2013

The Development and Applications of the HINT Scoring Function: Exploring Colchicine-Site Anticancer Agents and Tautomerism

Chenxiao Da
Virginia Commonwealth University

THE DEVELOPMENT AND APPLICATIONS OF THE HINT SCORING FUNCTION:

EXPLORING COLCHICINE-SITE ANTICANCER AGENTS AND TAUTOMERISM

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

by

Chenxiao Da
B. Eng.
East China University of Science and Technology, Shanghai, China, 2008

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

Virginia Commonwealth University
Richmond, Virginia

May 2013

长风破浪会有时，直挂云帆济沧海

The time will come for me to ride the wind and split the waves;
I will set the clouds as sails and cross the vast sea

--李白

Li, Bai

（701-762）

Acknowledgement

This work is dedicated to all the people who have been guiding and helping me through my research and my daily life in this foreign country and supporting me no matter what happens.

I would like to gratefully and sincerely thank my PhD advisor, Dr. Glen Kellogg. He gave me lots of freedom regarding to research so that I could generate hypothesis, look for existing studies and plan my own calculations independently. And when I encountered difficulties, he always provided me suggestions based on his invaluable experience and tried to walk me through the problems, no matter whether they were for specific projects, or for confusions about career. I felt very fortunate to join Merck for a summer internship in 2012. I would not be able to go there to understand how molecular modeling was done in industry without his recommendation and permission for me not being in the lab for three whole months.

I would also like to thank Dr. John Gupton, Dr. Martin Safo, Dr. Umesh Desai and Dr. Douglas Sweet, who served in my graduate committee, for their patience in reading this dissertation and for their valuable suggestions. I want to thank Dr. John Gupton and his postdoc Nakul Telang especially. Without their synthesis work, this majority of this disertation would be based on air.

I would like to thank our collaborator Dr. Susan Mooberry. It was her group who measured all the activities of the compounds. It was an honor to work with Dr. Gupton's and Dr. Mooberry's groups. I would also like to thank Dr. Hardik Parikh, Dr. Aurijit Sarkar, Dr. Ashutosh Tripathi, Dr. Vishal Koparde, Meng Zhang and all the colleagues from Dr. Kellogg's group.

I would like to express my special gratitude to my supervisor and mentor Dr. Edward Sherer, my manager Dr. Brad Sherborn and my colleagues at Merck, Rahway, NJ. The internship was really fun and fruitful. Ed taught me the basic quantum mechanics and how to use Gaussian. And it was the first time I was truly thinking about the future as a modeler.

I would like to thank all my friends in the department, especially Akul Mehta and Christina Camara who really helped me a lot. I cherish my friendship with my old roommates Li Wang and Kai Qian. They supported me through my good and bad times. I am also grateful towards Lingqing Zou. She helped me realize my strengths and weaknesses.

I thank my parents, my grandparents and all the other family members including my dog. They are the foundation of my life.

Finally I thank the department and the school for providing the financial support and resources.

Table of Contents

List of Tables

Table

List of Figures

Figure

List of Schemes

Abstract

THE DEVELOPMENT AND APPLICATIONS OF THE HINT SCORING FUNCTION: EXPLORING COLCHICINE-SITE ANTICANCER AGENTS AND TAUTOMERISM

By Chenxiao Da, B.Eng., 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
2013

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

Richmond VA

The overall aim of this work was to apply HINT, an empirical scoring function based on the understanding of hydrophobicity, to analyze and predict the binding affinities and biological activities of colchicine-site anticancer agents. The second, concurrent aim

was to improve the scoring function by incorporating tautomerism within the modeling process. Our belief is that proper evaluation of tautomeric forms for small molecules will improve performance of virtual screening.

The novel pyrrole-based compounds targeting the colchicine site were docked into the receptor using HINT as a rescoring function. Two distinct binding modes dictated by the size and shape of a subpocket were predicted to differentiate the highly active compounds from the weak ones. Of the residues predicted to participate in binding for the active binding mode, Cys241β was revealed to form a weak but critical hydrogen bond with the ligand. A larger collection of colchicine-site agents, biologically tested in the same laboratory including our pyrrole-based compounds were subject to 3D quantitative structure-activity relationship (QSAR) study. Using results on docking the pyrrole compounds as a guide, relative binding poses and QSAR models were built to facilitate ligand design and optimization. A new 3D modeling approach was introduced to visually highlight the unique features of highly active compounds and the commonality of all compounds in the dataset using HINT maps and successfully tested on the colchicine-site agents. These results will provide valuable guidance in the future design and development of new colchicine-site agents. To incorporate tautomerism within HINT, we proposed and developed two workflow approaches: a general search tool using a simple and intuitive algorithm analyzing hydrogen shift patterns to identify and enumerate tautomeric structures, and a database that contains commonly observed

tautomeric structures. The first approach was designed for small-scale docking studies and the second approach was designed for large-scale virtual screening. The tautomer module in HINT will give more accurate modeling results when the compound encountered is able to tautomerize.

CHAPTER 1

Introduction to the HINT (Hydropathic INTeractions) scoring function and related drug discovery applications

1.1 Hydrophobicity and free energy

Hydrophobicity is a commonly observed phenomenon that can be described as the tendency of nonpolar molecules to form aggregates in order to reduce their surface contact with polar molecules. The resulting phase separation from mixing oil with water is one simple example. It is easily understood that polar molecules tend to stay together in terms of classical attraction forces among atoms, such as hydrogen bonds, electrostatic interactions and van der Waals' interactions involving dipoles. The attractions between two nonpolar molecules, however, are not fully attributable to the weak London dispersion forces (also known as van der Waals') stemming from the instantaneous dipoles. Kirkwood noted the additional attractions between nonpolar groups as early as 1954, and other studies concerning the low solubility of nonpolar solutes indicated the same.[1] In fact, the energy that causes the association of nonpolar molecules is partly due to the increase in entropy when the (previously) ordered water molecules are scattered into solution.

The term "hydrophobic bond" was first used by Kauzmann in 1959.[2] Also, Hermann published three papers on hydrophobic bonds to explain how the loss of ordered water molecules causes the entropy increase. Clearly the description of this phenomenon as a "bond" is incorrect, but it is still used because of its pragmatic usefulness.[3-5]

Hydrophobic bonds or hydrophobic interactions are thus used to describe the tendency of two nonpolar atoms to associate together due to both enthalpy and entropy changes in the process. Comprehensive reviews of hydrophobic bonds can be found in Meyer *et al.*[1] and Sarkar and Kellogg.[6]

The change in Gibbs free energy ΔG is the universal determiner of the feasibility of a reaction or a binding process, including the association of nonpolar moieties due to hydrophobic interactions. The association constant K_a of a ligand to a receptor can be directly related to ΔG using

$$\Delta G = -RT \ln K_a$$

The change in Gibbs free energy of a system is calculated as:

$$\Delta G = \Delta H - T\Delta S$$

Where ΔG is the change in free energy; ΔH is the change in enthalpy, which is a measure of the formation and deformation of non-covalent bonds in the binding process; ΔS is the change in entropy, a measure for the change in randomness of the system; and T is the constant temperature.

An actual ligand-receptor binding process is fundamentally more complicated, and may be treated as being composed of the following steps (**Figure 1.1**):

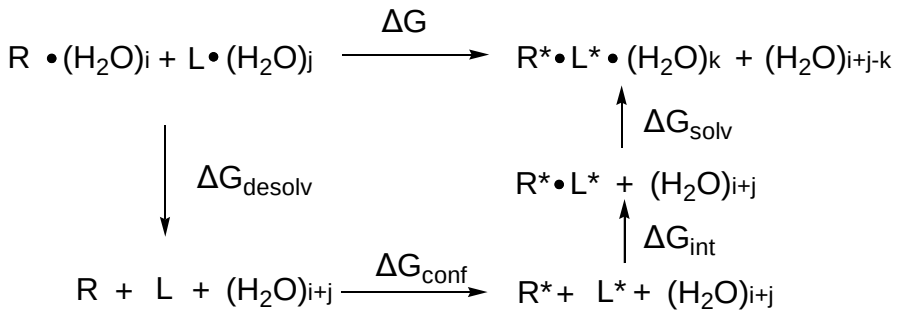


Figure 1.1. The process of a ligand binding to a receptor. R represents the receptor; L represents the ligand; i is the number of water molecules inside the binding pocket of the receptor; j is the number of the water molecules surrounding the ligand; k is the number of the water molecules left in the binding pocket when the ligand is bound. * indicates the conformational change associated with binding. delsov, conf, int and solv indicate the desolvation, conformational change, interaction and solvation respectively.

The total process is the interaction between R (the receptor/protein) and L (the ligand), both of which are solvated, to yield the R*•L* complex, also solvated. ΔG is the total change in Gibbs free energy for this process, which is directly related to the observed association constant K_a. However, computationally modeling this process directly is impossible, so the process is treated by separated steps. First, the ordered water molecules surrounding the ligand and the receptor are stripped off in the desolvation step. Then, the ligand and the receptor go through conformational changes to get ready for binding, because their most stable conformations in solution might be different from their bound conformations. The next step is binding, during which the ligand is anchored by the interactions between itself and the residues inside the binding pocket. The last step considered the possibility that there could be residual water molecules in the binding pocket that bridge between the ligand and the receptor, so these water molecules are placed back in the pocket. ΔG_{desolv}, $\Delta G_{conf,}$ ΔG_{int} and ΔG_{solv} are the free energy changes for these steps respectively.

Therefore, the change in Gibbs free energy of the binding process is

$$\Delta G = \Delta G_{desolv} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{solv}$$

Ideally, a computational function would calculate all these energies to get an accurate prediction of free energy and therefore an accurate association constant.

1.2 The development of HINT (Hydropathic INTeractions) for free energy prediction

HINT (Hydropathic Interactions)[7] is a program developed by Kellogg and Abraham to quantify and visualize hydrophobic and polar interactions and to predict free energy of binding. The basic idea of HINT is that, since the tendency of forming hydrophobic aggregates is directly related to the change in free energy, we can measure the tendency at the atomic level using a force field. Then, we would be able to count not just enthalpy but also entropy in the prediction of free energy compared to the traditional use of only hydrogen bonding, electrostatic interactions and van der Waals' interactions.

The HINT score, given for the interactions between a ligand and a protein or between a protein and another protein, is calculated as,

$$B = \sum \sum b_{ij} = \sum \sum (a_i S_i a_j S_j R_{ij} T_{ij} + r_{ij})$$

where B is the HINT score, b_{ij} represents the interactions between atoms i (from ligand or protein) and j (from the other binding moiety); a_i and a_j are the hydrophobic atom constants for atoms i and j derived from partition coefficient $LogP_{o/w}$, a measure of hydrophobicity; Nonpolar atoms generally have positive constants and polar atoms generally have negative constants; S_i and S_j are the solvent accessible surface areas

(SASAs) for atoms i and j; T_{ij} is a function that discriminates between favorable interactions and unfavorable interactions using either "+1" or "-1" (see **Table 1.1**); R_{ij} is the simple exponential e^{-r} and r_{ij} is an implementation of the Lennard-Jones potential function, which is partly used as a penalty function to identify van der Waals' violations (two atoms get too close). For reference, a higher HINT score indicates more favorable interactions.

Table 1.1. The matrix for the identification of T_{ij}'s value.

Atom Type	**H (apolar)**	**H (polar)**	**C (apolar)**	**Polar (N,O,etc.)**
H (apolar)	$+1^{1}$	-1^{2}	$+1^{1}$	-1^{2}
H (polar)	-1^{2}	-1^{3}	-1^{2}	$+1^{4}$
C (apolar)	$+1^{1}$	-1^{2}	$+1^{1}$	-1^{2}
Polar (N,O,etc.)	-1^{2}	$+1^{4}$	-1^{2}	-1^{5}

The value of T_{ij} is either "+1" or "-1" depending on the atom types of the two atoms under evaluation. T_{ij} is used because the hydrophobic atom constants for polar atoms (negative constants) can not differentiate unfavorable polar interactions from favorable polar interactions (i.e. if multiplying a negative constant with another negative constant obtains only positive values).

[1] hydrophobic-hydrophobic
[2] hydrophobic-polar
[3] acid-acid (two polar hydrogens)
[4] acid-base or hydrogen bond
[5] may depend on charge, but usually base-base and unfavorable (T_{ij} = -1)

How HINT categorizes different types of interactions is explained in **Table 1.1**. Hydrophobic interactions (favorable hydrophobic-hydrophobic interactions), unfavorable

hydrophobic-polar, hydrogen bonding, acid-base, unfavorable acid-acid and base-base are recognized by evaluating the polarities of the atoms involved in the interaction.

The hydrophobic atom constants, *a*, are the key parameters in HINT. Each type of atom is assigned an '*a'* constant from calculations. These constants are derived from Hansch and Leo's hydrophobic fragment constants,[8] which are constructed to predict the partition coefficient $LogP_{o/w}$ values of a variety of organic compounds. HINT breaks down the values of the fragments and assigns them on an atom basis within the molecule according to their connectivity factors, namely neighboring bonding, branching, rings and chains.[6]

While the experimental observation that free energy is proportional to hydrophobic surface area[4] supports the use of solvent accessible surface areas S_i and S_j, the selection of the hydrophobic distance function R_{ij} is open to exploration. We consistently use the simple exponential relationship based on the Leo Polar Proximity Factors[8] and the report stating the hydrophobic interaction is long range and decays exponentially with distance.[9]

The partition coefficient $LogP_{o/w}$ in the 1-octanol/water system is an experimental measure of hydrophobicity. It is in reality an equilibrium constant, so it is directly related to the free energy change (imagine 1-octanol as the receptor and note that 2.303 is for the conversion from Ln to Log):

$$\Delta G = -2.303\ R\ T\ LogP$$

HINT utilizes the thermodynamic information (hydrophobic atom constants) from LogP to predict the interactions between atom pairs. The effects of enthalpy and entropy in

the desolvation, solvation, conformational change and the actual binding steps (**Figure 1.1**) are inherently included in LogP measurement. Therefore, HINT scores are believed to be directly related to free energy changes.

The correlation between HINT score and free energy change was first determined by studying 53 protein-ligand complexes of known crystal structures.[10] The calculated HINT interactions scores were correlated with the free energy changes converted from experimentally measured binding constants K_i or K_d (**Figure 1.2**) and a linear relationship $\Delta G = -0.00195 \times \text{HINT Score} - 5.543$ was obtained with a standard error of ± 2.6 kcal/mol. Better correlations with standard errors approaching ±1.0 kcal/mol can be obtained for data sets reporting the binding of a series of ligands to the same protein. Later correlations were built using 76 protein-ligand complexes (**Figure 1.3**)[11] with resolutions less than 3.2 Å and the standard error was improved to ± 2.4 kcal/mol. If only considering the 56 complexes (**Figure 1.3**) with resolution less than 2.5 Å, the standard error further dropped down to 1.8 kcal/mol.

