. r). Generally the hydropathic-dependent function, R_{ij}, is the simple exponential e^{-r} and r_{ij} is an implementation of the Lennard- Jones potential function. The r_{ij} term is mostly a penalty function to prevent van der Waals violations. The double sum, $\sum\sum b_{ij}$, is the total interaction score for the system. The HINT convention is that favorable interactions are scored with $b_{ij} > 0$ and unfavorable interactions are scored with $b_{ij} < 0$. The logic function T_{ij} returns a value of 1 or –1 depending on the character of the interacting polar atoms (i.e. $a < 0$): there are three possibilities: acid–acid, acid–base, or base–base; only acid–base is scored favourably. Since polar moieties are markedly differentiated from hydrophobic moieties, the explicit interactions between a ligand and its receptor can be predicted. Besides this, the general microenvironment of the binding site can also be characterized. Likewise, and more exciting, complementary binding

points within the binding pocket can also be derived, giving a deep insight as to what an ideal ligand would look like. The HINT forcefield and its free energy scoring functions have been extensively applied and validated and have shown to be useful in a wide variety of biomolecular simulations[93,100-102]. To further extend the capability and applicability of HINT program, all the algorithms have been incorporated in a toolkit[103].

1.4 Amalgamation of Experimental and Theoretical Approaches: Scope and Limitations

The field of molecular modeling lies at the interface of experimental and theoretical approaches. Computational modeling techniques attempt to simulate an abstract model of a particular system and are often plagued by very drastic approximations. The problem is further compounded by inherent experimental errors and artifacts. While, structural data available from x-ray crystallography and NMR has undeniably encouraged the efforts towards understanding the structure and functions of protein, only partial agreement between experimental and theoretical data has been observed based on these crystal structures. Often the issue raised is how representative is a single crystal structure in a structure-based drug design project? In addition to the obvious correlation between crystal structure resolution and accuracy, there are other considerations such as the reliability of place water molecules and thermal motion of unconstrained residues. By and large it has been difficult to correlate solution binding data, crystallographic structural data with theoretical calculations. The main reason for this inconsistency could be attributed to inherent experimental constraints. The binding data is generally not of a

quality to enable accurate quantitative comparisons. Some spectroscopic experiments can give direct measurement of free energy while calorimetric experiments just yields the enthalpic or heat energy changes of system. In addition to this, ΔG computed via sophisticated simulations often do not correlate well with the experimental binding measurements as quite often computational chemists involved in pharmaceutical research interchangeably use K_i and IC_{50} with K_d data in calculation of free energy relationships. It is commonly becoming a practice in drug discovery community to have assays that generate IC_{50} values rather than equilibrium constants. These approximations severely undermine the accuracy of models and should be used with discretion.

One of the main aims in structure based drug design is explaining the binding interactions between a drug and its target. This is often achieved by virtual screening, more specifically, identifying a lead molecule out of thousands of molecules by docking and predicting affinity against a target protein. There are number of issues that currently are not adequately addressed. For example, in such calculations enthalpic contributions are often estimated by theoretical methods, knowledge-based potential functions, or parameters derived from experimental data. The force-fields used for calculations of free energy and intermolecular interactions assume that steric and electrostatic forces are sufficient to account for the observed biological interactions. However, these alone are never sufficient in accurate prediction of biomolecular interactions as they do not always include solvation/desolvation effects. The phenomenon of solvation/desolvation within the protein active site resulting in hydrophobic interactions has been particularly difficult to model computationally. Besides this, appropriate consideration of ionization and

tautomerization state of ligand and protein is also very important. Assumption of structural rigidity as an approximation may have severe entropic repercussions. In addition, binding may lead to protein adaptability and additional conformational changes that are not normally considered. These entropic contributions to binding are much less well-defined and often poorly quantitated or even ignored in many cases. Most approaches sum up these interactions separately as distinct enthalpic and entropic contributions, whereas, in reality, the ligand-protein recognition is a concerted event and thermodynamic quantities cannot be just simply summed. The interaction energies calculated without entropic contributions are not always expected to correlate well with the measured free energies of binding. Virtual screening has limited accuracy in exchange for a list of few promising candidates for further evaluation.

Pioneering studies in free energy calculation over the past few years have made significant progress towards this goal. Among the several methods available to calculate free energy, complex and time consuming molecular dynamics simulations with explicit water have been shown to correlate with free energy. Free energy perturbation (FEP) calculations based on statistical mechanics relates the free energy of a system calculating the ensemble averages and treating solvent molecules and ions explicitly. However, calculation of free energy using molecular dynamics simulation is plagued by errors from a variety of sources. Statistical errors arise from the limited sampling of configurations accompanied by errors arising from the imperfect equilibration of the system. Multiple levels of time and uncertainty are compounded for complex biomolecular systems with inclusion of solvents and counter ions. However, the size of the system makes

approaches of this sort computationally expensive and so simplified models are generally applied.

Though computational modeling techniques hold a great promise for future progress in drug discovery and development it is still an evolving technology and has number of limitations. However the tools and techniques should be used with great cautions. Any theoretical model generated should be validated with experimental methods.

1.5 Research Plan

A new methodology based on HINT paradigm[94] is applied to design novel inhibitors of tubulin protein and predict the binding mode of the protein-ligand complexes. The research utilizes the biological data, correlated with structure, to optimize the current (and succeeding) drug leads to design more effective candidate compounds with improved efficacy and minimal toxicity. The study also aims towards developing novel molecular modeling tools to design clinically viable molecules.

Given the utility of microtubule targeting as a strategy in the treatment of malignancy, the design and development of a new and clinically active class of microtubule-targeting compounds is highly desired. Pyrrole containing molecules derived from nature have proven to be particularly useful as lead compounds for drug development[104,105]. Highly functionalized pyrroles have previously been demonstrated to have potent cytotoxicity against a variety of human tumors with activity expressed at nM and μM concentrations in human breast tumor cell lines[106,107]. Collaborators in this research have recently

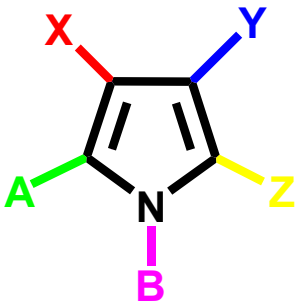
developed a series of brominated pyrroles, whose structure suggests that they might interfere with tubulin function (Figure 1a, b). Both cellular studies with one of the compounds, JG-03-14 (3,5-dibromo-4-(3,4-dimethoxyphenyl)-1H-pyrrole-2-carboxylic acid ethyl ester) and molecular modeling analysis tend to support the contention that these compounds function as microtubule poisons[108]. As described later, the development of additional synthetic or semi-synthetic pyrrole derivatives is greatly facilitated due to their relative ease of synthesis, which, in turn, facilitates manipulation of the molecular properties in adding and removing a broad range of functional groups[109]. Consequently, this research proposes to refine and develop pyrrole-containing alkaloids as new chemotherapeutic agents for the treatment of breast cancer.

1.5.1 Preliminary Research/Data

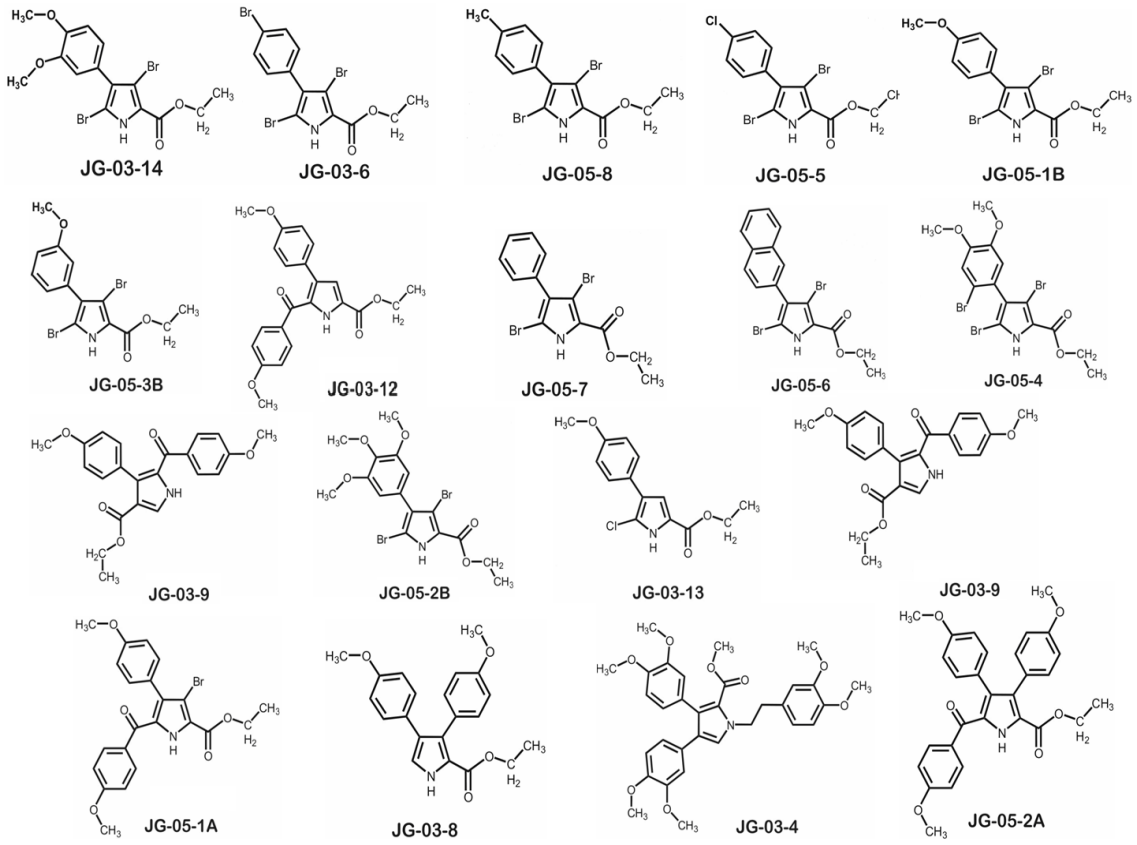
In the initial studies carried out by our collaborators, several substituted pyrroles (Figure 1b) were evaluated for their inhibitory activity on the growth of MCF-7 breast tumor cells[108]. The studies demonstrated that the highest degree of anti-proliferative activity was expressed by JG-03-14, while a lesser degree of activity was evident for compound JG-05-2. The other analogs examined failed to demonstrate significant anti-proliferative activity at concentrations up to 500 nM.

Figure 1: a) Highly functionalized pyrrole scaffold. **b)** Structure of substituted pyrroles.

a)



b)



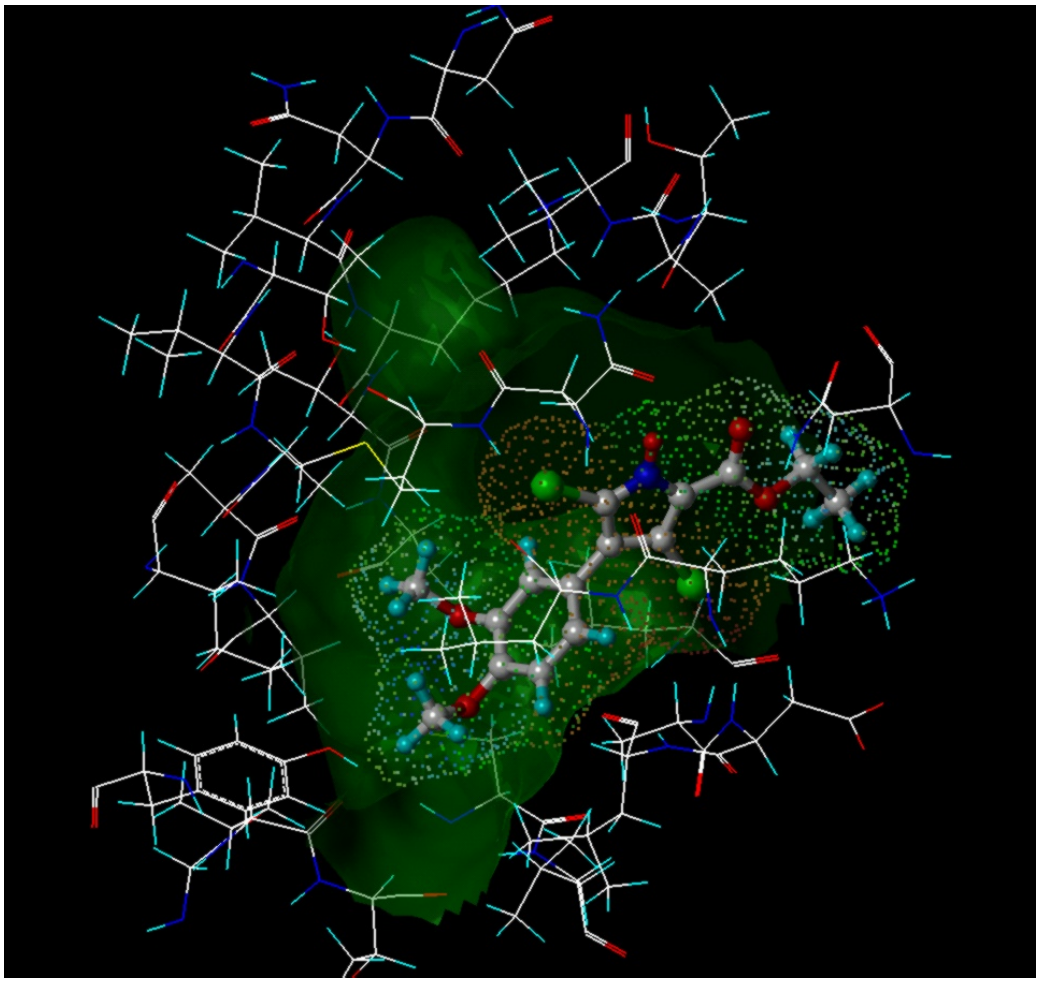
These observations provided insight into the structural requirements for the growth inhibitory activity/cytotoxicity of this class of agents and offer significant opportunity for structural alterations, which could lead to an improved drug candidate. All the subsequent experimental studies conducted by our collaborators suggested that JG-03-14 promoted extensive cell death during the initial period of exposure. These observations indicated that JG-03-14 abrogates breast tumor cell growth and survival through both the promotion of autophagic cell death and induction of a permanent growth arrested state. COMPARE analysis across NCI cancer cell line panel evaluated similar activity profile of colchicine and JG-03-14 and further corroborated mechanism of action of JG-03-14 to be similar to that of colchicine. Thus, there is strong evidence that this class of compounds interferes with the microtubule function. The effects of JG-03-14 along with the two positive controls, combretastatin A-4 and thiocolchicine, on tubulin assembly were studied (Table 1). All these agents inhibited the assembly of tubulin.

Examination of the initial docking model in detail illustrates (Figure 2), the pocket shape and size, as indicated by the green translucent surface. The occupancy of the JG-03-14 compound in this pocket is shown by the dot surface (color coded by brown=lipophilic, green=polar), regions of the site that do not have optimum interactions and consequently, locations at which molecular modifications of the lead compound might be effective, can be seen. For example, the left region of the pocket, which is quite hydrophobic, effectively accommodates the two methoxy group substituents to the aromatic ring of JG-03-14, but may be able to tolerate more steric bulk.

Table 1: JG-03-14 inhibits tubulin assembly and [^{3}H]colchicine binding to tubulin. The effects of JG-03-14, and the positive controls combretastatin A-4 and thiocolchicine, on tubulin assembly were studied. All these agents inhibited the assembly of tubulin and the EC_{50} values for inhibition (at 10 μM tubulin) were calculated. The effects of these compounds on 5 μM [^{3}H]colchicine binding to 1 μM tubulin were determined. ND =not determined.

	Inhibition of tubulin assembly (EC_{50} in μM)	% Inhibition of binding by inhibitor at	
		1 μM	5 μM
JG-03-14	1.5 ± 0.2	49 ± 3	84 ± 0.6
Combretastatin A-4	1.3 ± 0.1	83 ± 2	97 ± 2
Thiocolchicine	1.0 ± 0.1	ND	67 ± 0.07

Figure 2: Difference between pocket (green surface) and JG-03-14 structure volumes indicates regions of interest for designing new active analogs.



This model also suggests that the pyrrole substituent α to the nitrogen [$C(=O)OCH_2CH_3$ in JG-03-14] can be optimized to make better interactions with the residues at the entrance to the pocket. Highly selective novel analogs with minimal toxicity can be designed using the proposed HINT-based *de novo* design methodology. The primary aim of this research is to design a library of highly selective pyrrole-based analogues. The fragment-based approach can, therefore, be considered as a good working hypothesis for rationally designing novel analogs since the individual building blocks have to be linked in order to be a meaningful inhibitor. If the individual building blocks can be chemically linked easily in a chemical fashion, the approach has the advantage that the building blocks can be synthesized separately and linked in the final stage of the synthesis.

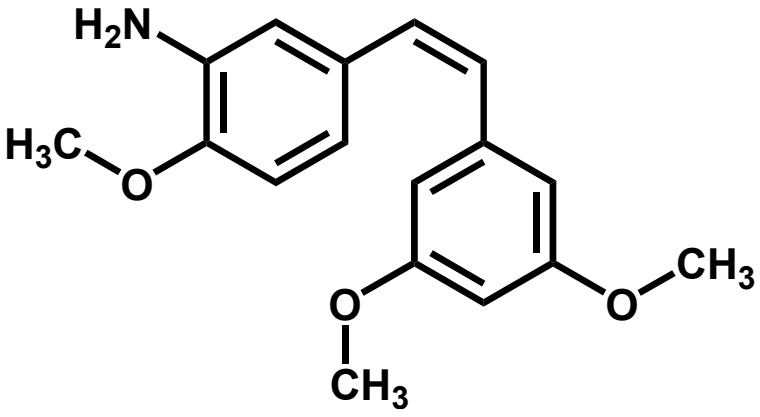
A very extensive library of compounds for the synthesis and evaluation of analogs of JG-03-14 can be created using fragment-based methodology. The rationale is to select the compounds which have demonstrated interesting bioactivity, and vary the A, B, X, Y and Z groups (Figure 1a) in a logical manner based on modeling studies and the use of standard QSAR considerations (lipophilic character, electronic effects and steric effects). Each of the target molecules can then be subsequently subjected to docking and QSAR evaluation prior to the synthesis. Structure activity information from other known agents, which have similar binding properties, can also be incorporated into the selection of the indicated target molecules. This study also appreciates the fact that the most active compound found may not be the ideal candidate due to toxicity effects on normal cells. It

serves little purpose for a drug to kill tumor cells if overall toxicity to host tissue is high (i.e., the drug has low selectivity and/or a narrow therapeutic index. One of the objectives of this research was to design novel analogues using structure-based/ligand based computational methods. This is only the starting point for one of several approaches to be explored as predictive tools for molecular design. Therefore, this methodology will allow generation of a large family of strong candidates, on which, further xenograft and clinical studies can be carried out.

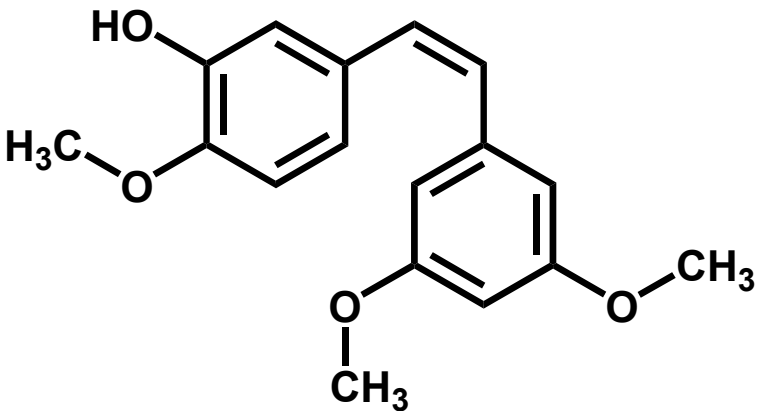
Another such attempt was carried out to design and develop novel stilbene analogs that also target the microtubules at the colchicine site and act as anti-leukemic agents. A series of novel stilbene derivatives have been synthesized and studied with the main goal to investigate SAR of the stilbene analogs, as well as to improve its water solubility, a potentially negative aspect of this molecule that could be a serious hindrance in the pre-clinical development. We attempted to optimize stilbene 5C using computer-based drug design and synthesize derivatives with benzimidazole or indole group. Derivatives with good cytotoxic activity, in particular, the derivatives 5C and 6C (Figure 3) were obtained and the study gave rise to two novel leads for further investigation. Alternative approach was also adopted to make prodrugs of stilbene 5C. A morpholino-carbamate derivative, prodrug of 5C, has a very good solubility in water, and is active in suppressing growth of tumor cells at a concentration of 5000 nM, which is a concentration 100 times higher than the parent stilbene 5C.

Figure 3: Stilbene 5C and 6C.

Stilbene 5C



Stilbene 6C



These data suggest that stilbene analogs could be a good candidate for further pre-clinical and clinical development as a new anti-tumor agent for cancer therapy.

A second synergistic, goal of this research was to design and develop novel molecular modeling software tools that would *de novo* identify and characterize the binding site and design analogs within the binding site. The available crystal structure of tubulin protein has a low resolution of 3.58 Å, undermining the ability to design highly selective ligands. Target-based generation of lead compounds requires that the three-dimensional structure of the ligand binding site be known as accurately as possible. Identification and characterization of active sites is imperative in the study of protein structure, particularly for the design of molecules that interfere with its function and modulate activity. However, in a low resolution crystal structure model, knowledge about the correct orientation of side chains is limited by ambiguities in position and orientation because the experimental electron density envelopes are generally featureless. Through the computational modeling studies, it was intended to elucidate the binding process by better characterizing the binding site.

This involved analyzing the hydropathic character of the binding site and generating synthetically viable functionalities on parent molecules to exactly complement the binding features in the active site. This feature effectively facilitated designing clinically viable drug candidate with minimal toxicity. The methodology can also be extended to other systems and may find extensive applications in *de novo* drug design projects. These

efforts are expected to give rise to a new class of selective and potent anti-tumor drugs for use in combination chemotherapy.

1.5.2 Expectations, Interpretation and Impact

A key success criterion for this research is to productively use the biological cytotoxicity data to design new and more potent analogs. Development and biological evaluation of diverse set of analogs will be accomplished by the collaborators. The synthetic methodology used by our collaborators is very flexible and rapid, making it possible to synthesize and test many compounds designed computationally and thus, allow significant diversity in pyrrole substituents. This research plans appropriate SAR (Structure Activity Relationship) guided structural alterations on the lead compounds to optimize the desired bioactivity and minimize toxicity. This will involve structural changes, which will probe steric, electronic and lipophilic character of the parent molecule. Once the targeted compounds are prepared, sensitivity will be compared in breast tumor cells and normal cells.

Favorable findings in terms of new compounds based on this research design would guide the synthetic efforts to develop analogs with a higher degree of selectivity. The success of this research project depends on the effectiveness of the multidisciplinary collaboration. For example, cytotoxicity studies will allow molecular models and QSAR approaches to be refined, which will then suggest the next round of pyrrole analogs to be synthesized and biologically evaluated with the ultimate goal of generating a significant clinical candidate.

The iterative cycles of modeling→synthesis→testing→modeling, etc. will be continuously monitored to improve the models and consequently, the predictions of the compound activities.

CHAPTER 2

Docking and Hydropathic Scoring of Polysubstituted Pyrrole Compounds with Anti-Tubulin Activity

2.0 Abstract

Compounds that bind at the colchicine site of tubulin have drawn considerable attention with studies indicating that these agents suppress microtubule dynamics and inhibit tubulin polymerization. Data for eighteen polysubstituted pyrrole compounds are reported, including antiproliferative activity against human MDA-MB-435 cells and calculated free energies of binding following docking the compounds into models of αβ-tubulin. These docking calculations coupled with HINT interaction analyses are able to represent the complex structures and the binding modes of inhibitors such that calculated and measured free energies of binding correlate with an r^2 of 0.76. Structural analysis of the binding pocket identifies important intermolecular contacts that mediate binding. As seen experimentally, the complex with JG-03-14 (3,5-dibromo-4-(3,4-dimethoxyphenyl)-1H-pyrrole-2-carboxylic acid ethyl ester) is the most stable. These results illuminate the binding process and should be valuable in the design of new pyrrole-based colchicine site inhibitors as these compounds have very accessible syntheses.

2.1 Introduction

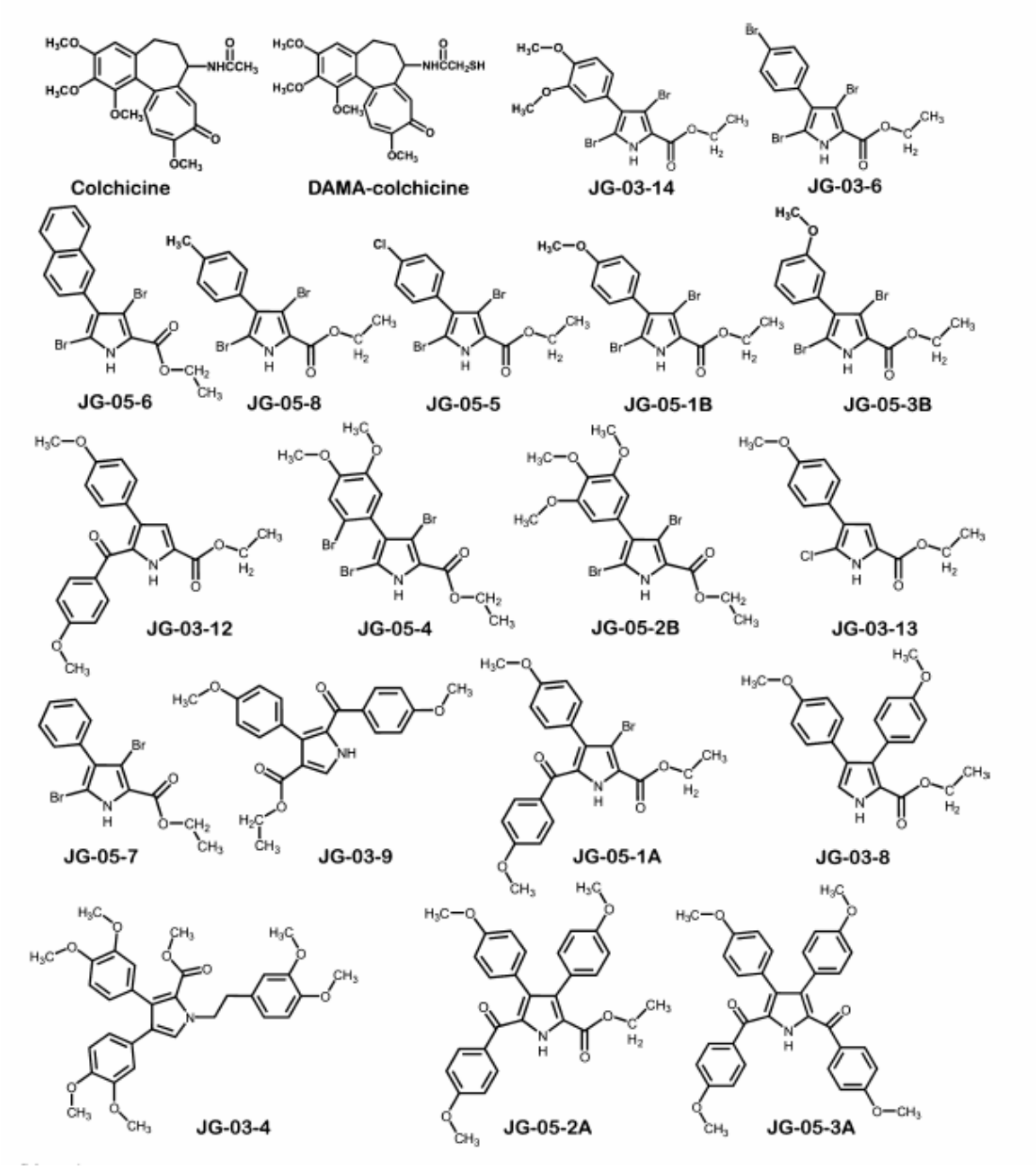
A large number of targets are under exploration for chemotherapeutic treatments for cancer. In the past several years, based on the efficacy and commercial successes of paclitaxel and the vinca alkaloids, there have been major efforts to design inhibitors that bind and interfere with the function of microtubules. Microtubules are essential elements of the cytoskeleton and extremely important in mitosis and cell division. Colchicine, the first drug known to bind to the tubulin protein[110,111], inhibits microtubule formation and at high concentrations causes loss of cellular microtubules. In contrast, paclitaxel and its analogues actually promote microtubule polymer formation[112-114], albeit by acting at a different site on tubulin than colchicine. A variety of small molecules with diverse molecular scaffolds have been shown to bind tubulin at the colchicine site[115-118]. One class of these compounds receiving particular attention has been that based on the natural product combretastatin A-4 discovered by Pettit[119,120]. Despite some successes, the discovery of new, more efficacious inhibitors is becoming increasingly important because of multi-drug resistance to tubulin-binding antimitotic agents[121].

Natural products containing pyrrole have diverse and interesting biological activities, and have proven to be particularly useful as lead compounds for drug development[105,109,122]. As part of long-term program to develop vinylogous iminum salt-based syntheses of natural products containing the pyrrole group, some early synthetic intermediates were evaluated against a panel of human and murine tumor cell lines[123]. Many of these synthetic compounds were highly active against cancer cell lines, and some inhibited

DNA synthesis without binding directly to DNA[124]. Earlier studies also provided clues to the design of pyrrole analogs that might have potent antiproliferative activities and the ability to bind to tubulin.

The marine natural product lukianol A contains a highly oxygenated 3,4-diphenylpyrrole motif, and it potently inhibited the growth of the human KB cancer cell line[125]. Banwell and colleagues suggested that lukianol A represented a configurationally stable hybrid of combretastatin A-4 and colchicine[126]. John Gupton from University of Richmond has synthesized another class of biologically interesting pyrroles, which are somewhat related to the pyrrolomycin natural products by virtue of their halogenated pyrrole backbone[127]. Brominated analogs of previously synthesized pyrroles were prepared, and they retained the cytotoxic activity exhibited by the non-brominated pyrroles[128]. While mechanistic studies indicated that two of the brominated pyrrole compounds bound directly to DNA, causing DNA cross-linking, the mechanisms of action of other brominated pyrrole analogs remained unknown. Based on the structural similarity of the compounds to combretastatin A-4 and Banwell's suggestion that several pyrrole containing natural products represent stable hybrids of combretastatin and colchicine, the effect of brominated pyrroles were examined on cellular microtubules. Several of the brominated pyrrole analogs had microtubule depolymerizing effects. The most potent was 3,5-dibromo-4-(3,4-dimethoxyphenyl)-1*H*-pyrrole-2-carboxylic acid ethyl ester (JG-03-14; structure in Figure 4).

Figure 4: Substituted Pyrroles



It is significant to note that JG-03-14 is structurally unique among combretastatin-like compounds. It possesses a single phenyl group which is highly oxygenated; other combretastatin analogs always possess two highly oxygenated phenyl groups. The mode of action and the cellular effects of this compound were evaluated in detail, and it was found to be a potent antimitotic agent. Binding studies show that JG-03-14 binds to tubulin within the colchicine site. Our initial studies indicate that this compound has antitumor effects and it represents a promising lead for the generation of new analogs with important biological properties. While investigating the cytotoxic activity of compounds in a series of synthetic polysubstituted pyrroles (Figure 4), our interest in the colchicine binding site of tubulin as a putative target for computational drug design studies was piqued after a COMPARE[129] analysis showed a correlation between one of the compounds (JG-03-14) and colchicine of 0.681 over the 45 cell lines that were assayed for both compounds. COMPARE evaluates similarities in activity profiles across the NCI cancer cell line panel, and has been used to elucidate modes of action for new anticancer agents[129]. COMPARE analysis and modeling studies suggest that JG-03-14 represents a novel chemotype for the colchicine site. However, in any case, the true therapeutic potential of the colchicine site on tubulin has not been fully explored because of the lack of truly *atomic level* knowledge of the site.

In 2000, Hamel and colleagues mapped the binding site of colchicinoids on β-tubulin[130]. Using molecular modeling and computational docking of colchicinoids into the electron crystallographic model of β-tubulin in protofilaments[130], they found two potential binding sites. The first was entirely encompassed within β-tubulin with the colchicinoids forming

adducts with Cys 356. The second potential site was located at the α/β interface and involved hydrogen bonding with Cys 241. More recently, Nguyen and colleagues[131] developed a comprehensive pharmacophore model for structurally diverse colchicine-like tubulin inhibitors using a structure-based approach on the newly available α/β-tubulin:DAMA-colchicine X-ray structure[132]. This crystal structure definitively identified a cleft at the α/β interface as the colchicine binding site, but has a resolution of only 3.58 Å and thus requires considerable computational effort before models derived from it can be considered "all-atom".

In this chapter we report the results of docking this set of putative ligands into the colchicine site of tubulin to build stereochemically reasonable models. We evaluated these docking models with the HINT free energy force field[91] and found a good correlation between HINT scores and measured IC_{50}s of cell proliferation by the compounds. While the measured IC_{50}s represent a downstream biological effect and we are making the pragmatic assumption that the modes of action for all compounds in this series are the same, these results do allow us to appropriately characterize the colchicine binding site and will also serve in design and validation of new compounds similar to JG-03-14 in later stages of this research. This is particularly relevant since these and other polysubstituted pyrrole compounds are synthetically accessible.

2.2 Materials and Methods

2.2.1 Synthesis of Pyrrole Compounds

The synthetic methods used to prepare the highly functionalized pyrroles and related derivatives depicted in Figure 4 can be found in previously reported work[128,133-136].

2.2.2 Antiproliferative Activity of Substituted Pyrroles against Human Tumor Cell Lines

The antiproliferative effects of the compounds were evaluated in MDA-MB-435 cells using the SRB assay as previously described[137]. All the biological assays were carried out by Dr. Susan L. Mooberry and colleagues at SFBR. A 48 hr exposure time was used. The IC_{50} values, i.e., the concentration that causes 50% inhibition of proliferation, was calculated from the log dose-response curves and represents the mean of three independent experiments. The effects of the compounds on cellular microtubules were evaluated using indirect immunofluorescence techniques. Briefly, A-10 cells were exposed to the compounds for 18 hr and then the cells were fixed and microtubules visualized using a β-tubulin antibody and the DNA was visualized using DAPI. The EC_{50}s for microtubule depolymerization were determined using visual observation as previously described[138]. A range of concentrations was tested for each compound and the percent microtubule loss determined for each concentration. The data from 3 independent experiments were averaged and plotted as percent microtubule loss vs. concentration and EC_{50} values calculated.

2.2.3 Model Building

The X-ray crystal structure (3.58 Å) of αβ-tubulin complexed with DAMA-colchicine[132] PDB code: 1SA0) was used in this study. The stathmin-like domain and the C and D subunits were removed from the model. After hydrogen atoms were added to the model, their positions were optimized to an energy gradient of 0.005 kcal-Å/mol with the Tripos force field (in Sybyl 7.1) while keeping heavy atom positions fixed. The models for pyrrole analogues were constructed using the Sybyl 7.3 (www.tripos.com) suite and optimized similarly[78].

2.2.4 Docking

Computational docking was carried out using the genetic algorithm-based ligand docking program GOLD 3.0[69]. GOLD explores ligand conformations fairly exhaustively and also provides limited flexibility to protein side chains with hydroxyl groups by reorienting the hydrogen bond donor and acceptor groups. The GOLD scoring function is based on favorable conformations found in Cambridge Structural Database and on empirical results of weak chemical interactions[139]. The active site was defined by a single solvent accessible point near the center of the protein active site, an approximate radius of 10 Å, and the GOLD cavity detection algorithm. GOLD docking was carried out without constraints to get an unbiased result and to explore all possible binding modes of the ligands. The tri-methoxy phenyl fragment of colchicine was used as the template for biasing the pose of all ligands. In this study we performed 100 GOLD genetic algorithm runs, as opposed to the default of 10 and early termination of ligand docking was

switched off. All other parameters were the defaults. To evaluate and validate GOLD performance the co-crystallized ligand DAMA-colchicine[132] was extracted and docked. GOLD accurately reproduced the experimentally observed binding mode of DAMA-colchicine in αβ-tubulin. All remaining ligands were docked using the same parameters.