Further improvement can also be achieved by considering more details during binding, such as pH effects[12] and explicit structural water molecules in the binding site.[13]

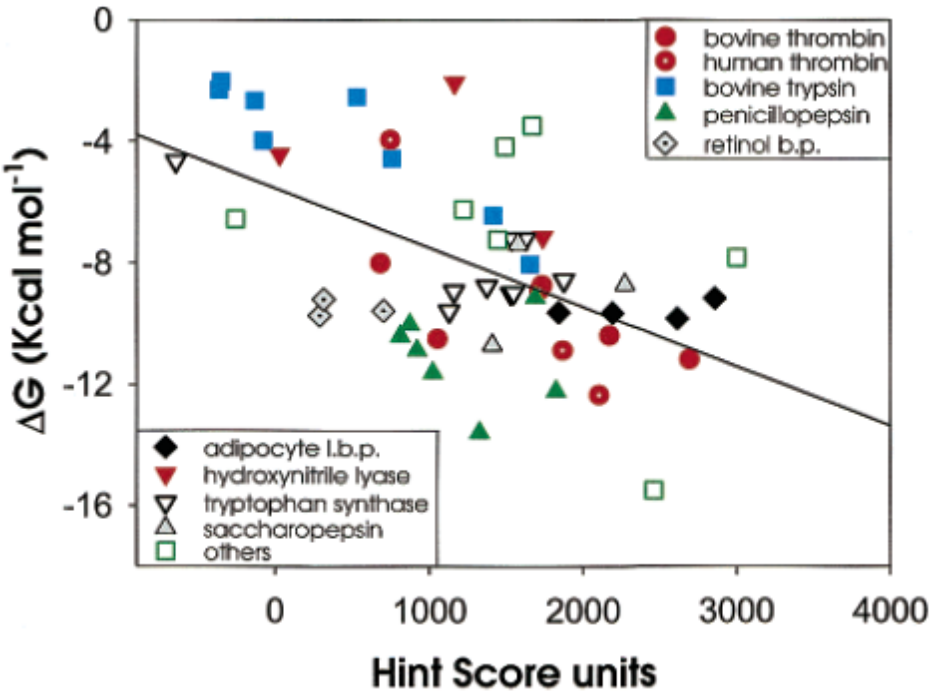


Figure 1.2. Experimental ΔG vs. HINT score units for 53 ligand-protein complexes.[10]

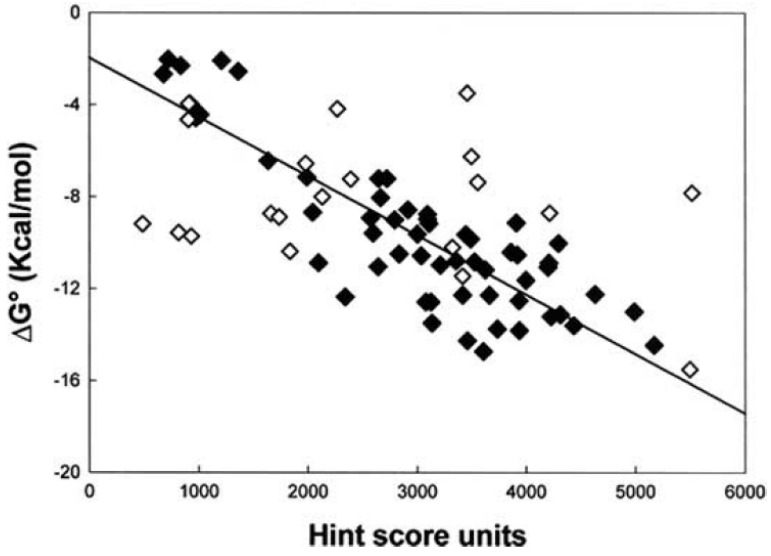


Figure 1.3. Correlation between the experimental free energy of binding and HINT score units. The 56 ligand-protein complexes with resolution less than 2.5 Å are represented by closed diamonds. The 76 ligand-protein complexes with resolutions less than 3.2 Å are represented by the open and closed diamonds.

1.3 HINT applications

1.3.1 Docking and scoring

HINT scoring in docking is the important application of HINT. HINT is designed to evaluate the interactions between two binding moieties, usually a ligand and a protein or two proteins. HINT interaction scores are used to explain this binding and in many cases, the activities of ligands. Another very powerful application is to predict the binding pose of the ligand. How a newly synthesized or designed ligand interacts with its active site or receptor is always of special interest in drug design and development. Being able to visualize the residues of the receptor around the ligand greatly facilitates the optimization of individual interactions and can lead to the design of more active ligands. Generally, the structure of the receptor is known from crystallization or homology modeling. External docking applications, such as GOLD,[14] are used to exhaustively generate possible conformations of the ligand inside the receptor's binding pocket. HINT is then used to calculate the HINT interaction scores for all the conformations. The conformation with the highest HINT score is generally recognized as the binding pose of the ligand. The activity of the ligand or its binding free energy change is then correlated with the highest HINT score.

HINT maps can be created after the HINT interaction scores are calculated. HINT maps provide direct visualization of the interactions. An example is shown in **Figure 1.4**, where hydrophobic interactions are indicated by green contours, favorable polar interactions (acid-base, hydrogen bonding and Coulombic) are indicated by blue contours and unfavorable polar interactions (acid/acid and base/base) by red.

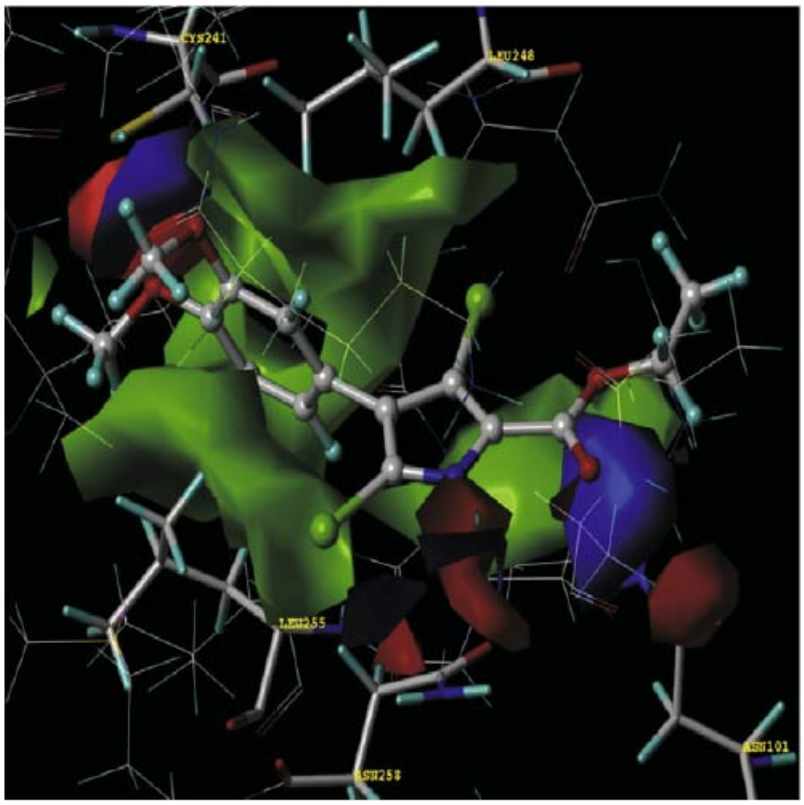


Figure 1.4. An example of a HINT map of a pyrrole analogue binding to the colchicine site on microtubules.

A number of studies have been published using this protocol to determine the binding poses of ligands and to correlate the corresponding HINT interaction scores with activities.[6] The details about on such application of this protocol will be discussed in Chapter 2.

1.3.2 3D-QSAR: HINT-CoMFA

3D-QSAR (Quantitative Structure-Activity Relationship) is an early application of HINT. The concept of QSAR is to create a regression model to relate a set of predictor variables to the response variable (the activity measurement). In 3D-QSAR, the predictor variables are the interactions between a probe atom placed around the ligand

and the ligand itself in a “field” map. As in all forms of QSAR, a large number of ligands need to be collected for 3D-QSAR modeling. Their “interactions” are calculated and the “interactions” that have significant impact on the activities of the ligands are identified by regression analysis. The positions of these important “interactions” indicate the fragments (functional groups) of the ligands that are important for activity (**Figure 1.5**). 3D-QSAR modeling is a ligand-based technique because, in principle, it does not require any information about the structure of the receptor.

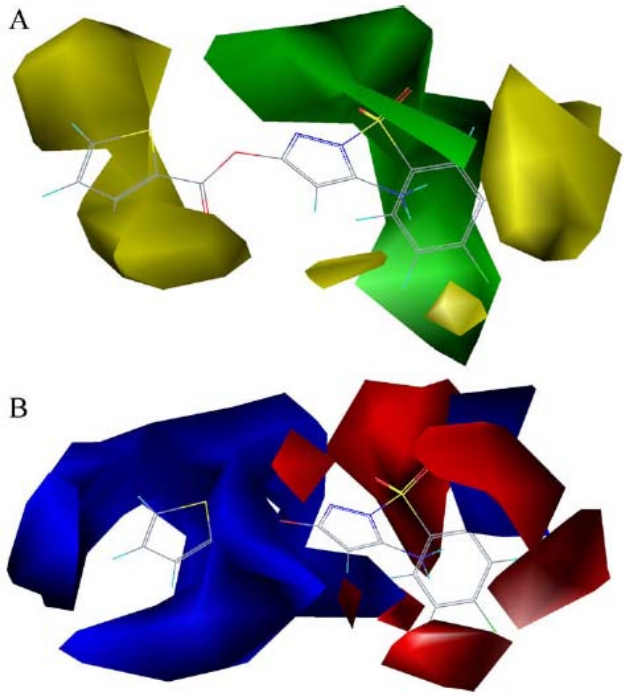


Figure 1.5. An example of 3D-QSAR maps of benzenesulfonyl-pyrazol-ester compounds as cathepsin B inhibitors.[15] Favorable steric regions are in green and unfavorable steric regions are in yellow, and favorable acidic regions are in blue and favorable basic regions are in red.

CoMFA (Comparative Moelcular Field Analysis) is the prototypical 3D-QSAR method, first reported by Cramer and colleagues in 1988, that calculates steric and electrostatic

fields (interactions) surrounding the ligands.[16] In 1991, our group showed that HINT can create a hydrophobic/polar field by measuring HINT interactions between the ligand and a probe atom, using the equation shown above, that provides information in addition to the two fields in CoMFA.[17] Thus these two methods are combined to form the HINT-CoMFA protocol. This combined method has been used in quite a few studies and showed significant statistical improvement compared to using only CoMFA, especially when the ligands and actives were particularly hydrophobic.[6] The use of HINT-CoMFA will be further discussed in Chapter 3.

1.3.3 Computational titration and structural water molecules

It is reasonable to state that improvements of free energy prediction and binding pose prediction can be achieved by considering more details of the binding process, such as the ionization states of functional groups on ligands and residues and water molecules in the binding site.

Computational Titration[11, 18-19] was developed to enumerate and evaluate multiple ionization states for both ligand and receptor in parallel (**Figure 1.6**). For each "state ensemble", the positions of hydrogen atoms involved in hydrogen bonding are optimized. In addition, alternate possibilities considering different ionization states for acidic and basic functional groups are evaluated. The "state ensemble" that receives the highest HINT score is proposed to be the optimum description of binding, which can be used to correlate with the corresponding pK_a and pH.

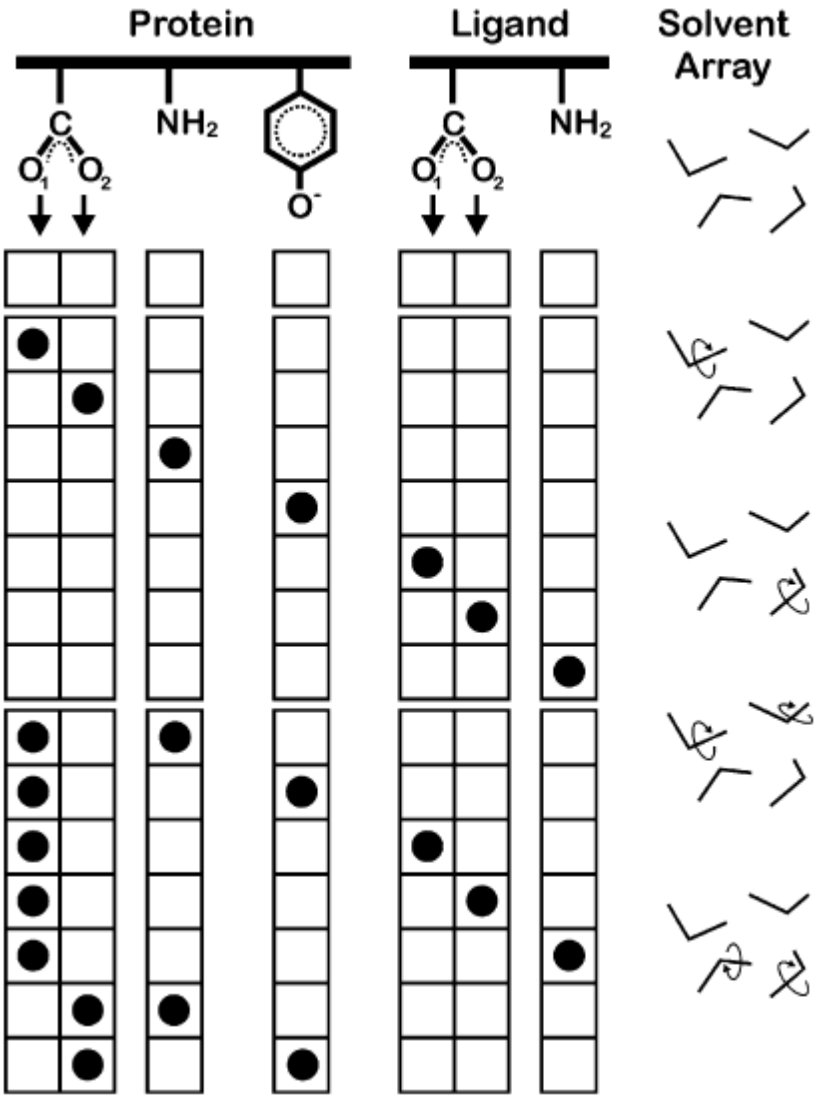


Figure 1.6. Schematic illustration of Computational Titration algorithm.[11] Each filled circle represents a protonation of that functional group.

The roles of water molecules in a binding site have been recognized as bridging the ligand and the receptor to stabilize or sometimes destabilize the binding. Incorporating their contribution in HINT scoring has been reported by our group to improve the correlation between HINT scores and experimentally determined binding constants.[12]

The incorporation is simply done by calculating the HINT scores between the explicit water molecules and the ligand/receptor. The equation is:

$$H_{Total} = H_{protein\text{-}ligand} + H_{ligand\text{-}water} + H_{protein\text{-}water}$$

where H_{Total} is the final total HINT score indicating the binding of the ligand and $H_{ligand\text{-}water}$ and $H_{protein\text{-}water}$ are HINT scores between the ligand and water and the protein and water.

Because water molecules are not well defined in low-resolution crystal structures, it is desirable to have a method to predict where the water molecules are in the binding site. Our group has developed a HINT function that can estimate the interactions between a potential water molecule and its surrounding atoms. Thus, a set of well-placed solvent molecules can be computationally generated for addition to the model.

HINT scoring of water molecules can be combined with the Rank algorithm, which calculates the number and geometric quality of hydrogen bonds for a potential water molecule. A statistical model[20] was built to correlate the existence of a water molecule at a specific position in protein binding sites with its interactions with surrounding atoms evaluated by HINT and Rank. The model was based on a training set of 125 water molecules in the binding sites of 13 proteins determined by crystallography. We then tested the model on an independent set of 68 water molecules in 9 proteins and 87% of them were correctly predicted.

The water molecules at protein-protein interfaces were also studied recently.[21] The understanding of their roles will be helpful for understanding these associations and for the design of protein-protein docking tools.

1.4 Specific Aims of study and overview of chapters

HINT is a powerful tool for molecular modeling based a hydropathic force field derived from experimental determination of $LogP_{o/w}$. The long-term goal of our group's research is to gain knowledge about biological association by probing the mere hydrophobic effect and hydropathic interactions. The goal of this study was to answer the questions of whether: 1) it was possible to explain the anticancer activities of colchicine-site agents using computational molecular modeling techniques built upon HINT, and 2) in the process, can we improve the HINT scoring function and protocols, by considering tautomerism.

To answer these questions, our specific aims were formed: 1) to apply existing HINT modeling tools to understand the binding of pyrrole-based anticancer agents synthesized by Dr. Gupton and co-workers to the colchicine site; 2) to generalize the binding features of a more comprehensive collection of colchicine-site agents including our pyrrole-based compounds and other structurally diverse compounds and 3) to develop new tools that incorporate tautomerism within HINT for the modeling of tautomeric structures. This latter aim is more general in scope.

Chapter 2 describes results obtained for Aim 1. Colchicine-site agents are the new rising stars of cancer treatments.[22] They possess newly discovered anti-vascular activities that can cut off a tumor's nutrients and have the potential to circumvent drug resistance related to the use of taxanes and vinca alkaloids.[23] We have been developing a series of pyrrole-based compounds targeting the colchicine site.[24] Understanding their binding modes was one critical step for the further optimization of

their activities. Our modeling protocol based on the HINT scoring function would be tested and parameterized for colchicine-site ligands. The background of microtubules, colchicine-site agents and the development of our in-house pyrrole-based compounds are introduced in Chapter 2. The molecular modeling protocol of docking and scoring with the HINT scoring function is elaborated. The modeled binding modalities, important features and corresponding residues on the receptor are discussed.

Chapter 3 describes results obtained for Aim 2. After we achieved success in modeling the pyrrole-based compounds, we identified other colchicine-site agents with different scaffolds. We wanted to apply a similar modeling protocol to understand their binding modes and to thus have a comprehensive view of the colchicine-site agents and the binding site. Such model would be valuable for ligand design and optimization. The colchicine-site agents that were included in modeling are shown in Chapter 3. The use of 3D-QSAR modeling with the CoMFA (Comparative Molecular Field Analysis), HINT-CoMFA and CoMSIA (Molecular Similarity Indices in a Comparative Analysis) methods is described. The generalized features of colchicine-site agents from the models and the interacting residues are shown. A new approach based on the HINT maps, and without using regression, is introduced to visualize the important features of a pool of compounds. Its application is demonstrated on the colchicine-site agents.

Chapter 4 describes results obtained for Aim 3. Tautomerism describes a situation in which one compound can have multiple structures different in the positions of their hydrogen bond donors and acceptors, i.e., the positions of hydrogens and double bonds shift within the molecule. Traditionally, tautomerism was not considered in virtual screening and only one structure for one compound was stored in the database and

used for docking. In the process of improving binding predictions, considering multiple tautomeric structures for each compound is one step forward and we have been adopting this idea within HINT. A new workflow plan considering tautomerism is shown in Chapter 4. We developed two approaches for the workflow: a general search tool for small-scale tasks such as docking of single ligands and use of a tautomer database for large-scale virtual screening. The algorithms for tautomer identification and enumeration and the structure of the tautomer database are detailed. The validation and use of the new tools are also presented.

Chapter 5 presents the overall conclusions of the study.

References

1. Meyer, E. E.; Rosenberg, K. J.; Israelachvili, J. Recent Progress in Understanding Hydrophobic Interactions. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15739-15746.
2. Kauzmann, W. Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem.* **1959**, *14*, 1-63.
3. Hermann, R. B. Theory of hydrophobic bonding. I. Solubility of Hydrocarbons in Water, within the Context of the Significant Structure Theory of Liquids. *J. Phys. Chem.* **1971**, *75*, 363-368.
4. Hermann, R. B. Theory of hydrophobic bonding. II. Correlation of Hydrocarbon Solubility in Water with Solvent Cavity Surface Area. *J. Phys. Chem.* **1972**, *76*, 2754-2759.
5. Hermann, R. B. Theory of hydrophobic bonding. III.Method for the Calculation of the Hydrophobic Interaction Based on Liquid State Perturbation Theory and A Simple Liauid Model. *J. Phys. Chem.* **1975**, *79*, 163-169.
6. Sarkar, A.; Kellogg, G. E. Hydrophobicity – Shake Flasks, Protein Folding and Drug Discovery. *Curr. Top. Med. Chem*. **2010**, *10*, 67-83.
7. Kellogg, E. G.; Abraham, D. J. Hydrophobicity: Is LogPo/w More than the Sum of Its Parts? *Eur. J. Med. Chem.* **2000**, *35*, 651-661.

8. Hansch, C.; Leo, A. J. *Substitutent Constants for Correlation Analysis in Chemistry and Biology*. John Wiley and Sons Inc.: NY, **1979**.
9. Israelachvili, J.; Pashley, R. The Hydrophobic Interaction is Long Range, Decaying Exponentially with Distance. *Nature*. **1982**, *300*, 341–342.
10. Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. 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.
11. Kellogg, G. E.; Fornabaio, M.; Chen, D. L.; Abraham, D. J.; Spyrakis, F.; Cozzini, P.; Mozzarelli, A. 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.
12. 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.
13. Fornabaio, M.; Spyrakis F.; Mozzarelli, A.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E. 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.
14. Jones, G; Willett, P; Glen, R. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *J. Mol. Biol.* **1995**, *245*, 43-53.
15. Zhou, Z.; Wang, Y.; Bryant, S. H. Multi-Conformation 3D QSAR Study of Benzenesulfonyl-Pyrazol-Ester Compounds and Their Analogues as Cathepsin B Inhibitors. *J. Mol. Graph Model.* **2011**, *30*, 135-147.
16. Cramer, R. D. III.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959-5967.
17. Kellogg, G. E.; Semus, S. F.; Abraham, D. J. HINT: a New Method of Empirical Hydrophobic Field Calculation for CoMFA. *J. Comput. Aid. Mol. Des.* **1991**, *5*, 545-552.
18. Spyrakis, F.; Fornabaio, M.; Cozzini, P.; Mozzarelli, A.; Abraham, D. J.; Kellogg, G. E. Computational Titration Analysis of a Multiprotic HIV-1 Protease-Ligand Complex. *J. Am. Chem. Soc.* **2004**, *126*, 11764-11765.
19. Bayden, A. S.; Fornabaio, M.; Scarsdale, J. N.; Kellogg, G. E. Web Application for Studying the Free Energy of Binding and Protonation States of Protein-Ligand Complexes Based on HINT. *J. Comput. Aided. Mol. Des.* **2009**, *23*, 621-632.

20. Amadasi, A.; Surface, J. A.; Spyrakis, F.; Cozzini, P.; Mozzarelli, A.; Kellogg, G. E. Robust Classification of "Relevant" Water Molecules in Putative Protein Binding Sites. *J. Med. Chem.* **2008**, *51*, 1063-1067.
21. Ahmed, M. H.; Spyrakis, F.; Cozzini, P.; Tripathi, P. K.; Mozzarelli, A.; Scarsdale, J. N.; Safo, M. A.; Kellogg, G. E. Bound Water at Protein-Protein Interfaces: Partners, Roles and Hydrophobic Bubbles as a Conserved Motif. PLoS One. **2011**, *6*, e24712.
22. Lu, Y.; Chen, J.; Xiao, M.; Li, W.; Miller, D. D. An Overview of Tubulin Inhibitors that Interact with the Colchicine Binding Site. *Pharm. Res.* **2012**, *29*, 2943-2971.
23. Stengel, C.; Newman, S. P.; Leese, M. P.; Potter, B. V.; Reed, M. J.; Purohit, A. Class III Beta-tubulin Expression and in vitro Resistance to Microtubule Targeting Agents. *Br. J. Cancer.* **2010**, *102*, 316-324.
24. Tripathi, A.; Fornabaio, M.; Kellogg, G. E.; Gupton, J. T.; Gewirtz, D. A.; Yeudall, W. A.; Vega, N. E.; Mooberry, S. L.; Docking and Hydrophobic Scoring of Polysubstituted Pyrrole Compounds with Antitubulin Activity. *Bioorg. Med. Chem*. **2008**, *16*, 2235-2242.

CHAPTER 2

Small Molecule Modeling in the Development of Pyrrole-Based Colchicine-site Anticancer Agents

2.1 Colchicine-site agents targeting microtubules to treat cancer

Microtubules are major cytoskeletal components in eukaryotic cells and participate in a variety of cell functions including maintenance of cell shape, intracellular transport, and forming mitotic spindles for segregating chromosomes during mitosis. Microtubules assemble and disassemble by a reversible process called dynamic instability involving discrete α/β tubulin heterodimers.[1] Diverse agents suppress microtubule dynamics; in rapidly dividing cells they induce mitotic arrest and initiate apoptosis.[2] Traditionally, microtubule-targeting agents are classified as microtubule stabilizing or destabilizing agents based on their effects on microtubule polymer mass at high concentrations. A more practical classification in terms of drug design divides them according to their binding sites on tubulin, which are the taxane domain, the vinca domain, the colchicine site and new sites discovered as more structurally diverse agents are developed.[2]

Unlike the taxanes (paclitaxel) and vinca alkaloids (vincristine), neither colchicine (**Figure 2.1**) nor any colchicine site agents have been successful in cancer chemotherapy due to their severe toxicity to normal tissues.[3] However, recent studies of one family of colchicine site agents, i.e., analogues of combretastatin A-4 (**Figure 2.1**), have reported antivascular actions leading to the rapid collapse of tumor vasculature.[4] A number of CA4 analogues are in clinical trials refueling the search for novel colchicine site agents. Emerging drug resistance due to the expression of the βIII-tubulin isotype

has compromised the clinical use of taxanes and vinca alkaloids.[5] Resistance to different types of microtubule targeting agents was recently suggested to be related to their binding sites and that βIII-tubulin mediated drug resistance might be circumvented by colchicine site agents.[6] Natural and synthetic compounds, e.g., podophyllotoxins, arylindoles, sulfonamides, 2-methoxyestradiols and flavonoids, bind within the colchicine site.[7] This structural diversity provides many possibilities for optimization and new scaffold design.

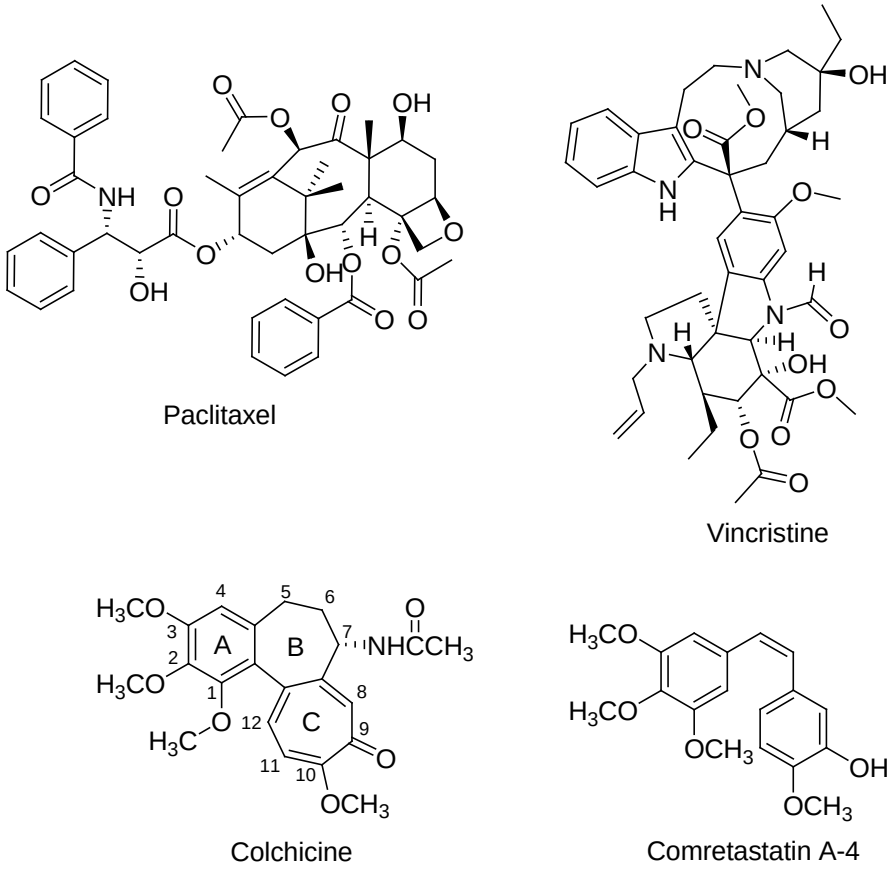


Figure 2.1. The structures of agents that target microtubules.

The colchicine site has been characterized with X-ray crystallography by co-crystallization of the protein with N-deacetyl-N-(2-mercaptoacetyl)-colchicine (DAMA-

colchicine);[8] the site is at the interface between α- and β-tubulin. A simplified mechanism was proposed to explain the destabilizing effect of tubulin on microtubules based on crystallographic analysis.[8-9] The dissembling conformation of microtubules is stabilized by colchicine binding so that the natural switch to the assembling conformation is blocked (**Figure 2.2**).