Dockings with different/optional constraints such as enforced hydrogen bonds, hydrophobic regions and scaffold match were also explored. For hydrogen bond constraints, docking was biased so that the ligands make hydrogen bonds with Asn258, Ser178, Asn101, and the backbone amides of Ala180 and Val181. For region hydrophobic constraints the ligand positions were constrained by defining a hydrophobic sphere where the tri-methoxy phenyl moiety of colchicine was positioned. Then specific ligand atoms to be docked in the hydrophobic region of the active site were defined. Alternatively, scaffold match constraints were used to place the ligand at a specific position within the active site.

2.2.5 Hydropathic Scoring

The HINT (Hydropathic INTeractions) scoring function[91] (version 3.11S β) was used to investigate the structural aspects of the interactions by analyzing and ranking the GOLD docking solutions. For selection of the optimum docked conformation and to further differentiate the relative binding efficacy of the pyrrole ligands, interaction scores were calculated for each pose found by docking. The protein and ligands were partitioned as distinct molecules. "Essential" hydrogen atoms, i.e., only those attached to polar atoms (N, O, S, P) were explicitly considered in the model and assigned HINT constants. The

inferred solvent model, where each residue is partitioned based on its hydrogen count, was applied. The solvent accessible surface area for the amide nitrogens of the protein backbone were corrected with the "+20" option. Finally, HINT scores were plotted against experimental binding free energies that were calculated using the standard Gibbs free energy equation:

$$\Delta G_{binding} = - R\,T \ln(K_{eq}), \qquad (7)$$

where R is Boltzmann's constant (1.9872 cal K^{-1} mol^{-1}) and T is 298 K; K_{eq} is an equilibrium binding constant, ideally K_D. In this work measured IC_{50} values are being used as approximations for equilibrium constants.

2.3 Results and Discussion

While the character of the colchicine binding site was investigated by Nguyen *et al.*[131], their study was directed at deriving a generalized pharmacophore for the site and consequently the data set included only two polysubstituted pyrroles. These compounds represent an emerging class of agents with potential activity against a variety of human tumors with activity expressed at nM or μM concentrations in human tumor cell lines[108,135], but having advantages over natural products in terms of drug design and development. In particular, we have been exploring a series of brominated pyrroles whose structure suggests that they might interfere with tubulin function. One member of this series (JG-03-14, 3,5-dibromo-4-(3,4-dimethoxyphenyl)-1H-pyrrole-2-carboxylic acid ethyl ester), for which NCI tumor panel activity had been obtained[128], was suggested

by COMPARE[129] to have an activity profile similar to colchicine. Cellular studies with JG-03-14 further support the contention that these compounds function as microtubule poisons[108]. In addition, JG-03-14 was found to promote autophagic cell death while retaining activity in tumor cells expressing the multidrug resistant pump[108,133]. Because the development of additional synthetic or semi-synthetic pyrrole derivatives in this class is facilitated with their relatively facile syntheses, including modification of the molecule by adding and removing functional groups[128], we have both a rather extensive collection of molecules in-hand (Figure 4) for building predictive molecular models and the potential for rational design and synthesis of many others.

2.3.1 Antiproliferative Activity of Polysubstituted Pyrroles

Results from a number of assays have previously appeared regarding the cytotoxic activities of the lead compound JG-03-14[108,128,140]. However, most of the compounds in this series have not been examined in detail. An important component of structure-activity relationships and/or computational activity predictions is having reproducible and comprehensive data for a relatively large series of compounds, even those with comparatively poor activity, because understanding why particular compounds are inactive is potentially just as valuable as data on active compounds. Table 2 sets out the experimental antiproliferative assay results against human MDA-MB-435 cancer cells for the compounds in Figure 4. While JG-03-14 remains the compound with the most potent (36 nM) activity, a few others (Table 2) have activities that are only about 7-10 fold less potent, thus suggesting that design of additional new compounds with desirable properties is possible.

Table 2: Experimental IC_{50}, EC_{50} and docking results for polysubstituted pyrrole compounds.

Compound	Activity Set	IC_{50} for anti-proliferation[a]	pIC_{50}	$\Delta G_{binding}$ (kcal mol^{-1})	HINT score	HINT LogP	EC_{50} for loss of microtubules[b]
JG-03-14	A	36 nM[c]	7.74	-10.14	626	2.60	490nM
JG-03-6		312 nM	6.51	-8.87	609	3.17	>50μM[f]
JG-05-1B		361 nM	6.44	-8.78	483	2.36	5.1 μM
JG-05-3B		278 nM	6.55	-8.94	410	2.36	2.4 μM
JG-05-5		919 nM	6.04	-8.23	559	3.02	>50 μM[f]
JG-05-8	B	2.2 μM	5.70	-7.71	433	2.97	>50 μM[f]
JG-03-12		2.6 μM	5.58	-7.61	221	4.88	>50 μM[f]
JG-05-4		1.1μM	5.95	-8.12	455	3.34	7.1 μM
JG-05-6		1.4 μM	5.85	-7.98	351	3.59	7.5 μM
JG-03-13		5 μM	5.30	-7.23	163	1.48	>50 μM[f]
JG-05-1A		1.9 μM	5.72	-7.80	508	5.62	>50 μM[f]
JG-05-2A		4.2 μM	5.37	-7.33	149	6.58	>50 μM
JG-05-7	C	10 μM	5.00	-6.82	152	2.43	>50 μM
JG-05-2B		13 μM	4.89	-6.66	136	2.90	>50 μM
JG-03-9		10 μM	5.00	-6.82	54	4.63	>50 μM
JG-03-4		10 μM	5.00	-6.82	27	6.66	>50 μM
JG-03-8[d]		>10 μM	4.00	-5.45	-241	3.69	>50 μM
JG-05-3A[d]		>20 μM	3.70	-5.04	296	9.02	>50 μM
Colchicine[e]	N/A	2 μM	5.70	-7.77	563	3.24	N/D
DAMA-colchicine[e]		2 μM	5.70	-7.77	455	3.70	N/D

[a]Antiproliferative activity against human MDA-MB-435 cells using the SRB assay.
[b]Microtubule depolymerizing activity for microtubule loss.
[c]From reference 110.
[d]pIC_{50} and $\Delta G_{binding}$ calculated for $10 \times IC_{50}$.
[e]Reported colchicine IC_{50} data was an average of values reported previously in the literature, references 110. For calculation purposes DAMA-colchicine was assumed to have same binding as colchicine.
[f]Mechanism of cytotoxicity appears to be unrelated to microtubule disrupting activity.
N/A – not applicable; N/D – not determined.

Results of a second assay, microtubule depolymerizing activity EC_{50}s for microtubule loss that serves as a partial check on mechanism of action, are also reported in Table 2.

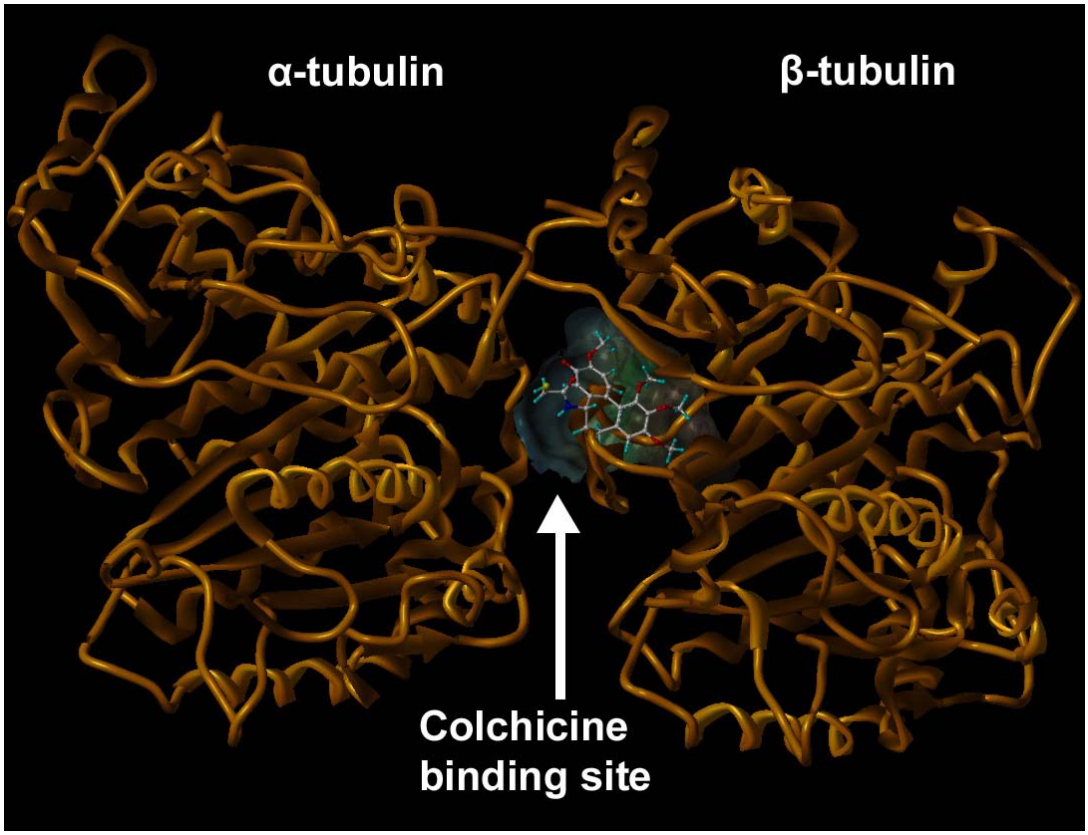
2.3.2 The Colchicine Binding Site

Binding models for each pyrrole analogue were investigated to delineate steric, electrostatic and hydropathic features of the colchicine binding site. Because we have focused on a series of eighteen compounds with IC_{50}s ranging over more than three orders of magnitude (see Table 2), we performed detailed docking studies with GOLD[69,139] followed by free energy scoring using the HINT protocol[91,102] to assess the binding modes. Without added constraints GOLD was found to reliably re-dock the crystallographic DAMA-colchicine ligand (RMSD = 0.76 Å) that was then used as the reference for all other docking experiments. The HINT score for co-crystallized DAMA-colchicine was 139; in the re-docked pose this score was 455. However, docking of the pyrrole analogues with GOLD produced a mixture of orientations that could not be rationalized with the GOLD docking score. Thus, as we have described in an earlier report[96], docked poses were re-scored with HINT and we chose the highest HINT-scored pose for further analysis (see Table 2). Docking poses created using a variety of constraints (see Methods) did not yield higher scoring models and were less interpretable than the "freely" bound models we are using. These docked models of substituted pyrroles fit within the pharmacophoric model proposed by Nguyen *et al.*[131], and for the structural features in common between the substituted pyrroles and in the Nguyen *et. al.* study, the docking models are in generally good agreement. Key is that the hydrophobic methoxy substituted ring of the pyrrole analogues sits at the hydrophobic center where

the TMP moiety of colchicine is found. Note that, although the pyrrole compounds have quite similar structures and are generally positioned in the binding pocket with essentially the same mode, the HINT scores are very sensitive and slight positional differences are detectable in the scores. This sensitivity combined with the number of compounds in the data set allowed us to analyze the site in considerable detail.

The focus of these computational investigations was on structural aspects of the interactions. The colchicine binding site lies at the interface between the α and β subunits of tubulin, mostly in the β subunit lined by helices 7 and 8 (see Figure 5). The funnel-shaped binding cavity has a volume of about 600 $Å^3$. Residues Tyr202β, Val238β, Thr239β, Cys241β, Leu242β, Leu248β, Leu252β, Leu255β, Ile378β, and Val318β form the narrow funnel end-like part and confer a strong hydrophobic character to this part of the cavity. At the wider portion, the cavity is surrounded by Ala250β, Asp251β, Lys254β, Asn258β, Met259β, Ala316β, Ala317β, Thr353β and Ala354β making it moderately polar/moderately hydrophobic. The open mouth end is surrounded by Asn101α, Thr179α, Ala180α, Val181α and Thr314β, Asn349β, Asn350β, Lys352β. The crystal structure for the complex indicates that DAMA-colchicine (and presumably colchicine) is positioned in the pocket such that its tri-methoxyphenyl (TMP) moiety sits snugly in the narrow hydrophobic pocket. Colchicine also forms hydrogen bonds with the backbone amides of Ala180α and Val181α.

Figure 5: Colchicine binding site at the interface between the α and β subunits of tubulin.

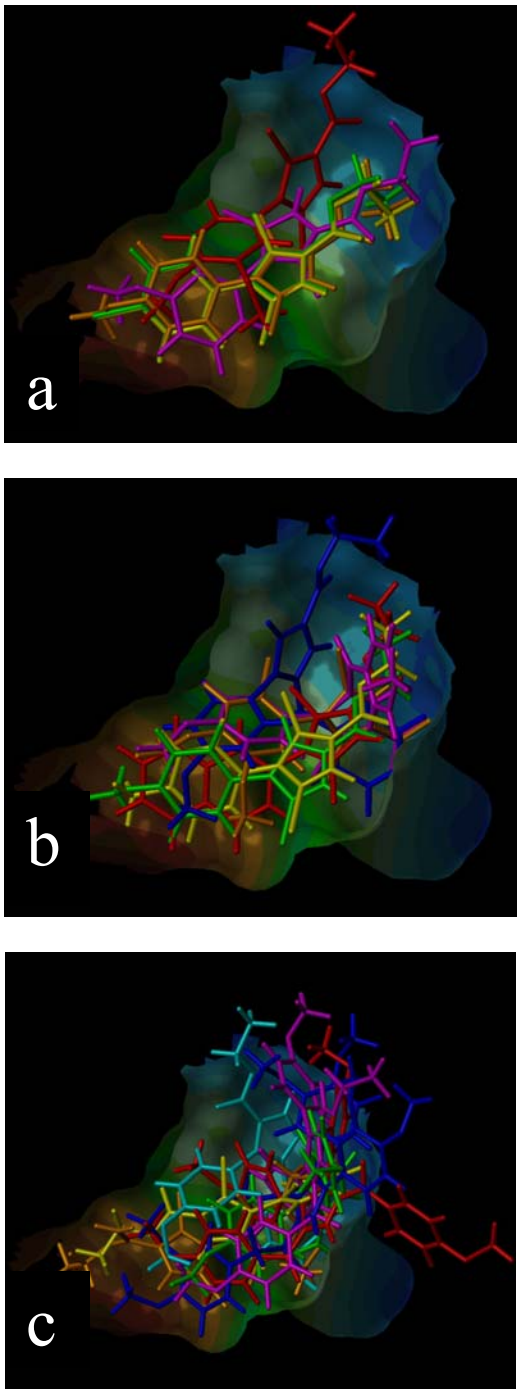


2.3.3 Structure-Activity Binding Relationships

The pyrrole analogues were clustered into three activity sets in order to study them in detail (see Table 2). The first set (A) was comprised of substituted pyrroles that showed antiproliferative activity with sub-μM IC_{50}s. The second set (B) consisted of ligands with IC_{50} values ranging from 1 μM to 5 μM. The remaining ligands, with IC_{50} values above 5 μM, comprised the third set (C). The analogs from subset A have noticeable similarity in their structures and are relatively simpler molecules than those in sets B and C. For all of these (set A) compounds the pyrrole ring is substituted by bromines at the 3 and 5 positions and an ethyl ester group at position 2. The differences among this group are substitutions to the phenyl ring at the 4 position of pyrrole. In these, the more potent compounds, most substituents to the phenyl ring, i.e., Cl, Br and methoxy, serve to make this portion of the ligand hydrophobic. Figure 6A illustrates the final docked orientations of the high-affinity pyrroles in the colchicine site of tubulin. The hydrophobic substituted phenyl ring fits snugly in the hydrophobic (narrow funnel) region of the binding pocket. The docked model for JG-03-14 is qualitatively similar to one reported earlier[108].

HINT hydropathic analysis reveals more detail concerning the forces orienting these ligands in the binding site. First, hydrophobic interactions are the dominating force contributing towards the stability of the complexes, with additional hydrogen-bonding interactions anchoring the ligands in the cavity. As listed in Table 2, the most potent-binding ligand has the highest HINT score (*vide infra*), i.e., JG-03-14 interacts with the binding site residues forming the most stable complex. The methoxy-substituted phenyls are positioned deep in the hydrophobic cavity surrounded by Cys241β, Leu242β,

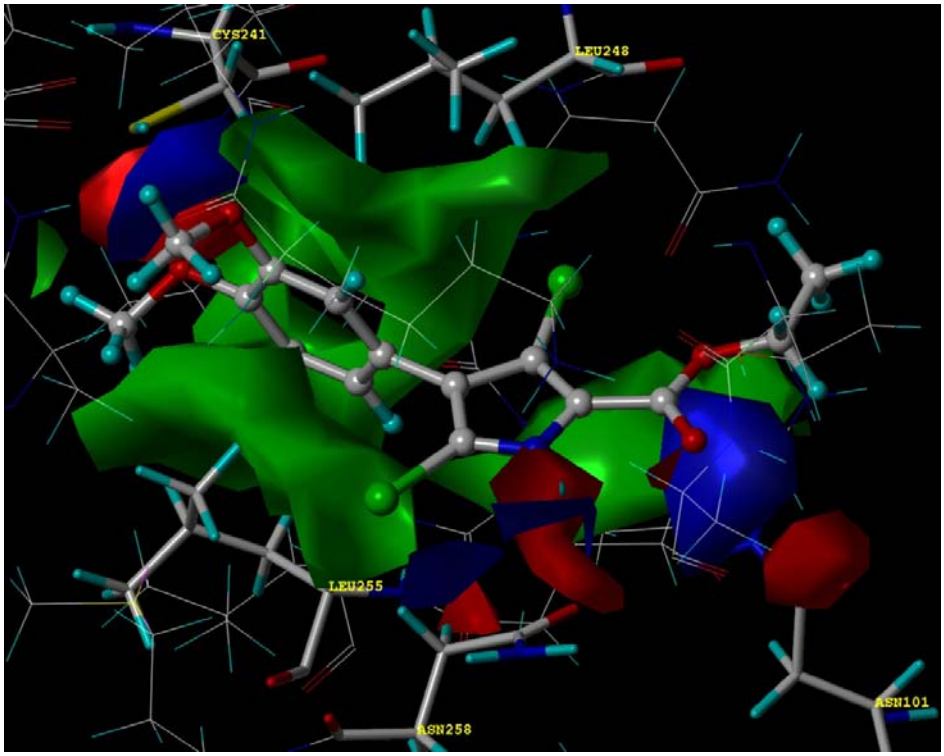
Figure 6: Pyrrole analogues docked at colchicine binding site. (A) Substituted pyrroles with activity in sub-μM IC_{50}. (B) Ligands with IC_{50} ranging from 1 μM to 5 μM. (C) Ligands with IC_{50} value above 5 μM.



Leu248β, Ala250β, Leu255β, Ala354β and Ile378β, all of which contribute to favorable hydrophobic-hydrophobic binding. Figure 7 illustrates these interactions in a HINT map, where the relative sizes of the displayed contours represent the strength, and the colors represent the character, of the interactions between JG-03-14 and the tubulin colchicine binding site. The phenyl ring of JG-03-14 fits in a hydrophobic glove formed by the Leu248 and Leu255. Favorable polar interaction with Asn101, Cys241 and Asn258 also contribute in binding. The ethyl ester tail of the ligand faces towards the polar opening and is stabilized with a strong hydrogen-bond to the amide oxygen of Asp258β with a length of 2.41 Å. Another set of strong hydrogen-bonds are formed between the amine of Asn101α and the carbonyl oxygen (1.48 Å) and alcoholic oxygen (2.72 Å) of the ligand's ethyl ester substituent. This latter feature, anchoring of the flexible ethyl ester tail, is somewhat different in our models as compared to those of Nguyen *et al*[131], probably due to the lack of steric constraints at the open polar end of the cavity.

On analyzing subset B, docked ligands in the low μM range, it can be seen that these ligands are somewhat similar to the subset A ligands, but with slightly bulkier groups overall as in JG-03-12, JG-05-6 and JG-05-1A, more highly substituted at the pyrrole ring as in JG-03-12, and/or with less hydrophobic substituents as in JG-03-13, JG-05-4, and JG-05-8. For example, in JG-03-13 the single chlorine substitution is less hydrophobic than the two bromines of JG-03-14 and having only one methoxy also reduces this compound's hydrophobicity. In the case of JG-05-8 only a single methyl group substitutes the phenyl ring at the para position.

Figure 7: HINT interaction maps for JG-03-14 (ball and stick rendering) at colchicine binding site. Blue contours represent regions of favorable polar interactions, e.g., hydrogen bonds, red contours represent unfavorable polar interactions and green contours represent favorable hydrophobic interactions.



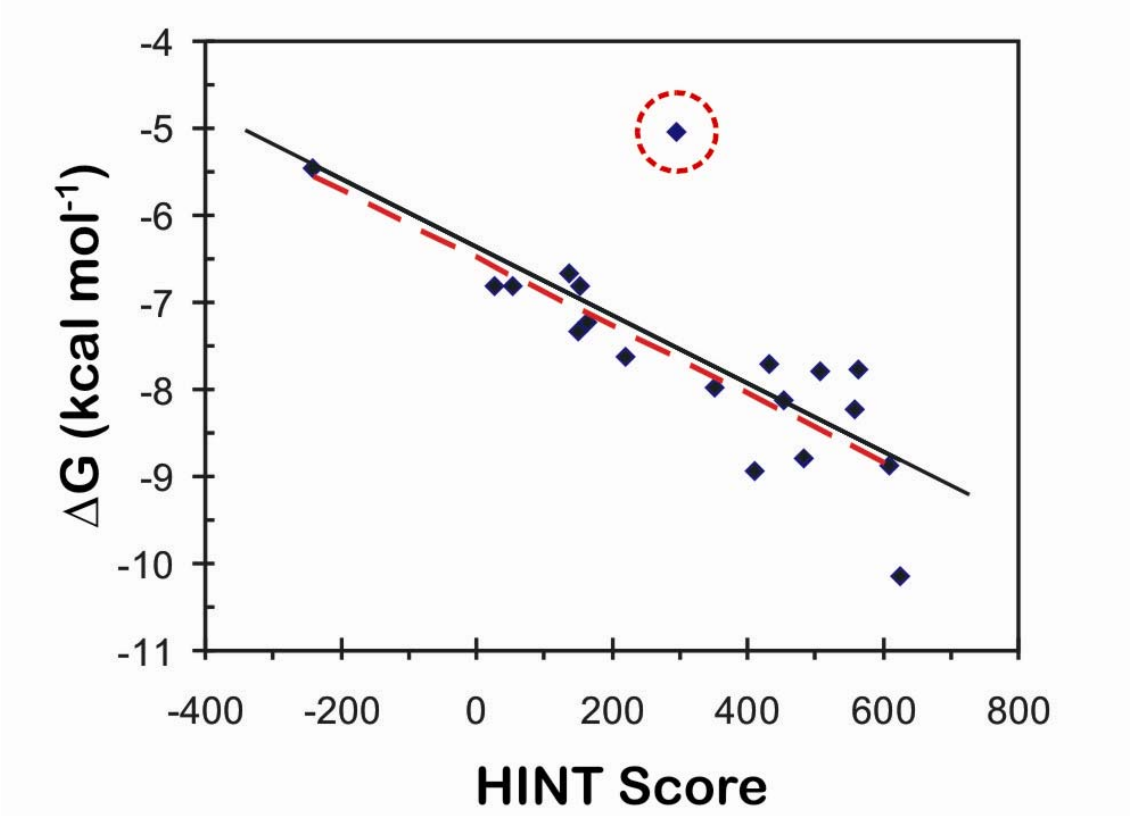
JG-03-12 and JG-05-1A have bulkier substitutions at the 5 position of the pyrrole, likely leading to their higher (less potent) IC_{50} values. The docked models (Figure 6B) and detailed HINT analysis confirm this SAR by showing a relatively poor fit in the active site for the bulkier ligands, and poorer hydrophobic HINT scores for the less optimally-substituted ligands. A single exception, the napthyl-substituted JG-05-6 produces a high HINT score inconsistent with its relatively low potency, but this may be due to this compound being too hydrophobic (Table 2) for solubility or transport to the site (*vide infra*).

Lastly, many of the subset C (inactive) ligands (Figure 6C) do not fit well in the site, while others are inappropriately decorated to make the required contacts with the site residues. Many of them have one or more bulkier substituents on the pyrrole ring, and only fit in the binding pocket with their side chains protruding out of the pocket.

2.3.4 Predictive Models for Ligand Binding

Figure 8 presents the correlation between the experimental binding ($\Delta G_{binding}$ as calculated from IC_{50}, see Methods) in kcal mol^{-1} and HINT scores for the eighteen synthetic pyrroles in this study. The IC_{50}s, antiproliferative activities of the compounds, are being taken in this work as approximations of binding affinity, with the implicit assumption that the antiproliferative activity is wholly due to tubulin binding. The consequences of this assumption will be discussed below.

Figure 8: Dependence of the experimental ΔG on HINT score units for Tubulin-pyrrole complexes. The solid black line represents the regression for ΔG vs. HINT score for all protein-ligand complexes. The red line represents the regression for ΔG vs. HINT score excluding the circled outlier (JG-05-3A).



The trend represented by the plot of Figure 8 indicates that higher scoring complexes are generally among those with more favorable free energies of binding, while lower scoring complexes are generally those with unfavorable binding. The correlation equation:

$$\Delta G = -0.0039\ H_{TOTAL} - 6.35, \quad (8)$$

has an $r^2 = 0.5818$ and a standard error of ± 0.52 kcal mol^{-1}. A better correlation is observed after omitting the outlier JG-05-3A from the correlation. Although this compound shows a high HINT score suggesting optimal intermolecular interactions within the tubulin colchicine site, it has a very high $logP_{o/w}$ value of 9.02 (Table 2) suggesting that this compound would likely not be transported to the binding site and may even be insoluble. The scoring function does not take into account cell permeability and completely ignores whether or not the compound could *in vivo* or *in vitro* be accessible to the binding site. Thus, the *unfavorable physiochemical properties* of JG-05-3A, and not statistical evaluation, warrant excluding it from the model and justify treating it as an outlier. Ignoring this outlier gives an $r^2 = 0.76$ and a standard error of ± 0.41 kcal mol^{-1}, with a very similar correlation equation:

$$\Delta G = -0.0039\ H_{TOTAL} - 6.51, \quad (9)$$

We believe that this model is predictive such that we can identify the active (subset A) ligands from the inactive (subset C) ligands with reasonable confidence and that further refinement of the model with additional data will improve its usefulness. However, it

must be noted that the EC_{50} for tubulin depolymerization data (Table 2) suggest that several of the compounds (two in set A) that dock in the colchicine binding site with high HINT scores do not appear to cause perturbations of cellular microtubules, i.e., their interactions within the colchicine binding site may not be the mechanism of cytotoxicity. Thus, while we cannot state that all antiproliferative activity is due to tubulin binding in the pyrrole compounds, there is enough experimental information for several of the more active compounds, and a compelling case for JG-03-14, to believe that designing ligands to bind with optimimum interactions in the tubulin colchicine binding site will produce compounds that will likely disrupt cellular microtubules and cause antimitotic actions.

2.4 Summary

The present communication demonstrates that the state-of-the-art molecular modeling calculations along with HINT interaction calculations are able to complement experimental studies of binding in many aspects, including accurate representation of the structure of the complex and the binding mode of putative drugs. The structural analysis of the binding pocket has identified important intermolecular contacts that mediate binding. The complex with JG-03-14 has the highest binding score corroborating the experimental data. In conclusion, the present series of pyrrole analogues have yielded representative compounds that are potent tubulin polymerization inhibitors and others that bind with less efficacy, but that still provide useful information for designing compounds with improved performance and selectivity.

CHAPTER 3

Hydropathic Analysis and Biological Evaluation of Stilbene Derivatives as Colchicine Site Microtubule Inhibitors with Anti-Leukemic Activity.

3.0 Abstract

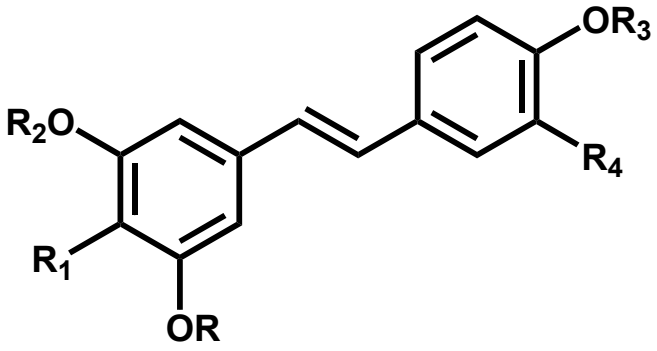
The crucial role of the microtubule in the cell division has identified tubulin as a target for the development of therapeutics for cancer; in particular tubulin is a target for antineoplastic agents that act by interfering with the dynamic stability of microtubules. A molecular modeling study was carried out to accurately represent the complex structure and the binding mode of a new class of stilbene-based tubulin inhibitors that bind at the αβ-tubulin colchicine site. Computational docking along with HINT score analysis fitted these inhibitors into the colchicine site and revealed detailed structure-activity information useful for inhibitor design. Quantitative analysis of the results was in good agreement with the *in vitro* antiproliferative activity of these derivatives (ranging from 3 nM to 100 μM) such that calculated and measured free energies of binding correlate with an r^2 of 0.89 (standard error ± 0.85 kcal mol^{-1}). This correlation suggests that the activity of unknown compounds may be predicted.

3.1 Introduction

Stilbene-based compounds are largely present in nature and have become of particular interest to chemists and biologists because of their wide range of biological activities[141,142]. Stilbene itself does not occur in nature, but hydroxylated stilbenes have been found in many medicinal plants. The hydroxylated stilbene trans-resveratrol (trans-3,40,5-trihydroxystilbene, Figure 9) is a phytoalexin present in grapes and plays a role in the prevention of coronary artery disease associated with red wine consumption[141,143-145]. Resveratrol has also antioxidant and anti-inflammatory properties and could be a potential chemopreventive and therapeutic agent in cancer[146,147]. *In vitro* inhibition of cell proliferation[148] and *in vivo* anti-neovascularization by resveratrol have been demonstrated[149]. On the other hand, the cis-stilbene motif represents a key structural feature of a broad class of natural and synthetic compounds endowed with an exceptionally strong tubulin polymerization inhibitor activity interfering mainly with microtubule formation at the tubulin colchicine binding site. In this context, the natural cis-stilbene combretastatin A-4 (CA-4) (Figure 9), an antimitotic agent isolated from the bark of the South African tree Combretum caffrum, and its 30-amino derivative (AC-7739) possess potent and interesting antitumor activity[150,151].

Combretastatins have been demonstrated to bind at the colchicine binding site and destabilize microtubule assembly and prevent spindle formation in mitotic cells[152]. The relatively simple structure and high affinity of combretastatins for the colchicine binding site has led to the synthesis and subsequent evaluation of a large number of

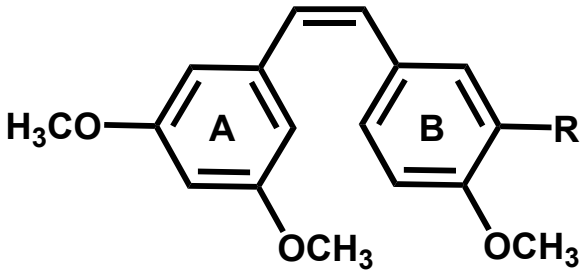
Figure 9. Natural and synthetic stilbenes



trans-Resveratrol **R=R_1=R_2=R_3=R_4=H**

Combretastatin A-4(CA-4) **R=R_2=R_3=CH_3, R_1=OCH_3, R_4=OH**

AC-7739 **R=R_2=R_3=CH_3, R_1=OCH_3, R_4=NH_2**



a: Stilbene 5c **R=$NH_3^+Cl^-$**

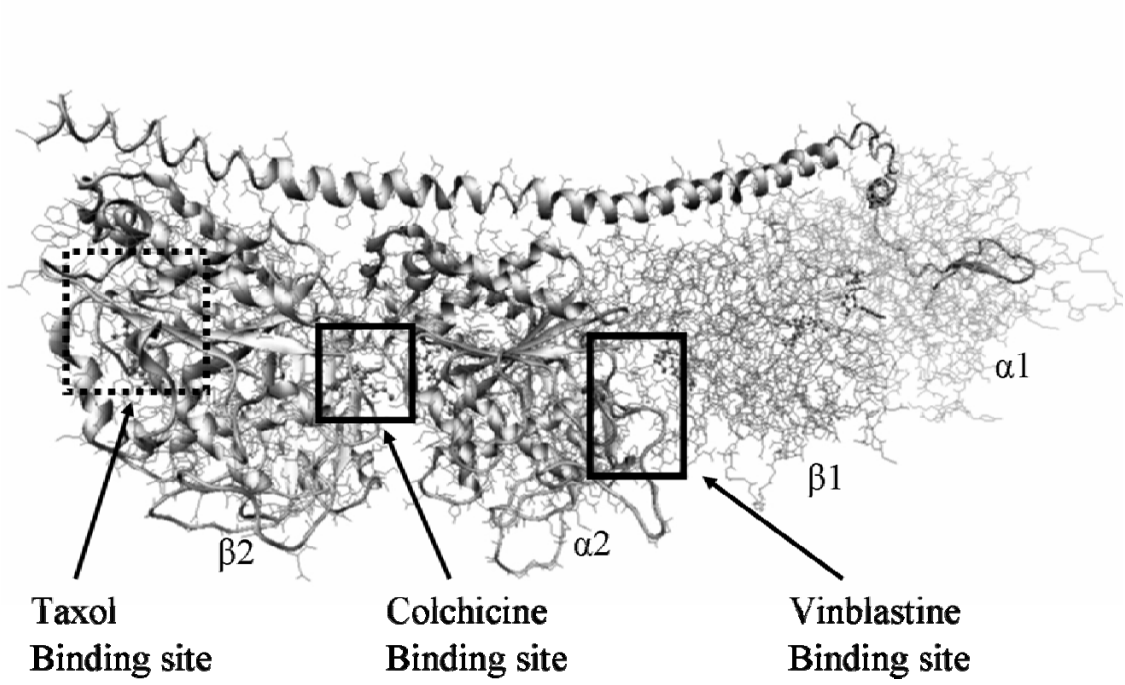
b: Stilbene 6c **R=OH**

analogues; novel compounds derived from this core continue to hold interest as potential therapeutics[152-154]. Stilbene and its analogues (Figure 9 a, b) are structurally similar to combretastatin and are also able to bind to microtubules, suppress microtubule dynamics, and arrest the cell cycle at a G2/M phase that is associated with induction of apoptosis[155,156].

Their mechanism of action is related to tubulin-binding properties that result in rapid tumor endothelial cell damage, neovascular shutdown and subsequent hemorrhagic necrosis[157]. Our collaborative group has recently synthesized a series of derivatives structurally related to both resveratrol and CA-4, in cis and trans orientations. Modifications were made by placing OH, NH2, or OCH3 groups at positions 30 and 40 and OCH3 at positions 3,5. The IC50 for each of each stilbene was tested in HL60 cells. Several active stilbenes were identified and, among them, cis-3,40,5-trimethoxy-30-aminostilbene (5C) and cis-3,40,5-trimethoxy-30-hydroxystilbene (6C) were found to induce HL60 apoptosis at nanomolar concentrations (IC50 = 30 nM)[158]. Examination of the structures of stilbenes 5C and 6C reveals similarity to colchicine, combretastatin CA4[158], ZD6126, and AVE-8062.

The new stilbenes interfere with microtubule formation at the tubulin colchicine binding site (Figure 10), in a similar manner as CA-4 and AC-7739, inducing apoptosis in HL60 leukemia cells, but not in normal c-kit-positive hematopoietic progenitor cells at similar concentrations[156]. These two stilbenes have a similar cytotoxic effect in HL60 cells

Figure 10. The Tubulin-colchicine:RB3-SLD complex, The complex includes alternating tubulin αβ heterodimers, with the colchicine binding site at the intradimer interface, the taxol binding site on the β subunit and the vinblastine binding site at the interdimer interface of the αβ subunit.



expressing multiple-drug resistant gene[158], which is responsible for chemotherapeutic resistance to anthracycline in leukemia. In addition to the effect in leukemia cells, stilbene 5C is also highly potent against various solid tumor cells and blocks cell cycle progression in G2-M phase[158]. Stilbene 5C is tolerated in mice up to 100 mg/kg by intraperitoneal injection without major organ toxicity. In particular, there was no bone marrow toxicity, and the ability of bone marrow engraftment was not affected by stilbene 5C treatment[156]. Similar to other colchicine site tubulin inhibitors, stilbene 5C selectively suppresses tumor vascular perfusion without damaging normal vascular perfusion based on a DCE-MRI study[159]. Most importantly, mice treated with five daily injections of 5C at 25 mg/kg/day did not show any compromise in heart function, indicating that it could be a colchicine site inhibitor with negligible cardiotoxicity[159].

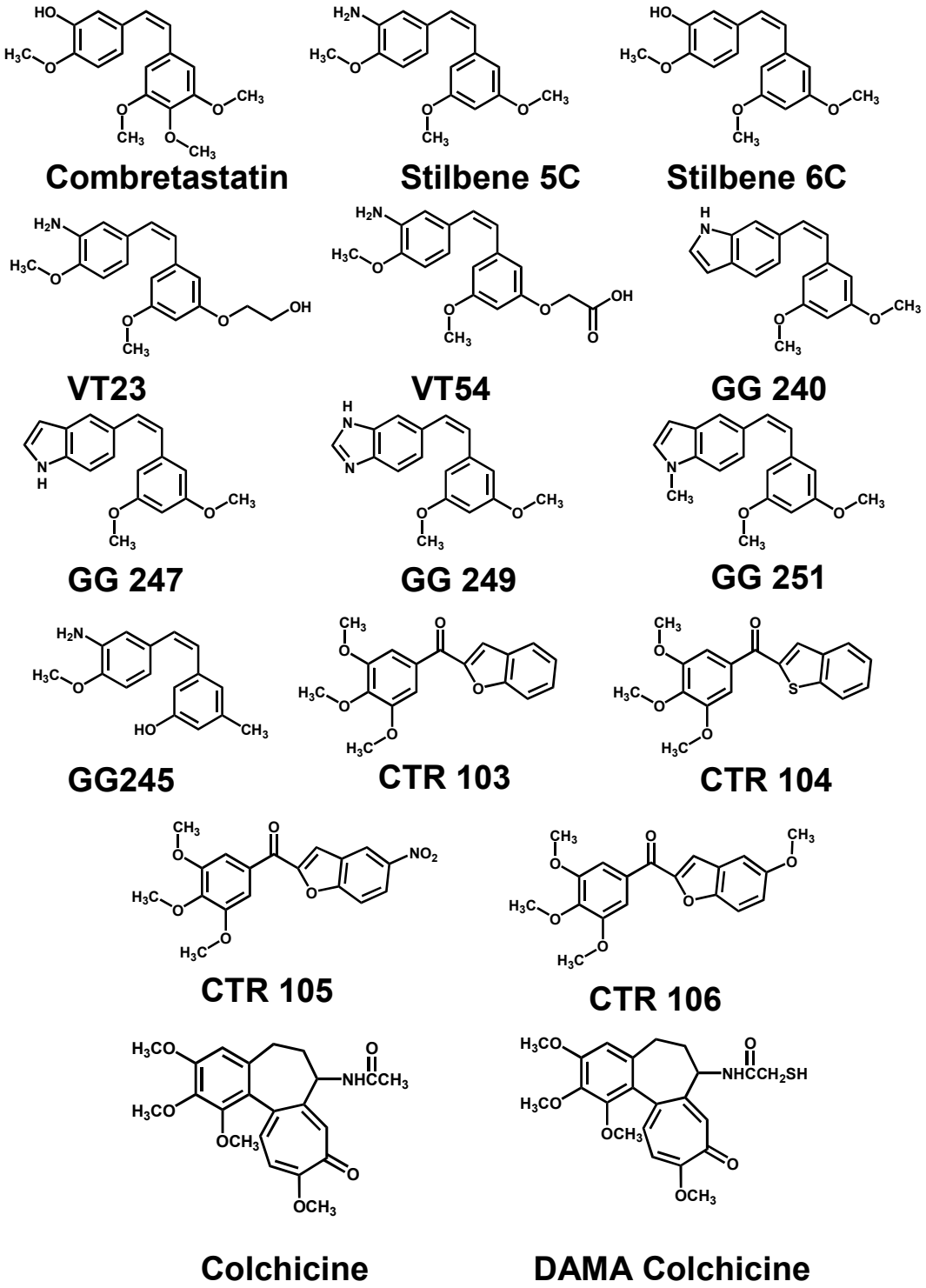
As the synthesis of stilbene analogues is relatively facile, our goal for this study was to create a quantitative structure-activity model of these compounds that would guide the synthesis of new, potentially more efficacious stilbene derivatives. Thus, we report here on the structural requirements for interaction between stilbene analogs and tubulin with computational docking and hydropathic scoring of multiple stilbene analogs into the colchicine-binding site of tubulin.

3.2 Materials and Methods

3.2.1 Synthesis

Synthesis of stilbenes 5c, 6c and its analogs (Figure 11) was performed as described by Roberti et al.[158] who provided all samples for bioactivity analysis.

Figure 11: Stilbene analogs



3.2.2 Antiproliferative Activity of Stilbenes against Human Tumor Cell Lines

The antiproliferative activity of stilbene derivatives was determined by Alamar Blue™ staining. The bioassays were carried out in the lab of Dr. Ray Lee at Massey Cancer Center, VCU. In brief, cells were grown in 96-well plates and treated with 0, 0.01, 0.03, 0.1, 0.3, 1.0, and 3.0 μM of the stilbene for 48 hours before being harvested for Alamar blue™ staining. In this staining, 1/10 volume of Alamar Blue™ solution was added to each well and optical density (OD) at 570 and 600 nm was determined by a microplate reader. The percentage of growth inhibition was calculated according to the manufacturer's formula as follows: $[(117216 \times A_{570}) - (80586 \times A_{600})]/[(117216 \times A^{o}_{570}) - (80586 \times A^{o}_{600})] \times 100$. In this formula, A_{570} is the absorbance of the treated samples at 570 nm; A_{600} is the absorbance of the treated samples at 600 nm; A^{o}_{570} is the absorbance of the untreated samples at 570 nm; and A^{o}_{600} is the absorbance of the untreated samples at 600 nm. The two constants, 117216 and 80586, are the extinction coefficients of Alamar Blue™ at 570 and 600 nm respectively. Each concentration was repeated in triplicates.

3.2.3 Model Building

The X-ray crystal structure (3.58 Å) of αβ-tubulin complexed with DAMA-colchicine[132] (PDB code: 1SA0) was used in this study. The stathmin-like domain and the C and D subunits were removed from the model. After hydrogen atoms were added to the model, their positions were optimized to an energy gradient of 0.005 kcal-Å/mol with the Tripos force field (in Sybyl 7.3) while keeping heavy atom positions fixed. The models for stilbene

analogues were constructed using the Sybyl 7.3 (www.tripos.com) suite and optimized similarly[78].

3.2.4 Docking

Computational docking was carried out using the genetic algorithm-based ligand docking program GOLD 3.0[69]. GOLD explores ligand conformations fairly exhaustively and also provides limited flexibility to protein side chains with hydroxyl groups by reorienting the hydrogen bond donor and acceptor groups. The active site was defined by using colchicine as the reference molecule in the protein active site creating an approximate radius of 10 Å around the reference molecule using the GOLD cavity detection algorithm. Because of the relatively poor resolution of the X-ray crystal structure and following the approach of Nguyen *et al.*[131], GOLD docking was carried out with template similarity constraints. This constraint biases the conformation of docked ligands towards a given solution. The trimethoxyphenyl fragment of colchicine was used as the template for biasing the pose of all ligands. In this study we performed 100 GOLD genetic algorithm runs, as opposed to the default of 10 and early termination of ligand docking was switched off. All other parameters were as the defaults. To evaluate and validate GOLD performance the co-crystallized ligand DAMA-colchicine[132] was extracted and docked. GOLD accurately reproduced the experimentally observed binding mode of DAMA-colchicine in αβ-tubulin. Combretastatin was docked first and the resulting model was scored and optimized. The remaining stilbene analogs were docked and optimized using combretastatin as a reference within 0.5 Å RMSD by using the same

parameters. Docked ligands were scored using the HINT force field scoring function and iteratively optimized for maximal interaction.

Dockings with different/optional constraints such as enforced hydrogen bonds, hydrophobic regions and scaffold match were also explored. For hydrogen bond constraints, docking was biased so that the ligands make hydrogen bonds with Asn258, Ser178, Asn101, and the backbone amides of Ala180 and Val181. For region hydrophobic constraints the ligand positions were constrained by defining a hydrophobic sphere where the tri-methoxy phenyl moiety of colchicine was positioned. Then specific ligand atoms to be docked in the hydrophobic region of the active site were defined. Alternatively, scaffold match constraints were used to place the ligand at a specific position within the active site. Generally, however, because the binding site is rather featureless, constraint or template-free docking was not successful.

3.2.5 Hydropathic Scoring

The HINT (Hydropathic INTeractions) scoring function[91] (version 3.11S β) was used to investigate the structural aspects of the interactions by analyzing and ranking the GOLD docking solutions. For selection of the optimum docked conformation and to further differentiate the relative binding efficacy of the stilbene ligands, interaction scores were calculated for each pose found by docking. The protein and ligands were partitioned as distinct molecules. "Essential" hydrogen atoms, i.e., only those attached to polar atoms (N, O, S, P) were explicitly considered in the model and assigned HINT constants. The inferred solvent model, where each residue is partitioned based on its hydrogen count,

was applied. The solvent accessible surface area for the amide nitrogens of the protein backbone were corrected with the "+20" option. Finally, HINT scores were plotted against experimental binding free energy.

3.3 Results and Discussion

3.3.1 Antiproliferative Activity of Stilbene Analogues

The biological activity of all compounds was tested in UCI101 ovarian cancer cells and qualitatively similar trends were observed in MDA-MB231 breast cancer cells[160]. Compounds could be separated into three groups according to their potency. Group A contains the most potent compounds, including combretastatin, stilbene 5C, GG251, colchicine, DAMA-colchicine, VT23 and stilbene 6C with IC_{50} less than 100 nM. Group B contains GG240, GG247, GG245, GG249, which has IC_{50} in the intermediate range between 0.3 and 1.0 μM. Group C is not active with IC_{50} above 5 μM (see Table 3).

3.3.2 The Colchicine Binding Site

A number of groups have performed modeling studies on the colchicine binding site because it is recognized as a potential target for anticancer drug development[131]. However, the low resolution of the available crystal structures for tubulin have made it difficult to fully delineate the essential structural and functional features involved in tubulin inhibition. In a low resolution crystal structure model our knowledge of the correct orientation of sidechains is limited by ambiguities in position and orientation because the experimental electron density envelopes are generally featureless. The crystal structure of tubulin protein used in this study has a fairly poor resolution of 3.58 Å, which

Table 3: Experimental IC_{50} and docking results for Stilbene and Campione derivatives.

Compound	Activity Set	IC_{50}[a]	pIC_{50}	$\Delta G_{binding}$ (kcal mol^{-1})	HINT score	HINT LogP
Combretastatin	A	3.3 ± 0.4 nM[b]	8.53	-11.62	1015	4.43
Stilbene 5C		32 ± 4 nM	7.53	-10.26	860	4.05
GG251		28 ± 3 nM	7.53	-10.26	861	4.73
Colchicine		30 ± 2 nM	7.53	-10.26	841	3.24
DAMA-colchicine		29 ± 2 nM	7.53	-10.26	810	3.70
Stilbene 6C		52 ± 7 nM	7.30	-9.96	673	3.94
VT23		65 ± 8 nM	7.22	-9.85	446	3.17
GG240	B	0.32 ± 0.02 µM	6.52	-8.89	579	2.85
GG247		0.30 ± 0.04 µM	6.52	-8.89	563	3.99
GG245		1.0 ± 0.08 µM	6.00	-8.18	285	3.44
GG249		1.1 ± 0.07 µM	6.00	-8.18	278	3.10
CTR104	C	5.2 ± 0.9 µM	5.30	-7.23	252	5.60
CTR105		10 ± 0.9 µM	5.00	-6.82	-202	2.44
CTR103		105 ± 6 µM	4.00	-5.45	49	5.44
CTR106		110 ± 12µM	4.00	-5.45	-446	5.37
VT54		107 ± 14µM	4.00	-5.45	-313	2.98

[a]Antiproliferative activity against UCI-101 ovarian cancer cell line using the Alamar Blue assay.
-pIC_{50} and $\Delta G_{binding}$ calculated for $10 \times IC_{50}$.
-Colchicine IC_{50} data was recorded against UCI-101 ovarian cancer cell line. For calculation purposes DAMA-colchicine was assumed to have same binding as Colchicine.

somewhat undermines our ability to structure-based design highly selective ligands[132]. In earlier studies with the colchicine binding site we carried out computational docking studies along with hydropathic interaction analysis on a family of substituted pyrroles and were able to represent the complex structure and the binding modes of the pyrrole class of inhibitors (see chapter 2)[100].

Binding modes for each stilbene analogue were investigated to understand the steric, electrostatic, and hydropathic features of the colchicine binding site. A molecular modeling study of docking these ligands into the colchicine site of αβ-tubulin was carried out in order to accurately represent the complex structure. The colchicine binding site is positioned at the interface between the α and β subunits of the tubulin protein, with the major part of it buried in the β subunit and lined by helices 7 and 8. The cavity, which is funnel shaped, has a volume of about 600 Å and opens up towards the α subunit of the interface surrounded by residues Asn101α, Thr179α, Ala180α, Val181α and Thr314β, Asn349β, Asn350β, Lys352β. The other, β subunit, end of the cavity is surrounded by residues Tyr202β, Val238β, Thr239β, Cys241β, Leu242β, Leu248β, Leu252β, Leu255β, Ile378β, and Val318β and forms the narrow funnel end-like part of the cavity. The predominance of hydrophobic residues confer a strong hydrophobic character to this part of the cavity. At the wider portion, the cavity is surrounded by Ala250β, Asp251β, Lys254β, Asn258β, Met259β, Ala316β, Ala317β, Thr353β and Ala354β making it moderately polar/moderately hydrophobic. DAMA-colchicine (and presumably colchicine) is snugly positioned in the crystal structure of the complex. The trimethoxyphenyl (TMP) moiety of colchicine is positioned in the pocket such that its sits snugly in the narrow hydrophobic region of the pocket with one of its methoxy oxygens

involved in hydrogen bonding with the thiol of Cys241β. Colchicine also forms hydrogen bonds with the backbone amides of Ala180α and Val181α.

The structural complex as reported in the X-ray structure was refined through ligand functional group and protein sidechain optimization, as incorporated in the HINT program. Iteratively, colchicine was translated and rotated and optimized for interactions. Taking colchicine as the template, combretastatin was computationally docked and scored. Similar to above, the combretastatin-tubulin complex was optimized. Next, the stilbene analogs were docked, this time taking combretastatin as the template, followed by HINT functional group and protein sidechain optimization. The docked models of the stilbene analogs fit within the pharmacophore model proposed by Nguyen *et al.*[131] and are similar to the models we reported earlier in chapter 2 for the pyrrole derivatives bound to αβ-tubulin[100]. These studies, coupled with HINT interaction analyses, are able to describe the complex structure and the binding modes of stilbene inhibitors. Note that HINT scores are very sensitive and slight positional differences are detectable in the scores. Qualitative analyses of the results showed general agreement with the experimental *in vitro* antiproliferative activity for these derivatives.

3.3.3 Structure-Activity Binding Relationships

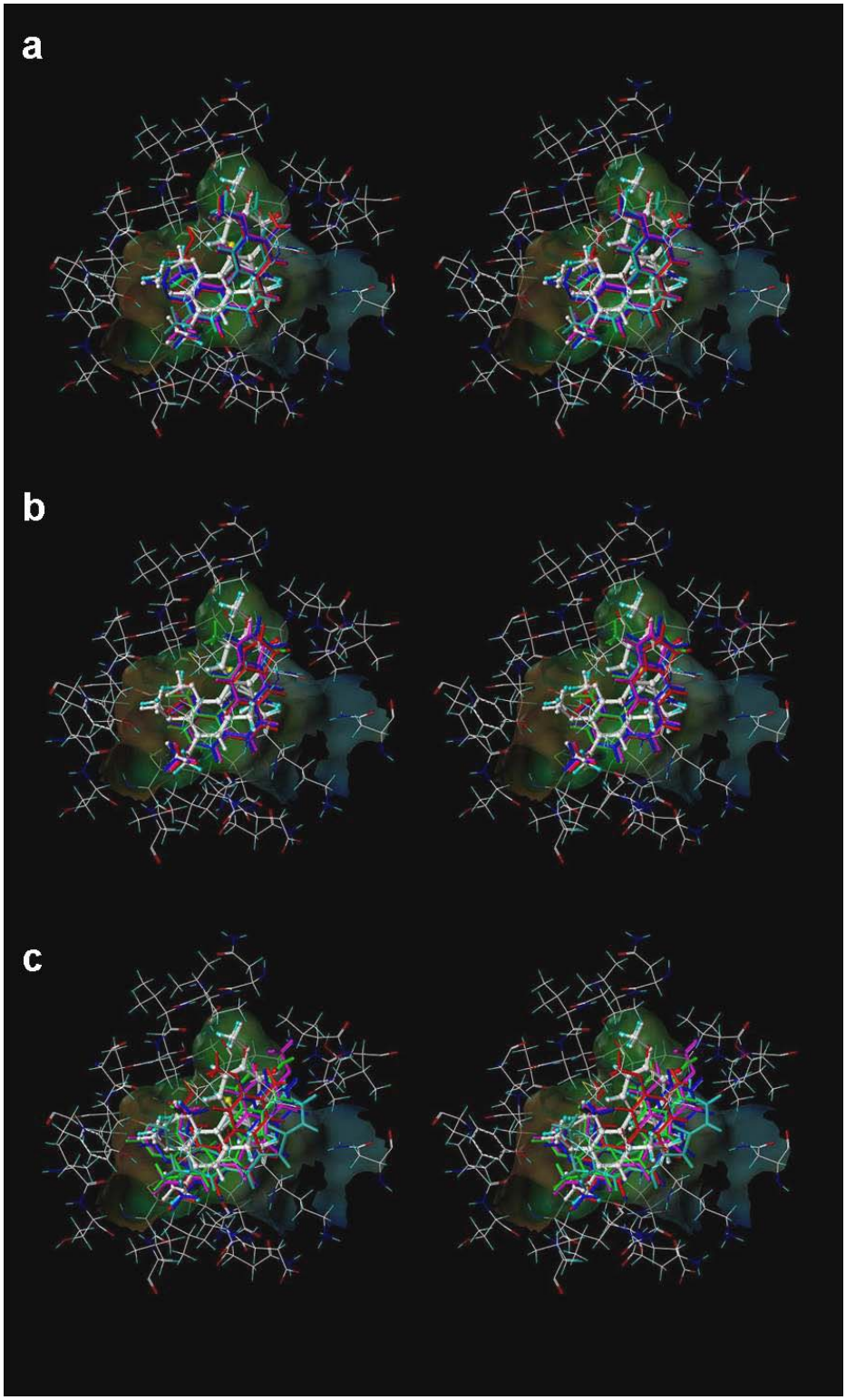
Structural analysis of the binding pocket identified important intermolecular interactions that mediate binding. The stilbene analogues were clustered into three activity sets in order to study them in detail (see Table 3). The first (A) was comprised of compounds that showed activity from 3 nM to 60 nM IC_{50}. The second (B) consisted of ligands with

IC_{50} values ranging from 0.3 μM to 1 μM. The remaining ligands, with IC_{50} values of 5 μM and above, comprised the third set (C). HINT hydropathic analysis reveals significant detail concerning the forces orienting these ligands in the binding site. First, hydrophobic interactions are the dominating force contributing towards the stability of the complexes, with additional hydrogen-bonding interactions anchoring the ligands in the cavity. The trimethoxyphenyl (TMP) moieties of colchicine and combretastatin are positioned within the narrow hydrophobic region of the cavity while the carbonyl oxygen on the unsaturated seven-member ring of colchicine and the hydroxyl group on combretastatin B ring are anchored through strong hydrogen bonding interactions with Asn258β and the Ala180α and Val181α backbone amides.

First, in examining the docked models for activity subset A (see Figure 12A), it is interesting to note that, although the stilbene compounds and combretastatin are quite dissimilar structurally to colchicine (excepting the TMP moiety), they are generally positioned in the binding pocket with essentially the same mode. In the case of stilbene 5C and GG251 the hydrophobic substituted phenyl ring fits snugly in the hydrophobic (narrow funnel) region of the binding pocket that superimposes on the TMP moiety of colchicine. In fact, stilbene 5C and GG251 have a very similar binding mode to that of combretastatin, the major contribution towards binding coming from hydrophobic interactions. The methoxy substituted phenyls are positioned deep in the hydrophobic cavity surrounded by Cys241β, Leu242β, Leu248β, Ala250β, Leu255β, Ala354β and Ile378β, all of which contribute to favorable hydrophobic-hydrophobic binding. The phenyl ring of Stilbene5C and GG251 fits in a hydrophobic glove formed by the Leu248β

and Leu255β. Favorable polar interaction with Asn101α, Cys241β and Asn258β also contribute to the tight binding. The NH_2 group on the B ring of Stilbene5C faces towards the polar opening and is stabilized with a strong hydrogen-bond to the amide oxygen of Asn258β with a length of 2.55 Å. Another set of strong hydrogen bonds is formed between the amide backbone of Val181α and the amine on the B ring of Stilbene5C with a hydrogen bond distance of 3.384 Å. However, in the case of GG251 the hydrogen bonding interaction is not observed with the Asn258β and Val181α amide backbone, instead the complex is predominantly stabilized through hydrophobic interactions. Similar interactions are observed with stilbene 6C and VT23 with some subtle differences.

Figure 12. Stilbene analogs docked at the colchicine binding site on αβ-tubulin. (a) Substituted stilbenes with activity in sub-μM IC_{50}. (b) Compounds with IC_{50} ranging from 0.3 μM to 1.0 μM. (c) Compounds with IC_{50} values above 5 μM.



In the case of stilbene 6C, although the hydroxyl group is retained on the B ring as in combretastatin, the removal of methoxy group from the position 4 of TMP moiety results in loss of activity. However, this loss is offset by introducing the NH_2 group on ring B in stilbene 5C, thus accounting for the slight differences in their activity. In the case of VT23 the extension of methoxy chain on TMP moiety results in further loss of activity probably due to geometric constraints enforced by the narrow hydrophobic region of the cavity.

On analyzing subset B (compounds GG240, GG247, GG245, and GG249), docked ligands in the low μM range, it can be seen that these ligands are relatively similar to the subset A ligands. In this set of compounds substitution on ring B, in the cases of GG240, GG247 and GG249 the indole-carrying ring and in GG245 the amine and methoxy carrying ring, is varied. These substitutions result in a 10-fold decrease in the activity in compounds GG240 and GG247 where the N-methyl substitution is removed from the indole ring of GG251. Flipping the ring, as in the 6- and 5- substituted indole ring of GG240 and GG247, does not have any significant affect on activity. Activity is further decreased if the indole ring is replaced by a benzimidazole ring as in compound GG249 and a similar loss in activity is noted in GG245 where the methoxy groups of the stilbene 5C methoxy phenyl moiety is replaced by a methyl group and a hydroxyl group, confirming the importance of the methoxy phenyl moiety on ring A in binding. The benzofuran and benzothiophene analogs in the CTR series of compounds (subset C) are similar to 2-aroylindoles, where the 2-aroylindole ring is replaced by benzofuran and benzothiophene. Our docked complexes with the CTR series of compounds agree with

the Nguyen *et al.* pharmacophore model[131], but the rings are inappropriately substituted to make the required contacts with the binding site residues – leading to poor activity of these compounds. Figure 13 represents the putative binding mode of Stilbene 5c, the most potent compound of the series.

3.3.4 Predictive Models for Ligand Binding

Figure 14 presents the correlation between the experimental binding ($\Delta G_{binding}$ as calculated from IC_{50}) in kcal mol^{-1} and HINT scores for the synthetic stilbene analogs in this study. The trend represented by this plot indicates that higher scoring complexes are generally among those with more favorable free energies of binding, while lower scoring complexes are generally those with unfavorable binding. The correlation equation:

$$\mathbf{\Delta G = -0.0041\ H_{TOTAL} - 6.877,} \qquad (10)$$

has an $r^2 = 0.898$ and a standard error of ± 0.85 kcal mol^{-1}.

The IC_{50}s, antiproliferative activities of the compounds, are being taken in this work as approximations of binding affinity. The scoring function does not take into account cell permeability and completely ignores whether or not the compound could *in vivo* or *in vitro* be accessible to the binding site.

Figure 13: Representation of interactions of stilbene 5C in the colchicine active site of the tubulin protein.

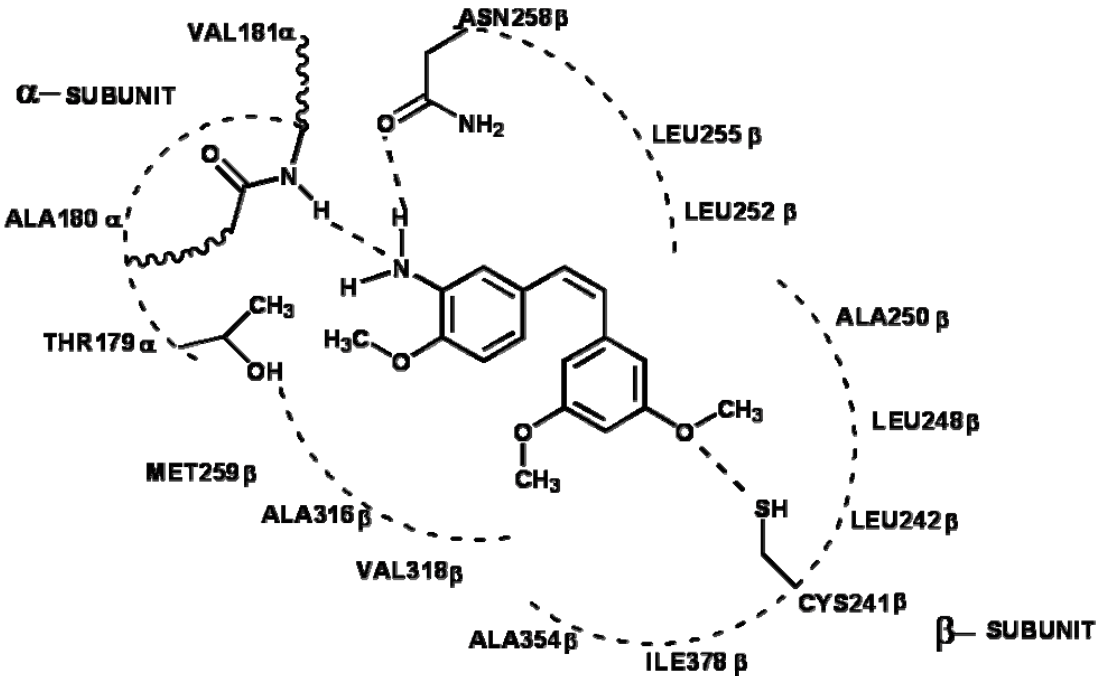
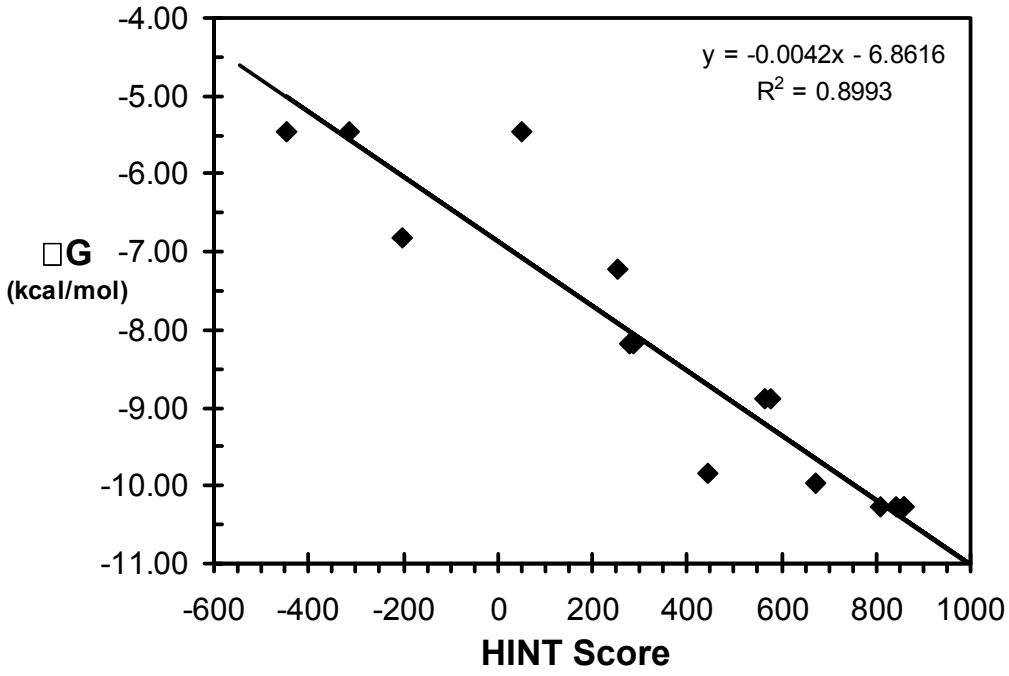


Figure 14. Correlation plot between free energy of binding, ΔG vs. HINT score. The line represents the regression for ΔG vs. HINT score for all protein-ligand complexes in this study.



However, the similar structures of these compounds and the fairly narrow range of LogPs suggest that these properties would be similar, if not the same, for all of these compounds, and thus can probably be ignored while comparing the compounds within the series. We believe that the model of Figure 14 is predictive such that it can distinguish the active (subset A) ligands from the inactive (subset B) ligands with reasonable confidence. Refinement of the model with additional data will further improve the understanding of binding process and predictive ability of the model.

3.4 Summary

The aim of this study was to accurately represent the complex structures and the binding mode of a new class of stilbene-based tubulin inhibitors. Both qualitative and quantitative analysis of the results suggested that the model was in general agreement with the *in vitro* antiproliferative activity observed experimentally for these compounds. A good correlation between the modeled interaction energies and estimated free energies of binding calculated from IC_{50} suggest that our model is able to represent the complex structures and the binding modes of inhibitors and under some circumstances be predictive with respect to new members of the stilbene series. We believe that we can identify active ligands from inactive ligands with reasonable confidence; our analysis has provided a rationale for selecting substituents that will yield more tightly binding analogues.

CHAPTER 4

A Novel and Efficient Tool for Identifying and Characterizing Protein Cavities and Binding Sites.

4.0 Abstract

Systematic investigation of a protein and its binding site characteristics are crucial for designing small molecules that modulate protein functions. However, fundamental uncertainties in binding site interactions and insufficient knowledge of the properties of even well-defined binding pockets have often made it difficult to design effective drugs with optimal activity. Herein, we report the development and implementation of a novel cavity detection algorithm that utilizes HINT toolkit functions that we are naming VICE (Vectorial Identification of Cavity Extents). This algorithm, which is based on geometric criteria applied to simple integer grid maps to delineate binding sites, is very efficient. In testing, we carried out a systematic investigation on a diverse set of proteins for locating and characterizing the indentations, cavities, pockets, grooves, channels and surface regions on proteins, protein-protein and protein-polynucleotide complexes. An interactive front-end provides a quick and simple procedure for identifying, displaying and manipulating cavities in a known protein structure. Information describing the cavity including its volume and surface area metrics, and lists of atoms, residues and/or chains

lining the binding pocket can be easily obtained and analyzed. For example, the relative cross-sectional surface area (to total surface area) of cavity openings in well-enclosed cavities is 0.06 ± 0.04 and in surface clefts or crevices is 0.25 ± 0.09.

4.1 Introduction

Modulation of the dynamics of a target protein binding site to elicit a pharmacological response is the major therapeutic approach for the treatment of a plethora of diseases. This is usually accomplished by developing small molecules that occupy a ligand recognition site. Drug development is a challenging process, owing to fundamental uncertainties in structural determination and associated issues such as structural and physicochemical characterization of the binding pockets, even under relatively static conditions such as in crystals subjected to x-ray analyses. Reliable, rational and efficient approaches to identifying and characterizing the binding sites of protein and other bioactive molecules should be valuable in the design of new drugs[161]. In recent years there has been a surge in the number of crystal structures deposited in Protein Data Bank[162]. Concomitantly, NMR and X-ray crystallography have played an increasingly crucial role in drug discovery through structure-based methods and virtual screening of extensive libraries of compounds. Facilitating this has been the design and development of many computational tools with a large range of functions. In particular, a number of programs have been developed to *de novo* identify the binding pockets in proteins[161,163]. Such tools have provided valuable information for better understanding protein binding site architecture. However, the accurate identification and quantitation of binding

pockets is not an entirely straightforward process, and the existing computational tools have numerous strengths and weaknesses.

Proteins have "pockets" for molecules to bind; however, these pockets may not be observed from an initial inspection. Protein surfaces are formed by numerous cavities and protrusions that are interlinked through small narrow channels and are often are interspersed with numerous holes or voids. However, amidst all of this surface irregularity, the complex positioning of residues creates specific microenvironments for ligand recognition, binding and catalysis. These ligand binding sites vary widely in shape and size and are often roughly classified according to their position on a protein and/or shape and described as deep pocket, grooves, indentations, surface concavities, branched pockets, voids, channels, etc. The size and shape of these protein cavities dictates the three-dimensional geometry of ligands that will bind within, and guides the important intermolecular contacts that mediate this binding. Binding sites that are formed by several neighboring pockets/cavities and auxiliary pockets near the active site are often suggested as providing additional ligand binding surface, which adds further to the complexity. Efficient analysis of the shape and size of protein pockets and cavities thus becomes important as structural changes due to side-chain rotations and backbone movements, loop motion and/or ligand-induced conformational changes may significantly alter the topography of the active site. A thorough structural analysis of the target binding site is critical to propel a drug discovery project forward. There has been significant progress in this endeavor in recent years[161,163,164].

4.1.1 Theoretical Approaches for Identifying Binding Sites on Proteins.

Identification and characterization of active sites is key in studying protein structure, particularly when designing molecules that interfere with function and modulate activity. There are a number of ways in which binding sites or cavities in proteins can be located and characterized, e.g., with several existing programs such as VOIDOO [165], LIGSITE [166], POCKET [167], POCKET-FINDER [168], CAST [169], PASS [170], APROPOS [171], SURFNET [172], Q-SITEFINDER [173] and others. These programs can be generally classified into categories according to the approach they take to identify and delineate the cavity: i) evolutionary methods (structure/sequence alignments); ii) probe/energy-based methods; and iii) geometric approaches.

Evolutionary methods use a heuristic approach to predict cavities in unknown proteins from known protein structures based on family and/or functional criteria. With the abundance of structural- and sequence-related data for many protein families, this approach has found increased application in identifying and characterizing protein target binding sites[174,175]. Structural similarity and three-dimensional templates are used to compare and classify putative binding sites in uncharacterized protein structures with unknown functions, e.g., with similarity searches over functional site databases like LigBase [176] and INTERPRO [177] that detect functional similarity when homology is minimal. The approach by Bickel *et al.* [178] uses statistical methods to identify active sites by residue identity within and outside functional subfamilies. Programs like ConSurf [179] identify functional regions of proteins by surface mapping of phylogenetic information, while Rate4Site [180] applies evolutionary determinants in mapping the functional regions

on a protein surface. These methods are likely to continually evolve with the increasing availability of structural and sequence data from structural genomics projects.