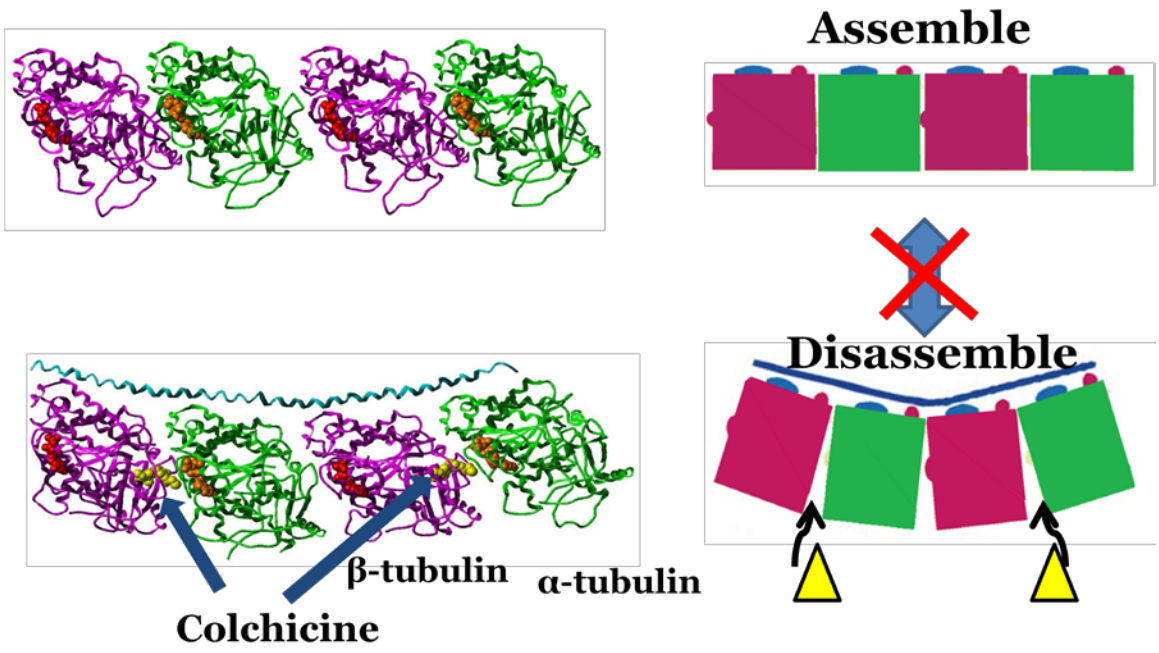


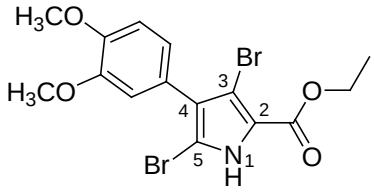
Figure 2.2. The simplified mechanism of chochicine destabilizing microtubules. The crystal structures of microtubule unbound (top) and bound with colchicine (bottom) are shown with schematic representations.

2.2 Identification of a pyrrole-based lead and previous studies

JG-03-14 (3,5-dibromo-4-(3,4-dimethoxyphenyl)-1H-pyrrole-2-carboxylic acid ethyl ester, **Figure 2.3**) was synthesized along with a pool of brominated pyrrole compounds and it was found that the compound demonstrated potent antiproliferative activity against a wide range of cancer cell lines, strong microtubule-destabilizing activity and is

a poor substrate of the multidrug-resistant P-glycoprotein pump that effluxes taxanes and vinca alkaloids.[10] Further studies showed that the compound disrupts multiple endothelial cell functions suggesting the potential for vascular-disrupting activities.[11] Since JG-03-14 inhibited the binding of [^{3}H]colchicine,[10] and a COMPARE analysis, which evaluates the similarity between two compounds with respect to the NCI 60-cell line assay,[12] showed correlation between JG-03-14 and colchicine, it is highly likely that the compound also binds at the colchicine site.

A previous study conducted in our group modeled JG-03-14 and a diverse set of analogues.[13] A quantitative linear QSAR relationship between a free energy like quantity based on IC_{50} and HINT score was obtained. The HINT score, which considers hydrophobic and polar interactions as well as entropic effects, as described in Chapter 1, has been shown to correlate with binding free energy for small molecule-biomacromolecular complexes.[14] Most importantly, a binding pose of JG-03-14 was proposed in the study. In this respect, JG-03-14 has become a valuable lead candidate and the five atoms on its pyrrole scaffold can be easily modified for structural-activity relationships (SAR) according to the proposed binding pose, providing a basis for future optimization and development.



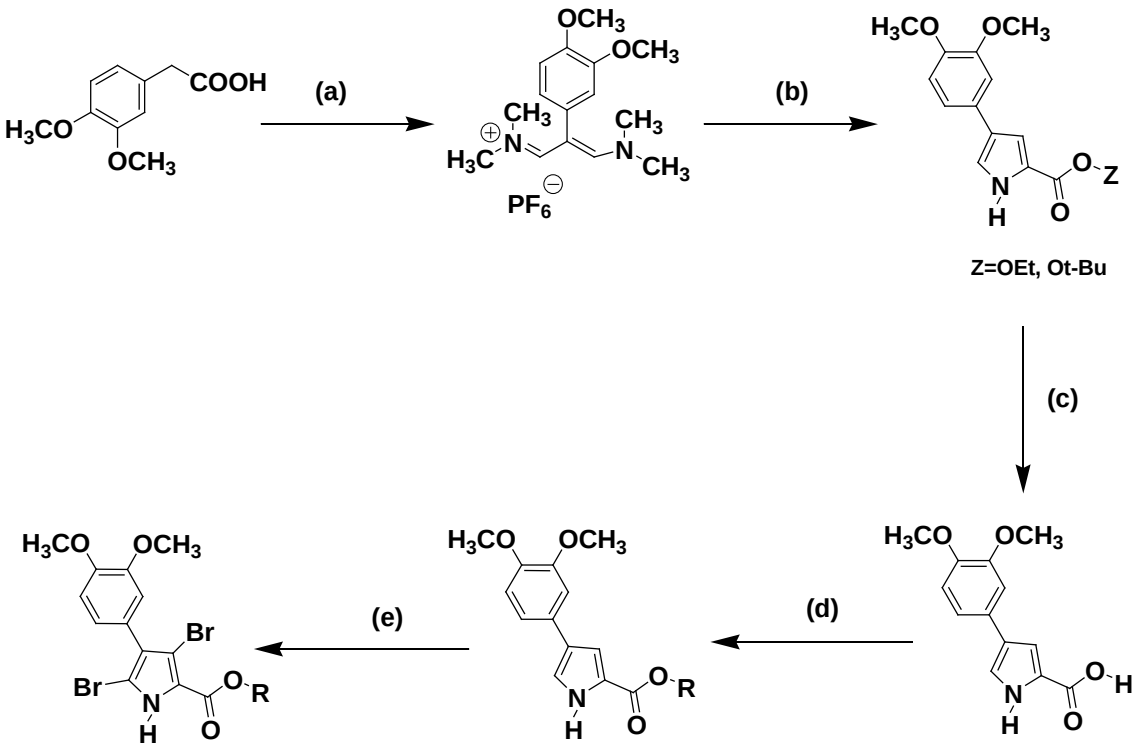
JG-03-14

Figure 2.3. The structure of JG-03-14.

2.3 Improvement to the previous binding model by studying the C-2 compounds

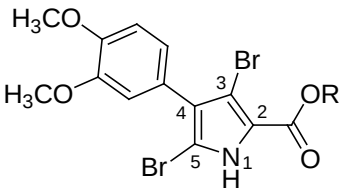
2.3.1 Introduction

In this study, we retain the 3,4-dimethoxylphenyl at C-4 and the two bromine groups at C-3 and C-5 of JG-03-14 and focus on modifications to the ester at the C-2 position of the pyrrole core. Our collaborators Gupton and co-workers have previously reported the synthesis of JG-03-14[15] and have utilized a similar sequence of reactions as outlined in **Scheme 2.1** to prepare the new analogues listed in **Table 2.1**.



Scheme 1. Preparation of JG-03-14 analogues with dodifications at the 2 position. (a) $POCl_3$, DMF and Heat, followed by $H_2O/NaPF_6$ (b) Glycine ethyl ester or glycine t-butyl ester and NaOt-Bu, DMF and Heat (c) NaOH,EtOH/H_2O and Heat (d) ROH, 1,1'-carbonyldiimidazole, DBU and DMF (e) Dibromodimethylhydantoin, $CHCl_3$ and Heat.

Table 2.1. Structures, biological activity and properties of C-2 pyrrole compounds.



Cmpd	R	Antiproliferation[a] IC_{50} (μM)	Cellular microtubule loss[b]	Binding Mode	HINT score[c]	HINT logP	ALOGPs[d]
Colchicine	-	0.014	100% loss at 0.5 μM	-	549	3.24	1.59
JG-03-14 (2.1)	Ethyl	0.036[e]	100% loss at 0.5 μM	I	418	2.60	4.44
2.1a	methyl	0.168	50% loss at 5 μM	I	524	2.06	3.87
2.1b	*n*-propyl	<0.050	75% loss at 5 μM	I	157	3.14	4.74
2.1c	*i*-propyl	0.108	70% loss at 5 μM	I	-179	3.14	4.70
2.1d	*t*-butyl	2.0	No loss up to 10 μM	II	187	3.24	5.02
2.1e	*n*-butyl	1.3	15% loss at 10 μM	II	530	3.68	5.05
2.1f	*n*-hexyl	3.5	35% loss at 10 μM	II	256	4.76	5.83
2.1g	benzyl	5.3	No loss up to 10 μM	II	713	3.61	5.39
2.1h	$-(CH_2)_3NMe_2$	4.6	10% loss at 10 μM	II	293	2.52	4.10
2.1i	$-(CH_2)_2NMe_2$	5.2	10% loss at 10 μM	II	358	2.57	3.82
2.1j	$-(CH_2)_3NMe_2H^+Cl^-$	8.0	No loss up to 10 μM	II	631	0.27	0.39
2.1k	$-(CH_2)_2NMe_2H^+Cl^-$	10.7	No loss up to 10 μM	II	774	0.78	0.27
2.1l	4-methoxylphenyl	18.3	No loss up to 10 μM	II	957	4.37	5.48

[a]Experiments were performed using human MDA-MB-435 cancer cells; [b]Loss of interphase microtubules was evaluated in A-10 cells; [c]515 HINT score units≈ 1 kcal mol -1 (Ref 16); [d]ALOGPs was calculated at Virtual Computational Chemistry Laboratory, http://www.vcclab.org; [e]Ref 10.

Antiproliferative activities were measured by our collaborators Mooberry *et al.* in MDA-MB-435 cancer cells using the sulforhodamine B assay and effects on cellular microtubules were evaluated in A-10 cells using immunofluorescence as previously described.[10] Results are presented in **Table 2.1**.

2.3.2 Materials and methods

The X-ray crystal structure of αβ-tubulin complexed with DAMA-colchicine (pdbid: 1SA0) was prepared with Sybyl 8.1.[16] The stathmin-like domain, the C and D subunits, were deleted. Hydrogen atoms were added and their orientations were optimized by the Tripos force field to a gradient of 0.005 kcal mol^{-1} $Å^{-1}$. The docking studies were performed using GOLD 5.0.[17] The ligands were docked in the active site, which was defined by the space in a 6 Å radius around DAMA-colchicine. Docking conformations generated with GOLD and filtered initially by GoldScore were further analyzed with HINT. The most active ligand JG-03-14 (**2.1**) was docked first with GOLD without constraints. The resulting conformations were rescored with HINT and the best docking pose of **2.1** was defined as its binding mode. This pose was then used to define a similarity constraint in GOLD such that other ligands (**2.1a-2.1l**) were docked in the way that best matches this shape. For these ligands, the conformations/binding modes were chosen based on the highest calculated HINT score.

2.3.3 Results and discussion

For this study, the SAR is analyzed with respect to the antiproliferative activities of compounds **2.1** (JG-03-14) and **2.1a–l**. Antitubulin activity generally trends with antiproliferative activity. **2.1** remains the most active compound (36 nM). Compared to **2.1**, **2.1a** had a 4-fold decrease in activity likely due to its one-carbon shorter ester. Similarly, the longer and bulkier alkyl substitutions *n*-propyl (**2.1b**) and *i*-propyl (**2.1c**) decreased antiproliferative activity. Larger groups, *t*-butyl (**2.1d**), *n*-butyl (**2.1e**) or *n*-hexyl (**2.1f**), were tolerated but with a significant activity loss of more than 30-fold. A dramatic loss was also observed for aromatic substitutions (**2.1g**, **2.1l**). The

incorporation of a comparatively polar amine did not increase the activity significantly (**2.1h**–**k**), suggesting that activity drop is related to sterics, and not to solubility.

The observation that the protonated amines (**2.1j**, **2.1k**) had a further 2-fold drop in activity compared to their free base analogues (**2.1h**, **2.1i**) may be due to their weaker ability to penetrate the cell membrane. Moreover, no microtubule effects were observed up to 10 μM for the amine derivatives, suggesting that a different mechanism of action of antiproliferation might be at play. The SAR suggests that only the properly sized group would be favorable for activity and the ethyl group of **2.1** provides that optimum.

To rationalize the SAR from a structure-based perspective, we performed docking studies with the X-ray crystal structure of DAMA-colchicine/tubulin co-crystal (pdbid: 1sa0).[8] It should be noted that the resolution of the 1sa0 structure for αβ–tubulin is poor (3.58 Å) and resulting modeling studies have a higher degree of uncertainty than in other systems. The colchicine site is mostly buried in the β-subunit surrounded by helices H7 and H8, loop T7, and strands S8 and S9. The T5 loop of the α-subunit also contributes to the pocket (see **Figure 2.4**). DAMA-colchicine occupies the pocket such that ring A fits deep within a subpocket close to H7, ring C fits into another subpocket close to T5, ring B is centered within the main pocket and the DAMA chain is pointing to the pocket‘s entrance. For convenience, we will refer to the subpockets where rings A and C bind as subpockets A and C. The compound **2.1** and its analogues **2.1a–l** were docked to the colchicine site using GOLD[17] and re-scored using HINT. The compounds can be divided into two sets based on their computationally predicted binding modes (**Figure 2.4**). In both modes, the dimethoxyphenyl ring locates in the subpocket A, overlapping the trimethoxyphenyl ring (ring A) of DAMA-colchicine. The positions of the

C-2 ester chain differ between the two modes. In mode I, the R group of the ester has "acceptable" size (i.e., **2.1**, **2.1a–c**), and fits within subpocket C and thus overlaps well with ring C of DAMA-colchicine, while in mode II, the entire molecule is reoriented with the bulkier **2.1d**–**l** R groups extending out from the main pocket towards its opening.