The idea of *in silico* mapping of protein surfaces was first conceptualized by Lee & Richards (1971) [181] based on the idea of an "accessible" surface area. Connolly (1982, 1983) [182] suggested the concept of "solvent excluded surface" and developed the eponymous algorithm for calculating molecular surfaces with a rolling spherical probe. Later, Kuntz *et al.* developed an algorithm that fills all pockets and grooves on the surface of receptor molecule with a set of balls [183]. While the probe sphere radius is generally 1.4 Å to approximate a water molecule surveying the solvent accessible surface of the protein, this sphere radius can be varied to map other representations such as the van der Waals surface. Kuntz *et al*. used this approach to define the binding site in the first implementations of the DOCK automated docking program [33,184]. Another novel approach of using spherical probes on a regular Cartesian point grid was implemented by Peter Goodford in GRID [185] and by Martin Karplus in MCSS (multiple copy simultaneous searches) [186]. In GRID, a binding region on a protein is mapped by calculating interaction energies between a (functional) probe group placed at each grid point and the atoms of the protein. In MCSS, about 1000 to 5000 small functional groups (probes) are interacted with the protein surface simultaneously and energy minima are calculated to define favorable interaction sites. The generated functional maps of the binding site indicate the most favorable regions for placing ligand groups with properties similar to the probes. A number of cavity detection algorithms based on this approach have been reported: Voorintholt *et al.* adopted an approach where grids are used to store

the distance to the nearest atom [187]; a similar approach was taken by Del Carpio *et al*. [188] in searching for pocket regions in a protein; the POCKET program by Levitt and Banaszak [167] uses a 3D Cartesian grid and spherical probes to map protein surfaces and pockets using a modification of the marching cubes algorithm; and the CHANNEL algorithm [189] uses a sphere of radius R to probe a node space that fills the unit cell of a crystal lattice.

Some probe/energy-based approaches to detect cavities overlap with geometric approaches in a way that a probe of a specified volume is only used to exclude van der Waals overlap as the protein surface is surveyed. The VOIDOO program reported by Kleywegt and Jones [165] uses atom fattening or a flood fill algorithm on a regular 3D grid to identify and delineate cavities. Another such widely used algorithm for cavity detection is LIGSITE developed by Hendlich, Rippmann and Barnickel [166]. This algorithm is similar to POCKET, but circumvents many of its drawbacks: first, grid points within a protein atom's van der Waals sphere are discarded; next, the remaining lattice points are scored according to their degree of burial by scanning grid points along the three Cartesian axes and the four cubic diagonals; and finally, the area delineating these grid points is clustered to describe contiguous cavities. Another method totally relying on geometric criteria is the PASS algorithm developed by Brady and Stouten [170] where the cavities in a protein are filled with a set of spheres.

Cavity detection based on alpha shape theory [190] incorporates a different, purely algorithmic, approach. The Automatic PROtein Pocket Search (APROPOS) method

developed by Peters, Fauck and Frömmel [171] is based on purely geometric criteria for identifying binding sites using atomic coordinates. Atoms are represented as a set of points in 3D Euclidean space and the envelope or surface is derived by Delaunay triangulation [191] of these points. The alpha shape algorithm describes these surfaces as lists of adjacent triangles and, depending on the value of alpha, delineates the cavity shape. The program CAST developed by Liang and Woodward [169] also applies alpha shape principles along with discrete flow theory to determine the shape of the binding pocket as a negative image of cavity derived from Delaunay tetrahedrons [191]. Alpha shapes and Delaunay triangulations are rich in geometric information from which area and volumes of inaccessible cavities can also be calculated. Most of these algorithms can fairly easily identify binding pockets and can be used in combination with other drug design tools to provide valuable information for structure-based drug design projects.

4.1.2 Vectorial Identification of Cavity Extents (VICE)

The present chapter describe our work in developing computational tools for drug design [192,193]. In this chapter we describe a new computer algorithm called VICE for identifying and delineating the active site in proteins or other biomacromolecules based on geometric criteria applied to simple integer grid maps using minimal floating point mathematics. Our objective in this chapter is to find pockets and shallow binding regions that have the characteristics of receptor sites, identify the amino acid residues surrounding them, and calculate descriptive metrics regarding the sites. The algorithm was applied to a diverse set of over 60 proteins in order to locate, investigate and characterize their various indentations, cavities, pockets, grooves, channels and surface regions. This is a starting

point towards comprehensive analysis of protein topography with respect to its function and an efficient and robust method for finding active sites that would be compatible with other tools and protocols we have developed based on our HINT empirical force field model [91,96,102].

4.2 Materials and Methods

The dataset of protein complexes in this study consisted of examples from the literature having binding pockets of diverse shapes, sizes and types. Table 3 lists the proteins evaluated by their PDB code and the associated cavity type for which the binding sites were calculated. All protein structure coordinates, in PDB format, were retrieved from the RCSB (Brookhaven) Protein Data Bank [162]. Molecular modeling was performed using the Sybyl 7.3 program suite (www.tripos.com) on Irix and Linux workstations [78]. The protein structures were prepared for this study by removing all the water molecules, ions, and any cofactors associated with the structure. Hydrogen atoms were added to the structures using the "Add Hydrogens" tool within the Sybyl Biopolymer module before further analysis.

The cavity detection and analysis programs were constructed using subroutines from the HINT toolkit [103]. Several new subroutines were composed for 3D map manipulation and analysis. Of particular value were an enhanced suite of functions for Boolean maps (where each grid value can only be zero or one) that forms the basis of the search algorithm as described in the Results and Discussion section. The algorithm provides several user-adjustable options to optimize the cavity calculation. With these parameters

it is possible to change the focus from the entire protein to a small region for a detailed investigation. For the initial surveys in this study, the grid boxes were defined as the molecular extents of each biomacromolecule with a grid resolution of 1 Å and margin of 3-5 Å. Most importantly, the cavity definition was set at 0.5, which is the fraction of vectors reaching a protein "wall" instead of the box edge. The maximum unrestrained path-length (vector length) was set to 20 Å by default but was increased to 40-60 Å to explore very large cavities or tunnels and channels. The minimum closed contour volume was set to 100 $Å^2$ to eliminate small clusters or irrelevant voids. The shaping factor was usually set to be 0.50, but was varied from 0.35 to 0.6 to interactively smooth some pockets that presented small and inaccessible sub-pocket regions. In the figures shown in this chapter, the surface of the pocket was displayed by contouring the cavity map at a value of 0.5, i.e., matching the cavity definition.

4.3 Results and Discussion

Protein binding regions provide a microenvironment for substrates, inhibitors, other proteins or biomacromolecules to interact and modulate the protein's activity. This chapter describes a computational tool for locating and investigating the binding regions of protein from a standard PDB file. This section describes and illustrates the algorithm, outlines the quantitative cavity metrics that can be derived through this algorithm, and highlights in some detail several of the more than sixty cases we have used to validate the methodology for this work. The rather remarkable variation that is observed in shapes and sizes of binding cavities is evident even from this small number of examples.

4.3.1 The VICE Algorithm

The VICE (Vectorial Identification of Cavity Extents) algorithm is schematically illustrated through Figure 15 to 21. After the region of interest, which can be the entire target protein or portions thereof, is defined, a grid cage with user selectable resolution is created. While 1 Å resolution is typical, larger or smaller values may be appropriate depending on computational requirements. The key advantage of this algorithm is that many of the calculations are performed on integers and on integer (Boolean) grid maps so that the method is very efficient. In the first step grid points occupied by atoms in the target molecule are set to zero, while those unoccupied are set to one. These latter points are potentially in the cavity; each will be examined by the algorithm. The search tools are sets of vectors whose directions are determined by the grid nodes (Figure 15). In the first shell the set of 2D vectors are {(1,0);(1,1);(0,1);(-1,1);(-1,0);(-1,-1);(0,-1);(1,-1)}, while in the second shell set the unique 2D vectors are {(2,1);(1,2);(-1,2);(-2,1);(-2,-1);(-1,-2);(1,-2);(2,-1)}. Each vector is projected until it reaches an edge of the grid box (Figure 16) and the nodes that the vector passes through constitute a path list.

Figure 15: Vector representations of direction: red = shell 1, green = shell 2, blue = shell 3

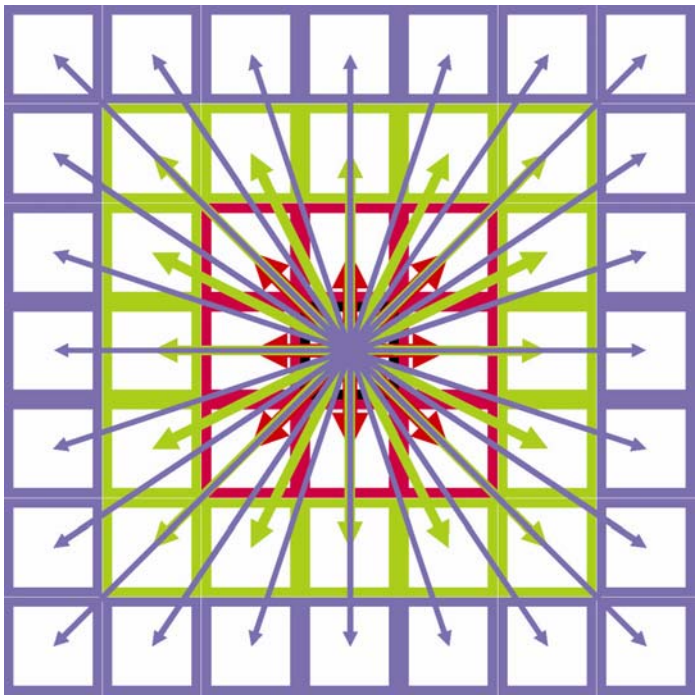
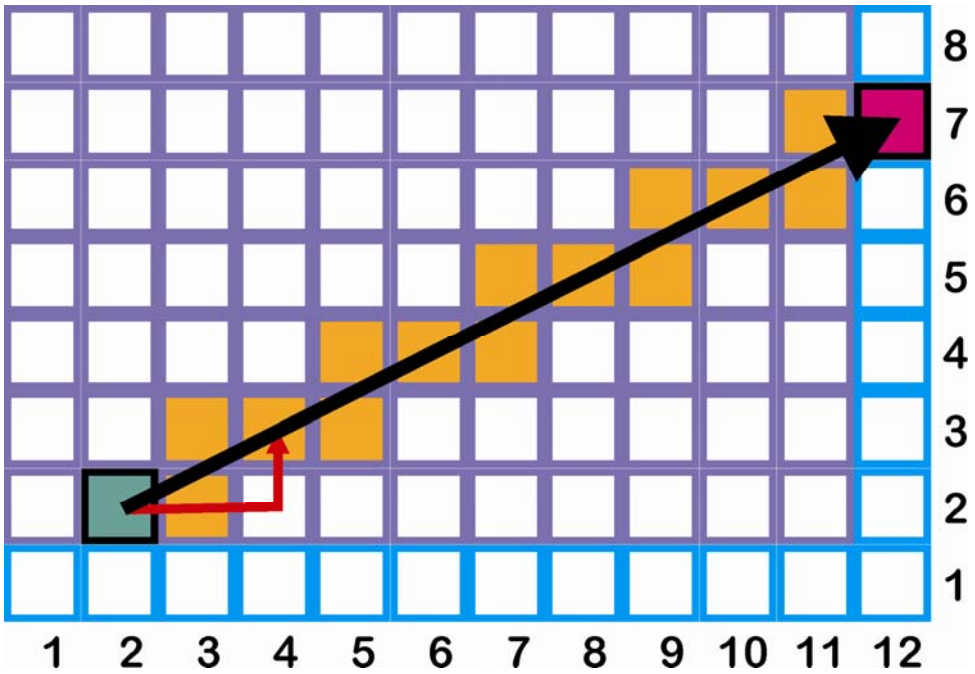


Figure 16: Vector (starting in green) continues until reaching grid box edge (red) and all nodes in path (orange shading) are tested.



Each vector is classified through analysis of its path list (Figure 17) as having: a clear path to edge, i.e., it does not pass through an occupied node; a blocked path; or is "stalled", i.e., it has neither reached the box edge nor has it passed through an occupied node. These latter vectors are treated as having clear path; their purpose is to ameliorate the possibility that a very long vector may inadvertently pass through occupied nodes belonging to another biomacromolecular subunit or because of a slightly curved pocket entrance. The stalled vector length is a parameter that may be adjusted depending on the anticipated dimensions of the cavity. The fraction of vectors classified as blocked is evaluated for each grid point. Thus, each grid point is classified as "inside" or "outside" the putative cavity based on a parameter with nominal cutoff value of 0.5 (Figure 18). A few grid points, mostly at the cavity mouth, are ambiguous (e.g., 0.5 ± 0.05); these are recalculated with additional shells of vectors and tightening criteria until a final disposition can be determined.

Two steps are applied to refine the cavity definition. First, narrow channels, i.e., one grid node in width, and tendrils are eliminated by forcing a requirement that each "inside" point have a minimum of "inside" neighbors (Figure 19). This can be applied recursively to "shape" the cavity. Lastly, to eliminate stray irrelevant pockets, each enclosed surface must have a minimum volume. While these steps can be performed automatically without user input, the algorithm is designed to display the intermediate raw maps and allow interactive application of the refinement.

Figure 17: Each grid point is surveyed with set of vectors that: a) are blocked by molecule (black), b) have clear path to box edge (green), or c) are stalled (pink) because with their finite length they do not reach box edge and thus are considered as having a clear path. Node 1 is clearly outside the cavity (more clear than blocked paths), node 2 is clearly inside (more blocked than clear), while node 3 is ambiguous requiring further examination with shell 2 vectors.

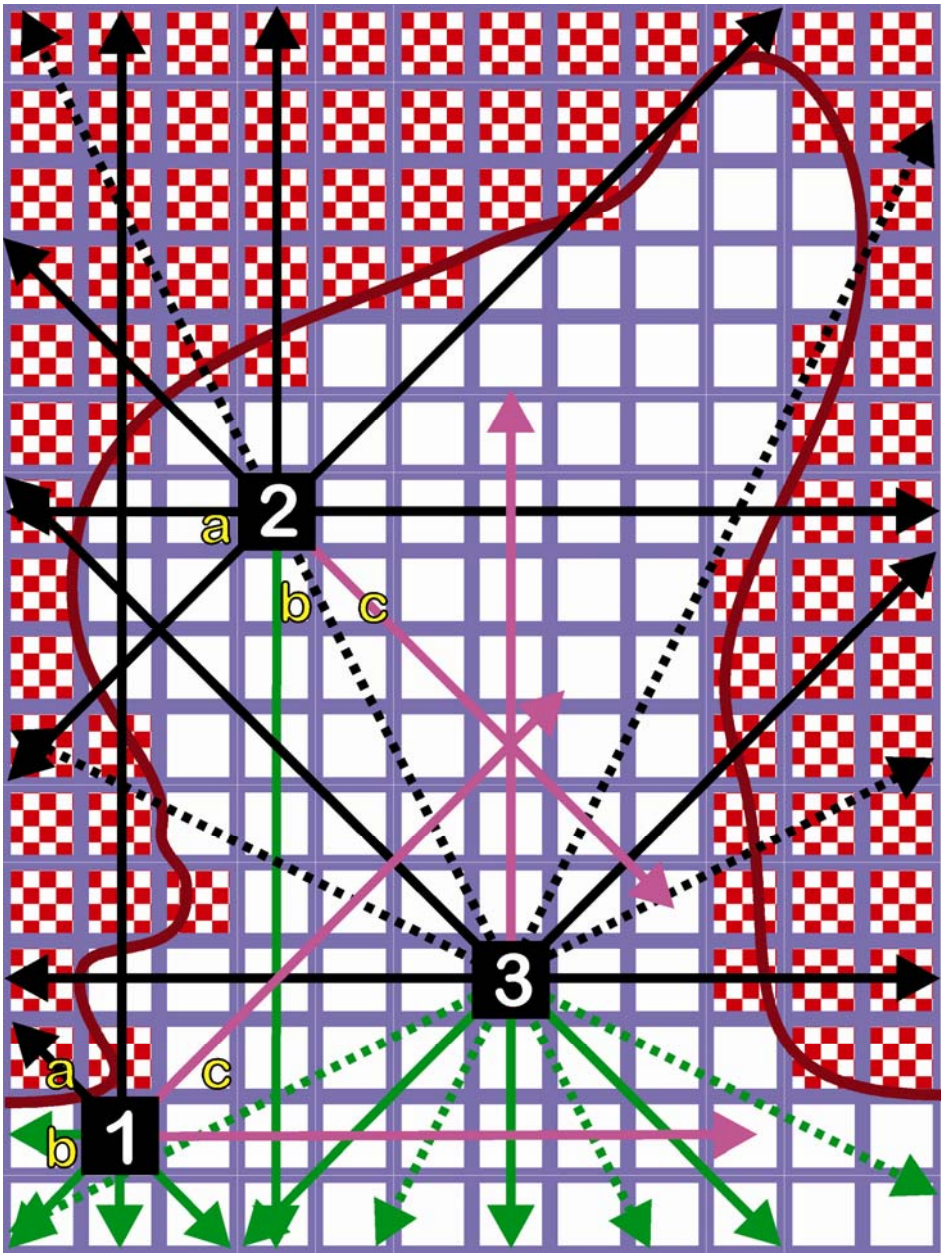


Figure 18: The fraction of blocked vectors is represented as a contourable scalar quantity that most impacts the definition of "cavityness" at the mouth.

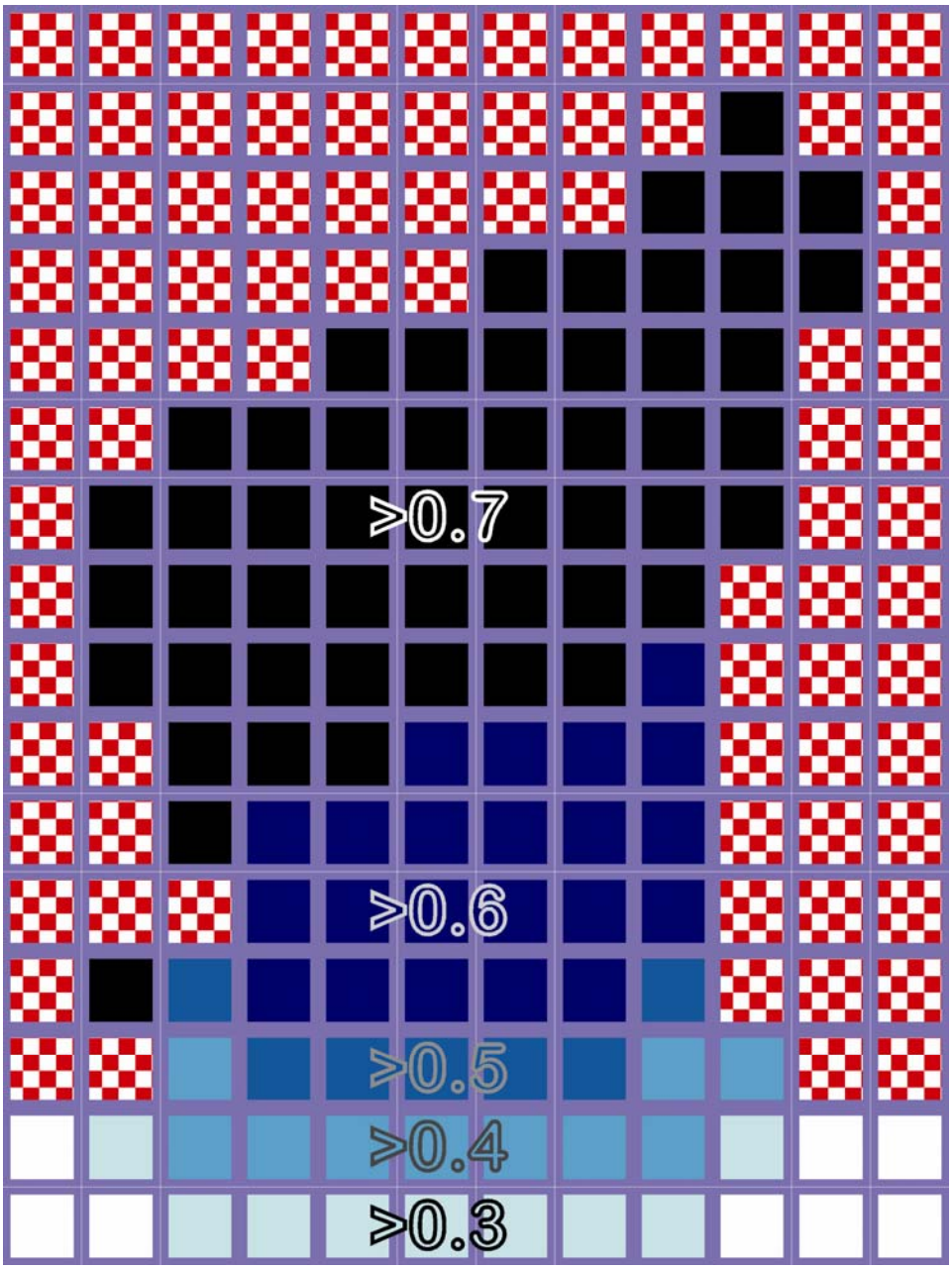
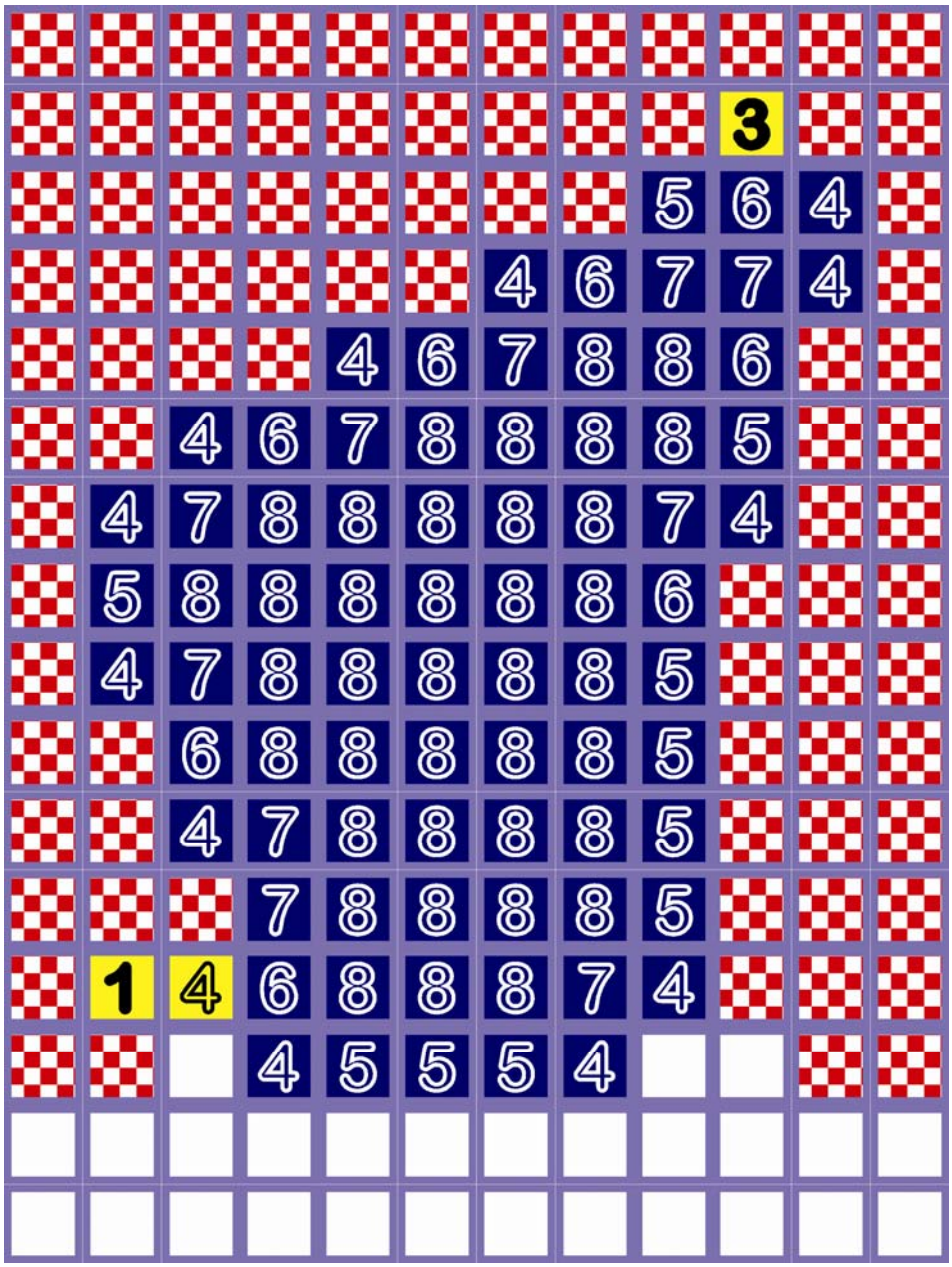


Figure 19: Tendrils, very narrow channels and other vague regions are tested with neighbor count that requires each node to have a minimum number of neighbors defined to be inside the cavity. The nodes indicated in yellow are subject to this filter, which may be applied recursively. Not shown: each closed solid contour must have a minimum volume or it will be deleted.



4.3.2 Overview of Protein Structure Studies

We carried out a systematic investigation of VICE on a diverse set of proteins to locate and investigate the indentations, cavities, pockets, grooves, channels and surface regions on these proteins. The dataset consisted of examples of proteins from the literature having binding pockets of diverse shapes and sizes. All protein structure coordinates, retrieved from the RCSB (www.rcsb.org) [162], were prepared as described in the Methods section. Our test set included: 16 cases where the binding pocket is a well-defined, well-enclosed, deeply buried pocket; 9 cases where the cavity or groove is on the protein's surface; 10 cases where the cavity is created by a protein–protein interface (more challenging since protein–protein dimers do not often show deep well-defined cavities that are putative binding sites for small molecules); 10 cases of cavities at DNA- or RNA-protein interfaces; 5 cases of protein structure pairs with very flexible binding pockets due to movements of flexible loops resulting in both open and closed cavities; 5 cases of proteins with channels or tunnels, i.e., ion channels, porins, and ligand gated channels; and lastly, 4 cases of proteins with multiple and/or allosteric sites including some with adjacent auxillary sub-pocket sites that may have additional biochemical roles.

A variety of metrics can be obtained or calculated for protein cavities. Of the most potential interest is the cavity volume that can be reported in terms of both its ligand-occupied and unoccupied fractions. Figure 20 illustrates how these metrics are calculated through manipulation of integer grid maps. We have also derived an automated algorithmic method (Figure 21) to estimate the cavity cross-sectional entrance areas.

Figure 20. Cavity Volume Metrics. The volume of the cavity (V_C) is indicated by yellow shading, the volume of the ligand (V_L) is indicated by vertical green bars, the volume of the ligand occupying the cavity (V_O) is the intersection of V_C and V_L, i.e., yellow shading + green bars. The unoccupied cavity volume is $V_C - V_O$, and the volume of the ligand outside the cavity is $V_L - V_O$.

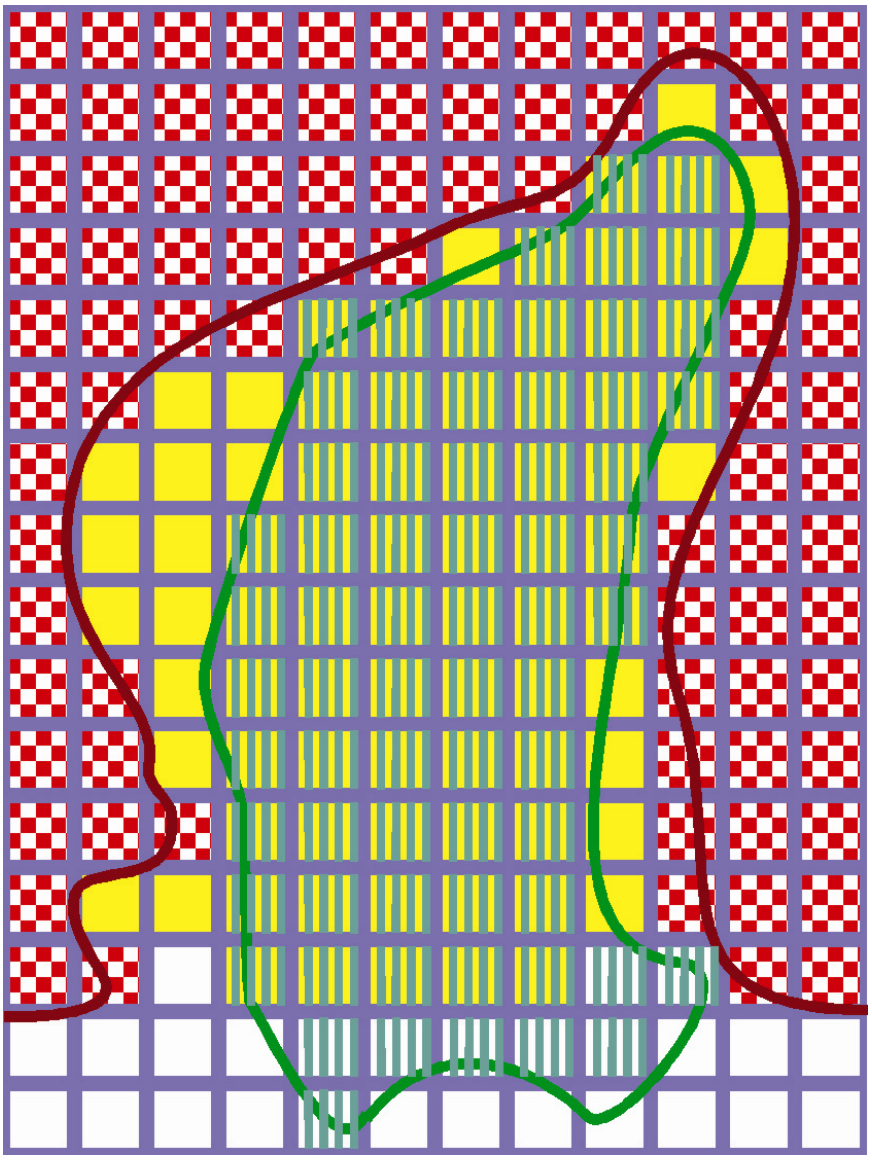
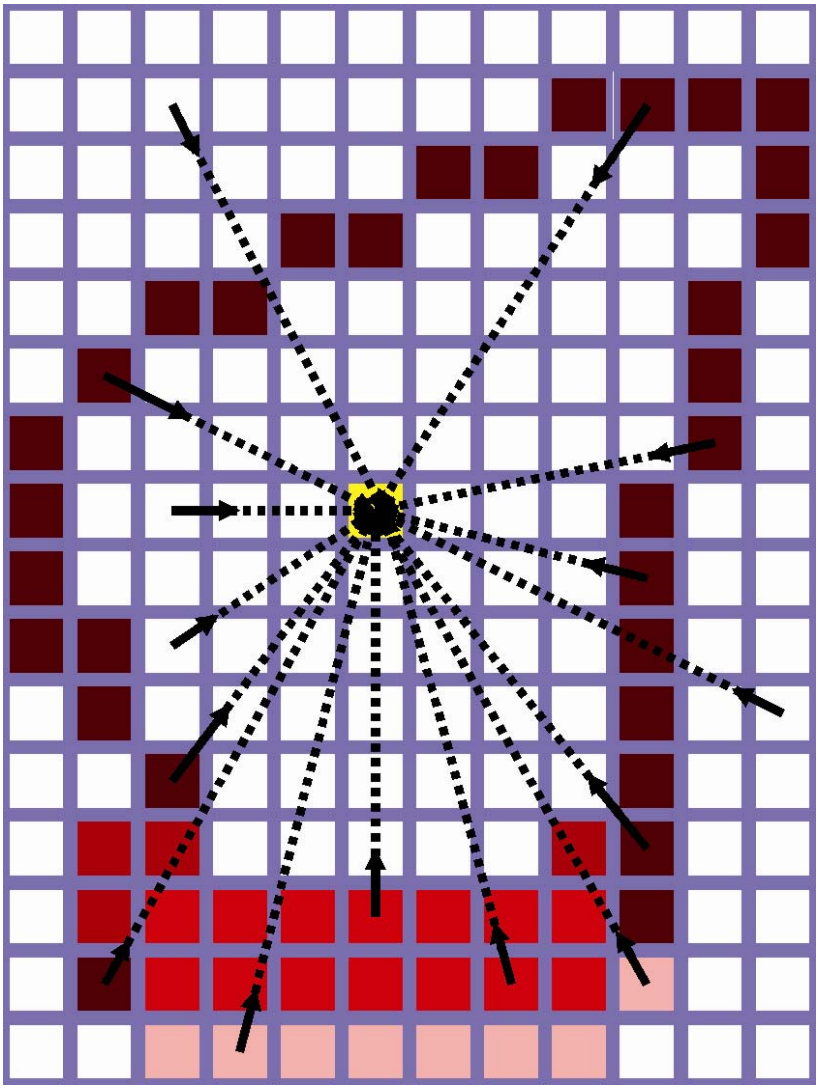


Figure 21: Cavity Entrance Calculation. The cavity entrance is calculated from the derivative of the map illustrated in Figure 18. Vectors are projected from each grid node toward the center-of-gravity of the cavity (dashed lines); the path (as in Figure 16) is determined and the absolute value of the difference between the starting grid point and the first node on that path is calculated as the derivative. Paths completely inside or outside will have close to zero slope (white), paths clearly crossing from outside to inside will have slope values close to one (dark red), while the ambiguous cavity mouth grid points will have intermediate slope values.



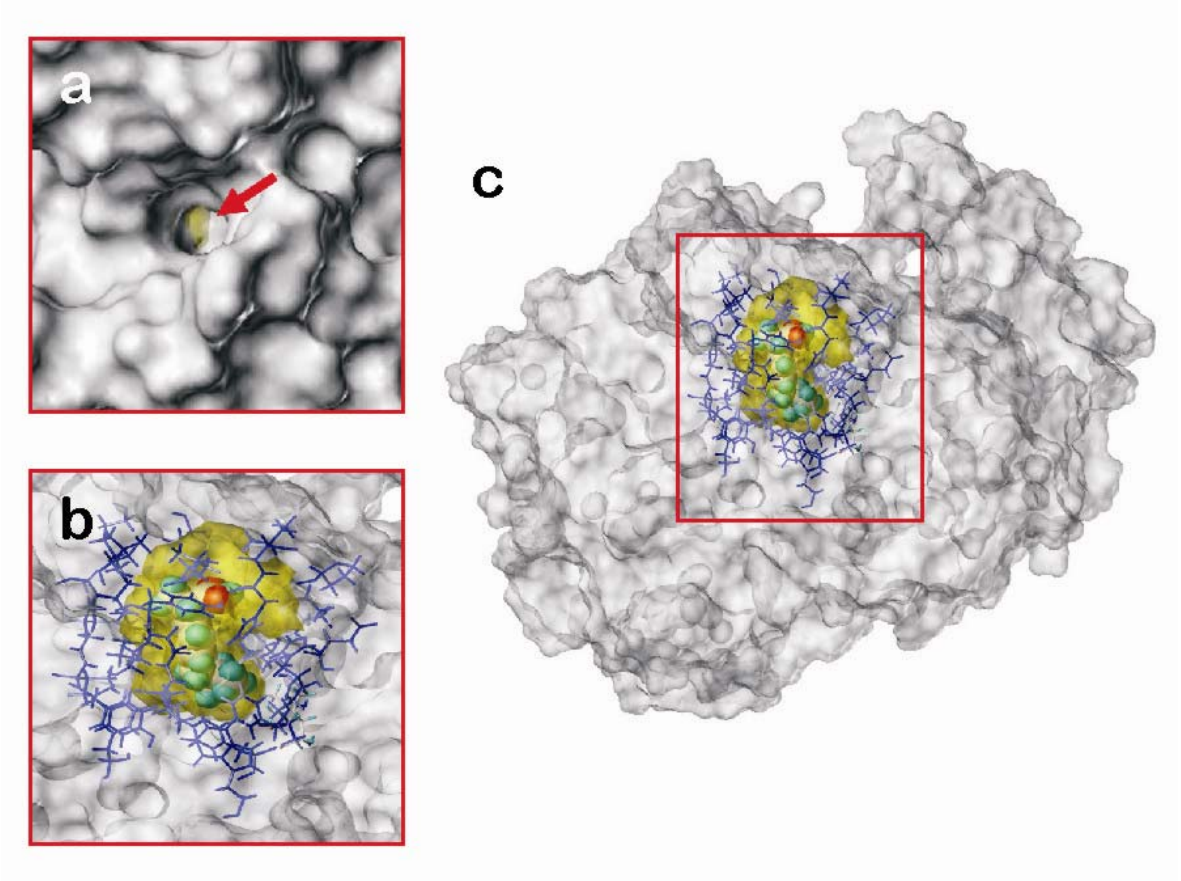
These volume and area metrics for the 64 biomacromolecules, some with multiple pockets or symmetry-related sites, in this study are set out from Table 4 to 10. Lastly, identification of protein residues and/or atoms lining the cavity may also be useful information for drug design and/or site-directed mutagenesis studies. These data are indicated below for a few cases, but are readily available from the analysis module in the algorithm. In the following paragraphs we focus on several examples, and present, somewhat qualitatively, the level of success the VICE algorithm has obtained in describing these cavities for a broad range of variations in the architecture of binding pocket viz. deeply buried binding pockets, cavities at protein-protein dimer, and with DNA/RNA interface. The method is able to detect cavities, shallow grooves, cleft and channels within a protein. The program also addresses the problem of defining the limits of a cavity, especially its boundary with free space, i.e., at the entrance (vide infra).

4.3.3 Well-enclosed Cavities/Deeply Buried Pockets

In the initial examples, we characterized deeply buried binding pockets that are, in other terms, well-enclosed cavities. These cases also may be thought of as essentially closed continuous volumes in the interior of protein molecules. While these binding pockets, which might bind small molecules, are sometimes not obvious from initial inspection, most available cavity detection software can effectively detect them. Although there may be a number of these voids inside a protein, it has been observed that the active site is usually the largest cavity in a protein [168,173] because a large pocket provides increased surface area and hence increased opportunity for small molecule binding. Thus, one of the problems faced by these algorithms is identifying the primary binding pocket amongst (often) numerous small clefts and voids. In addition, the boundary of the active site is often not well demarcated and numerous snake-like tendrils can project from the binding envelope. An important success factor of a cavity detection algorithm is in presenting a single, clean well-bounded cavity.

Prostaglandin H_2 synthase (PDB 1eqg) is an example of this class of cavity. A detailed structural analysis of NSAID binding with prostaglandin H_2 synthase is discussed by Selinsky et *al.* [194]. Figure 22 illustrates this protein and its detected cavity. The inset at the upper left shows the relatively small opening (calculated as 22 $Å^2$ by our algorithm) while the inset at the lower left extracts the cavity, ligand and surrounding residues (Pro86, Ile89, His90, Leu93, Met113, Val116, Arg120, Phe205, Val344, Ile345, Tyr348, Val349, Leu352, Ser353, Tyr355, Leu357, Leu359, Phe381, Leu384, Tyr385, Trp387, His513, Phe518, Glu520, Met522, Ile523, Glu524, Gly526, Ala527, Ser530,

Figure 22. Well-enclosed Cavity: Prostaglandin H_2 synthase. Prostaglandin H_2 synthase (1eqg) examined with the VICE algorithm and displayed with MOLCAD and Sybyl. a) The protein Connolly surface is displayed with opaque rendering. The small opening to the cavity is indicated by the red arrow; b) the ligand, ibuprofen rendered in CPK (space-filled), and the residues lining the cavity are shown. The yellow translucent surface illustrates the extents of the cavity. The protein is rendered with a translucent Connolly surface; c) shown as in b) but displaying the entire protein.



Leu531 and Leu 534). The cavity volume is estimated at 814 $Å^3$ of which only 214 $Å^3$ are occupied by ligand. We have not included any volume contribution from water in calculated volume estimates as the number of water molecules detected by x-ray crystallography varies greatly with crystallographic resolution [195].

Similarly, the anti-malarial compound fosmidomycin binds to IspC (PDB 1onp) [196] and the detected cavity is well-defined (Figure 23), surrounded by residues Ser151, Glu152, Gly185, Ser186, Gly187, Gly188, Trp212, Ser213, Ile218, Ser222, Asn227, Lys228, Glu231, Ser254, Met276 and a Mn ion. Here, the binding site is deeply buried with a volume of 342 $Å^3$, while the volume of fosmidomycin is 136 $Å^3$ of which 127 $Å^3$ occupies the active site. Most cavities in this class have opening surface areas that are about 10% or less of the total cavity surface area and have occupancy factors of around 35-50% (See Table 4).

Figure 23. Well-enclosed Cavity: IspC. IspC (1onp) examined with the VICE algorithm and displayed with MOLCAD and Sybyl. a) The protein Connolly surface is displayed with opaque rendering. The relatively small opening to the cavity can be seen; b) the ligand, the anti-malarial compound fosmidomycin rendered in CPK, and the residues lining the cavity are shown. The yellow translucent surface illustrates the extents of the cavity. The protein is rendered with a translucent Connolly surface and the space-filling magenta sphere is the manganese ion; c) shown as in b) but displaying the entire protein.

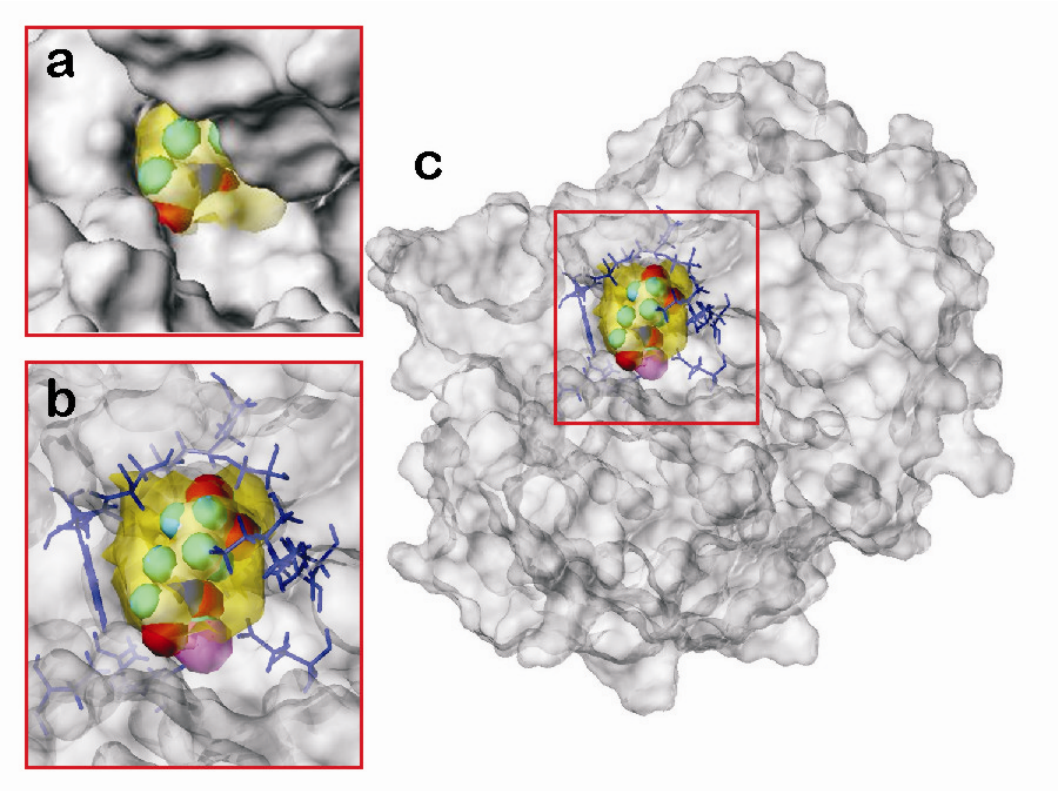


Table 4: Protein cavity data for deeply buried pockets

PDB Code	Cavity Type	Subcavity ID	Cavity vol. (Å^3)	Ligand volume (Å^3)			Occupied fraction	Cavity surface area (Å^2)		Opening fraction
				Total	Inside	Outside		Total	Opening(s)	
3cox	Deeply buried pockets		1451	575	565	10	0.39	1834	42	0.022
1onp			342	136	127	9	0.37	405	49	0.121
1aes			179	61	61	0	0.34	209	0	0.000
1mbd			903	521	418	103	0.46	1400	154	0.110
1piv			564	330	314	16	0.59	823	35	0.043
1gf1			426	321	250	71	0.59	458	2	0.004
1tys			1499	576	549	27	0.37	2079	194	0.093
1asc			380	198	182	16	0.48	477	33	0.069
1eqg			814	215	214	1	0.26	1230	22	0.018
1m6y			596	295	269	26	0.45	695	29	0.042
1ju3			203	113	101	12	0.50	208	10	0.048
1ydd			273	147	121	26	0.54	308	11	0.036
1hor			402	183	176	7	0.44	510	11	0.022
121p			924	344	333	11	0.36	1116	107	0.096
2acs			947	656	547	109	0.58	1147	97	0.085
1njs			578	431	305	126	0.53	791	71	0.090

4.3.4 Groove/Cleft on the Surface of a Protein

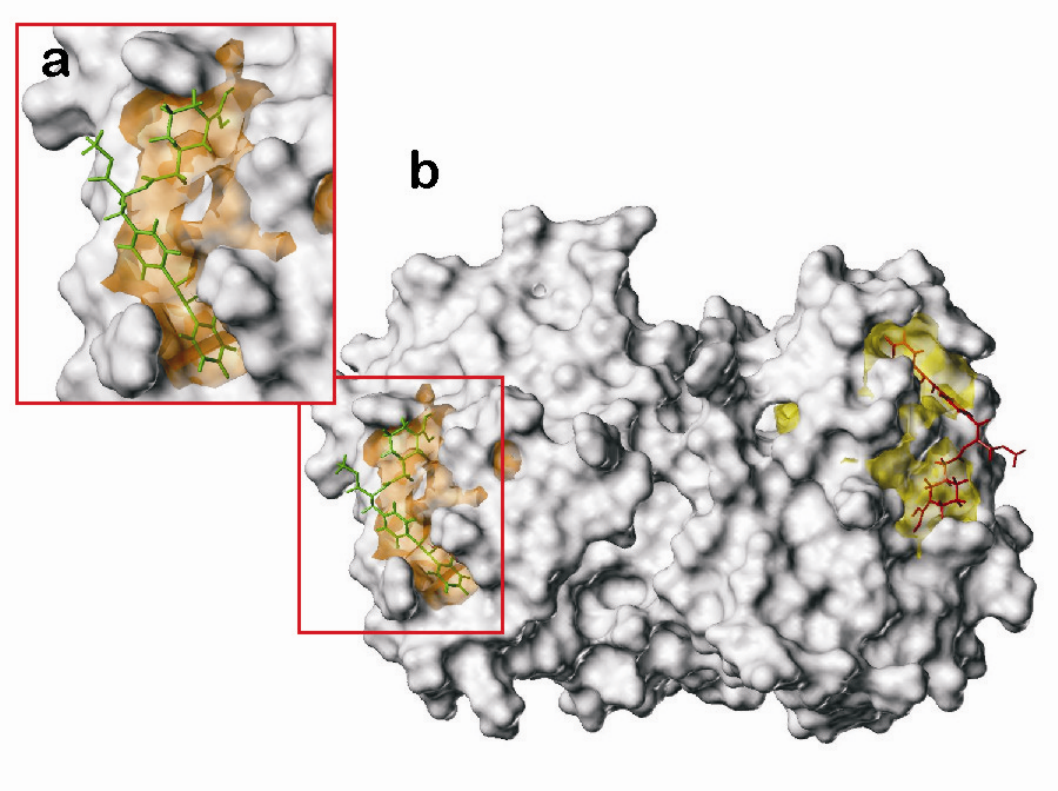
The more shallow cavities and surface grooves are also potential sites for binding of drugs, ligands, proteins and other biomacromolecules. Identification and size characterization of surface pockets and occluded cavities are often the initial steps in protein structure-based drug design. The most important of these binding pockets are generally found to be particularly large and deep clefts. While internal cavities are fairly easy to define as they generally correspond to well-enclosed regions completely bounded by surrounding atoms, in many cases interactions between protein and small molecule tend to involve what can appear to be a nearly planar surface on the surface of the protein. However, on the nano-scale protein surfaces are irregular with many clefts and grooves of varying shapes and sizes, and it is often difficult to define the boundaries of these shallow pockets. In particular the "open" boundary at the mouth is ambiguously defined even in the best of circumstances with this class of protein cavity. Our algorithm, as described in Figure 15-21, defines this boundary in terms of a user-adjustable parameter that represents the ratio of vectors finding the cavity wall over all vectors for each grid point. For this work we used the default value of 0.5 for this parameter. Most shallow cavities can be characterized one key metric: they generally have opening cross-sectional areas (Table 5) of about 30% of the total cavity surface area.

Table 5: Protein cavity data for cavity from surface depression

PDB Code	Cavity Type	Subcavity ID	Cavity vol. (Å^3)	Ligand volume (Å^3)			Occupied fraction	Cavity surface area (Å^2)		Opening fraction
				Total	Inside	Outside		Total	Opening(s)	
1dr1	Cavity from surface depression		2010	686	673	13	0.33	2728	348	0.128
1m48			400	418	141	277	0.35	558	190	0.341
1fmp			396	238	173	65	0.44	510	120	0.236
1k2v			395	426	247	179	0.63	604	203	0.336
1b12			293	244	174	70	0.59	309	62	0.201
2ngr			856	282	264	18	0.31	964	164	0.170
1e3e			2352	474	464	10	0.20	3261	734	0.225
1bxl			1498	1476	518	980	0.35	2065	791	0.383
1adg			1288	457	422	35	0.33	1597	435	0.272

One example of a shallow cavity on the surface of protein is illustrated with cytokine interleukin-2 (1m48) [197] in Figure 24. Here, the binding site is mapped to a shallow groove on the surface of protein. This particular protein is a symmetric homodimer so that there are two essentially identical binding sites. Cytokine interleukin-2 has been implicated as one of the principal mediators in proliferation and differentiation of activated cells in an immune response. It attaches through its surface to the trimeric IL-2R receptor, thereby triggering an immune response. Although the binding pocket is actually present as a surface cleft divided by a ridge, the cavity detection algorithm was able to capture both sides of the pocket. Interestingly, while a large portion of the ligand hangs out of the pocket, the two terminal ends are buried within the pocket.

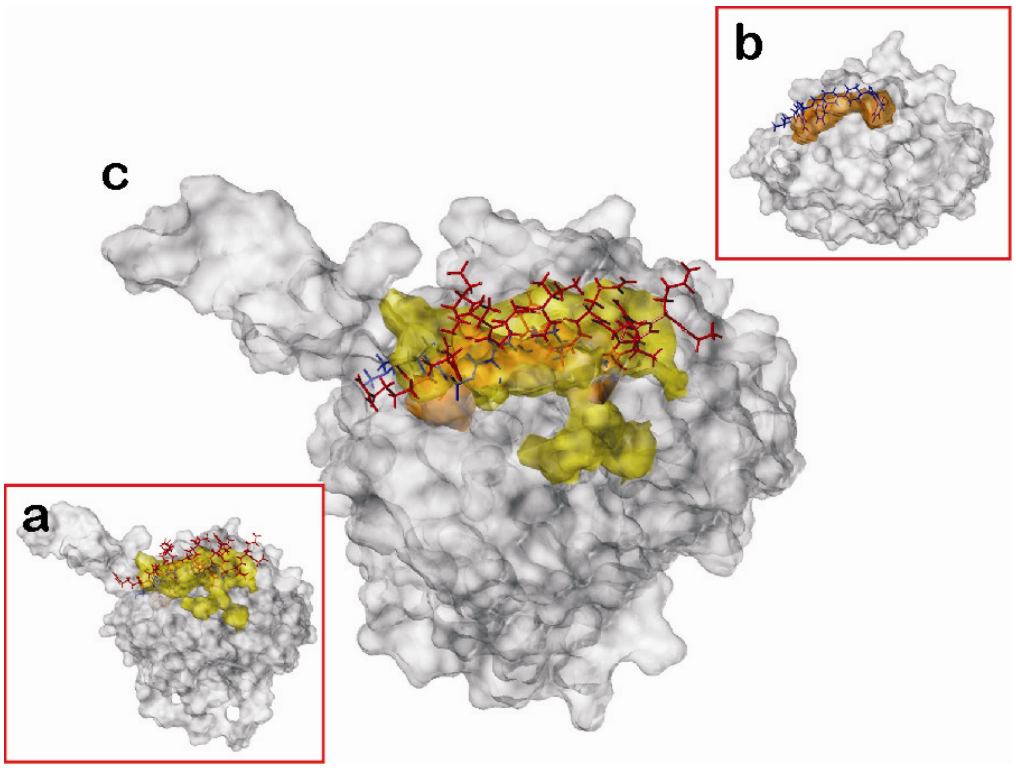
Figure 24. Shallow Cavity on Protein Surface: Cytokine interleukin-2. The cytokine interleukin-2 dimer (1m48) has one essentially identical shallow cavity binding site on each of the two chains. a) The inhibitor Ro26-4550 is bound in the cavity of chain A: the cavity extents are displayed as the orange contour volume. Both ends are well-bound but much of the middle of the ligand is external to the cavity; b) both sites are displayed in this view of the entire protein.



In another example, as illustrated in Figure 25, a cavity was identified on the surface of the BCL-X_L protein (1bxl, 2yxj) [198,199], a pro-survival protein whose function is regulated by the binding of anti- or pro-apoptotic factors. Several anti-apoptotic proteins can bind to the BH3 domain of BCL-X_L in tumor cells where it is overexpressed. These interactions increase the survival rate of the cancer cell and may contribute to drug resistance. In contrast, pro-apoptotic proteins such as BAK can induce apoptosis by their binding to the BH3 domain; thus, the BH3 domain on BCL-X_L could be exploited as an attractive drug target in cancer chemotherapy. The BH3 domain has a largely hydrophobic surface with an estimated volume of 1300 $Å^3$. The lower left inset of Figure 25 shows BAK bound to the BH3 domain of BCL-X_L (1bxl). The associated cavity is indicated in yellow. However, a smaller sub-pocket (indicated in orange) can also be identified on the BH3 domain that binds small molecule modulators such as ABT-737 (2xyj) as shown in the upper right inset of Figure 25. The overlap of these two sites is shown in the central portion of Figure 25, and suggests that the bound ABT-737 ligand would block the binding of BAK.

Exploitation of such cavities and sub-pockets at the interface between proteins could have important implications in drug discovery as more is learned about the role of protein-protein interactions in biological processes.

Figure 25. Shallow Cavity on Protein Surface: BCL-X_L. Two structures of the BCL-X_L protein with BAK protein and inhibitor ABT-737 bound within its binding cavity. a) BCL-X_L protein (1bxl) with sixteen residue BAK protein (red capped stick representation) bound within the surface cavity (yellow translucent envelope); b) BCL-X_L protein (2yxj) with ABT-737 inhibitor (blue capped sticks) bound in a relatively smaller sub-pocket (orange translucent surface); c) overlap superposition of 1bxl and 2yxj structures showing the correspondence of the two pockets. Cavity extents illustrated with yellow and orange translucent envelopes.



4.3.5 Cavity Formed at a Protein-Protein Interface

Next, we consider examples of cavities at protein–protein interfaces. These interactions have an important role in many biological processes and cavities at the interface of protein-protein dimers offer particularly attractive, but as yet largely unrealized, opportunities for therapeutic intervention. However, uncertainties owing to the structural changes due to domain movement upon binding and the often insufficient knowledge of well-defined binding pockets, coupled with the irregular shape and size of typical protein–protein interfaces, have made it difficult to design inhibitory ligands that can modulate protein-protein interactions. Although a large surface area is usually buried on each side of the actual interface, there is often only a relatively small cavity or groove where a small molecule can fit and thus inhibit the protein-protein interaction. Our cavity detection algorithm is able to identify and delineate these binding sub-pockets on large interfaces as described in the section above. Detailed information related to volume, surface area and chain, residues that line the pocket is also illustrated (see Table 6).

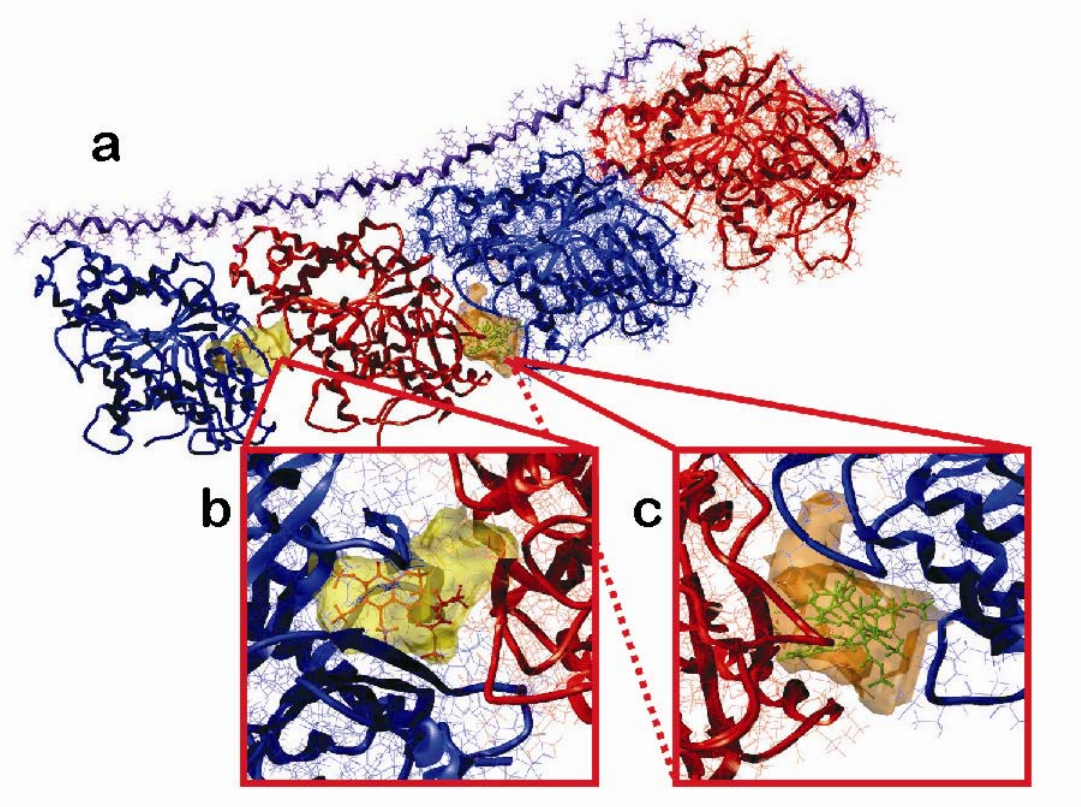
Table 6: Protein cavity data for cavity formed at protein-protein interface

PDB Code	Cavity Type	Subcavity ID	Cavity vol. (Å^3)	Ligand volume (Å^3)			Occupied fraction	Cavity surface area (Å^2)		Opening fraction
				Total	Inside	Outside		Total	Opening(s)	
1cqp	Cavity formed at protein-protein interface	Chain A	982	396	380	16	0.38	1200	191	0.159
		Chain B	941	389	359	30	0.36	1175	198	0.168
1dn2		Chain E	620	1253	231	1022	0.14	1086	245	0.225
		Chain F	578	1264	198	1066	0.12	863	280	0.324
1i8l		Chain C	923	10475	79	10396	0.007	1378	190	0.137
		Chain D	807	10302	63	10239	0.005	1395	221	0.158
1jhl			527	11955	95	11860	0.18	507	196	0.387
1mlc			423	11927	167	11760	0.39	483	138	0.286
1rv1			692	516	328	188	0.47	881	218	0.247
1t4f			834	1011	402	609	0.48	208	676	0.308
1vfb			413	20770	61	20710	0.15	389	130	0.334
3ink			215	12306	18	12228	0.08	217	67	0.309
1z2b		Colchicine	842	374	343	31	0.39	1015	28	0.027
		Vincristine	1457	705	615	90	0.39	1956	381	0.194

However, in some cases, cavities at protein-protein interfaces can be observed, either at the joint between two subunits of the same protein or for a protein-protein complex. In one example, for αβ-tubulin (1z2b) (Figure 26) [200], our cavity detection algorithm defined the binding envelope at the wide interface between protein-protein units. Tubulin is the basic building block of microtubules, critical for mitosis and cell division, and an important target for anti-cancer drugs.

Tubulin exists as a heterodimer and joins end-to-end to form a protofilament with alternating α and β subunits. The staggered assembly of 13 protofilaments forms hollow, cylindrical microtubule filaments. Three distinct binding sites have been identified on tubulin heterodimers for the taxol, colchicinoids and vinca classes of drugs. Although Taxol binds wholly on the β subunit, the colchicine binding site lies at the intradimeric interface of α and β subunits of tubulin and the vinblastine binding site is located at the interdimeric interface of αβ-subunits. The colchicine and vinblastine binding sites have been difficult to map as these binding pockets are poorly demarcated between the subunit interfaces and the crystallographic resolution is rather poor at 3.58 Å. However, our algorithm was able to clearly identify and delineate binding envelopes at these subunit interfaces: the colchicine binding site (Figure 26, left inset) has a volume of 842 $Å^3$ with an opening directly at the α-β interface with a estimated opening area of 28 $Å^2$; and the vinblastine site cavity has an estimated volume of 1457 $Å^3$ and an opening of 381 $Å^2$.

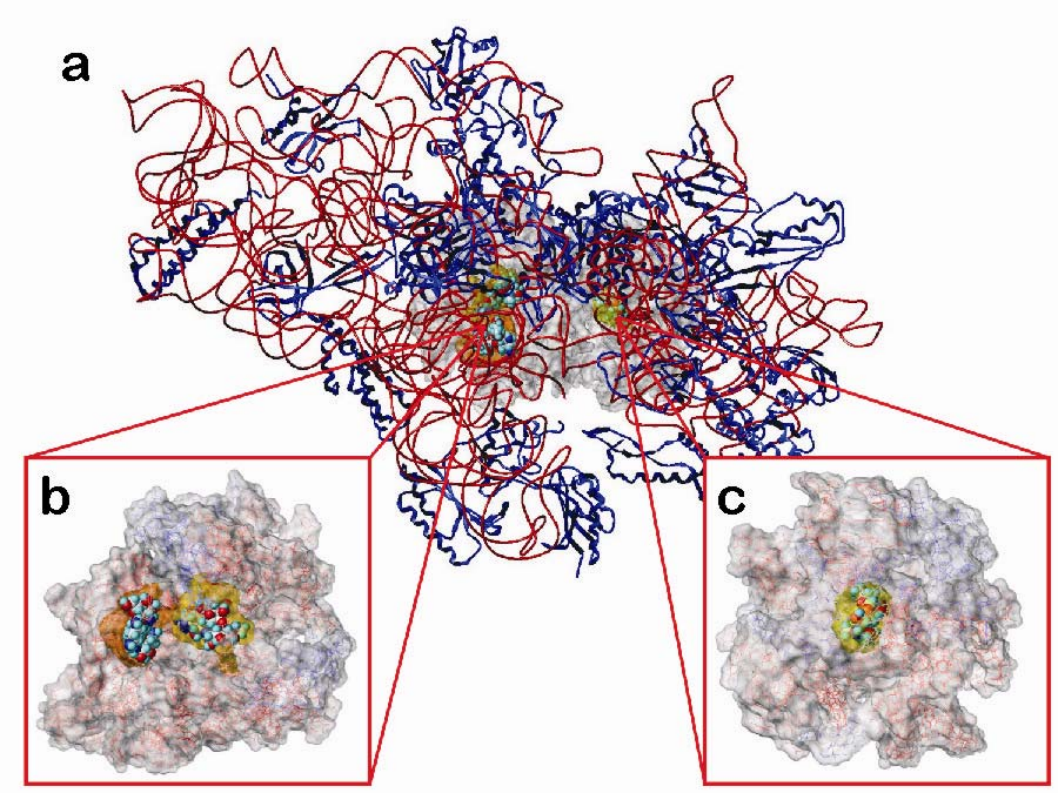
Figure 26. Cavity at Protein-Protein Interface. a) The tubulin protein (1z2b) with colchicine and vinblastine binding sites at interfaces between the α and β subunits. The tubulin polymer is rendered in ribbon and tube with the α subunits shown in red and β subunits shown in blue; b) inset shows the colchicine binding pocket (yellow contour) at the intra-dimeric interface of the αβ-subunit; c) inset shows the vinblastine binding site (orange contour) at the inter-dimeric interface between αβ-subunits.



4.3.6 Cavity Formed at a Protein-Polynucleotide Interface

Protein-DNA/RNA interactions primarily are related to regulation of gene expression and are thus associated with important functions. Cavities or pockets formed by proteins at protein-nucleic acid interfaces are designed to mediate interactions and allow sequence-specific recognition of a gene. Each nucleic acid binding motif on a protein consists of a specific binding pocket that recognizes and stabilizes the DNA/RNA. To bind in this fashion a protein must make contact with the nucleic acid in such a way that the nucleotide sequence can be recognized. Ligands that can interfere with this recognition, either by occupying the putative nucleic acid binding site and blocking DNA/RNA binding, or by exploiting cavities formed in the protein-polynucleotide complex, may be therapeutically significant. As an example of the latter strategy, Figure 27 shows binding pockets detected on the 30S ribosomal subunit (1fjg) [201]. Three well-defined major cavities are detected indicating the binding sites for the antibiotics spectinomycin, paromomycin and streptomycin. The binding pocket for spectinomycin, which inhibits elongation factor G catalyzed translocation of the peptidyl-tRNA from the A-site to the P-site, has a volume of 633 $Å^3$ with spectinomycin completely enclosed within the cavity. The majority of interactions are with RNA bases C1063, G1064, C1066, G1068, C1069, A1191, C1192, G1193, U1194, G1386, G1387, with protein residues Ala121 & Gly 122 lining the cavity envelope. Paromomycin, an aminogycoside, binds in the major groove at the decoding center on H44 and induces errors in translation by increasing the affinity and stability of tRNA for the A-site.

Figure 27. Cavity at Protein/Polynucleotide Interface. a) The 30S ribosomal subunit (1fjg) is rendered as ribbon and tube, except within 20 Å of binding region where a MOLCAD surface display is shown to highlight the binding pockets for the antibiotics spectinomycin, paromomycin and streptomycin; b) the binding site for paromomycin (orange envelope) and streptomycin (yellow envelope) are illustrated. The antibiotic drugs are rendered in spacefill; c) the binding pocket for spectinomycin (yellow envelope) is illustrated.



The volume of this cavity is 1605 $Å^3$ and it is lined by bases C1404, G1405, U1406, C1407, A1408, C1409, G1410, G1488, G1489, C1490, G1491, A1492, A1493, G1494, U1495, C1496, G1497 and protein residue Lys47. Adjacent to this binding pocket is a third cavity which binds streptomycin, a drug that inhibits protein synthesis by interfering with the initial selection and proofreading of tRNA. The volume of the predicted binding pocket is 988 $Å^3$ with numerous nucleotides from 16S RNA and residues from the S12 protein lining the binding envelope. While a limited numbers of base pairs are involved in recognition and stabilization, designing an inhibitor that binds at an interface must involve sufficient nucleic acid and protein contact so that the ligand fits snugly.

Detailed analysis of cavities formed at the protein-polynucleotide was carried out and information retrived is tabulated in table 7.

Table 7: Protein cavity data for cavity formed at a protein-polynucleotide interface.

PDB Code	Cavity Type	Subcavity ID	Cavity vol. (Å^3)	Ligand volume (Å^3)			Occupied fraction	Cavity surface area (Å^2)		Opening fraction
				Total	Inside	Outside		Total	Opening(s)	
1aud	Cavity formed at a protein-polynucleotide interface		881	6104	270	5834	0.31	1056	285	0.270
1esg			2848	3131	100	3031	0.04	4337	2804	0.647
1euq			924	343	324	19	0.34	1097	154	0.140
1x9n			12267	7761	2491	5270	0.20	18884	6207	0.329
2c62			1791	3002	486	2516	0.27	2985	1017	0.341
2cdm			4197	4797	1388	3409	0.33	8018	2467	0.308
2euv			1662	4958	487	4471	0.29	1742	414	0.238
2hhs			4160	5171	705	4466	0.17	6377	3025	0.474
2hw8			1435	7379	466	6913	0.32	2159	1391	0.644
2i0q			2672	2214	914	1300	0.34	3862	1263	0.343

4.3.7 Flexible Cavities with Loop or Domain Movements

All proteins have an intrinsic flexibility that is required for a wide range of biochemical processes in catalysis, regulation, and protein assembly. However, in some cases experimental evidence has indicated that the shape and size of the ligand binding envelope may change due to domain movements; e.g., molecular recognition and ligand binding is induced by large loop movements where flexibility in the protein main chain influences the ligand binding[202]. Ligand binding may involve a wide range of structural changes in the receptor protein, from hinge movement of entire domains to small side-chain rearrangements in the binding pocket residues. Many protein functions in fact involve conformational transitions that involve opening and closing of relatively rigid parts of that protein about flexible joints. The analysis of side chain flexibility gives insight valuable for improving docking algorithms and for ligand design when domain movement and/or loop flexibility opens and closes the binding pocket. Instead of well-defined binding pockets, most proteins that have 'induced' domain movement lack deep clefts or clearly shaped binding pockets. Thus, this is an interesting case study for cavity detection – where the change in the size and shape of binding pocket due to domain movement is calculated by comparison between pairs of *holo* and *apo* proteins (See Table 8).

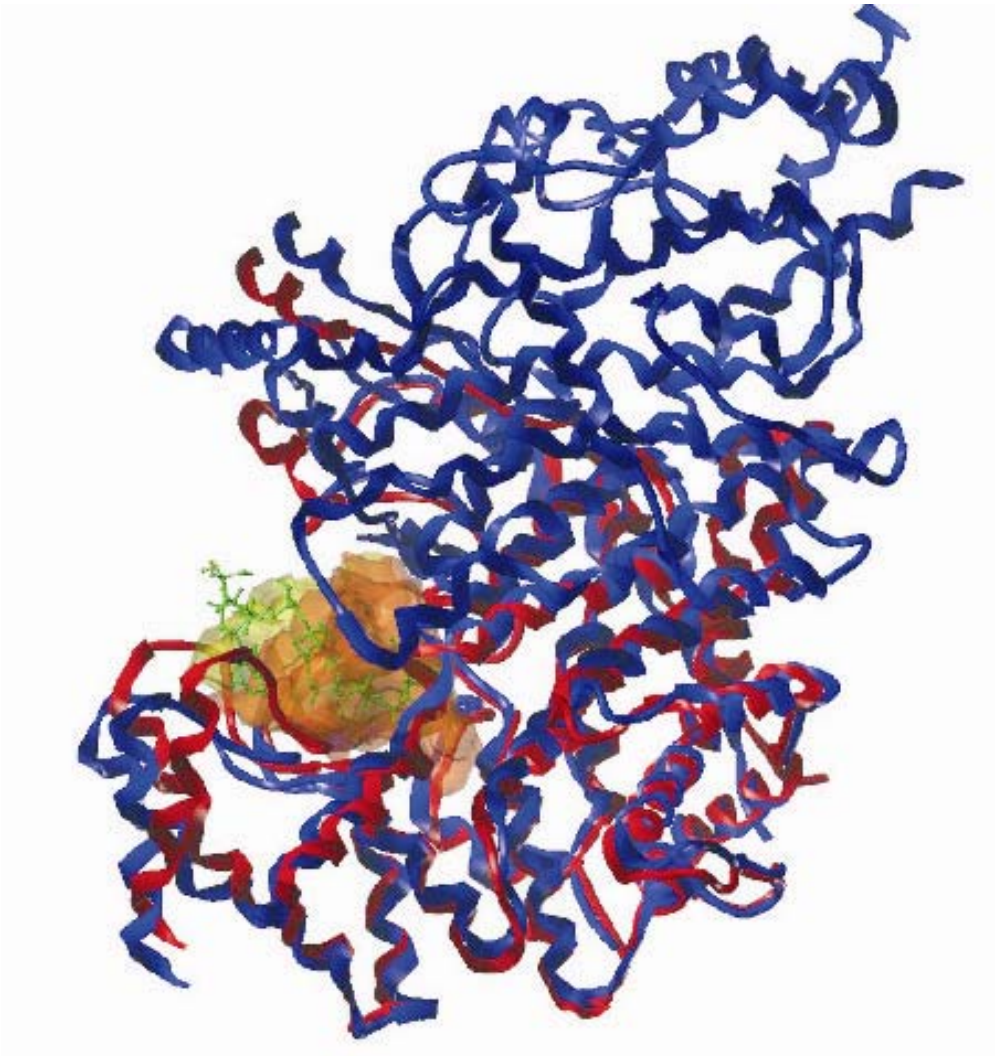
Table 8: Protein cavity data for cavites with loop or domain movements.