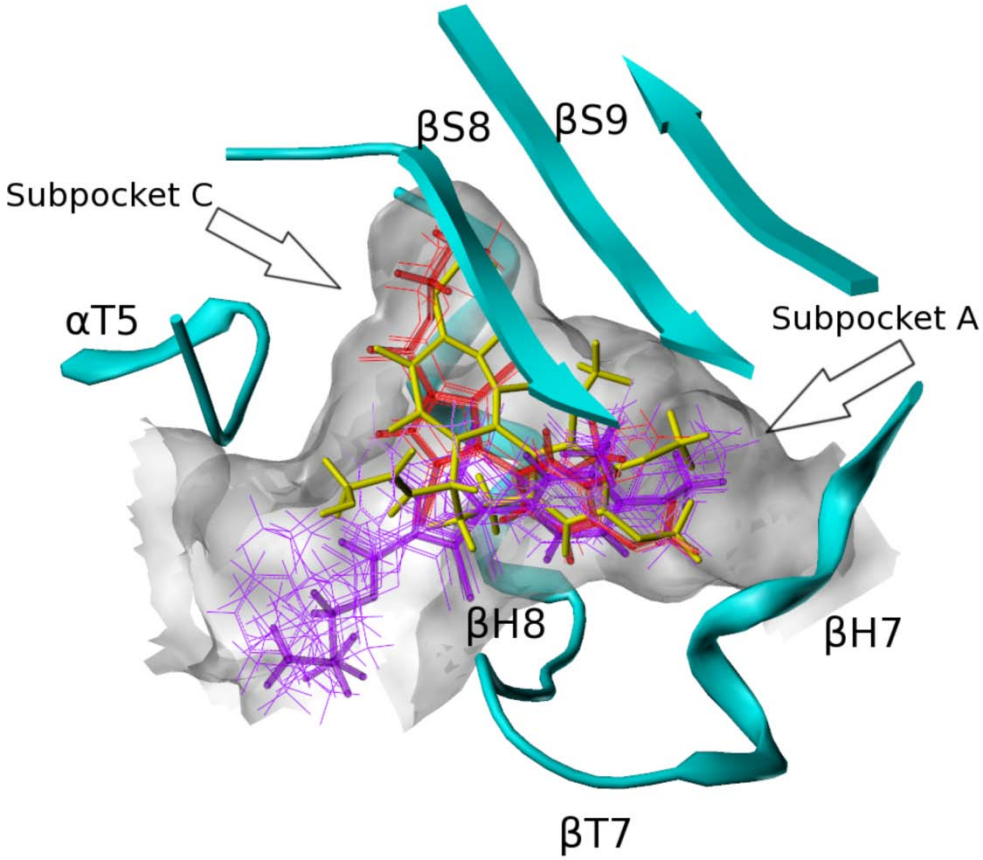


Figure 2.4. Colchicine (yellow) and binding modes of pyrrole-based C-2 analogues (mode I: red; mode II: purple). The extents of the colchicine site, as illustrated by MOLCAD, are shown in white.

To illustrate the specific interactions between the ligands and site, we calculated intermolecular HINT interaction maps[18-19] using **2.1** (**Figure 2.5A**) as representative of

mode I and **2.1e** of mode II (**Figure 2.5B**). First, subpocket A, which fits the dimethoxyphenyl ring in both modes, is quite hydrophobic. In both modes, the four-carbon side chains of Leu248β and Leu255β clamp the phenyl ring in place while deeper in the pocket, other residues lock the ligand's methoxys. Polar interactions also play a part, as Cys241β is in proximity to these two methoxys, with distances between the cysteine's sulfur and the oxygens of 3.06 Å and 3.45 Å, thus likely forming at least one hydrogen bond to support the binding. Also in both modes, there is a favorable interaction in the main pocket between the backbone oxygen of Asn258β and the ligand's pyrrole nitrogen.

Both hydrophobic and polar residues characterize subpocket C, which fits the esters in mode I binding. The alkyl ends reach the hydrophobic bottom, while the carboxyl oxygens anchor the ester by forming hydrogen bonds with the backbone nitrogen of Val181α. The main pocket includes its funnel opening and is much more spacious than subpocket C. It easily tolerates the size of the longer esters binding with mode II by flipping the pyrrole core – thus exposing the ester tail to the solvent while keeping the dimethoxylphenyl ring in subpocket A. Our models suggest that a new hydrogen bond, stabilizing the ester tail in mode II, is formed between the amide nitrogen of Asn101α and the ligand's carbonyl oxygen. The interactions for the various R groups of **2.1d**–**l** are poorly defined as the pocket entrance broadens and has a large solvent exposure.

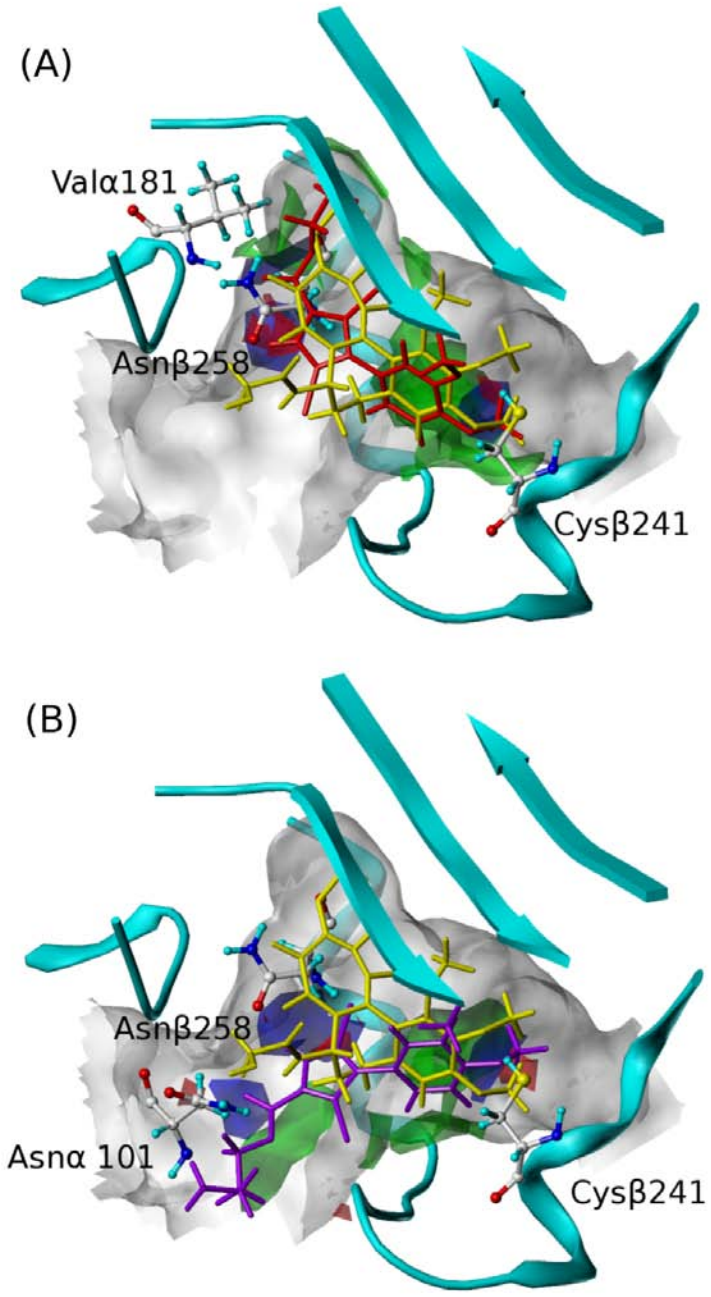


Figure 2.5. HINT interaction maps of (A) **2.1** (binding mode I) and (B) **2.1e** (binding mode II). Green contours represent favorable hydrophobic interactions; blue contours represent favorable polar interactions (hydrogen bonds, acid/base, Coulombic); red contours represent unfavorable polar interactions. **2.1** is shown in red, **2.1e** in purple and colchicine in yellow.

The compounds in mode I displayed notably higher antiproliferative activity and antitubulin activity than the compounds in mode II. It is clearly important to effectively occupy both subpockets A and C in the colchicine site. The SAR within the mode I set is size related: the methyl of **2.1a**, *n*-propyl of **2.1b** and the *i*-propyl of **2.1c** may not position the ester carbonyl (hydrogen bonded to Val181β) as well as the ethyl of **2.1**. In contrast, in the mode II set, the ester R extends from the pocket into (and possibly out of) the pocket's entrance. The SAR for these analogues simply may not be interpretable as these tails are highly flexible and thus subject to interactions with a wide array of residues as well as solvent.

It is also instructive to compare, in detail, the binding of colchicine and the pyrrole-based compounds **2.1** and **2.1a–l**: 1) depletion of ring B of colchicine retains activity, while rings A and C, which adopt a similar conformation as in mode I, are necessary for high affinity binding;[20] and 2) residues Cys241β (subpocket A) and Val181α (subpocket C) appear to be important for antitubulin activity since the removal of any A ring methoxy group close to Cys241β weakens the binding to tubulin and microtubule inhibition.[21] In the next section of this chapter we will explore the SAR of subpocket A. Also, isocolchicine, whose structural difference to colchicine is in the C ring (methoxy at C-9 and keto at C-10) binds weakly and only poorly inhibits microtubule assembly,[22] probably because of a loss of hydrogen bonding to Val181α. Both residues anchor the ligand in the more active mode I, while only Cys241β does so in the less active mode II. This may largely explain the difference in activity between the binding modes.

The HINT scores of **Table 2.1** were poor in distinguishing between binding in mode I and II. The reasons for this failure are instructive. First, the poor resolution of the tubulin

crystal structure and the flexibility of the pocket, especially the T5 and T7 loops, are a partial explanation. However, the binding modes themselves and the nature of the pocket are larger factors. **Table 2.1** lists the HINT scores in terms of two fragments – the common dimethylphenyl plus pyrrole (ring) and the ester. Interaction types further differentiate the latter. The total ring score is largely invariant (580±70), excluding **2.1b** and **2.1c**, where it is lower by > 200. The ester's H_{HH} for mode I (750±130) is much higher than for mode II (280±90). Interestingly, H_{HH} is highest for **2.1b** and **2.1c**, but accommodation of these longer esters was penalized by poorer ring interactions. For **2.1** and **2.1a-c**, hydrophobic binding in subpocket C is the key. Although the esters of mode II compounds appear to make productive contacts, these are in the very open funnel-like entrance of the pocket where dynamic solvent effects that can disrupt polar interactions must be assumed.

Table 2.2. HINT scores by fragment and interaction type.

Cmpd	Mode	HINT score[a]				
		Ring	Ester			
		H_{TOTAL}	$H_{HB} + H_{AB}$	H_{HH}	$H_{AA} + H_{BB}$	H_{HP}
2.1 (JG-03-14)	I	492	858	711	-499	-1328
2.1a	I	518	680	583	-427	-1048
2.1b	I	363	1069	855	-593	-1681
2.1c	I	321	766	833	-478	-1821
2.1d	II	570	778	226	-398	-1169
2.1e	II	626	781	225	-364	-1015
2.1f	II	435	665	387	-369	-1046
2.1g	II	589	812	284	-341	-868
2.1h	II	557	549	268	-336	-1071
2.1i	II	581	726	237	-400	-1027
2.1j	II	650	375	213	-236	-634
2.1k	II	596	847	196	-403	-755
2.1l	II	587	827	494	-240	-924

[a]Interaction types: favorable polar (hydrogen-bond, H_{HB}, and acid/base, H_{AB}), hydrophobic (H_{HH}), unfavorable polar (acid/acid, H_{AA}, and base/base, H_{AB}) and unfavorable hydrophobic-polar (H_{HP}).

2.3.4. Summary: SAR of C-2 analogues

In summary, mode I is a new binding motif observed for pyrrole compounds based on JG-03-14 (**2.1**) that is different from the previously reported binding mode,[13] which was actually mode II in this study (**Figure 2.6**). The ester chain in mode I overlaps with the C-10 substituents of colchicine and the SAR of colchicine C-10 analogues also shows that increasing length of the alkyl chain causes a concomitant decrease in activity.[23] We propose that the deeper burial of mode I ligands is more disruptive to the association of α- and β-tubulin subunits than is binding with mode II. The results of the study have been published.[24] We are continuing design and development of additional JG-03-14 (**2.1**) analogues by focusing on other positions of the pyrrole core as we attempt to gain a full view of the SAR.

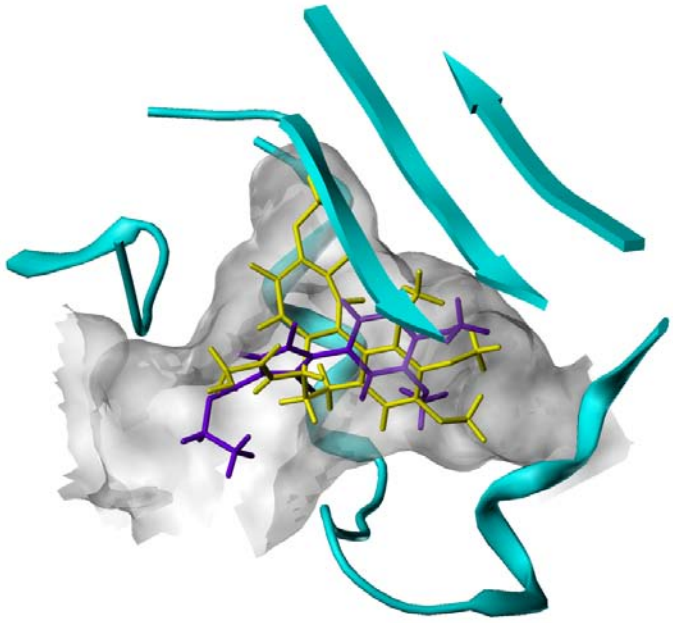


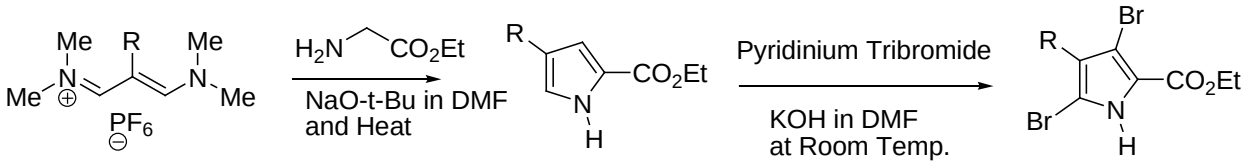
Figure 2.6. The previously identified binding mode of JG-03-14 (purple). Colchicine is shown in yellow.

2.4 A weak but critical hydrogen bond identified by studying the C-4 compounds

2.4.1 Introduction

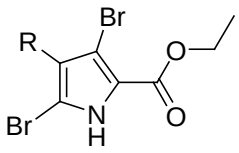
In this study, we retained the two bromine groups at C-3 and C-5 and the ethyl ester at the C-2 position and focused on modifications to the 3,4-dimethoxylphenyl ring at the C-4 position. Previously in this chapter and in publication,[24] we showed that JG-03-14's ethyl ester at C-2 is an ideally suited substituent for that position and this induces the 3,4-dimethoxylphenyl moiety of the compound to overlap with ring A of colchicine in the colchicine site and bind in a subpocket formed mainly by hydrophobic residues and one polar residue, Cysβ241. Here, we explore the electronic, hydrogen bonding and hydrophobic characteristics of substituents at C-4 to enrich our understanding of the SAR of these compounds.

Gupton *et al.* have previously reported the synthesis of JG-03-14 (Compound **2.1**)[15] and have utilized a similar strategy (**Scheme 2.2**) to prepare analogues **2.2a-2.2i** (**Table 2.3**).



Scheme 2.2. Preparation of pyrrole C-4 analogues.

Table 2.3. Structures, antiproliferative and microtubule inhibitory activities of pyrrole C-4 compounds.