PDB Code	Cavity Type	Subcavity ID	Cavity vol. (Å^3)	Ligand volume (Å^3)			Occupied fraction	Cavity surface area (Å^2)		Opening fraction
				Total	Inside	Outside		Total	Opening(s)	
1ama	Flexible cavities with loop or domain movements	*Holo*	330	290	222	68	0.67	425	42	0.099
9aat		*Apo*	300	290	211	79	0.70	402	64	0.159
1ank		*Holo*	950	572	438	134	0.46	1488	21	0.014
4ake		*Apo*	358	572	94	478	0.26	472	171	0.362
1atp		*Holo*	707	332	316	16	0.45	914	9	0.010
1ctp		*Apo*	531	332	188	144	0.35	669	129	0.193
2ohx		*Holo*	722	464	390	74	0.54	910	125	0.137
8adh		*Apo*	590	464	278	186	0.47	700	182	0.260
5cts		*Holo*	967	704	507	197	0.52	1181	167	0.141
5csc		*Apo*	439	704	167	537	0.38	662	328	0.495

Figure 28 shows the example of citrate synthase, 5cts [203] and 5csc [204], which are the *apo* (unliganded) and *holo* (ligand-bound) forms with cavity volumes of 439 $Å^3$ and 967 $Å^3$, respectively. The bound ligand, oxaloacetate, which has a volume of 704 $Å^3$, appears to induce this large domain movement in the enzyme and causes binding pocket residues to undergo side-chain conformational changes as well as changes in overall shape.

Residues His238, Asn242, Leu273, His274, Val314, Val315, Gly317, Tyr318, Gly319, His320, Ala321, Arg329, Gln364, Ala367, Ala368, Asn373, Asp375, Phe397 surround the binding pocket in the closed structure, while only residues His238, His274, His320, Arg329, Asp375, and Phe397 line the unliganded pocket.

Figure 28. Flexible Cavity with Loop or Domain Movement. The citrate synthase protein, 5cts (red) and 5csc (blue), the *apo* (unliganded) and *holo* (ligand-bound) forms, respectively, is illustrated. A relatively smaller binding pocket is detected in 5cts (orange envelope); however, the native ligand oxaloacetate (green capped sticks) induces a domain movement that significantly alters the shape and size of the binding pocket (yellow envelope) in 5csc.



4.3.8 Multi-Domain Proteins with Channels or Tunnels

Understanding the structure and function of channels and pores within biomolecules is important, e.g., to a large number of critical disease states and in compensating for drug resistance due to efflux. Channels and pores and other passages across cell membranes facilitate the movement of small molecules and ions. These transmembrane proteins, such as ion channels, transporters and G-protein coupled receptors, are exceptionally significant drug targets. Apart from this, channels and tunnels also facilitate the access and exit for substrates/products in some catalytic processes. Channels/pores are often dynamic in nature and can be relatively flexible in size and shape and access through them is often regulated by small molecules binding to an active site. Thus, while many of the available algorithms and associated programs developed to detect and characterize binding pockets are successful with well-enclosed pockets and surface grooves, for the most part these procedures fail to detect long, twisted tunnels connecting the interior of a binding pocket to the exterior environment. In fact, it is surprisingly difficult mathematically to differentiate between true channels and tunnels and random voids if the tunnel has a narrow diameter or constriction point(s).

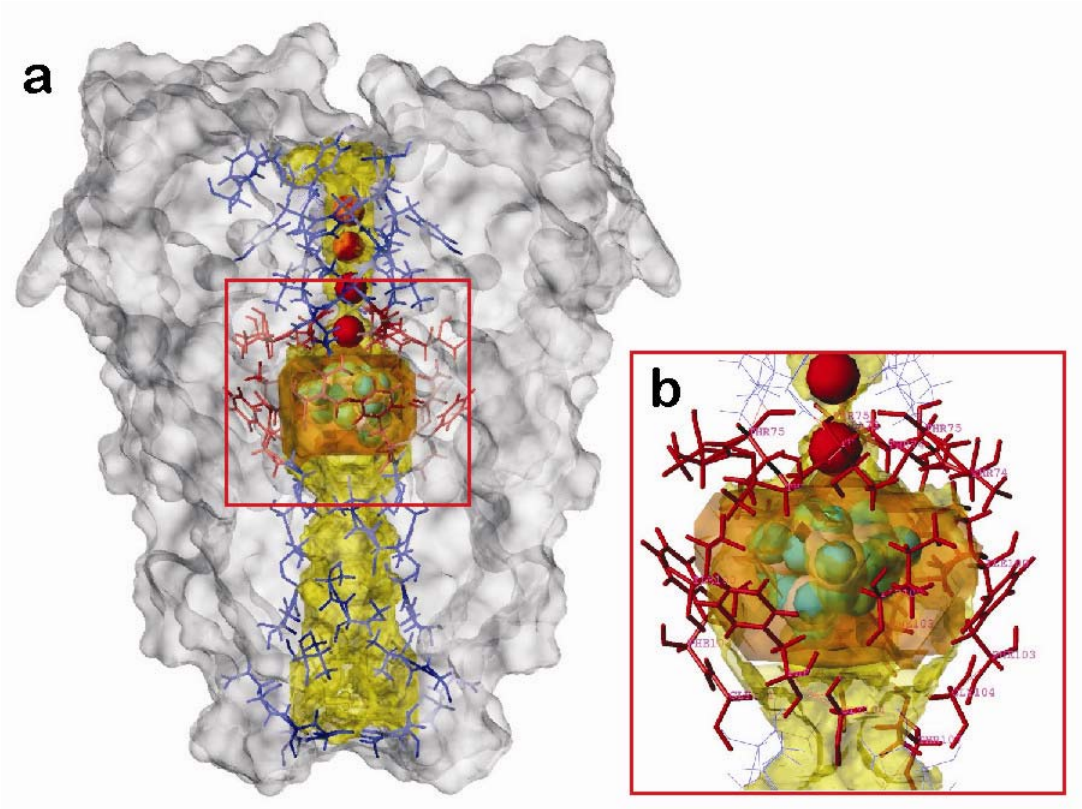
With the recent availability of crystal structures for large membrane-bound proteins, detection and mapping of the interior of these channels can give insight into the binding process for design and development of more selective drugs. Our cavity detection algorithm provides sufficient flexibility and interactivity to map binding sites as well as the channels and tunnels through a protein (See Table 9).

Table 9: Protein cavity data for proteins with channels or tunnels.

PDB Code	Cavity Type	Subcavity ID	Cavity vol. (Å^3)	Ligand volume (Å^3)			Occupied fraction	Cavity surface area (Å^2)		Opening fraction
				Total	Inside	Outside		Total	Opening(s)	
1av2	Proteins with channels or tunnels		305	-	-	-	-	509	92	0.180
1j95[a]		channel	1575	319	260	58	0.15	995	3	0.003
		TBA ligand	658	168	165	3	0.24	850	2	0.002
1okc			6192	608	592	16	0.09	7592	608	0.080
2al5			3266	518	501	17	0.15	4526	271	0.063
2byq			37512	-	-	-	-	51168	2816	0.055

In the example illustrated in Figure 29, the KcsA potassium channel (1j95) [205], we applied our cavity detection algorithm to first identify the binding pocket inside the channel, which is occupied by tetrabutylammonium in this structure. This was easily detected with the program's default parameters, e.g., a grid resolution of 1.0 Å. Next, by decreasing the grid spacing to 0.3 Å, the program successfully identified and delineated a long, narrow porous channel traversing the entire length of the protein's transmembrane axis. It should be noted that this latter calculation was resource intensive due to the very large number of surveyed grid points, but this level of computation was necessary in order to adequately sample the protein structure. The total volume of channel was calculated to be 1342 $Å^3$ with the binding cavity of 615 $Å^3$, while the tetrabutylammonium occupies 168 $Å^3$ inside the binding cavity of the channel and is well-enclosed by hydrophobic residues.

Figure 29. Channels and Tunnels. a) The KscA K^+ ion channel (1j95) plotted with translucent MOLCAD surface. The binding pocket at the center of the channel is illustrated with an orange contour map; its tetrabutylammonium inhibitor is rendered in CPK (space-fill). The channel, traversing the entire length of the protein, is highlighted with the yellow contour map. Detection of the channel required calculations with a very large number of grid map points and high resolution. The potassium ions are rendered as the red spheres; b) expanded view of the inhibitor binding cavity.



4.3.9 Multiple Cavities and Allosteric Binding Pockets

The detection of auxiliary binding sites is becoming increasingly crucial as many proteins have more than one biochemical role and are likely to employ separate binding sites in performing these distinct biochemical tasks. Allosteric binding pockets may offer additional recognition sites that modulate the catalytic function of a protein. These auxiliary binding pockets may be located far away from the catalytic site, as in case of glycogen phosphorylase, or may overlap with the active site. Traditionally, allosteric sites were considered to be distal binding sites for molecules that modulate the function of a protein by a feedback mechanism.

While the mechanisms of allosteric modulation of proteins have been extensively studied, discovery efforts to efficiently identify and characterize these binding sites continue as exploiting them may lead to development of entirely new classes of drugs. However, it can be a non-trivial matter to find and characterize allosteric binding sites when these sites are present as auxiliary pockets overlapping with the main active site. Figure 30 illustrates an example of an allosteric site on glycogen phosphorylase b (1c50) [206]. The crystal structure shows an allosteric binding site for the co-crystallized molecule CP320626. Our program identified this binding site with a volume of 431 $Å^3$ close to the AMP binding site with a volume of 728 $Å^3$. The main PLP catalytic site, with a volume of 849 $Å^3$, is about 30 Å distant from the allosteric site.

Table 10 shows detailed analysis of other proteins with auxillary and allosteric pockets.

Figure 30. Auxiliary and allosteric sites. The glycogen phosphorylase b (1c50) with multiple binding pockets. a) A close-up view of the allosteric/auxiliary site. The AMP cofactor (red sticks, green cavity contour) and allosteric site (yellow contour) with inhibitor CP320626 (blue sticks) are in separate subsites of the overall surface groove (orange contour); b) the main catalytic site (cyan contour) is bound with PLP and is quite deeply buried in the protein.

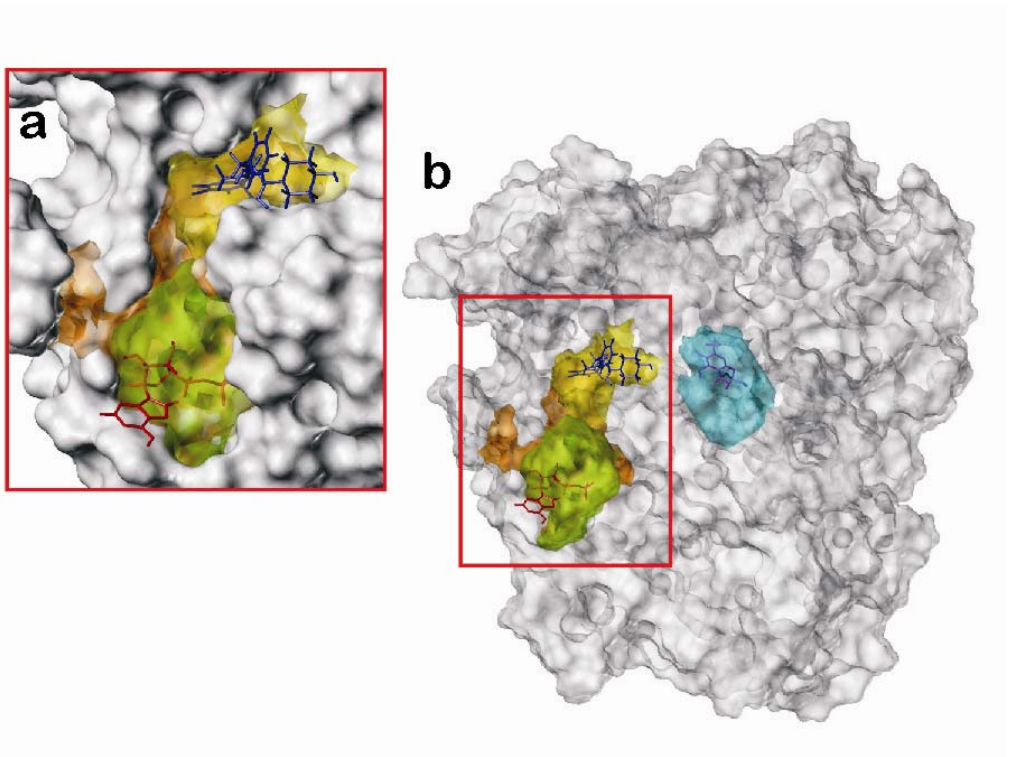


Table 10: Proteincavity data for proteins with multiple cavities and allosteric binding pockets.

PDB Code	Cavity Type	Subcavity ID	Cavity vol. (Å^3)	Ligand volume (Å^3)			Occupied fraction	Cavity surface area (Å^2)		Opening fraction
				Total	Inside	Outside		Total	Opening(s)	
1c50	Multiple cavities and allosteric binding pockets	PLP	849	184	169	15	0.19	1114	9	0.008
		AMP	728	237	142	95	0.17	945	256	0.270
		CHI	431	371	112	259	0.16	681	164	0.240
1kv2		allosteric	1293	484	458	26	0.34	1676	261	0.155
		main	639	327	178	149	0.22	787	35	0.044
1q0b		allosteric	515	256	221	35	0.40	667	85	0.127
		main	558	282	219	63	0.35	915	281	0.307
1v4s		allosteric	722	264	261	3.0	0.36	960	62	0.064
		main	282	138	133	5.0	0.46	330	4	0.012

4.4 Summary and Outlook

The identification, delineation and visualization of protein active sites is a critical facet of drug design. These site topographies play crucial roles in molecular recognition. Proteins may have many pockets and cavities of various sizes, some of which whose function, e.g., protein-protein interaction, is unknown; it is possible that some may be usefully exploited with selective molecules that bind and modulate that protein's function. Thus it is important to be able to characterize these binding pockets, quantitatively and qualitatively. This algorithm and program provides a simple and interactive tool for locating and delineating all the cavities, pockets, grooves, clefts and tunnels of a protein whose structure is known. In most cases the default parameters produce good results very rapidly because the majority of the calculations are performed with integer arithmetic. Thus, we believe that this tool could be a useful starting point for virtual screening by automatically identifying potential binding sites in a first pass. Of course, more reliable and biologically meaningful results will be obtained if the user can focus on particular regions or features by selecting one or more of the pockets and investigating them in detail by adjusting a relatively small number of calculational parameters.

A second major advantage of this program is that it calculates a fairly extensive set of metrics describing a binding pocket and its occupants. Cavity volumes, cavity surface areas, and lists of atoms, residues and/or chains lining the binding pocket can be retrieved. The estimated cross-section surface area of cavity openings is particularly

interesting as it may suggest methods to describe the maximum size of ligands to enter a site.

With this rapid and robust cavity algorithm in place, we are exploring virtual screening and docking algorithms that use property-encoded cavities, e.g., the HINT complement map, as first stage targets. Such cavity maps would be inherently suited for grid-based pose generation and scoring.

CHAPTER 5

Complexity in Modeling and Understanding Protonation States: Computational Titration of HIV-1 Protease Inhibitor Complexes

5.0 Abstract

The computational titration algorithm based on the "natural" HINT (Hydropathic INTeractions) force field is described. The HINT software model is an empirical, non-Newtonian force field derived from experimentally measured partition coefficients for solvent transfer between 1-octanol and water ($LogP_{o/w}$). The titration algorithm allows the identification, modeling and optimization of the multiple protonation states of residues and ligand functional groups at the protein-ligand active site. The importance of taking into account pH and ionization states of residues, which strongly affect the process of ligand binding, for correctly predicting the binding free energy is discussed. The application of the computational titration protocol to a set of six HIV-1-cyclic inhibitor complexes is presented and the advance of HINT as a virtual screening tool is outlined.

5.1 Introduction

A typical problem in modeling biomolecular systems or molecular recognition systems is accurately modeling the energetics of binding. Thermodynamic analyses of protein-ligand interactions not only give vital insight into the free energy changes of the system, but also elaborate whether an interaction is enthalpy or entropy driven[207-209]. Calculation of binding free energy involves evaluation of both enthalpic and entropic contributions and forms an integral part of structure-based drug design protocols[210,211]. To this end, computer simulations have strived to predict binding free energy and concomitantly interpret experimental data. The fundamental idea behind development of robust computational models was to incorporate as much of the physiochemical parameters defining protein-ligand interactions as possible[212]. These models can contribute significantly to the understanding of the structural and energetic basis of biomolecular interactions, with the goal of *de novo* predicting the binding free energy of protein-ligand complexes.

Generally, enthalpic contributions are estimated by theoretical methods, knowledge-based potential functions, or parameters derived from experimental data[213]. The force-fields used for calculations of free energy and intermolecular interactions assume that steric and electrostatic forces are sufficient to account for the observed biological interactions[213]. However, these terms alone are never sufficient for accurate prediction of biomolecular interactions as they do not always include solvation/desolvation effects. The entropic contributions to binding are much less well-defined and often poorly quantitated or even ignored in most cases. Most approaches sum up these interactions

separately as distinct enthalpic and entropic contributions, whereas, in reality, the ligand-protein recognition is a concerted event and thermodynamic quantities cannot be just simply added up[214]. The phenomenon of solvation/desolvation within the protein active site, resulting in hydrophobic interactions and other phenomena, has been particularly difficult to model computationally. However, developments in free energy calculation over the past several years have made significant progress towards this goal[215,216]. Among the methods currently available to calculate free energy, those based on complex and time-consuming molecular dynamics simulations, with explicit consideration of water molecules, have been shown to correlate fairly well with free energy. Free energy perturbation (FEP) calculations, based on statistical mechanics, can predict the free energy of a system by analyzing ensemble averages (calculated by molecular dynamics (MD) or Monte Carlo (MC) simulations) and treating solvent molecules and ions explicitly[217]. Unfortunately, besides being computationally expensive, calculation of free energy using MD or MC is plagued by errors and problems from a variety of sources[218]. Thus, computational time and uncertainty is compounded for complex biomolecular systems with solvents and counterions[218,219].

$\Delta G_{binding}$ values computed via sophisticated simulations often do not correlate well with experimental binding measurements. This could be due to several reasons. First, it is becoming common practice in the drug discovery community to use assays generating IC_{50}s rather than equilibrium constants, and computational scientists often interchangeably use IC_{50} instead of K_i or K_d to calculate free energies. Furthermore, the experimental binding data itself present uncertainties, i.e., there are often differences of

one order of magnitude (corresponding to about 1 kcal mol^{-1}) between inhibition data collected in different laboratories, and thus they are not always of a quality to enable accurate quantitative correlations with computational data. On the other hand, structural data available from x-ray crystallography and NMR has undeniably fostered an understanding of the biological complexity of molecular recognition. However, often only partial agreement between experimental and theoretical binding energy data based on crystal structures has been observed.

It is difficult to correlate solution binding data and crystallographic structural data because of a number of inherent experimental limitations. Biomolecular interactions are sensitive to subtle changes in experimental conditions such as pH, buffer, ionic strength, and temperature under which data are collected[220]. While the pH used for making the binding measurements is generally perceived as the pH at which binding takes place, that pH may not actually be the optimum pH for binding. This discrepancy is magnified when crystals for x-ray analysis are grown under a still different pH condition. Unless at very high resolution, an x-ray structure by itself reveals little direct information regarding the protonation states of the active site residues in a protein. As protein-ligand specificity and stability are known to be sensitive to structural details, presumably because protonation states of ionizable residues and the details of the hydrogen-bonding network are very important for optimum interaction, it is imperative to identify, characterize and understand the protonation states of residues. Even a change in the ionization state of a single residue or ligand functional group may have a profound effect on the results for structure-based energy calculations.

In general, molecular association depends on the ionic strength and pH of the solution[220]. Protein crystals may contain between 30% and 70% solvent[221], which includes the buffer solution as well. These solvent molecules and ions are distributed among the protein molecules of the crystal lattice. However, due to their inherent electrostatic properties, they may distribute themselves in different solvent channels depending on the nature of the residues lining these solvent pockets. Thus, pH influences ligand binding both directly, by changing the hydrogen-bonding character of ionizable site residues, and indirectly, by altering the shape and properties of the site with specifically bound solvent molecules[222,223]. Another significant factor in quantitative estimation of interaction strength is the microscopic dielectric constant, which is almost never known[224]. While many approximations can be made for the dielectric constant of the protein interior[225,226], the most accurate modeling of dielectric effects within proteins requires consideration of the atomic polarizabilities of the heterogeneous protein and the solvent (both water and counter ions)[227,228]. However, the size of a typical protein-ligand system renders approaches of this nature very computationally expensive; thus, simplified models that use macroscopic dielectric models – uniform or distance dependent – are generally applied[229].

In order to accurately predict binding free energies, it is fundamental to take into account pH, ionization and entropic contributions for virtual screening experiments. However, in many biomolecular systems the lack of extensive binding and inhibition data as a function of pH limits the likelihood of good correlation between calculated and measured

binding data. In this chapter we have examined x-ray structures of HIV-1 protease in complex with several cyclourea and cyclic sulfamide inhibitor analogs. The binding of ligands to a biological macromolecule is made even more complicated with the presence of multiple ionizable groups, and the HIV-1 protease/ligand system has several[230]. Existing modeling techniques are insufficient to characterize atomic level details of binding and do not often consider the multiple protonation states and ensemble of protonation states that can exist in these systems, many of which are quite similar in energy. We have applied our computational titration protocol, based on HINT (Hydropathic INTeraction), to analyze and identify the best protonation models for these complexes. Our main interest in developing the computational titration algorithm is its ability to identify and optimize all possible protonation states so that rational models with atomic details can be constructed and applied to modeling ligand binding energetics.

5.2 Materials and Methods

Optimized molecular models of the HIV-1 protease-ligand complexes in this work were taken from our previous study on the contribution of water molecules to the energetics of HIV-1/ligand binding[101]. In that study, the GRID program was used [185] for identifying and placing water in favorable locations that were unoccupied due to crystallographic uncertainty. In the cyclic urea complexes, only complex 1AJX had crystallographically detected waters 313 and 313', while these waters were not experimentally reported for complexes 1DMP and 1QBT and were positioned with GRID. However, in case of cyclic sulfamide complexes, crystallographic water 313 and 313' were experimentally

confirmed for the three complexes 1AJV, 1G35 and 1G2K except for water 313' on 1G2K, where it was positioned using GRID.

The modeling programs Sybyl 7.2 (www.tripos.com)[78] suite and HINT 3.11S were used in this study. In the HINT computational titration protocol, the protein, ligand and water were partitioned as distinct molecules: only hydrogen atoms deemed "semi-essential", i.e., only those attached to polar atoms (N, O, S, P) and those attached to unsaturated carbons were explicitly considered in the model and assigned HINT constants. The inferred solvent model, where each residue is partitioned based on its hydrogen count, was applied. The solvent accessible surface area for the amide nitrogens of the protein backbone were corrected with the "+20" option. All Asp, Glu and His residues within an 8 Å radius of the ligand and the ionizable functional groups on the ligand (amine, phosphine, carboxylate groups) were selected for titration. In this study only aspartates were present at the HIV-1 active site and only one ligand had a titratable amine.

Optimization of each protonation state model focuses on exhaustive optimization of the R–XH_n bond dihedral angles that are exhaustively optimized by forcing rotation through the entire 360°. This rotation positions the polar hydrogens for optimum hydrogen bonding and intermolecular HINT energy score. The procedure targets primary amine, hydroxyl and thiol groups on both protein and ligand. Note that the –OH on protonated carboxylate groups are also optimized in this way. In His the imidazole ring is flipped to optimize hydrogen bonding, as are the terminal amides of Asn and Gln. All the possible models were composed and scored using equation 1. Finally, HINT scores were plotted

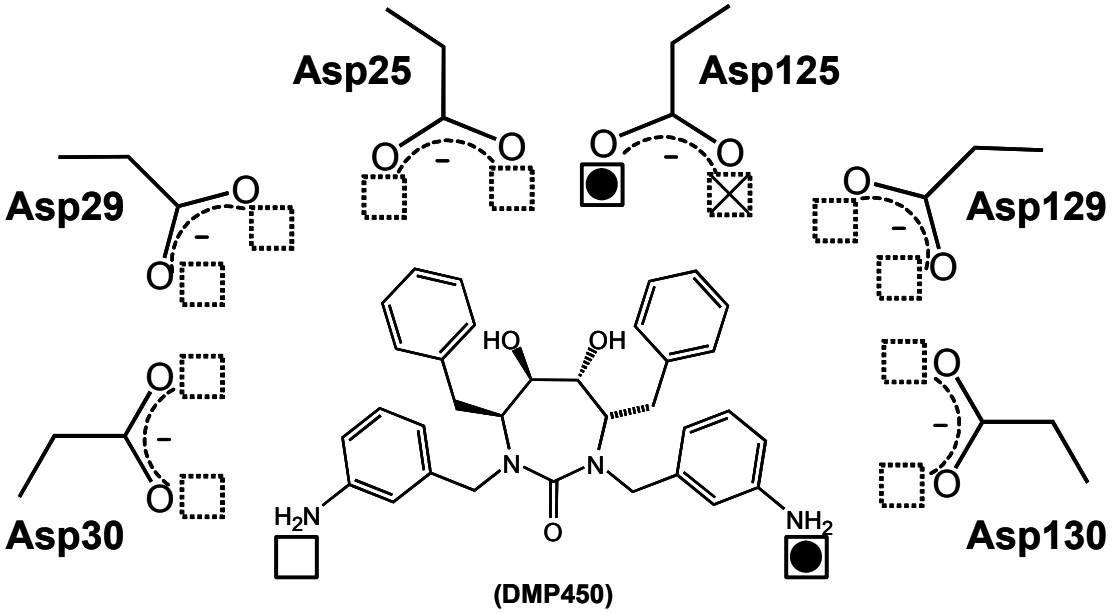
as a function of pH for all the models. Normal and Boltzmann-weighted averages were then obtained for each protonation level.

5.3 Results and Discussion

5.3.1 The Computational Titration Algorithm

The Computational Titration (CT) algorithm is based on the empirical HINT (Hydropathic INTeractions) free energy non-covalent forcefield[92], and involves modeling of optimum ionization conditions at the binding site[99,192]. The Computational Titration methodology allows exploration of the ionization states of active site residues and ligand functional groups as a function of protonation. This novel method, implemented in the HINT software, involves building and scoring of distinct protonation models, i.e., a defined ionization state for each relevant residue or ligand functional group. First, all the possible ionization states of residues and ligand functional groups are enumerated with user choices of residue types[192,231,232]. The illustration of Figure 31 indicates how this is accomplished. Each box represents a potential position for a proton. With no added protons there is only 1 model, but with one added proton there are 14 unique models, and so on. In the example of Figure 31, one of the ligand's amines is protonated and one of the two carboxylate oxygens of Asp125 is protonated. The acidic residues Asp and Glu are usually subjected to titration; the basic residues Lys and Arg are usually kept in their protonated states, while His is normally selected for titration. The CT algorithm further allows selection of Tyr and Cys for titration in cases where exploring the ionization of these residues is desired.

Figure 31. Protonation model. Schematic representation of active site ionizable residues and ligand functional groups subjected to computational titration; in this case for complex 1DMP. Boxes indicate potential protonation sites, i.e., potential positions for protons. The sole protonation site on amines is indicated by a solid box, whereas the two potential protonation sites on aspartates are indicated by dotted boxes. (Protons can be placed on either carboxylate oxygen, but not both). The illustrated model has two added protons: one of the two amine groups on the ligand and one of the two carboxylate oxygens on Asp125 are protonated ("occupied" solid boxes); the other carboxylate oxygen is unavailable for protonation.



The analogous functional groups on the ligands, i.e., carboxylic acids, amines, aromatic alcohols and thiols, are user-selectable for titration. Cofactors or water molecules are also taken into account. The HINT Titration algorithm identifies and exhaustively optimizes the water molecules that are in potentially bridging positions, i.e., within a 4Å radius from both protein and ligand, and thus able to interact with both[98]. During the computation, each model corresponding to a particular protonation state is built and then optimized. Here, all rotatable bonds involving polar hydrogens ($R\text{-}XH_n$) – including those newly created via protonations such as the -OHs of carboxylic acids – at both the protein active site and on the ligand will be examined and exhaustively optimized[192] to maximize hydrogen bonds formation, i.e., by rotation of these bonds through the entire 360°. In addition, Asn and Gln side chain amides are rotated ± 180° and oriented for optimal interaction. Simply, this algorithm creates rotameric models that are isocrystallographic in that all of the models should fit within the experimental electron density envelopes and be indistinguishable. Next, the algorithm calculates the HINT score of each rotameric model and creates a table of HINT score as a function of site charges. From that a "titration curve" can be calculated. To translate the HINT scores to free energy, we used the equation[102]:

$$\Delta G = -0.00195\, H_{TOT} - 5.543 \qquad (11)$$

Finally, the statistical thermodynamics Boltzmann energy partition function is applied to each site charge to calculate population-weighted averages of binding free energy for

each site charge. The Maxwell–Boltzmann distribution from statistical mechanics forms the basis of understanding classical molecular phenomena in terms of how energy is distributed in an average sample of states. Mathematically, the Boltzmann distribution can be expressed in the form:

$$f(E) = Ae^{-E/kT} \tag{12}$$

It gives the probability of any molecule existing in an energy state E as a function of its free energy. In our case, it elucidates the possibility of an ionization state existing in a particular energy state as a function of that state's free energy. The probability decreases exponentially as the free energy of the state increases. The Boltzmann average energy weights the energies of each state by these probabilities and is more representative of the overall ensemble energy than a normal average.

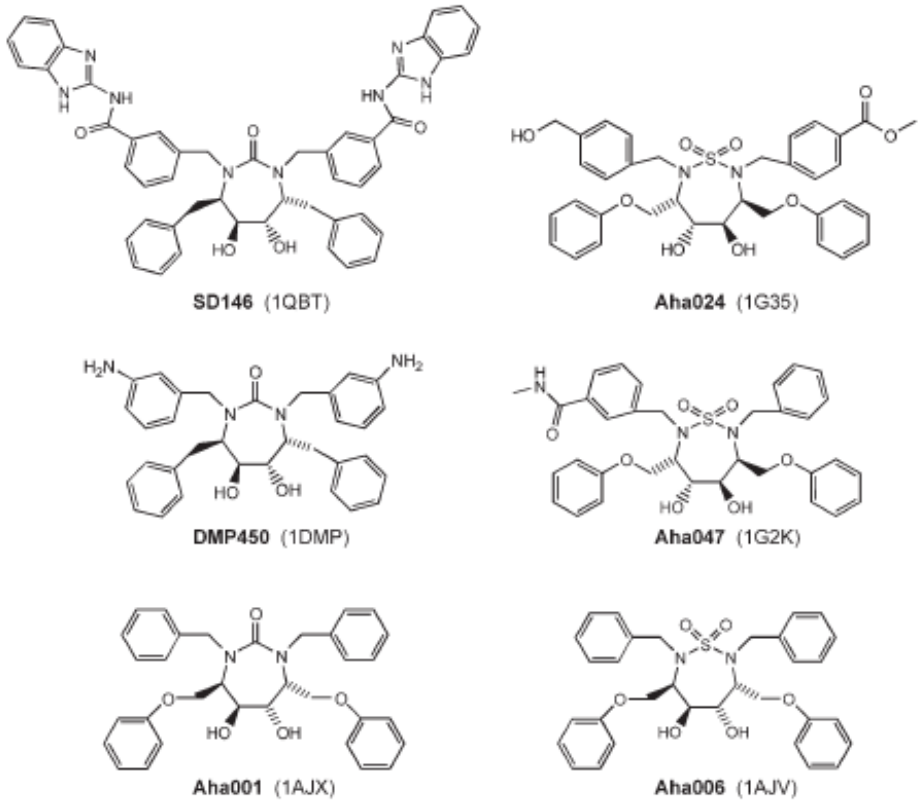
5.3.2 Ionization State Ensemble of HIV-1 Protease.

HIV-1 protease, which has been widely studied because of its crucial role in propagation of the AIDS virus, but also as a prototypical target enzyme vital for proteolytic cleavage of viral proteins, is a homodimeric aspartyl protease with 99 amino acid residues in each subunit[230]. The active site of the enzyme is situated at the interface between the two monomers. A conserved water molecule, water 301, located at the HIV-1 protease symmetry axis and bridging the two subunits, has been observed in all HIV protease-ligand complexes hydrogen bonded to residues Ile50 and Ile150 and to specific inhibitors. Two other water molecules, water 313[233] and its pseudo-symmetrical 313',

are largely conserved and crystallographically detected in several HIV-1 protease-ligand complexes. Initially, protease inhibitors, mostly linear peptide analogues, were designed such that they coordinated with water 301, linking them to the amide hydrogens of Ile50 and Ile150 on the flaps of the protease dimer. It was observed from several studies that this water molecule was crucial for the binding of these inhibitors[234]. Thus, it was hypothesized that incorporation of the binding features of this structural water molecule *into* an inhibitor would energetically favor the binding interactions, increasing both binding affinity as well as specificity, since this structural water is unique to the aspartyl proteases and its substitution would be thermodynamically favorable at least partly due to increased entropy.

Meticulous design of cyclic ureas as HIV-1 protease inhibitors specifically designed to displace water 301 was reported by Lam *et al.* in 1994[235]. An essential feature of this class of analogues was the carbonyl oxygen that mimicked the hydrogen bonding features of the key water molecule. Furthermore, the conversion of the flexible linear peptidic inhibitors into rigid, cyclic structures with restricted conformations provided additional favorable entropic benefits. The preferred conformations and stereochemistry of these inhibitors were confirmed by x-ray crystallography[235,236] (see Figure 32). The crystal structures also revealed the hydrogen bonds between the ligand diol groups and the carboxylates of the catalytic aspartates (Asp25 and Asp125) that serve to anchor the scaffold in the active site. However, since protons are normally not detected by x-ray crystallography, there is experimental uncertainty in the correct assignment of the protonation state of the catalytic dyad.

Figure 32 . Cyclic inhibitors of HIV-1 protease.



Careful analysis of *all* the possible protonation states can reveal more about the hydrogen bonding, including the possible existence of alternative networks of hydrogen bonds. This understanding could lead to the design of better inhibitors with greater binding affinity.

In the current study we examined six HIV-1 protease-inhibitor complexes (see Figure 32) representing cyclic urea derivatives, SD146 (pdb code 1QBT[237]), DMP450 (pdb code 1DMP[238]) and Aha001 (pdb code 1AJX[239]), and cyclic sulfamide derivatives, Aha024 (pdb code 1G35[240]), Aha047 (pdb code 1G2K[240]) and Aha006 (pdb code 1AJV[239]). For the cyclic urea analogs the carbonyl oxygen substitutes the position occupied by the oxygen of water 301 and thus forms hydrogen bonds with Ile50 and Ile150. For the sulfamide analogs both oxygens of the sulfamide group are engaged in hydrogen bonding, with one oxygen hydrogen-bonded to the amide nitrogen of Ile50, and the other to the amide nitrogen of Ile150. The diols are engaged in hydrogen bond networks with the catalytic aspartates.