Cmpd	R	Antiproliferation IC_{50} (μM)	Microtubule inhibition EC_{50} (μM)	pEC_{50}	HINT score
Colchicine	—	0.016 ± 0.002	0.030	7.52	549
2.1 (JG-03-14)	3,4-dimethoxylphenyl	0.036 ± 0.002	0.490	6.31	643
2.2a	Phenyl	10.3 ± 1.3	> 75	3.82[b]	170
2.2b	4-methylphenyl	2.24 ± 0.2	> 75	3.82[b]	579
2.2c	4-chlorophenyl	0.919 ± 0.020	> 75	3.82[b]	754
2.2d	4-bromophenyl	0.312 ± 0.020	~ 94[a]	4.03	815
2.2e	4-methoxylphenyl	0.843 + 0.090	7.0	5.15	563
2.2f	3-methoxylphenyl	0.633 ± 0.01	2.4	5.62	558
2.2g	3,4,5-trimethoxylphenyl	12.9 ± 1.9	> 75	3.82[b]	124
2.2h	1-napthyl	3.24 ± 0.20	7.0	5.15	805
2.2i	3-indolyl	1.98 ± 0.20	17.8	4.75	271
2.2j	4-trifluoromethoxylphenyl	1.70 ± 0.10	27.1	4.57	649
2.2k	4-thiomethylphenyl	0.626 ± 0.020	18.5	4.73	541
2.2l	3,4-dichlorophenyl	0.806 ± 0.060	9.9	5.00	1012
2.2m	3-fluoro-4-methoxylphenyl	0.539 ± 0.040	14.1	4.85	567
2.2n	6-ethoxyl-2-napthyl	1.99 ± 0.20	> 75	3.82[b]	577
2.2o	1,3-benzodioxol-6-yl	1.80 ± 0.20	29.7	4.53	428
2.2p	1,4-benzodioxan-6-yl	4.36 ± 0.3	20.9	4.68	590
2.2q	2-bromo-4,5-dimethoxylphenyl	2.64 ± 0.30	14.0	4.85	781

[a] 40% microtubule loss at 75 μM, EC_{50} ~ 75 / (2 × 0.4) = 94 μM. [b]Assumed EC_{50} = 150 μM.

Antiproliferative activities were measured by Mooberry *et al.* in MDA-MB-435 cancer cells using the sulforhodamine B assay and effects on cellular microtubules were evaluated in A-10 cells using immunofluorescence as previously described.[10] Results are presented in **Table 2.3**.

2.4.2 Materials and methods

Sybyl 8.1[16] was used to prepare the X-ray crystal structure models of αβ-tubulin complexed with different ligands (pdbid: 1sa0, 1sa1, 3hkc, 3hkd and 3hke). For each structure, the procedure was the same as previously reported. The stathmin-like domain, the C and D subunits, were deleted. Hydrogen atoms were added and their orientations were optimized by the Tripos force field to a gradient of 0.005 kcal mol^{-1} $Å^{-1}$. GOLD 5.1[17] was used for docking studies. The ligands were docked in the active site, which was defined by the space in a 6 Å radius around the complexed small molecule. One hundred GA runs generated one hundred docking conformations for each ligand with GOLD and filtered initially by GoldScore. They were further analyzed with HINT. First, the ligands were docked to all five tubulin structures with GOLD without constraints. The resulting conformations were rescored with HINT and the best docking poses were indicated by the highest HINT score. Next, the differences between ligand binding to the five receptors for each ligand were checked and it was found that binding with 3hkc generally gave higher scores. The resulting 3hkc-ligand complexes were further minimized with the Tripos force field and rescored again. The minimized conformation and the new HINT score were defined as the docking conformation/binding mode and binding score of the ligand.

2.4.3 Results and discussion

The structural-activity relationship is discussed first. All structural modifications for this study were at the C-4 position of the pyrrole core. Antiproliferative activities as well as microtubule depolymerizing activities were measured (**Table 2.3**). Compounds **2.2a**-**2.2d** showed very weak or barely any effect on microtubule polymerization with EC_{50} values of 75 μM or higher. Compound **2.2a**, the unsubstituted ring analog, showed negligible antiproliferative activity, while this activity for **2.2b**-**2.2d** (especially **2.2c** with an IC_{50} of 0.919 μM) likely indicates a different mechanism of action, although some form of microtubule inhibition may still be responsible. For the rest of the compounds, **2.1** and **2.2e**-**2.2q** (and colchicine), the microtubule inhibitory activity correlates well with the antiproliferative activity ($pEC_{50} = 1.10 \times pIC_{50} - 1.57$, $r^2 = 0.79$, **Fig. 2.7**). Interestingly, for these compounds, $pEC_{50} - pIC_{50} = 1.00 \pm 0.43$ μM, which indicates that microtubule inhibition is consistently one order of magnitude weaker than overall inhibition of proliferation.

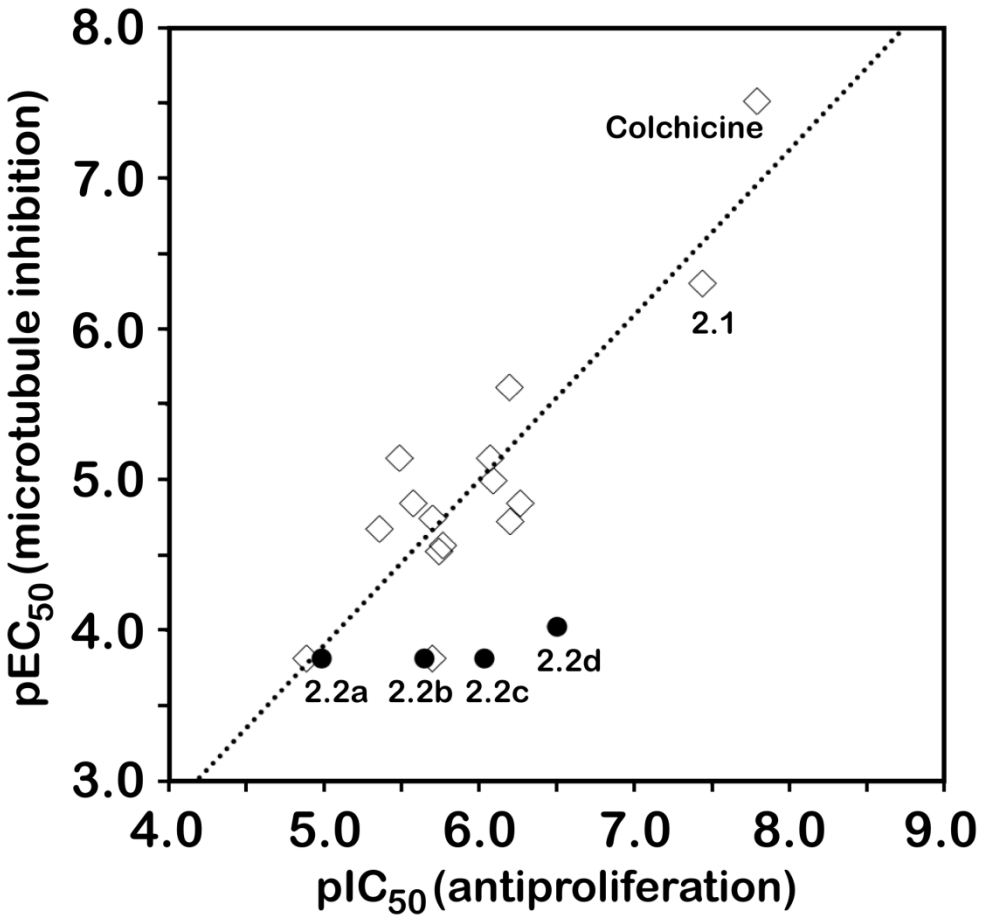


Figure 2.7. Correlation of pEC_{50} and pIC_{50}. Compounds indicated by closed circles are not included in correlation. They have high antiproliferative activity but negligible microtubule inhibition, which may indicate an alternative mechanism of action.

The SAR was analyzed for C-4 analogues with respect to the EC_{50}. The active lead compound **2.1** (0.490 µM) bore two methoxy groups on the phenyl ring attached at the C-4 position. Removing either of the methoxys showed a significant decrease in activity by 14-fold (**2.2e**, 7.0 µM) and 5-fold (**2.2f**, 2.4 µM), respectively; as noted above, a complete loss of microtubule inhibitory activity was observed when all ring substitutions were eliminated (**2.2a**), suggesting the significance of both methoxys with a particular preference for the meta-methoxy group. When compared to **2.2e**, replacing the

hydrophobic methyl with a more polar trifluoromethyl while retaining the acceptor ether oxygen (**2.2j**) or with the weaker sulfur acceptor (**2.2k**) resulted in minor losses in activity by 4-fold and 2.5-fold, respectively. Furthermore, attempting to recover activity with hydrophobic groups at the para-position with **2.2b** (methyl), **2.2c** (chloro) and **2.2d** (bromo) was completely ineffective with respect to microtubule inhibition, although the antiproliferative activity for these analogues increases with substituent hydrophobicity. Overall, these results suggest that the hydrogen bonding properties of the C-4 ring substituents play the more critical role in microtubule inhibition, although clearly the ether oxygen in –OMe may also serve to place the hydrophobic methyl in a more ideal position.

Addition of a second chlorine at the meta-position recovered activity (**2.2l**, 9.9 μM). This may be partially explained by the weak hydrogen bond accepting character of chlorine, but also its placement in the meta-position is a factor – as was seen in the comparison between **2.2f** and **2.2e**. Probably because fluorine is less hydrophobic and smaller than chlorine (although a stronger acceptor), the fluorinated compounds, **2.2m**, was no more effective as a microtubule inhibitor than its des-fluoro analogue **2.2e**.

The inhibitory activity observed for large aromatic rings as C-4 substituents (**2.2h** and **2.2i**) can be attributed to their hydrophobicities and also the hydrogen bond acceptor character of the aromatic π-clouds in napthyl and indolyl. To further investigate this putative hydrogen bonding, the H-bond acceptor was repositioned with 6-ethoxyl-2-napthyl at C-4 (**2.2n**) with negative effect.

Restriction of the rotation of two methyl groups was achieved by first forming a methylene bridge between two oxygens (**2.2o**), which led to a 60-fold decrease in activity compared to **2.1**. Secondly, an ethylene bridge (**2.2p**) fared somewhat better with only a 40-fold activity decrease. Interestingly, addition of a third methoxy to the phenyl ring at its 5-position, as in **2.2g**, did not lead to the expected increase, but, instead, a total loss in activity; however, placing a bromine at the ring's 2-position and removing the 3-methoxy (**2.2q**) produced a >5-fold activity increase over **2.2g**.

Modeling was performed to rationalize the observed SAR. The colchicine site is located at the interface of α- and β-tubulin and mostly buried in β-tubulin. It is surrounded by helices H7 and H8, loop T7 and strands S8 and S9 of β-tubulin and loop T5 of α-tubulin. Comparison of crystal structures of αβ-tubulin heterodimers complexed with different ligands reveals the flexibility of the colchicine site, especially for loops T7 and T5. To understand the movement of the sidechains and backbones surrounding the site, we performed docking studies with five crystal structures (PDBIDs: 1sa0, 1sa1, 3hkc, 3hkd and 3hke). Docking poses were generated by GOLD and the resulting complexes were minimized in Sybyl with the Tripos forcefield and rescored with HINT. These results showed that the compounds tended to bind most favorably to the 3hkc model as indicated by higher HINT scores. While 1sa0 is complexed with colchicine and is thus frequently used for docking colchicine site agents, 3hkc, is co-crystallized with the structurally unrelated N-{2-[(4-hydroxylphenyl)amino]pyridin-3-yl}-4-methoxybenzene-sulfonamide (**Figure 2.8**). In docking compounds **2.1** and **2.2a**-**2.2q**, however, the T5 loop of 3hkc appears to adapt and benefit from hydrogen bonding between the backbone carbonyl of Thr179α and the pyrrole nitrogen, while in colchicine binding, T5

yields to colchicine's amide chain as seen in 1sa0 (**Figure 2.9**). This binding mode is the same as we previously reported.[24] The ester chain of **2.1** and **2.2a**-**2.2q** partially overlaps with ring C of colchicine, fitting into subpocket C with the carbonyl oxygen forming a hydrogen bond with the backbone nitrogen of Val181α. The pyrrole core locates in the center of the site, forming hydrogen bonds with Asn258β and Thr179α. The phenyl moiety overlaps with ring A of colchicine, inserting into the hydrophobic subpocket A, which is formed by Tyr202β, Val238β, Thr239β, Leu242β, Leu248β, Leu252β, Ile378β and Val318β, with Leu248β and Leu255β clamping the phenyl moiety. One polar residue, Cys241β, donates to the ligand in the presence of an appropriately positioned acceptor. The presence of this latter residue in subpocket A explains the importance of hydrogen bond accepting character in C-4 substituents observed in the SAR studies.

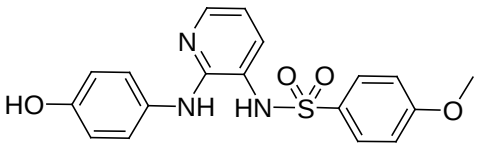


Figure 2.8. N-{2-[(4-hydroxylphenyl)amino]pyridin-3-yl}-4-methoxybenzenesulfonamide.

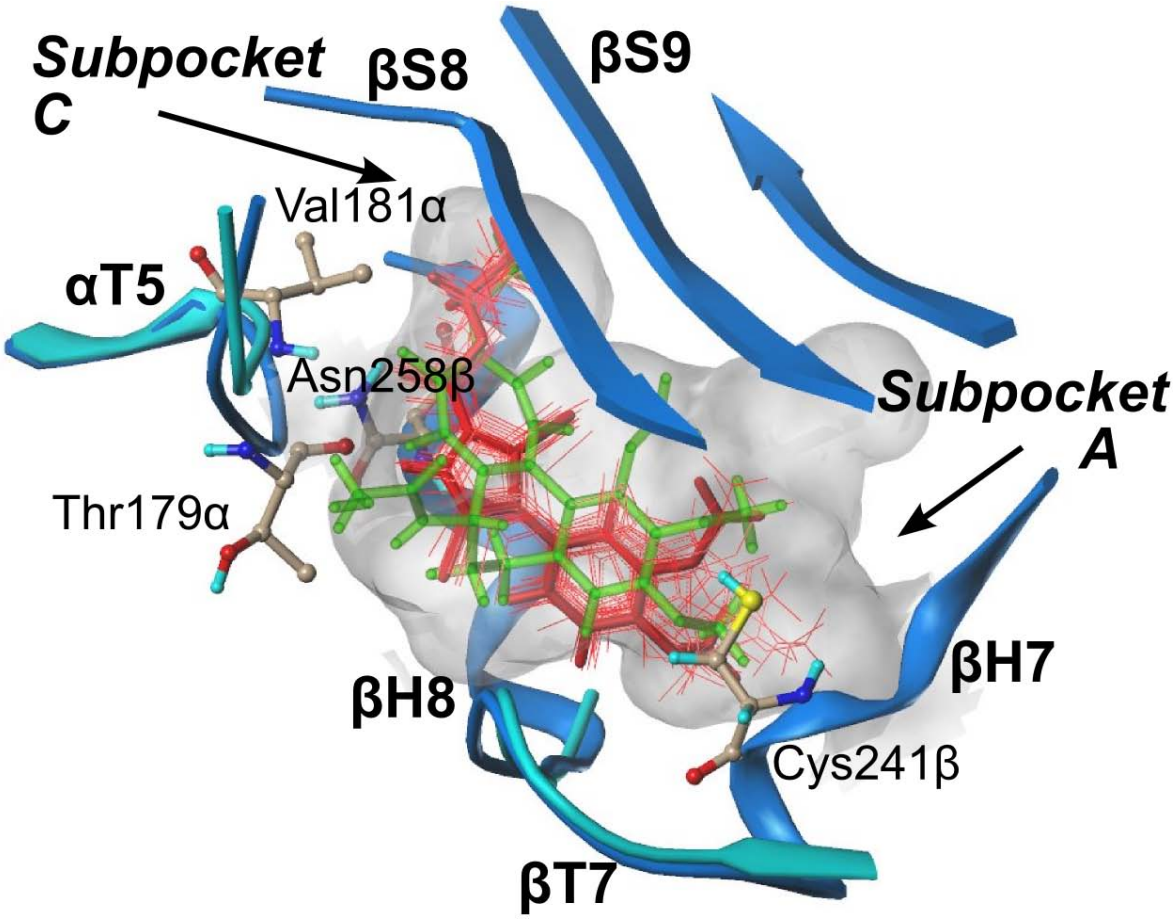


Figure 2.9. Colchicine (green) and binding modes of pyrrole-based C-4 analogues in red (JG-03-14 in heavy sticks). The extents of the colchicine site, as illustrated by MOLCAD, are shown in grayish white.

Detailed analysis of the binding conformations assists further interpretation of the SAR (**Figure 2.10**). In the case of **2.1**, the thiol hydrogen of Cys241β is pointed towards the methoxy at the *meta*-position and away from the *para*-position. This was observed for all other cases owing to the steric clashes that the thiol hydrogen could encounter if oriented in the other direction. The better hydrogen bonding for a *meta*-position substituent explains the activity of **2.2f** compared to **2.2e** and other similar cases. As for **2.2h** and **2.2i**, the distal (from the pyrrole core) rings were located directly beneath the

thiol hydrogen, thus acting as acceptors for the weak but critical hydrogen bond, but pocket steric issues cancelled this advantage. The fluorine atom of **2.2m** is also located at the *meta*-position, but the docking study suggested that a 180° ring flip shifted its position in space such that, although the fluorine was anchored by the backbone NH of Leu252β, it provided no additional bonding to Cys241β compared to **2.2e**. In the case of **2.2g**, the detrimental effect of the third methoxy is visually apparent: the tight distance (3.58 Å) between the backbone of Leu252β and the phenyl ring of the ligand can lead to significant steric clashes with a large substituent such as the 5-methoxy.

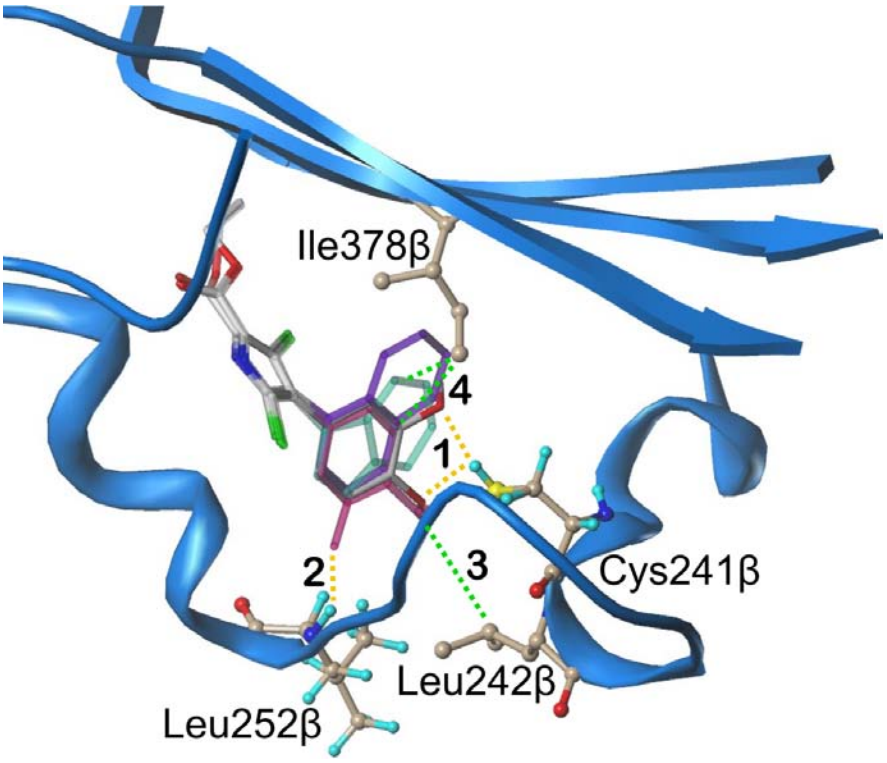


Figure 2.10. Specific hydrogen bonding (yellow) and hydrophobic (green) interactions in Subpocket A. Compounds **2.1** JG-03-14 (solid white), **2.2h** (translucent purple), **2.2i** (translucent green) and **2.2m** (translucent red) are shown. Notes: 1) the key H-bond interaction is with Cys241β, which is strongest with the O of methoxy in the ring's *meta* position; 2) some analogues, e.g., with F, can weakly H-bond with the NH of Leu252β; 3) the CH_3 of *p*-methoxy has key hydrophobic interactions with Leu242β; 4) Ile378β has hydrophobic interactions with *m*-methoxy or the rings of **2.2h** or **2.2i**.

The total HINT scores of C-4 analogues fail to show a tight relationship with pEC_{50} (**Figure 2.11**). However, isolating the HINT score for hydrogen bonding interactions involving Cys241β for a subset of analogues (**2.1**, **2.2e**, **2.2f**, **2.2j**, **2.2k**, **2.2m** and **2.2q**) that place, as separate entities, appropriately positioned hydrophobic groups *and* a hydrogen bond acceptor in the subpocket (while not inducing steric clashes), reveals a linear relation with respect to these compounds' pEC_{50}s (**Figure 2.12**). The implications are two-fold; first, the hydrogen bonding interaction with Cys241β is the key predictor, absent of steric clashes, for the microtubule inhibitory activity for this set of analogues; second, other interactions in the pocket, i.e., hydrophobic, are also necessary, but competitive with this weakly scored hydrogen bonding.

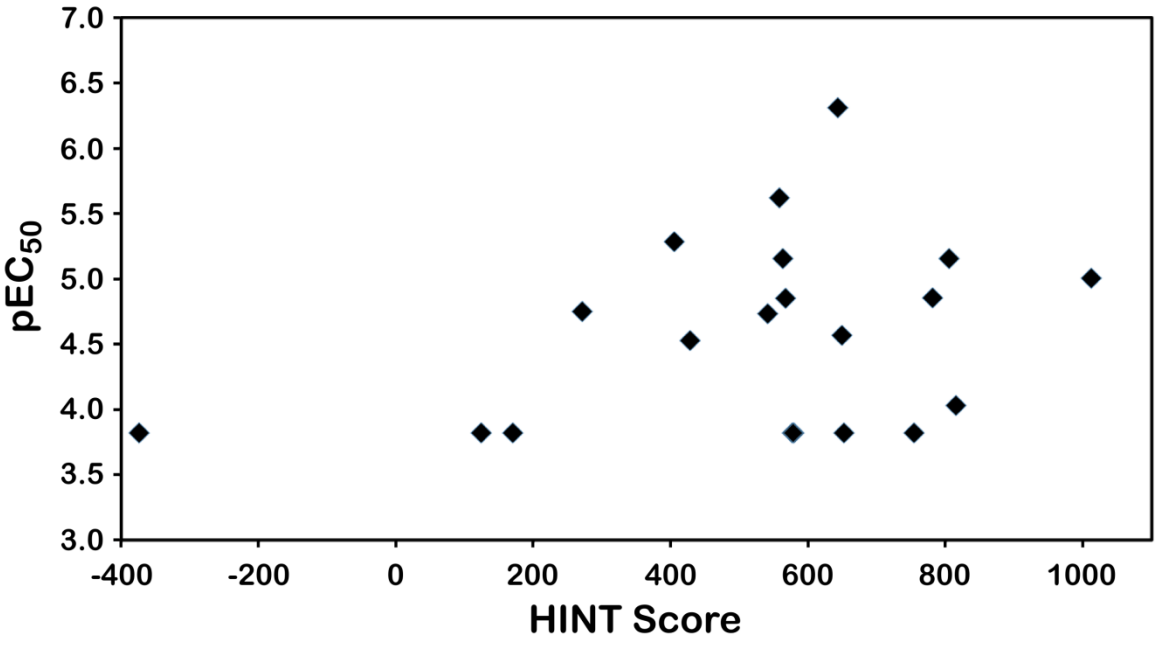


Figure 2.11. Plot of pEC_{50} vs. total HINT score. The total HINT scores of C-4 analogues fail to show a tight relationship with pEC_{50}.

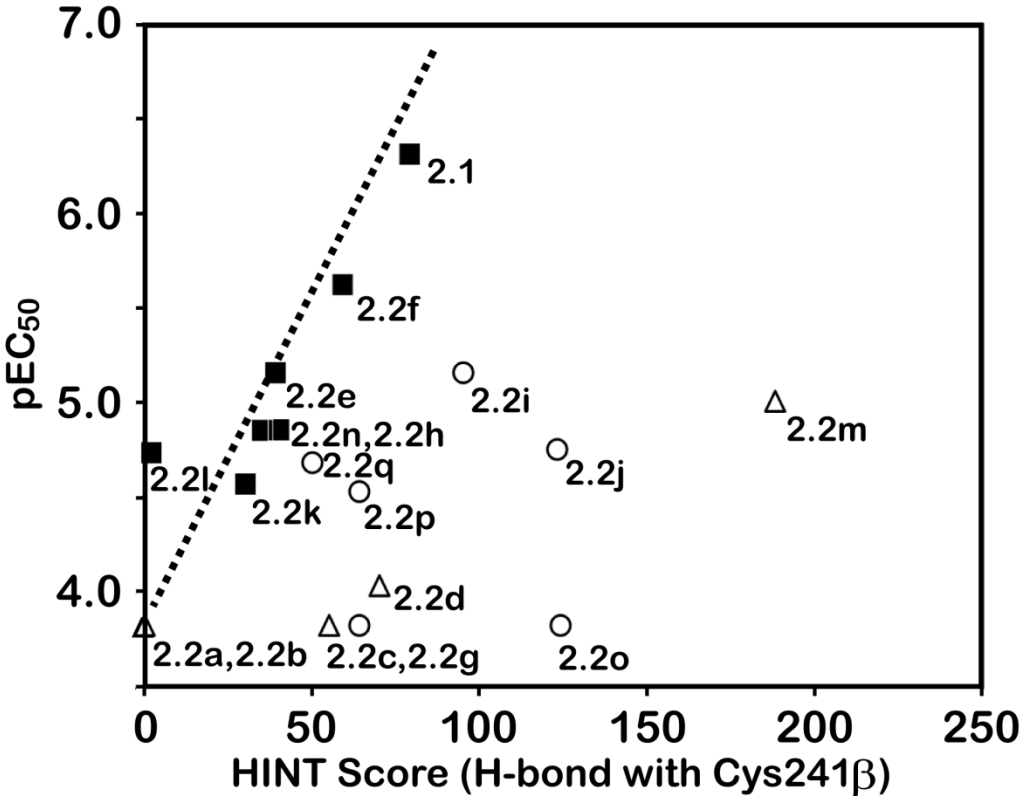


Figure 2.12. HINT H-bond component score for ring interactions with Cys241β. Closed squares represent compounds possessing both H-bond acceptors and appropriately placed hydrophobic groups. These compounds generally possess superior pEC_{50}s. Open triangles represent compounds with weak or no acceptors. Open circles represent compounds with steric issues and/or lacking key hydrophobic interactions.

The importance of both hydrophobic interactions and hydrogen bonding in subpocket A was seen in the SAR analysis and modeling studies. The latter dictates whether the C-4 analogues of pyrrole-based antitubulin agents display microtubule inhibitory activity and the strength of that activity, while the character of the pocket requires predominantly hydrophobic moieties. Underestimation of the Cys241β interaction was one probable reason that the total HINT score was a poor predictor of microtubule inhibitory activity. This thiol group acts as a hydrogen bond donor and while this type of hydrogen bonding interaction is generally regarded as weak and is thusly parameterized by HINT, it is not

even considered by many other scoring functions. For the downstream biological effect, inhibition of microtubules, the interaction assumed to be weak surprisingly stands out as a key factor. In fact, its absence might produce a different mechanism of action even when other portions of the structure are exactly the same, as shown particularly by **2.2d** with potent antiproliferative activity (0.312 μM) but weaker microtubule depolymerization activity (~94 μM). Cys241β has been previously identified as an important target residue for colchicine site agents.[7] In a study of 15 structurally diverse colchicine site inhibitors, the docked binding modes of all included hydrogen bonding to Cys241β (Cys239β in that study).[25] Our combined SAR and modeling study confirms the importance of that cysteine. Interestingly, βIII-tubulin has a mutation of Cys241β to Ser241β, and our scoring regimen would score H-bond donation from –OH more favorably than the -SH donation from Cys.

It should be noted that there is potentially a systematic error in our procedure. As GOLD optimizes ligand placement with a different forcefield (set of rules) than used by HINT in scoring, subtle structural effects, or in this case, the interplay of several of them, are not well scored post-docking as none of the models generated by GOLD capture the set of features in a single model that HINT would score highest. This is likely to be a general observation in docking/rescoring studies, irrespective of utilized scoring functions, when subtle effects are at play.

2.4.4 Summary: SAR of C-4 analogues

We reported the modeling studies of C-4 analogues of pyrrole-based antitubulin agents targeting the colchicine site. For compounds that depolymerized microtubules, a linear correlation was observed between the antiproliferative activity and microtubule inhibitory

activity, Molecular modeling results explained the SAR very well and they both revealed that a weak hydrogen bond involved with Cysβ241 was the key determiner of microtubule inhibitory activity, but the ideal ligand must incorporate (and properly position) this acceptor within an otherwise hydrophobic framework. Surprisingly, just the loss of that particular hydrogen bonding interaction appears to shift the antiproliferative mechanism of action away from microtubule inhibition. This study has fairly exhaustively probed subpocket A; the 3,4-dimethoxyphenyl substituent at the pyrrole C-4 is – to date – the most ideal. This study has been published.[26] The development of analogues focusing on other positions on the pyrrole core is in progress.