The HIV-1 protease active site presents a number of ionizable residues ideally suited for the computational titration protocol. In the six HIV-1-ligand complexes ionizable residues located within 8 Å of the ligand were chosen for titration. The contributions of waters 313 and 313' were also included in the calculations. Note that, in a sense, water, being both a potential hydrogen bond donor and acceptor, may act as a pH buffer by re-orientating after a protonation change on a neighboring functional group. Thus, allowing it to freely rotate is an important component of the CT algorithm. For all complexes, the

active site residues Asp25, Asp29, and Asp30 on chain A, and Asp125, Asp129 and Asp130 on chain B were selected for titration, while only the aminic groups on the 1DMP ligand were subjected to titration. Thus, complex 1DMP has a total of eight ionizable functional groups – six on the protein and two on the ligand – yielding 2916 protonation models for building and scoring. All the remaining complexes have six ionizable residues at the protein active site – yielding 729 protonation models.

The results of the titration are shown in Figure 33 for each of the six HIV-1-ligand complexes investigated, where the HINT scores for each protonation model are plotted as a function of site charge. The normal (arithmetic) average and the Boltzmann average (statistically weighted average based on site populations) are calculated for each site charge and the corresponding "titration curves" are obtained. Details of the CT calculations, i.e., number of models, normal average and Boltzmann average free energies, for each site charge are listed in Table 11 and 12. With the exception of complex 1QBT, the titration curves for all the complexes show a bell-like shape, particularly evident in complex 1DMP (Figure 33b). The 1QBT titration curve (Figure 33a) has a different trend and while the free energy diminishes with protonation, the curve does not reach a minimum. The ligands considered in this study have mostly similar chemical structure and size, but the ligand in 1QBT is bulkier with two substituted benzoimidazole groups. This complex has the highest experimental binding free energy, but its calculated energy is underestimated by the HINT scoring function. This could be due to the structure of the ligand itself: possibly the protonation state of the benzomidazole groups that are not

Table 11. Computational Titration results for the HIV-1 protease-cyclic inhibitor complexes: cyclic urea ligands.

Cyclic Urea Ligands				
PDB	**Site Charge**	**Model Count**	**Normal Average [a]) [kcal mol^{-1}]**	**Boltzmann Average [a]) [kcal mol^{-1}]**
1QBT	-6	1	-4.67	-4.67
	-5	12	-5.02	-5.11
	-4	60	-5.39	-5.55
	-3	160	-5.72	-5.95
	-2	240	-6.01	-6.32
	-1	192	-6.29	-6.65
	0	64	-6.53	-6.91
1DMP	-6	1	-8.28	-8.28
	-5	14	-8.56	-10.09
	-4	85	-8.83	-11.50
	-3	292	-9.05	-11.79
	-2	620	-9.23	-11.82
	-1	832	-9.37	-11.61
	0	688	-9.49	-11.23
	1	320	-9.59	-10.69
	2	64	-9.71	-10.02
1AJX	-6	1	-7.40	-7.41
	-5	12	-7.68	-7.79
	-4	60	-7.81	-7.98
	-3	160	-7.93	-8.13
	-2	240	-7.99	-8.21
	-1	192	-8.00	-8.22
	0	64	-7.95	-8.10

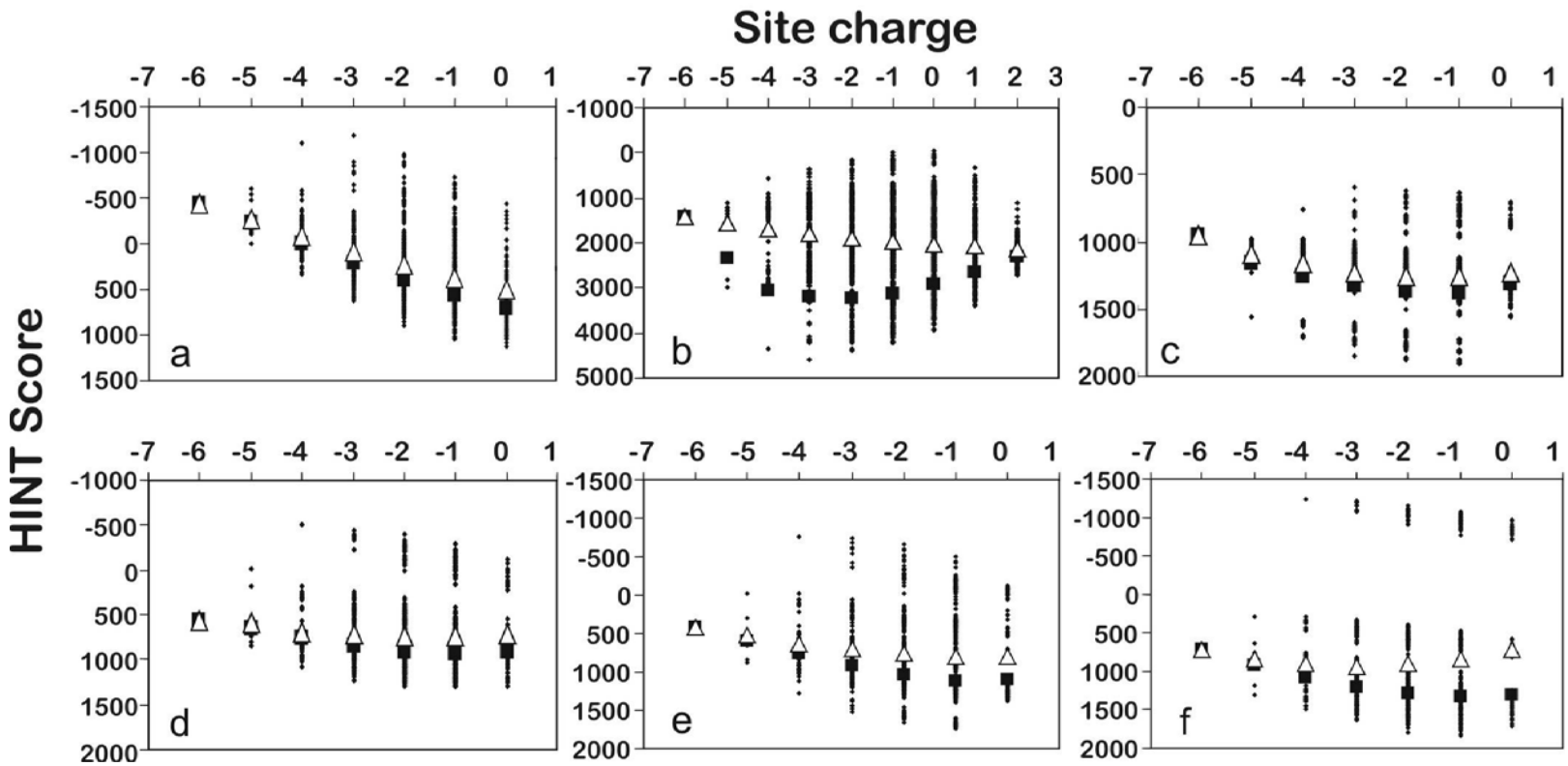
[a]) HINT scores are converted into free energies using eq. (11).

Table 12. Computational Titration results for the HIV-1 protease-cyclic inhibitor complexes: cyclic sulfamide ligands.

Cyclic Sulfamide Ligands				
PDB	**Site Charge**	**Model Count**	**Normal Average [a]) [kcal mol^{-1}]**	**Boltzmann Average [a]) [kcal mol^{-1}]**
1G35	-6	1	-6.65	-6.65
	-5	12	-6.69	-6.82
	-4	60	-6.89	-7.04
	-3	160	-6.95	-7.19
	-2	240	-6.98	-7.30
	-1	192	-6.97	-7.34
	0	64	-6.93	-7.31
1G2K	-6	1	-6.36	-6.36
	-5	12	-6.57	-6.71
	-4	60	-6.78	-7.04
	-3	160	-6.92	-7.34
	-2	240	-7.03	-7.59
	-1	192	-7.10	-7.74
	0	64	-7.11	-7.67
1AJV	-6	1	-6.94	-6.94
	-5	12	-7.17	-7.35
	-4	60	-7.32	-7.66
	-3	160	-7.36	-7.89
	-2	240	-7.31	-8.05
	-1	192	-7.17	-8.15
	0	64	-6.93	-8.11

[a]) HINT scores are converted into free energies using eq. (11).

Figure 33. Computational titration results for the HIV-1 inhibitor complexes analyzed: a) 1QBT, b) 1DMP, c) 1AJX, d) 1G35, e) 1G2K, f) 1AJV. All HINT scores are plotted as a function of site charge; in addition, normal averages (open triangles) and Boltzmann averages (closed squares) are depicted.



subjected to titration with our current algorithm. The lowest free energy point determined by Boltzmann statistics – the minimum of the Boltzmann average titration curve – is what we define to be the "optimal" calculated free energy of binding representing the most favorable site charge, and by inference pH, for the specific ligand binding. Figure 34 illustrates the best model for HIV-1 protease in complex with the cyclic urea inhibitor 1DMP, for which the titration curve is shown in Figure 33b. While the "best" model is the one with the highest HINT score and presumably represents the protonation model corresponding to the best binding, it must be highlighted that there are many models with similar energy, and that it is likely that many of them coexist, especially at room temperature or above where binding assays are performed. For 1DMP the best model corresponds to a -3 site charge, where the two amine groups on the ligand and only Asp 125 on the protein are protonated. As expected, Asp25 and Asp125 are engaged in a complex network of hydrogen bonds with the ligand's diols. The two ammonium groups on the ligand make hydrogen bonds with Asp29/Asp30/water313, and Asp129/Asp130/water313', respectively. The deprotonated Asp29 and Asp129 residues are also involved in hydrogen bonds with waters 313 and 313', respectively.

Experimental and calculated binding free energies for the examined HIV-1 ligand complexes are reported in Table 13. The "optimal" site charges, i.e., the optimal protonation states that correspond to the optimal binding energies calculated by the Boltzmann statistical analysis, are also reported. The results are encouraging, giving an average error of ± 2.5 kcal mol^{-1} in the prediction of binding energy, excluding the outlier

Figure 34. Complex 1DMP. Best model for HIV-1 protease in complex with a cyclic urea inhibitor (complex 1DMP). Ligand (ball and stick representation, colored by atom type), key active site residues (titratable aspartates in stick representation, colored by atom type; Ile50/150 stick representation, orange) and waters 313/313' (stick representation, magenta) are highlighted. Asp125 and the two amine groups on the ligand are protonated in this model. The network of hydrogen bonds at the active site is illustrated with yellow dashed lines.

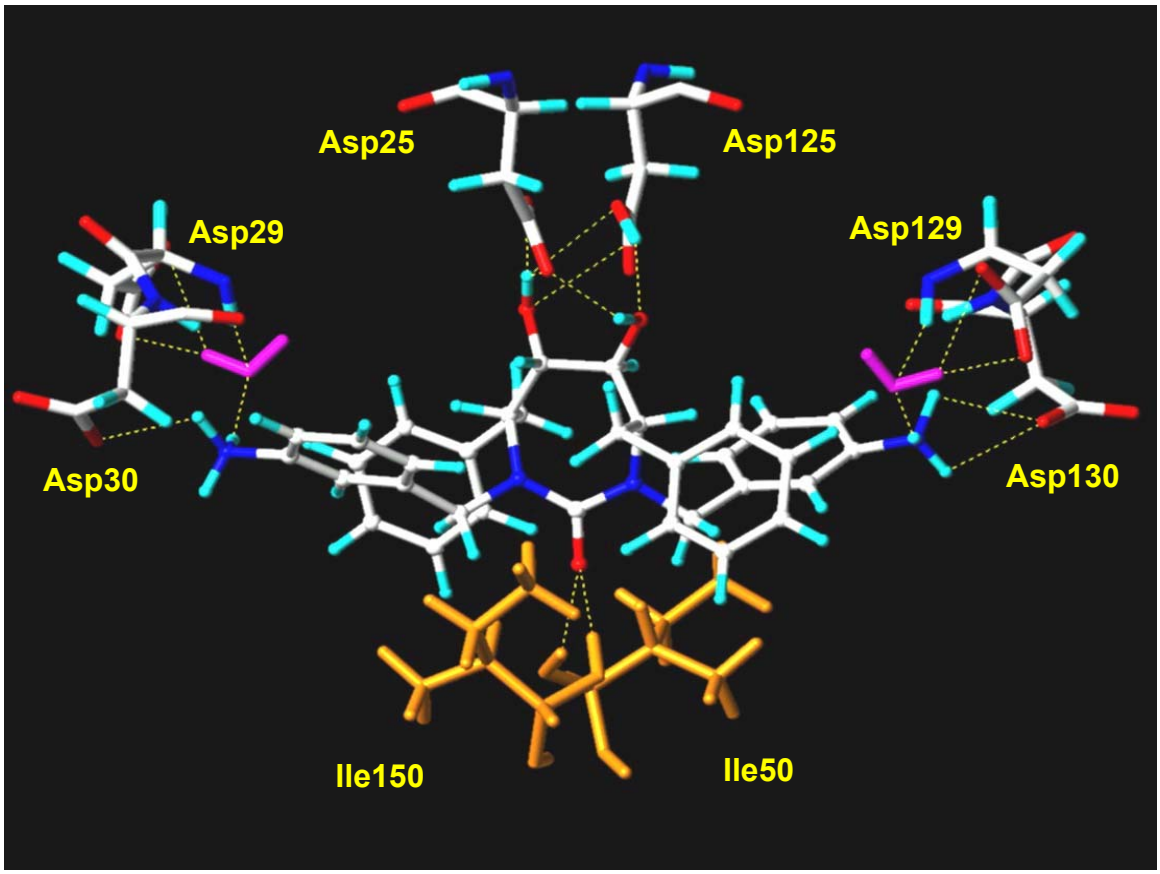


Table 13: Experimental and calculated optimal (Boltzmann-weighted average) binding free energies for HIV-1 ligand complexes.

PDB	Experimental $\Delta G_{binding}$[a]) [kcal mol^{-1}]	Optimal Calculated $\Delta G_{binding}$[b]) [kcal mol^{-1}]	Optimal site charge
1QBT	-14.44	-6.91	0[c])
1DMP	-12.99	-11.82	-2
1AJX	-10.79	-8.22	-1
1G35	-11.06	-7.34	-1
1G2K	-10.82	-7.74	-1
1AJV	-10.52	-8.15	-1

[a]) Data from.[185]

[b]) HINT scores are converted into free energies using eq. (11).

[c]) 1QBT has not reached a true titration curve minimum. This is the value at the curve's lowest point, thus the "optimal" site charge may be at a more positive value (see Figure 33a).

complex 1QBT where the calculated energy is underestimated by the computational titration and which displayed an atypical titration profile (Figure 33a). It is important to recognize that using an equation correlating HINT scores with binding energies that is calibrated for the specific analyzed system, instead of a general equation, would likely improve free energy prediction.

5.4 Summary

The computational protocol described here allows modeling of the multiplicity of protonation states, an often overlooked aspect of structure that has implications for drug discovery. This approach allows generating hypotheses on the best model for binding, i.e., the model with protonation corresponding to the optimal binding energy. The binding energy is evaluated with the HINT scoring function, which has previously been shown to fairly accurately predict binding free energies [101]. In considering all the ionizable residues at the active site and modeling all the possible protonation states of residues and functional groups at this site, the computational titration algorithm represents fairly realistically the fluxional behavior of hydrogens in solvated biological systems. Hydrogens, in fact, are not static elements of a real biomolecular system. However, the computational titration procedure remains computationally expensive. The number of models increases rapidly as the number of ionizable residues/functional groups in the binding pocket increases. The inclusion of water molecules in the calculations makes these procedures even more time consuming. Nevertheless, we are committed to making this tool available and are currently developing a web-enabled version that will be publicly accessible.

CHAPTER 6

Conclusions

Computational tools have been extensively used in drug design and development from exploration of targets to corroborate experimental data. Computer aided drug design has been used to design new bioactive compounds with the aim of optimizing the overall pharmacodynamic and pharmacokinetic profile of a drug candidate and expediting the process of drug discovery. These techniques have substantially reduced the time and economic resources needed to discover novel drug candidates. This dissertation discussed the computational modeling aspects of drug design for anti-cancer agents with a focus on their interactions with potential targets and an emphasis on explaining and suggesting each drug candidate's mechanism of action at the molecular level. Molecular modeling studies and development of computational tools that aid in drug design were also discussed. A novel *de novo* program for cavity detection and its application were described. The software tool identifies and delineates the active site within the protein. Another computational tool that allows identification, modeling and optimization of the multiple

protonation states of residues and ligand functional groups at the protein-ligand active site was also described and discussed.

The compounds which bind to the colchicine site of tubulin have drawn considerable attention and studies indicated that these agents may influence microtubule dynamics by inhibiting polymerization. However, fundamental uncertainties in many aspects of microtubule biology and insufficient knowledge of the binding site interactions have undermined the possibility of identifying effective drugs with minimal toxicity. Nevertheless, the compounds which bind to the colchicine site of tubulin have recently emerged as attractive targets for cancer therapy. Pyrrole-containing molecules derived from nature have proven to be particularly useful as lead compounds for drug development. We designed and developed a series of substituted pyrroles that inhibit growth and promote death of breast tumor cells at nM and μM concentrations in human breast tumor cell lines. Subsequent experimental studies demonstrated that the highest degree of antiproliferative activity was expressed by JG-03-14 (3,5-dibromo-4-(3,4-dimethoxyphenyl)-1Hpyrrole-2-carboxylic acid ethyl ester). COMPARE analysis across the NCI panel of cancer cell lines, along with molecular modeling studies, showed that JG-03-14 had a similar mechanism of action to colchicine and combretastatin.

The current work demonstrated that molecular modeling docking calculations along with HINT interaction analysis were able to complement experimental studies of binding in many aspects, including accurate representation of the complex structure and the binding

mode of inhibitors. HINT scoring function was used to investigate the structural aspects of the interactions. On the basis of calculations, the complexes were ranked according to their decreasing HINT scores. These results, calculated for 22 pyrrole compounds, provide insights into the structural requirements for the growth inhibitory activity/cytotoxicity of this class of agents and offer significant opportunity for structural alterations, which could lead to an improved drug candidate. Complex with JG-03-14 is the most stable corroborating the experimental data. These results are important for the understanding the binding process and valuable in the design of new Pyrrole-based colchicine site inhibitors. Compounds of this type have potential for further development. Hydropathic interaction analysis has provided a rationale for selecting the substituent on a parent ligand which will yield more tightly binding analogues.

Stilbenes are a group of natural compounds with many biological activities. The mechanism of action of stilbenes is by interfering with microtubule polymerization through the colchicine-binding site. Two highly potent stilbenes, cis-3,40,5-trimethoxy-30-aminostilbene (stilbene 5c) and cis-3,40,5-trimethoxy-30-hydroxystilbene (stilbene 6c) induce G2/M cell-cycle arrest and leukemic cell death in nanomolar range without affecting normal bone marrow progenitor cells. Animal studies showed that stilbenes are well tolerated and suppresses tumor growth in mice. Further experimental results indicated that stilbene 5c is a microtubule- interfering agent and can be potentially useful in leukemic therapy. A molecular modeling study was carried out to accurately represent the complex structure and the binding mode of a new class of stilbene-based tubulin inhibitors

that bind at the αβ-tubulin colchicine site. Computational docking along with HINT score analysis fitted these inhibitors into the colchicine site and revealed detailed structure-activity information useful for inhibitor design. Quantitative analysis of the results was in good agreement with the *in vitro* antiproliferative activity of these derivatives (ranging from 3 nM to 100 μM) such that calculated and measured free energies of binding correlate with an r^2 of 0.89 (standard error ± 0.85 kcal mol^{-1}). This correlation suggests that the activity of unknown compounds may be predicted. The results are important for understanding the binding process and valuable in the design of new stilbene-based colchicine site inhibitors.

Apart from the application of computational methodologies, design and development of new computational tools has also been discussed in the manuscript. The development and implementation of a novel cavity detection algorithm is also reported and discussed. The algorithm named VICE (Vectorial Identification of Cavity Extents) utilizes HINT toolkit functions to identify and delineate binding pocket in a protein. This algorithm, which is based on geometric criteria applied to simple integer grid maps to delineate binding sites, is very efficient. The program was applied to a representative set of proteins from different classes having binding pockets of different shapes and sizes. It was demonstrated that the application is capable of detecting and delineating indentations, cavities, pockets, voids, grooves, channels, tunnels, pores and surface regions on protein. The interactive front-end provides a quick and simple procedure for identifying, displaying and manipulating cavities in a known protein structure. In addition to computing volumes, our method also

provides information on atoms, residues and chains which contribute to the cavities. These observations have been most thoroughly characterized and correlate well with the experimental data.

The study also implemented the computational titration algorithm to understand the complexity of ligand binding and protonation state in the active site of HIV-1 protease. The importance of taking into account pH and ionization states of residues, which strongly affect the process of ligand binding, for correctly predicting the binding free energy is discussed. The application of the computational titration protocol to a set of six HIV-1-cyclic inhibitor complexes was carried out. The binding energy was evaluated with the HINT scoring function. The results were encouraging, giving an average error of ± 2.5 kcal mol^{-1} in the prediction of binding energy. The titration algorithm allowed the identification, modeling and optimization of the multiple protonation states of residues and ligand functional groups at the protein-ligand active site. This approach allows generating hypotheses on the best model for binding, i.e., the model with protonation corresponding to the optimal binding energy. In considering all the ionizable residues at the active site and modeling all the possible protonation states of residues and functional groups at this site, the computational titration algorithm represents fairly realistically the fluxional behavior of hydrogens in solvated biological systems.

To conclude, the overall purpose of this multidisciplinary endeavour was to design and develop novel therapeutics for cancer and understand the molecular mechanism involved

in drug action using computational tools. A second synergistic goal was to develop new computational tools that will aid in design bioactive molecules and understanding the molecular mechanism of protein-ligand binding process. To this end, application of molecular modeling has facilitated the design and development of the new anti-cancer compounds and has served to improve the understanding of the underlying mechanisms of microtubule depolymerizing agents. In its most general terms, the overall design and refinement of the novel antitumor compounds proposed herein is a fundamental step towards establishing a knowledge base that will enable the synthesis and testing of effective chemotherapeutic agents. The HINT force field can provide atomic level details of interactions and may help in the design of more selective drug molecules. The HINT program finds extensive application in drug design project and the knowledge acquired will prove to be productive and of scientific significance. This research will further provide the scientific community with additional knowledge that will help them in the *de novo* identification and characterization of binding site and to understand the relevant molecular-level interactions.

Literature Cited

(1) Garcia, M.; Jemal, A.; Ward, E. M.; Center, M.; Hao, Y. et al. Global Cancer Facts & Figures 2007. *American Cancer Society, 2007* **2007**.
(2) American Cancer Society. Cancer Facts & Figures 2008. *Atlanta: American Cancer Society; 2008.* **2008**.
(3) World Health Organization Fact Sheet N^0 297. **2009**.
(4) The World Health Organization's Fight Against Cancer: Strategies That Prevent, Cure and Care. **2007**.
(5) Ltadani, H.; Mizuarai, S.; Kotani, H. Can systems biology understand pathway activation? Gene expression signatures as surrogate markers for understanding the complexity of pathway activation. *Current Genomics* **2008**, *9*, 349-360.
(6) Xiao, G. G.; Recker, R. R.; Deng, H.-W. Recent advances in proteomics and cancer biomarker discovery. *Clinical Medicine: Oncology* **2008**, *2*, 63-72.
(7) Croce, C. M. Oncogenes and cancer. *New England Journal of Medicine* **2008**, *358*, 502-511.
(8) Hanahan, D.; Weinberg, R. A. The hallmarks of cancer. *Cell* **2000**, *100*, 57-70.
(9) Oliff, A. Farnesyltransferase inhibitors: targeting the molecular basis of cancer. *Biochim Biophys Acta* **1999**, *1423*, C19-30.
(10) Cerutti, P. A. Prooxidant states and tumor promotion. *Science* **1985**, *227*, 375-381.
(11) Griswold, D. P., Jr.; Harrison, S. D., Jr. Tumor models in drug development. *Cancer Metastasis Rev* **1991**, *10*, 255-261.
(12) Zarbl, H. Cellular oncogenes and carcinogenesis. *Molecular Carcinogenesis and the Molecular Biology of Human Cancer* **2006**, *103*, 103-129.
(13) Weiss, G. R.; Burris, H. A., 3rd; Eckhardt, S. G.; Rodriguez, G. I.; Sharma, S. et al. New anticancer agents. *Cancer Chemother Biol Response Modif* **1997**, *17*, 178-194.
(14) Weiss, G. R. Drugs in the treatment of cancer chemotherapy. *Appleton and Lange: East Norwalk* **1997**.
(15) Isoldi, M. C.; Visconti, M. A.; Castrucci, A. M. Anti-cancer drugs: molecular mechanisms of action. *Mini Rev Med Chem* **2005**, *5*, 685-695.
(16) Pervaiz, S. Anti-cancer drugs of today and tomorrow: are we close to making the turn from treating to curing cancer? *Curr Pharm Des* **2002**, *8*, 1723-1734.
(17) Wilson, L.; Jordan, M. A. New microtubule/tubulin-targeted anticancer drugs and novel chemotherapeutic strategies. *J Chemother* **2004**, *16 Suppl 4*, 83-85.
(18) Jordan, M. A.; Wilson, L. Microtubules as a target for anticancer drugs. *Nat Rev Cancer* **2004**, *4*, 253-265.
(19) Jordan, A.; Hadfield, J. A.; Lawrence, N. J.; McGown, A. T. Tubulin as a target for anticancer drugs: agents which interact with the mitotic spindle. *Med Res Rev* **1998**, *18*, 259-296.

(20) Houston, J. G.; Banks, M. The chemical-biological interface: developments in automated and miniaturised screening technology. *Curr Opin Biotechnol* **1997**, *8*, 734-740.
(21) Gallop, M. A.; Barrett, R. W.; Dower, W. J.; Fodor, S. P.; Gordon, E. M. Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries. *J Med Chem* **1994**, *37*, 1233-1251.
(22) Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P.; Gallop, M. A. Applications of combinatorial technologies to drug discovery. 2. Combinatorial organic synthesis, library screening strategies, and future directions. *J Med Chem* **1994**, *37*, 1385-1401.
(23) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening- an overview. *Drug Discovery Today* **1998**, *3*, 160-178.
(24) Verlinde, C. L. M. J.; Hol, W. G. J. Structure-based drug design: progress, results and challenges. *Structure* **1994**, *2*, 577-587.
(25) Kapetanovic, I. M. Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. *Chemico-Biological Interactions* **2008**, *171*, 165-176.
(26) Tajkhorshid, E.; Suhai, S. The dielectric effect of the environment on the pKa of the retinal Schiff base and on the stabilization of the ion pair in bacteriorhodopsin. *Journal of Molecular Structure (Theochem)* **2000**, *501-502*, 297-313.
(27) Murphy, K. P. Predicting binding energetics from structures: Looking beyond δG°. *Medicinal research reviews* **1999**, *19*, 333-339.
(28) Lee, C. Calculating binding energies. *Curr. Opin. Struct. Biol.* **1992**, *2*, 217-222.
(29) Anderson, A. C. The Process of Structure-Based Drug Design. *Chemistry & Biology* **2003**, *10*, 787-797.
(30) Campbell, S. J.; Gold, N. D.; Jackson, R. M.; Westhead, D. R. Ligand binding: functional site location, similarity and docking. *Curr Opin Struct Biol* **2003**, *13*, 389-395.
(31) Kleywegt, G. J. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr D* **1994**, *50*, 178-185.
(32) Selway, C. N.; Terrett, N. K. Parallel-compound synthesis: methodology for accelerating drug discovery. *Bioorg Med Chem* **1996**, *4*, 645-654.
(33) DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D. et al. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J Med Chem* **1988**, *31*, 722-729.
(34) Murcko, M. A. Recent advances in ligand design methods. *Reviews in Computational Chemistry* **2007**, *11*, 1-66.
(35) Dean, P.; Harris, B.; Willems, H. De novo design of focused libraries of drug-like compounds. *DrugPlus international* **2005**.
(36) Schneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat Rev Drug Discov* **2005**, *4*, 649-663.
(37) Beavers, M. P.; Chen, X. Structure-based combinatorial library design: methodologies and applications. *J Mol Graph Model* **2002**, *20*, 463-468.

(38) Hindle, S. A.; Rarey, M.; Buning, C.; Lengaue, T. Flexible docking under pharmacophore type constraints. *J Comput Aided Mol Des* **2002**, *16*, 129-149.
(39) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* **2004**, *3*, 935-949.
(40) Honma, T. Recent advances in de novo design strategy for practical lead identification. *Med Res Rev* **2003**, *23*, 606-632.
(41) Lengauer, T.; Rarey, M. Computational methods for biomolecular docking. *Current Opinion in Structural Biology* **1996**, *6*, 402-406.
(42) Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.,* **1968**, *33*, 491-497.
(43) Gilson, M. K. Theory of electrostatic interactions in macromolecules. *Curr. Opin. Struct. Biol.* **1995**, *5*, 216-223.
(44) Vajda, S.; Guarnieri, F. Characterization of protein-ligand interaction sites using experimental and computational methods. *Curr Opin Drug Discov Devel* **2006**, *9*, 354-362.
(45) Burgoyne, N. J.; Jackson, R. M. Predicting protein interaction sites: binding hot-spots in protein-protein and protein-ligand interfaces. *Bioinformatics* **2006**, *22*, 1335-1342.
(46) RosenfieId, R. E.; Swanson, J. S. M.; Meyer, E. F.; CarrelI, J. H. L.; Murray-Rust, P. Mapping the atomic environment of functional groups: turning 3D scatter plots into pseudo-density contours. *Journal of Molecular Graphics* **1984**, *2*, 43-46.
(47) Patel, Y.; Gillet, V. J.; Bravi, G.; Leach, A. R. A comparison of the pharmacophore identification programs: Catalyst, DISCO and GASP. *Journal of Computer-Aided Molecular Design* **2002**, *16*, 653-681.
(48) Barillari, C.; Marcou, G.; Rognan, D. Hot-spots-guided receptor-based pharmacophores (HS-Pharm): a knowledge-based approach to identify ligand-anchoring atoms in protein cavities and prioritize structure-based pharmacophores. *J Chem Inf Model* **2008**, *48*, 1396-1410.
(49) Ebalunode, J. O.; Ouyang, Z.; Liang, J.; Zheng, W. Novel approach to structure-based pharmacophore search using computational geometry and shape matching techniques. *J Chem Inf Model* **2008**, *48*, 889-901.
(50) Ghose, A. K.; Wendoloski, J. J. Pharmacophore Modelling: Methods, Experimental Verification and Applications. *Perspectives in Drug Discovery and Design* **1998**, 253–271.
(51) Landon, M. R.; Lancia, D. R., Jr.; Yu, J.; Thiel, S. C.; Vajda, S. Identification of hot spots within druggable binding regions by computational solvent mapping of proteins. *J Med Chem* **2007**, *50*, 1231-1240.
(52) Chen, J.; Lai, L. Pocket v.2: Further Developments on Receptor-Based Pharmacophore Modeling. *J. Chem. Inf. Model.* **2006**, *46*, 2684-2691.
(53) Bohm, H. J. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61-78.

(54) Gohlke, H.; Klebe, G. Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angew. Chem. Int. Ed* **2002**, *41*, 2644-2676.
(55) Spyrakis, F.; Cozzini, P.; Kellogg, G. E. Docking and scoring in drug discovery. *Burger's Medicinal Chemistry and Drug Discovery Seventh Edition* **2009**, *1*, Chapter 7.
(56) Leach, A. A Survey of Methods for searching the Conformational Space of Small and Medium-Sized Molecules. *Reviews in Computational Chemistry. 1991 VCH, New York.* **1991**, 1-55.
(57) Brooijmans, N., Kuntz, ID. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* **2003**, *32*, 335-373.
(58) Perola, E., Xu, K, Kollmeyer, TM, Kaufmann, SH, Prendergast, FG, Pang, YP, Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *J Med Chem* **2000**, *43*, 401-408.
(59) Jiang, F.; Kim, S. Soft docking": matching of molecular surface cubes. *J Mol Biol* **1991**, *219*, 79-102.
(60) Kuntz, I., Blaney, JM, Oatley, SJ, Langridge, R, Ferrin, TE. A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **1982**, *161*, 269-288.
(61) Sandak, B., Nussinov, R, Wolfson, HJ. An automated computer vision and robotics-based technique for 3-D flexible biomolecular docking and matching. *Comput Appl Biosci* **1995**, *11*, 87-99.
(62) DesJarlais, R., Sheridan, RP, Dixon, JS, Kuntz, ID, Venkataraghavan, R. Docking flexible ligands to macromolecular receptors by molecular shape. *J Med Chem* **1986**, *29*, 2149-2153.
(63) Sandak, B., Nussinov, R, Wolfson, HJ. A method for biomolecular structural recognition and docking allowing conformational flexibility. *J Comput Biol* **1998**, *5*, 631-654.
(64) Metropolis, N., Rosenbluth, A, Rosenbluth, M, Teller, A, Teller, E. Equation of State Calculations by Fast Computing Machines. *J Chem Phys* **1953**, *21*, 1087.
(65) Hart, T., Read, RJ. A multiple-start Monte Carlo docking method. *Proteins* **1992**, *13*, 206-222.
(66) Clark, D., Westhead, DR. Evolutionary algorithms in computer-aided molecular design. *Comput Aided Mol Des* **1996**, *10*, 337-358.
(67) Judson, R. Genetic Algorithms and their use in Chemistry. *Reviews in Computational Chemistry. 1997 VCH Publishers, New York.* **1997**, 1-73.
(68) Morris, G., Goodsell, DS, Halliday, RS, Huey, R, Hart, WE, Belew, RK, Olson, AJ. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Comput Chem* **1998**, *19*, 1639-1662.
(69) Jones, G., Willett, P, Glen, RC, Leach, AR, Taylor, R, Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **1997**, *267*, 727-748.
(70) Vieth, M., Hirst, J, Dominy, B, Daigler, H, Brooks, C. Assessing search strategies for flexible docking. *J Comput Chem* **1998**, *19*, 1623-1631.

(71) Wu, G., Robertson, DH, Brooks, CL, 3rd, Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER-A CHARMmbased MD docking algorithm. *J Comput Chem* **2003**, *24*, 1549-1562.
(72) Baxter, C., Murray, CW, Clark, DE, Westhead, DR, Eldridge, MD. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins* **1998**, *33*, 367-382.
(73) Price, M., Jorgensen, W. Analysis of Binding Affinities for Celecoxib Analogues with COX-1 and COX-2 from Combined Docking and Monte Carlo Simulations and Insight into the COX-2/COX-1 Selectivity. *J Am Chem Soc* **2000**, *122*, 9455-9466.
(74) Hoffmann, D., Kramer, B, Washio, T, Steinmetzer, T, Rarey, M, Lengauer, T. Two-stage method for protein-ligand docking. *J Med Chem* **1999**, *42*, 4422-4433.
(75) Bohm, H., Stahl, M. The use of scoring functions in drug discovery applications. *Reviews Comput Chem. 2002 Wiley-VCH, John Wiley and Sons, Inc., New York.* **2002**, 41-87.
(76) Kaminski, G., Jorgensen, WL. Performance of the AMBER94, MMFF94, and OPLS-AA Force Fields for Modeling Organic Liquids. *J Phys Chem* **1996**, *100*, 18010-18013.
(77) Weiner, S., Kollman, PA, Nguyen, DT, Case, DA. An all atom force field for simulations of proteins and nucleic acids. *J Comput Chem* **1986**, *252*, 230-252.
(78) The SYBYL software, TI, http://www.tripos.com/. 1995; Tripos Inc.: St. Louis, MO.
(79) Halgren, T. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem* **1996**, *17*, 490-519.
(80) Dinur, U., Hagler, A. New Approaches to Empirical Force Fields. *Reviews in Computational Chemistry. 1991 VCH, New York.* **1991**, 99-164.
(81) Nemethy, G., Gibson, KD, Palmer, KA, Yoon, CN, Paterlini, G, Zagari, A, Rumsey, S, Scheraga, HA. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *J Phys Chem* **1992**, *96*, 6472-6484.
(82) Naim, M., Bhat, S, Rankin, KN, Dennis, S, Chowdhury, SF, Siddiqi, I, Drabik, P, Sulea, T, Bayly, CI, Jakalian, A, et al. Solvated interaction energy (SIE) for scoring protein-ligand binding affinities. 1. Exploring the parameter space. *J Chem Inf Model* **2007**, *47*, 122-133.
(83) Zou, X., Yaxiong, S, Kuntz, ID. Inclusion of Solvation in Ligand Binding Free Energy Calculations Using the Generalized-Born Model. *J. Am. Chem. Soc.* **1999**, *121*, 8033-8043.
(84) Verdonk, M., Chessari, G, Cole, JC, Hartshorn, MJ, Murray, CW, Nissink, JW, Taylor, RD, Taylor, R. Modeling water molecules in protein-ligand docking using GOLD. *J Med Chem* **2005**, *48*, 6504-6515.
(85) Koppensteiner, W., Sippl, MJ. Knowledge based potentials-back to the roots. *Biochemistry (Mosc)* **1998**, *63*, 247-252.

(86) Muegge, I. PMF scoring revisited. *J Med Chem* **2006**, *49*, 5895-5902.
(87) Clark, R., Strizhev, A, Leonard, JM, Blake, JF, Matthew, JB. Consensus scoring for ligand/protein interactions. *J Mol Graph Model* **2002**, *20*, 281-295.
(88) McCammon, J. A. Free energy from simulations. *Current Opinion in Structural Biology* **1991**, *1*, 196-200.
(89) Ajay; Murcko, M. A. Computational Methods to Predict Binding Free Energy in Ligand-Receptor Complexes. *Journal of Medicinal Chemistry* **1995**, *38*, 4953-4967.
(90) Kollman, P. Free energy calculations-applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395-2417.
(91) Kellogg, G. E. Hydrophobicity: is $LogP_{o/w}$ more than the sum of its parts? *Eur J Med Chem* **2000**, *35*, 651-661.
(92) Kellogg, G. E.; Burnett, J. C.; Abraham, D. J. Very empirical treatment of solvation and entropy: a force field derived from log Po/w. *J Comput Aided Mol Des* **2001**, *15*, 381-393.
(93) Kellogg, G. E.; Abraham, D. J. Development of Empirical Biomolecular Interaction Models that Incorporate Hydrophobicity and Hydropathy. The HINT Paradigm. *Analysis* **1999**, *27*, 19-22.
(94) Kellogg, G. E.; Joshi, G. S.; Abraham, D. J. New Tools for Modeling and Understanding Hydrophobicity and Hydrophobic Interactions. *Med. Chem. Res.* **1992**, *1*, 444-453.
(95) Kellogg, G. E.; Semus, S. F.; Abraham, D. J. HINT: a new method of empirical hydrophobic field calculation for CoMFA. *J Comput Aided Mol Des* **1991**, *5*, 545-552.
(96) Spyrakis, F.; Amadasi, A.; Fornabaio, M.; Abraham, D. J.; Mozzarelli, A. et al. The consequences of scoring docked ligand conformations using free energy correlations. *Eur J Med Chem* **2007**, *42*, 921-933.
(97) Hansch, C.; Leo, A. J. Substituent Constants for Correlation Analysis in Chemistry and Biology. *John Wiley and Sons, Inc. New York, NY, 1979.* **1979**.
(98) Kellogg, G. E.; Chen, D. L. The importance of being exhaustive. Optimization of bridging structural water molecules and water networks in models of biological systems. *Chem Biodivers* **2004**, *1*, 98-105.
(99) Kellogg, G. E.; Fornabaio, M.; Spyrakis, F.; Lodola, A.; Cozzini, P. et al. Getting it right: modeling of pH, solvent and "nearly" everything else in virtual screening of biological targets. *J Mol Graph Model* **2004**, *22*, 479-486.
(100) Tripathi, A.; Fornabaio, M.; Kellogg, G. E.; Gupton, J. T.; Gewirtz, D. A. et al. Docking and hydropathic scoring of polysubstituted pyrrole compounds with antitubulin activity. *Bioorg Med Chem* **2008**, *16*, 2235-2242.
(101) Fornabaio, M.; Spyrakis, F.; Mozzarelli, A.; Cozzini, P.; Abraham, D. J. et al. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 3. The free energy contribution of structural water molecules in HIV-1 protease complexes. *J Med Chem* **2004**, *47*, 4507-4516.

(102) Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E. et al. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 1. Models without explicit constrained water. *J Med Chem* **2002**, *45*, 2469-2483.
(103) Kellogg, G. E.; Fornabaio, M.; Chen, D. L.; Abraham, D. J. New Application Design for a 3D Hydropathic Map–Based Search for Potential Water Molecules Bridging between Protein and Ligand. *Internet Electron. J. Mol. Des.* **2005**, *4*, 194-209.
(104) Wilson, E. K. Plumbing the Ocean Depths for Drugs. *Chem. and Eng. News* **2003**, *37-38*.
(105) Bailly, C. Lamellarins from A to Z: A Family of Anticancer Marine Pyrrole Alkaloids. *Curr. Med. Chem.-Anti-Cancer Agents* **2004**, *4*, 363-378.
(106) Evans M. A. , S. D. C., Holub J. M., Argenti A., Hoff M , Dalglish G. A., Wilson D. L., Taylor B. M., Berkowitz J. D., Burnham B. S., Krumpe K., Gupton J. T., Scarlett T. S., Durham, Jr. R. W., Hall I. H. Synthesis and Cytotoxicity of Substituted Ethyl 2-Phenacyl-3-phenylpyrrole-4 carboxylate. *Arch. Pharm. Pharm. Med. Chem.* **2003**, *336*, 181-190.
(107) Lee, M. Pyrrole Natural Products with Antitumor Properties. *Heterocyclic Antitumor Antibiotics: Topics in Heterocyclic Chemistry, Vol.2, Ed. by , Springer Verlag, Berlin/Heidelberg* **2006**, 53-92.
(108) Mooberry, S. W., K.; Dakshanamurthy, S.; Hamel, E.; Banner, E.; Kharlamova, A.; Hempel, J.; Gupton, J.; Brown, Identification and characterization of a new tubulin-binding tetrasubstituted brominated pyrrole. *M. Mol. Pharmacol.* **2007**, *72*, 132-140.
(109) Handy, S. T.; Zhang, Y. Approaches to the synthesis of the Lamellarins and related natural products. *Org. Prep. Proc. Int.* **2005**, *37*, 411-445.
(110) Owellen, R. J.; Owens, A. H., Jr.; Donigian, D. W. The binding of vincristine, vinblastine and colchicine to tubulin. *Biochem Biophys Res Commun* **1972**, *47*, 685-691.
(111) Jordan, A.; Hadfield, J. A.; Lawrence, N. J.; McGown, A. T. Tubulin as a target for anticancer drugs: agents which interact with the mitotic spindle. *Med Res Rev* **1998**, *18*, 259-296.
(112) Kumar, N. Taxol-induced polymerization of purified tubulin. Mechanism of action. *J Biol Chem* **1981**, *256*, 10435-10441.
(113) Sengupta, S.; Boge, T. C.; Liu, Y.; Hepperle, M.; Georg, G. I. et al. Probing the environment of tubulin-bound paclitaxel using fluorescent paclitaxel analogues. *Biochemistry* **1997**, *36*, 5179-5184.
(114) Altmann, K. H.; Gertsch, J. Anticancer drugs from nature--natural products as a unique source of new microtubule-stabilizing agents. *Nat Prod Rep* **2007**, *24*, 327-357.
(115) Sackett, D. L. Podophyllotoxin, steganacin and combretastatin: natural products that bind at the colchicine site of tubulin. *Pharmacol Ther* **1993**, *59*, 163-228.

(116) Verdier-Pinard, P.; Lai, J. Y.; Yoo, H. D.; Yu, J.; Marquez, B. et al. Structure-activity analysis of the interaction of curacin A, the potent colchicine site antimitotic agent, with tubulin and effects of analogs on the growth of MCF-7 breast cancer cells. *Mol Pharmacol* **1998**, *53*, 62-76.
(117) Kruse, L. I.; Ladd, D. L.; Harrsch, P. B.; McCabe, F. L.; Mong, S. M. et al. Synthesis, tubulin binding, antineoplastic evaluation, and structure-activity relationship of oncodazole analogues. *J Med Chem* **1989**, *32*, 409-417.
(118) Mejillano, M. R.; Shivanna, B. D.; Himes, R. H. Studies on the nocodazole-induced GTPase activity of tubulin. *Arch Biochem Biophys* **1996**, *336*, 130-138.
(119) Pettit, G. R.; Singh, S. B.; Hamel, E.; Lin, C. M.; Alberts, D. S. et al. Isolation and structure of the strong cell growth and tubulin inhibitor combretastatin A-4. *Experientia* **1989**, *45*, 209-211.
(120) Simoni, D.; Romagnoli, R.; Baruchello, R.; Rondanin, R.; Rizzi, M. et al. Novel combretastatin analogues endowed with antitumor activity. *J Med Chem* **2006**, *49*, 3143-3152.
(121) Ferguson, R. E.; Jackson, S. M.; Stanley, A. J.; Joyce, A. D.; Harnden, P. et al. Intrinsic chemotherapy resistance to the tubulin-binding antimitotic agents in renal cell carcinoma. *Int J Cancer* **2005**, *115*, 155-163.
(122) Gupton, J. T. Pyrrole natural products with antitumor properties. *Heterocyclic Antitumor Antibiotics: Topics in Heterocyclic Chemistry, Vol.2, Ed. by , Springer Verlag, Berlin/Heidelberg* **2006**, 53-92.
(123) Burnham, B. S., Gupton, J. T., Krumpe, K., Webb, T., Shuford, J., Bowers, B., Warren, A. E., Barnes, C., Hall, I. H. Cytotoxicity of substituted alkyl-3,4-bis(4-methoxyphenyl)pyrrole-2-carboxylates in L1210 lymphoid leukemia cells. *Arch. Pharm.* **1998**, *331*, 337-341.
(124) Gupton, J. T., Burham, B. S., Byrd, B. D., Krumpe, K. E., Stokes, C., Shuford, J., Winkle, S., Webb, T., Warren, A. E., Barnes, C. R., Henry, J., Hall, I. H. The cytotoxicity and mode of action of 2,3,4-trisubstituted pyrroles and related derivatives in human Tmolt4 leukemia cells. *Pharmazie* **1999**, *54*, 691–697.
(125) Yoshida, W. Y., Lee, K. K., Carroll, A. R.,Scheuer, P. J. A complex pyrrolooxazinone and its iodo derravative isolated from a tunicate. *Helv. Chim. Acta* **1992**, *75*, 1271-1275.
(126) Banwell, M. G., Flynn, B. L., Hamel, E., Hockless, D. C. R. Convergent synthesis of the pyrrolic marine natural products lamellarin-O, lamellarin-Q, lukianol-A and some more highly oxygenated congeners. *Chem. Commun.* **1997**, *2*, 207-208.
(127) Charan, R. D., Schlingmann, G., Bernan, V. S., Feng, X., Carter, G. T. Dioxapyrrolomycin biosynthesis in Streptomyces fumanus. *J. Nat. Prod.* **2006**, *69*, 29-33.
(128) Gupton, J. T.; Burham, B. S.; Krumpe, K.; Du, K.; Sikorski, J. A. et al. Synthesis and cytotoxicity of 2,4-disubstituted and 2,3,4-trisubstituted brominated pyrroles in murine and human cultured tumor cells. *Arch Pharm (Weinheim)* **2000**, *333*, 3-9.

(129) Cleaveland, E. S.; Monks, A.; Vaigro-Wolff, A.; Zaharevitz, D. W.; Paull, K. et al. Site of action of two novel pyrimidine biosynthesis inhibitors accurately predicted by the compare program. *Biochem Pharmacol* **1995**, *49*, 947-954.
(130) Bai, R.; Covell, D. G.; Pei, X. F.; Ewell, J. B.; Nguyen, N. Y. et al. Mapping the binding site of colchicinoids on beta -tubulin. 2-Chloroacetyl-2-demethylthiocolchicine covalently reacts predominantly with cysteine 239 and secondarily with cysteine 354. *J Biol Chem* **2000**, *275*, 40443-40452.
(131) Nguyen, T. L.; McGrath, C.; Hermone, A. R.; Burnett, J. C.; Zaharevitz, D. W. et al. A common pharmacophore for a diverse set of colchicine site inhibitors using a structure-based approach. *J Med Chem* **2005**, *48*, 6107-6116.
(132) Ravelli, R. B.; Gigant, B.; Curmi, P. A.; Jourdain, I.; Lachkar, S. et al. Insight into tubulin regulation from a complex with colchicine and a stathmin-like domain. *Nature* **2004**, *428*, 198-202.
(133) Gupton, J. T. K., Keith E.; Burnham, Bruce S.; Webb, Tammy M.; Shuford, Jordan S.; Sikorski, James A. The application of vinylogous iminium salt derivatives to a regiocontrolled and efficient relay synthesis of lukianol A and related marine natural products. *Tetrahedron* **1999**, *55*, 14515-14522.
(134) Gupton, J. T. C., Stuart C.; Miller, Robert B.; Lukens, John R.; Henry, Charlotte A.; Kanters, Rene P. F.; Sikorski, James A. The application of vinylogous iminium salt derivatives to the synthesis of Ningalin B hexamethyl ether. *Tetrahedron* **2003**, *59*, 207-215.
(135) Gupton, J. T. M., Robert B.; Krumpe, Keith E.; Clough, Stuart C.; Banner, Edith J.; Kanters, Rene P. F.; Du, Karen X.; Keertikar, Kartik M.; Lauerman, Nicholas E.; Solano, John M.; Adams, Bret R.; Callahan, Daniel W.; Little, Barrett A.; Scharf, Austin B.; Sikorski, James A. The application of vinylogous iminium salt derivatives to an efficient relay synthesis of the pyrrole containing alkaloids polycitone A and B. *Tetrahedron* **2005**, *61*, 1845-1854.
(136) Gupton, J. T. B., Edith J.; Scharf, Austin B.; Norwood, Bradley K.; Kanters, Rene P. F.; Dominey, Raymond N.; Hempel, Jonathan E.; Kharlamova, Anastasia; Bluhn-Chertudi, Itta; Hickenboth, Charles R.; Little, Barrett A.; Sartin, Melissa D.; Coppock, Matthew B.; Krumpe, Keith E.; Burnham, Bruce S.; Holt, Herman; Du, Karen X.; Keertikar, Kartik M.; Diebes, Anthony; Ghassemi, Shahnaz; Sikorski, James A. The application of vinylogous iminium salt derivatives to an efficient synthesis of the pyrrole containing alkaloids Rigidin and Rigidin E. *Tetrahedron* **2006**, *62*, 8243-8255.
(137) Clark, E. A.; Hills, P. M.; Davidson, B. S.; Wender, P. A.; Mooberry, S. L. Laulimalide and synthetic laulimalide analogues are synergistic with paclitaxel and 2-methoxyestradiol. *Mol Pharm* **2006**, *3*, 457-467.
(138) Rao, P. N.; Cessac, J. W.; Tinley, T. L.; Mooberry, S. L. Synthesis and antimitotic activity of novel 2-methoxyestradiol analogs. *Steroids* **2002**, *67*, 1079-1089.
(139) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* **2000**, *43*, 4759-4767.

(140) Arthur, C. R. G., John T.; Kellogg, Glen E.; Yeudall, W. Andrew; Cabot, Myles C.; Newsham, Irene F.; Gewirtz, David A. Autophagic cell death, polyploidy and senescence induced in breast tumor cells by the substituted pyrrole JG-03-14, a novel microtubule poison. *Biochemical Pharmacology* **2007**, *74*, 981-991.
(141) Burns, J.; Yokota, T.; Ashihara, H.; Lean, M. E.; Crozier, A. Plant foods and herbal sources of resveratrol. *J Agric Food Chem* **2002**, *50*, 3337-3340.
(142) Simoni, D.; Invidiata, F. P.; Eleopra, M.; Marchetti, P.; Rondanin, R. et al. Design, synthesis and biological evaluation of novel stilbene-based antitumor agents. *Bioorg Med Chem* **2009**, *17*, 512-522.
(143) Soleas, G. J.; Diamandis, E. P.; Goldberg, D. M. Resveratrol: a molecule whose time has come? And gone? *Clin Biochem* **1997**, *30*, 91-113.
(144) Zhou, H. B.; Chen, J. J.; Wang, W. X.; Cai, J. T.; Du, Q. Anticancer activity of resveratrol on implanted human primary gastric carcinoma cells in nude mice. *World J Gastroenterol* **2005**, *11*, 280-284.
(145) Pace-Asciak, C. R.; Hahn, S.; Diamandis, E. P.; Soleas, G.; Goldberg, D. M. The red wine phenolics trans-resveratrol and quercetin block human platelet aggregation and eicosanoid synthesis: implications for protection against coronary heart disease. *Clin Chim Acta* **1995**, *235*, 207-219.
(146) Jang, M.; Cai, L.; Udeani, G. O.; Slowing, K. V.; Thomas, C. F. et al. Cancer chemopreventive activity of resveratrol, a natural product derived from grapes. *Science* **1997**, *275*, 218-220.
(147) Le Corre, L.; Chalabi, N.; Delort, L.; Bignon, Y. J.; Bernard-Gallon, D. J. Resveratrol and breast cancer chemoprevention: molecular mechanisms. *Mol Nutr Food Res* **2005**, *49*, 462-471.
(148) Larrosa, M.; Tomas-Barberan, F. A.; Espin, J. C. The grape and wine polyphenol piceatannol is a potent inducer of apoptosis in human SK-Mel-28 melanoma cells. *Eur J Nutr* **2004**, *43*, 275-284.
(149) Kimura, Y. New anticancer agents: in vitro and in vivo evaluation of the antitumor and antimetastatic actions of various compounds isolated from medicinal plants. *In Vivo* **2005**, *19*, 37-60.
(150) Chaudhary, A.; Pandeya, S. N.; Kumar, P.; Sharma, P. P.; Gupta, S. et al. Combretastatin a-4 analogs as anticancer agents. *Mini Rev Med Chem* **2007**, *7*, 1186-1205.
(151) Nam, N. H. Combretastatin A-4 analogues as antimitotic antitumor agents. *Curr Med Chem* **2003**, *10*, 1697-1722.
(152) Tron, G. C.; Pirali, T.; Sorba, G.; Pagliai, F.; Busacca, S. et al. Medicinal chemistry of combretastatin A4: present and future directions. *J Med Chem* **2006**, *49*, 3033-3044.
(153) Bellina, F.; Cauteruccio, S.; Monti, S.; Rossi, R. Novel imidazole-based combretastatin A-4 analogues: evaluation of their in vitro antitumor activity and molecular modeling study of their binding to the colchicine site of tubulin. *Bioorg Med Chem Lett* **2006**, *16*, 5757-5762.

(154) Ji, Y.; Tian, R.; Lin, W. QSAR and molecular docking study of a series of combretastatin analogues tubulin inhibitors. *LNCS* **2007***, 4689*, 436.
(155) Belleri, M.; Ribatti, D.; Nicoli, S.; Cotelli, F.; Forti, L. et al. Antiangiogenic and vascular-targeting activity of the microtubule-destabilizing trans-resveratrol derivative 3,5,4'-trimethoxystilbene. *Mol Pharmacol* **2005***, 67*, 1451-1459.
(156) Cao, T. M.; Durrant, D.; Tripathi, A.; Liu, J.; Tsai, S. et al. Stilbene derivatives that are colchicine-site microtubule inhibitors have antileukemic activity and minimal systemic toxicity. *Am J Hematol* **2008***, 83*, 390-397.
(157) Lippert, J. W., 3rd Vascular disrupting agents. *Bioorg Med Chem* **2007***, 15*, 605-615.
(158) Roberti, M.; Pizzirani, D.; Simoni, D.; Rondanin, R.; Baruchello, R. et al. Synthesis and biological evaluation of resveratrol and analogues as apoptosis-inducing agents. *J Med Chem* **2003***, 46*, 3546-3554.
(159) Durrant, D.; Corwin, F.; Simoni, D.; Zhao, M.; Rudek, M. A. et al. cis-3, 4', 5-Trimethoxy-3'-aminostilbene disrupts tumor vascular perfusion without damaging normal organ perfusion. *Cancer Chemother Pharmacol* **2009***, 63*, 191-200.
(160) Durrant, D. E.; Richards, J.; Tripathi, A.; Kellogg, G. E.; Marchetti, P. et al. Development of water soluble derivatives of cis-3, 4', 5-trimethoxy-3'-aminostilbene for optimization and use in cancer therapy. *Invest New Drugs* **2009***, 27*, 41-52.
(161) Sotriffer, C.; Klebe, G. Identification and mapping of small-molecule binding sites in proteins: computational tools for structure-based drug design. *Farmaco* **2002***, 57*, 243-251.
(162) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N. et al. The Protein Data Bank. *Nucleic Acids Res* **2000***, 28*, 235-242.
(163) Campbell, S. J.; Gold, N. D.; Jackson, R. M.; Westhead, D. R. Ligand binding: functional site location, similarity and docking. *Curr Opin Struct Biol* **2003***, 13*, 389-395.
(164) Pazos, F.; Sternberg, M. J. Automated prediction of protein function and detection of functional sites from structure. *Proc Natl Acad Sci U S A* **2004***, 101*, 14754-14759.
(165) Kleywegt, G. J.; Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr D Biol Crystallogr* **1994***, 50*, 178-185.
(166) Hendlich, M.; Rippmann, F.; Barnickel, G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J Mol Graph Model* **1997***, 15*, 359-363, 389.
(167) Levitt, D. G.; Banaszak, L. J. POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *J Mol Graph* **1992***, 10*, 229-234.
(168) An, J.; Totrov, M.; Abagyan, R. Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol Cell Proteomics* **2005***, 4*, 752-761.

(169) Liang, J.; Edelsbrunner, H.; Woodward, C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci* **1998**, *7*, 1884-1897.
(170) Brady, G. P., Jr.; Stouten, P. F. Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aided Mol Des* **2000**, *14*, 383-401.
(171) Peters, K. P.; Fauck, J.; Frommel, C. The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. *J Mol Biol* **1996**, *256*, 201-213.
(172) Laskowski, R. A. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph* **1995**, *13*, 323-330, 307-328.
(173) Laurie, A. T.; Jackson, R. M. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics* **2005**, *21*, 1908-1916.
(174) Lichtarge, O.; Bourne, H. R.; Cohen, F. E. An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol* **1996**, *257*, 342-358.
(175) Aloy, P.; Querol, E.; Aviles, F. X.; Sternberg, M. J. Automated structure-based prediction of functional sites in proteins: applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking. *J Mol Biol* **2001**, *311*, 395-408.
(176) Stuart, A. C.; Ilyin, V. A.; Sali, A. LigBase: a database of families of aligned ligand binding sites in known protein sequences and structures. *Bioinformatics* **2002**, *18*, 200-201.
(177) Apweiler, R.; Attwood, T. K.; Bairoch, A.; Bateman, A.; Birney, E. et al. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res* **2001**, *29*, 37-40.
(178) Bickel, P. J.; Kechris, K. J.; Spector, P. C.; Wedemayer, G. J.; Glazer, A. N. Inaugural Article: finding important sites in protein sequences. *Proc Natl Acad Sci U S A* **2002**, *99*, 14764-14771.
(179) Armon, A.; Graur, D.; Ben-Tal, N. ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J Mol Biol* **2001**, *307*, 447-463.
(180) Pupko, T.; Bell, R. E.; Mayrose, I.; Glaser, F.; Ben-Tal, N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics* **2002**, *18 Suppl 1*, S71-77.
(181) Lee, B.; Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* **1971**, *55*, 379-400.
(182) Connolly, M. L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983**, *221*, 709-713.
(183) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **1982**, *161*, 269-288.
(184) Meng, E.; Shoichet, B.; Kuntz, I. D. Automated docking with grid based energy evaluation. *J. Comput Chem* **1992**, *13*, 505-524.

(185) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* **1985**, *28*, 849-857.
(186) Miranker, A.; Karplus, M. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* **1991**, *11*, 29-34.
(187) Voorintholt, R.; Kosters, M. T.; Vegter, G.; Vriend, G.; Hol, W. G. A very fast program for visualizing protein surfaces, channels and cavities. *J Mol Graph* **1989**, *7*, 243-245.
(188) Del Carpio, C. A.; Takahashi, Y.; Sasaki, S. A new approach to the automatic identification of candidates for ligand receptor sites in proteins: (I). Search for pocket regions. *J Mol Graph* **1993**, *11*, 23-29, 42.
(189) Kisljuk, O. S.; Kachalova, G. S.; Lanina, N. P. An algorithm to find channels and cavities within protein crystals. *J Mol Graph* **1994**, *12*, 305-307, 296.
(190) Edelsbrunner, H.; Mucke, E. Three-dimensional alpha-shapes. *ACM Trans Graph* **1994**, *13*, 43-72.
(191) Lee, D. T.; J., S. B. Two Algorithms for constructing a Delaunay triangulation. *Int J Comput Inf Sci* **1980**, *9*, 219-242.
(192) Kellogg, G. E.; Fornabaio, M.; Chen, D. L.; Abraham, D. J.; Spyrakis, F. et al. Tools for building a comprehensive modeling system for virtual screening under real biological conditions: The Computational Titration algorithm. *J Mol Graph Model* **2006**, *24*, 434-439.
(193) Chen, D. L.; Kellogg, G. E. A computational tool to optimize ligand selectivity between two similar biomacromolecular targets. *J Comput Aided Mol Des* **2005**, *19*, 69-82.
(194) Selinsky, B. S.; Gupta, K.; Sharkey, C. T.; Loll, P. J. Structural analysis of NSAID binding by prostaglandin H2 synthase: time-dependent and time-independent inhibitors elicit identical enzyme conformations. *Biochemistry* **2001**, *40*, 5172-5180.
(195) Carugo, O.; Bordo, D. How many water molecules can be detected by protein crystallography? *Acta Crystallogr D Biol Crystallogr* **1999**, *55*, 479-483.
(196) Steinbacher, S.; Kaiser, J.; Eisenreich, W.; Huber, R.; Bacher, A. et al. Structural basis of fosmidomycin action revealed by the complex with 2-C-methyl-D-erythritol 4-phosphate synthase (IspC). Implications for the catalytic mechanism and anti-malaria drug development. *J Biol Chem* **2003**, *278*, 18401-18407.
(197) Arkin, M. A.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N. et al. Binding of small molecules to an adaptive protein-protein interface. *Proc.Natl.Acad.Sci.USA* **2003**, *100*, 1603-1608.
(198) Sattler, M.; Liang, H.; Nettesheim, D.; Meadows, R. P.; Harlan, J. E. et al. Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis. *Science* **1997**, *275*, 983-986.
(199) Lee, E. F.; Czabotar, P. E.; Smith, B. J.; Deshayes, K.; Zobel, K. et al. Crystal structure of ABT-737 complexed with Bcl-xL: implications for selectivity of antagonists of the Bcl-2 family. *Cell Death Differ* **2007**, *14*, 1711-1713.

(200) Gigant, B.; Wang, C.; Ravelli, R. B.; Roussi, F.; Steinmetz, M. O. et al. Structural basis for the regulation of tubulin by vinblastine. *Nature* **2005**, *435*, 519-522.
(201) Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T. et al. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **2000**, *407*, 340-348.
(202) Hayward, S. Identification of specific interactions that drive ligand-induced closure in five enzymes with classic domain movements. *J Mol Biol* **2004**, *339*, 1001-1021.
(203) Karpusas, M.; Branchaud, B.; Remington, S. J. Proposed mechanism for the condensation reaction of citrate synthase: 1.9-A structure of the ternary complex with oxaloacetate and carboxymethyl coenzyme A. *Biochemistry* **1990**, *29*, 2213-2219.
(204) Liao, D. I.; Karpusas, M.; Remington, S. J. Crystal structure of an open conformation of citrate synthase from chicken heart at 2.8-A resolution. *Biochemistry* **1991**, *30*, 6031-6036.
(205) Zhou, M.; Morais-Cabral, J. H.; Mann, S.; MacKinnon, R. Potassium channel receptor site for the inactivation gate and quaternary amine inhibitors. *Nature* **2001**, *411*, 657-661.
(206) Oikonomakos, N. G.; Skamnaki, V. T.; Tsitsanou, K. E.; Gavalas, N. G.; Johnson, L. N. A new allosteric site in glycogen phosphorylase b as a target for drug interactions. *Structure* **2000**, *8*, 575-584.
(207) Ajay; Murcko, M. A. Computational methods to predict binding free energy in ligand-receptor complexes. *J Med Chem* **1995**, *38*, 4953-4967.
(208) Bohm, H. J.; Klebe, G. What can we learn from molecular recognition in protein-ligand complexes for the design of new drugs? *Angew Cbem Int Ed* **1996**, *35*, 2588-2614.
(209) Burt S. K., H. C. W., Greer J. Predicting receptor-ligand interactions. *Curr. Opin. Struct. Biol.* **1991**, *1*, 213-218.
(210) Lee C. Calculating binding energies. *Curr. Opin. Struct. Biol.* **1992**, *2*, 217-222.
(211) Murphy, K. P. Predicting binding energetics from structure: looking beyond DeltaG degrees. *Med Res Rev* **1999**, *19*, 333-339.
(212) Kollman, P. Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. *Acc Chem Res* **1996**, *29*, 461-469.
(213) Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chem Int Ed Engl* **2002**, *41*, 2644-2676.
(214) Dill, K. A. Additivity principles in biochemistry. *J Biol Chem* **1997**, *272*, 701-704.
(215) Reddy, M. R.; Krzysztof, A. Free Energy Calculations in Rational Drug Design. *Springer, Kluwer Academic/Plenum Publishers, New York.* **2001**, 317.
(216) Luque, F. J.; Curutchet, C.; Muriedas, J. M.; Bidon-Chanal, A.; Soteras, I. et al. Continuum solvation models: Dissecting the free energy of solvation. *Physical Chemistry Chemical Physics* **2003**, *5*, 3827-3836.

(217) Thomas, S. Free energy calculations. *Computational Biochemistry and Biophysics* **2001**, 169-197.
(218) Reddy, M. R., Erion M. D., Agarwal A. Free energy calculations: Use and limitations in predicting ligand binding affinities. *Rev. Comput. Chem.* **2000**, *16*, 217-307.
(219) Wang W., D. O., Reyes C. M., Kollman, P. A. Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annual Review of Biophysics and Biomolecular Structure* **2001**, *30*, 211-243.
(220) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. Prediction of pH-dependent properties of proteins. *J Mol Biol* **1994**, *238*, 415-436.
(221) Matthews, B. W. Solvent content of protein crystals. *J Mol Biol* **1968**, *33*, 491-497.
(222) Tajkhorshid E., S. S. The dielectric effect of the environment on the pKa of the retinal Schiff base and on the stabilization of the ion pair in bacteriorhodopsin. *Journal of Molecular Structure (Theochem)* **2000**, *501-502*, 297-213.
(223) Williams, R. J. P. The symbiosis of metal ion and protein chemistry. *Pure and Appl. Chem.* **1983**, *55*, 35-46.
(224) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144-1149.
(225) Åqvist, J. Calculation of absolute binding free energies for charged ligands and effects of long range electrostatic interactions. *J. Comp. Chem.* **1996**, *17*, 1587-1597.
(226) Mazur, J.; L., J. R. Distance-dependent dielectric constants and their application to double-helical DNA. *Biopolymers* **2005**, *31*, 1615-1629.
(227) Sternberg, M. J.; Hayes, F. R.; Russell, A. J.; Thomas, P. G.; Fersht, A. R. Prediction of electrostatic effects of engineering of protein charges. *Nature* **1987**, *330*, 86-88.
(228) Feig, M.; Brooks, C. L., 3rd Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr Opin Struct Biol* **2004**, *14*, 217-224.
(229) Mehler, E. L.; Solmajer, T. Electrostatic effects in proteins: comparison of dielectric and charge models. *Protein Eng* **1991**, *4*, 903-910.
(230) Navia, M. A.; Fitzgerald, P. M.; McKeever, B. M.; Leu, C. T.; Heimbach, J. C. et al. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **1989**, *337*, 615-620.
(231) Fornabaio, M.; Cozzini, P.; Mozzarelli, A.; Abraham, D. J.; Kellogg, G. E. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 2. Computational titration and pH effects in molecular models of neuraminidase-inhibitor complexes. *J Med Chem* **2003**, *46*, 4487-4500.
(232) Spyrakis, F.; Fornabaio, M.; Cozzini, P.; Mozzarelli, A.; Abraham, D. J. et al. Computational titration analysis of a multiprotic HIV-1 protease-ligand complex. *J Am Chem Soc* **2004**, *126*, 11764-11765.

(233) Jhoti, H.; Singh, O. M.; Weir, M. P.; Cooke, R.; Murray-Rust, P. et al. X-ray crystallographic studies of a series of penicillin-derived asymmetric inhibitors of HIV-1 protease. *Biochemistry* **1994**, *33*, 8417-8427.
(234) Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu Rev Biophys Biomol Struct* **1998**, *27*, 249-284.
(235) Lam, P. Y.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y. et al. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263*, 380-384.
(236) Ala, P. J.; DeLoskey, R. J.; Huston, E. E.; Jadhav, P. K.; Lam, P. Y. et al. Molecular recognition of cyclic urea HIV-1 protease inhibitors. *J Biol Chem* **1998**, *273*, 12325-12331.
(237) Jadhav, P. K.; Ala, P.; Woerner, F. J.; Chang, C. H.; Garber, S. S. et al. Cyclic urea amides: HIV-1 protease inhibitors with low nanomolar potency against both wild type and protease inhibitor resistant mutants of HIV. *J Med Chem* **1997**, *40*, 181-191.
(238) Hodge, C. N.; Aldrich, P. E.; Bacheler, L. T.; Chang, C. H.; Eyermann, C. J. et al. Improved cyclic urea inhibitors of the HIV-1 protease: synthesis, potency, resistance profile, human pharmacokinetics and X-ray crystal structure of DMP 450. *Chem Biol* **1996**, *3*, 301-314.
(239) Backbro, K.; Lowgren, S.; Osterlund, K.; Atepo, J.; Unge, T. et al. Unexpected binding mode of a cyclic sulfamide HIV-1 protease inhibitor. *J Med Chem* **1997**, *40*, 898-902.
(240) Schaal, W.; Karlsson, A.; Ahlsen, G.; Lindberg, J.; Andersson, H. O. et al. Synthesis and comparative molecular field analysis (CoMFA) of symmetric and nonsymmetric cyclic sulfamide HIV-1 protease inhibitors. *J Med Chem* **2001**, *44*, 155-169.

VITA

Ashutosh Tripathi was born on 23rd February, 1977 in Lucknow, Uttar Pradesh, India. He received his Bachelor of Pharmacy degree from Institute of Engineering and Technology, M.J.P. Rohilkhand University, Bareilly in 2002. He received his Master of Science degree in Cheminformatics from University of Manchester, United Kingdom in 2004. He joined Dr. Glen E. Kellogg's research group in the Department of Medicinal Chemistry & Institute for Structural Biology and Drug Discovery, Virginia Commonwealth University in January 2005. Ashutosh Tripathi was recipient of 2005-06 Graduate and First Professional Student Technology Research Grant, Virginia Commonwealth University, in 2005. He received J. Doyle Smith Award for achieving the greatest distinction in the areas of research, teaching, scholarship and service in the Department of Medicinal Chemistry at Virginia Commonwealth University, in 2008. Ashutosh also received American Association of Pharmaceutical Scientists (AAPS) Graduate Student Symposium Award from the Drug Design and Discovery section sponsored by Bristol-Myers and Squibb, in 2008.