2.5 Conclusions

Microtubules have been recognized as a target for cancer treatment for a long period of time. Colchicine-site agents, which target microtubules, although not successful to date in cancer chemotherapy due to toxicity issues, are showing intriguing newly discovered properties such as vascular disruption and possibility of circumventing βIII-tubulin-related drug resistance. Our lead compound, JG-03-14, showed valid evidence of being a colchicine-site agent. This pyrrole-based compound demonstrated potent antiproliferative activity and strong microtubule-destabilizing activity. Here, we studied its C-2 and C-4 analogues using molecular modeling techniques to explore the structure-activity relationship (SAR) and to understand and optimize their binding to the colchicine-site. We docked the analogues into different crystal structures of the colchicine-site and evaluated the interactions using the HINT scoring function. By studying the C-2 analogues, we improved the previous binding mode of the pyrrole-

based compounds. Two distinct binding modes were identified and they clearly differentiate the highly active analogues from the weak ones. The generalized SAR agrees well with the SAR of colchicine, supporting the rationality of the two modes. The residues that participate in binding were also identified. By studying the C-4 analogues, a critical hydrogen bonding interaction involving Cys241β was revealed. Although this interaction was supposed to be weak, loss of it appears to shift the antiproliferative mechanism of action away from microtubule inhibition. We have fairly extensively explored the C-2 and C-4 positions and identified specific features for an optimum ligand and corresponding residues in the colchicine-site. The future work will be focused on other positions for an even more comprehensive understanding of pyrrole-based analogues.

References

1. Jordan, M. A.; Wilson, L. Microtubules as a Target for Anticancer Drugs. *Nat. Rev. Cancer.* **2004**, *4*, 253-265.
2. Dumontet, C.; Jordan, M. A. Microtubule-Binding Agents: A Dynamic Field of Cancer Therapeutics. *Nat. Rev. Drug Discovery*. **2010**, *9*, 790-803.
3. Stanton, R. A.; Gernert K. M.; Nettles, J. H.; Aneja, R. Drugs That Target Dynamic Microtubules: A New Molecular Perspective. *Med. Res. Rev*. **2011**, *31*, 443-481.
4. Siemann, D. W. The Unique Characteristics of Tumor Vasculature and Preclinical Evidence for Its Selective Disruption by Tumor-Vascular Disrupting Agents. *Cancer Treat. Rev.* **2011**, *37*, 63-74.
5. Seve, P.; Dumontet C. Is Class III Beta-tubulin a Predictive Factor in Patients Receiving Tubulin-binding Agents? *Lancet Oncol.* **2008**, *9*, 168-175.
6. Stengel, C.; Newman, S. P.; Leese, M. P.; Potter, B. V.; Reed, M. J.; Purohit, A. Class III Beta-tubulin Expression and in vitro Resistance to Microtubule Targeting Agents. *Br. J. Cancer.* **2010**, *102*, 316-324.

7. Chen, J.; Liu, T.; Dong, X.; Hu, Y. Recent Development and SAR Analysis of Colchicine Binding Site Inhibitors. *Mini-Rev. Med. Chem.* **2009**, *9*, 1174-1190.
8. Ravelli, R. B.; Gigant, B.; Curmi, P. A.; Jourdain, I.; Lachkar, S.; Sobel, A.; Knossow, M. Insight into Tubulin Regulation from a Complex with Colchicine and a Stathmin-like Domain. *Nature.* **2004**, *428*, 198-202.
9. Löwe, J.; Li, H.; Downing, K. H; Nogales, E. Refined Structure of Alpha Beta-Tubulin at 3.5 A Resolution. *J. Mol. Biol*. **2001**, 1045-1057.
10. Mooberry, S. L.; Weiderhold, K. N.; Dakshanamurthy, S.; Hamel, E.; Banner, E. J.; Kharlamova, A.; Hempel, J.; Gupton, J. T. Brown, M. L. Identification and Characterization of a New Tubulin-Binding Tetrasubstituted Brominated Pyrrole. *Mol. Pharmacol.* **2007**, *72*, 132-140.
11. Dalyot-Herman, N.; Delgado-Lopez, F.; Gewirtz, D. A.; Gupton, J. T.; Schwartz, E. L. Interference with Endothelial Cell Function by JG-03-14, An Agent that Binds to the Colchicine Site on Microtubules. *Biochem Pharmacol.* **2009**, *78*, 1167-1177.
12. Cleaveland, E. S.; Monks A.; Vaigro-Wolff, A.; Zaharevitz, D. W.; Paull, K.; Ardalan, K.; Cooney, D. A.; Ford, H. Jr. Site of Action of Two Novel Pyrimidine Biosynthesis Inhibitors Accurately Predicted by the COMPARE Program. *Biochem. Pharmacol.* **1995**, *49*, 947-954.
13. Tripathi, A. Fornabaio, M.; Kellogg, G. E.; Gupton, J. T.; Gewirtz, D. A.; Yeudall, W. A.; Vega, N. E.; Mooberry, S. L.; Docking and Hydrophobic Scoring of Polysubstituted Pyrrole Compounds with Antitubulin Activity. *Bioorg. Med. Chem*. **2008**, *16*, 2235-2242.
14. Spyrakis, F.; Amadasi, A.; Fornabaio, M.; Abraham, D.J.; Mozzarelli, A.; Kellogg, G. E.; Cozzini, P. The Consequences of Scoring Docked Ligand Conformations using Free Energy Correlations. *Eur. J. Med. Chem.* **2007**, *42*, 921-933.
15. Gupton, J.; Burnham, B.; Krumpe, K.; Du, K.; Sikorski, J.; Warren, A.; Barnes, C.; Hall, I. Synthesis and Cytotoxicity of 2,4-Disubstituted and 2,3,4-Trisubstituted Brominated Pyrroles in Murine and Human Cultured Tumor Cells. *Arch. Pharm. Pharm. Med. Chem.* **2000**, *333*, 3-9.
16. SYBYL 8.1, Tripos Interactional, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
17. Jones, G; Willett, P; Glen, R. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *J. Mol. Biol.* **1995**, *245*, 43-53.
18. Kellogg, E. G.; Abraham, D. J. Hydrophobicity: Is LogPo/w More than the Sum of Its Parts? *Eur. J. Med. Chem.* **2000**, *35*, 651-661.
19. Wireko, F. C.; Kellogg, G. E.; Abraham, D. J. Allosteric Modifiers of Hemoglobin. 2. Crystallographically Determined Binding Sites and Hydrophobic

Binding/Interaction Analysis of Novel Hemoglobin Oxygen Effectors. *J. Med. Chem.* **1991**, *34*, 758-767.
20. Bhattacharyya, B.; Panda, D.; Gupta, S.; Banerjee, M. Anti-Mitotic Activity of Colchicine and the Structural Basis for Its Interaction with Tubulin. *Med. Res. Rev.* **2008**, *28*, 155-183.
21. Andreu, J. M.; Perez-Ramirez, B.; Gorbunoff, M. J.; Ayala, D.; Timasheff, S. N. Role of the Colchicine Ring A and Its Methoxy Groups in the Binding to Tubulin and Microtubule Inhibition. *Biochemistry*. **1998**, *37*, 8356-8368.
22. Hastie, S. B.; Williams, R. C. Jr; Puett, D.; Macdonald, T. L. The Binding of Isocolchicine to Tubulin. Mechanisms of Ligand Asso-ciation with Tubulin. *J. Biol. Chem.* **1989**, *264,* 6682-6688.
23. Staretz, M. E.; Hastie, S. B. Synthesis and Tubulin Binding of Novel C-10 Analogues of Colchicine. *J. Med. Chem.* **1993**, *36*, 758-764.
24. Da, C.; Telang, N.; Barelli, P.; Jia, X.; Gupton, J. T.; Mooberry, S. L.; Kellogg, G. E. Pyrrole-Based Antitubulin Agents: Two Distinct Binding Modalities are Predicted for C-2 Analogues in the Colchicine Site. *ACS Med. Chem. Lett.* **2012**, *3*, 53-57.
25. Nguyen, T. L.; McGrath. C.; Hermone, A. R.; Burnett, J. C.; Zaharevitz, D. W.; Day, B. W.; Wipf. P.; Hamel, E.; Gussio, R. A Common Pharmacophore for a Diverse Set of Colchicine Site In-hibitors Using a Structure-Based Approach. *J. Med. Chem.* **2005**, *48*, 6107-6116.
26. Da, C.; Telang, N.; Hall, K.; Kluball, E.; Barelli, P.; Finzel, K.; Jia, X.; Gupton, J. T.; Mooberry, S. L.; Kellogg, G. E. Developing Novel C-4 Analogues of Pyrrole-Based Antitubulin Agents: Weak but Critical Hydrogen Bonding in the Colchicine site. *Medchemcomm*. **2013**, *4*, 417-421.

CHAPTER 3

Mapping of the Colchicine Site with Docking and 3D-QSAR Analysis of Structurally Diverse Binders

3.1 Introduction

Microtubules have been treated as a target for cancer therapies for a long period of time, due to the fact that they are one of the major cytoskeletal components in eukaryotic cells and their critical functions, such as maintenance of cell shape, protein trafficking, signaling and segregation of chromosomes during mitosis.[1] Microtubule-targeting agents function as interference with microtubule dynamics, a process that controls the balance between microtubule assembly and microtubule disassembly.[2] Four major binding sites for these agents have been identified: the taxane site and the laulimalide / peloruside A site, both for microtubule-stablizing agents, and the vinca site and the colchicine site for microtubule-destablizing agents.[2-3]

Compared to taxanes and vinca alkaloids, which have been used successfully in clinical therapies for cancer; colchicine is restrained by its toxicity to normal tissues at effective drug concentrations and has only been approved for the treatment of familial Mediterranean fever and acute gout flares.[4] However, owing to the fact that microtubules are important regulators of endothelial cells, recently colchicine-site agents or colchicine-site inhibitors (CSI) are being intensively developed as angiogenesis inhibitors (prevent new blood vessel formation) and vascular disrupting agents (destroy existing vasculature) for cancer treatment.[5] Combretastatins, one family of colchicine-

site agents, are progressing through clinical trials for this purpose.[1,4] In addition, colchicine-site agents might be able to circumvent βIII-tubulin overexpression, which compromises the clinical use of taxanes and vinca alkaloids.[6-7]

A large number of CSIs including natural and synthetic compounds, have been reported and they possess a significant structural diversity. So far, the compounds under clinical investigation cover at least 26 different scaffolds such as colchicine, combretastatin, podophyllotoxin and steganacin, and there are even more in preclinical studies.[5,8] The ability of the colchicine site to accommodate such diversity is due to the inherent flexibility of the site, which has been demonstrated by X-ray crystal structures of the protein complexed with different agents[9-11] and molecular dynamics simulations.[12]

We have been developing pyrrole-based compounds as colchicine-site agents and identified their binding modes through ensemble docking with HINT[13] rescoring and detailed SAR comparison to colchicine (Chapter 2 of this work and in the literature.[14-17] While we continued exploring the modifications on the pyrrole scaffold, we expanded our computational analyses to other scaffolds and tried to consolidate all known information in order to have a more comprehensive understanding of the ligands and the binding pocket. We used 3D-QSAR (Quantitative Structure-Activity Relationship), a statistical technique that identifies the significant features that affect activity from a pool of compounds. 3D-QSAR models also reflects the electrostatic and the topological features of the pocket. Here, we collected all the data for colchicine-site agents that were tested in the same laboratory, by Mooberry and co-workers.[14-22] We performed ensemble docking with different crystal structures of the colchicine site to identify the docked modes of all agents. We then performed 3D-QSAR analyses to obtain a detailed

view of ligand binding. In this chapter, we also describe a new method that mines additional useful information from a pool of compounds. This new method takes advantage of HINT maps,[13] which are calculated to represent the hydrophobic and polar features of compounds, and was applied in this case to our collection of colchicine-site agents.

3.2 Materials and Methods

3.2.1 Dataset selection

The compounds used in this study were reported by Dr. Mooberry and co-workers as antitubulin agents.[14-22] We set two criteria to select compounds for the study, including compounds for both a training set and a test set. First, to ensure the consistency of the activity measurements, we only selected compounds with reported antiproliferative IC_{50} values measured in the MDA-MB-435 cancer cells using the SRB assay.[14] Second, we consider a compound as a valid colchicine-site binder for modeling if: 1) the compound showed inhibition of radiolabeled [^{3}H]colchicine binding to tubulin, or 2) its structurally similar parent compound showed the inhibition of [^{3}H]colchicine.

The reported cellular microtubule loss experiments evaluated in A-10 cells were verified to ensure that the antiproliferative activity for each compound corresponds to tubulin binding. In the cases where no microtubule effect was observed up to 50 μM (40 μM in some publications), the compounds were considered as very poor binders and the IC_{50} values were arbitrarily assigned to have IC_{50}s of 100 μM even though their true IC_{50} values might be higher or lower. The reason was that these experimentally measured

IC50 values were probably more related to other receptors than microtubules, and they should not be modeled in this study.

3.2.2 Identification of bioactive conformations and alignment

In general, Sybyl 8.1[23] was used to prepare the X-ray crystal structures of αβ-tubulin complexed with different ligands (pdbid: 1SA0, 1SA1, 3HKC, 3HKD and 3HKE) for docking. The preparation and docking procedure was the same as reported previously.[14,15] The ligands were docked using GOLD 5.1[24] into the active site, which was defined by the space in a 6 Å radius around the complexed small molecules. One hundred conformations were generated for each compound. They were initially analyzed by GoldScore and further rescored by the HINT[13] scoring function. To select the final "active" conformation, we considered binding to all five receptor structures, and we picked the conformation with both a high HINT score and a high degree of similarity to the conformation of the complexed small ligand in the crystal. Because the resolution of the tubulin crystal structures is so poor, around 3.5 Å, the crystallographic models for the bound ligands are only guides to their actual conformations.

The semi-ligand-based approach was performed by the "Fit Atoms" and the "Align Database" functions in Sybyl 8.1. First, for different scaffolds, the most active compounds (compound **3.1** DAMA-colchicine, compound **3.2** for pyrrole analogues, compound **3.39** combretastain A-4 and compound **3.54** for pyriminide analogues) were docked to the receptor. The pharmacophores were identified by analyzing the common functional features of the docked conformations of the highly active compounds and the complementary receptor structures. Then these representatives (the most active

compound) of the different scaffolds were superimposed on each other according to the pharmacophores. Other compounds were aligned to the corresponding representatives based on the common substructure. In the alignment step, all conformations, including those for the representatives, were optimized by the Tripos force field with Gasteiger-Hückel charges to a gradient of 0.005 kcal mol^{-1} $Å^{-1}$.

3.2.3 3D-QSAR modeling

Sybyl 8.1 was used to perform the 3D-QSAR analysis. The basic concept of 3D-QSAR is to correlate activity with the interaction fields surrounding the ligands. Such fields are calculated by measuring the interactions between a probe atom placed on a grid point around the ligand and the atoms of the ligand. Machine learning techniques are then applied to find a set of interaction values that are related to the activity changes. The partial-least-squares (PLS) regression method was used to derive the model in this study. The leave-one-out (LOO) cross-validation with the sample-distance PLS (SAMPLS) algorithm was used to identify the optimum number of components. The leave-one-out (LOO) method predicts the activity of each compound using the QSAR model built by all the other compounds except the predicted one. It evaluates the predictability and over-fitting of a regression model, and indicates the quality of the model with a cross-validated correlation coefficient called q^2. The optimum number of components was the smallest number that gave the largest value of q^2 as long as there was an increase of least 5% from the previous q^2 value. The non-cross-validated model was then built with the identified optimum number of components and using the entire training set. The statistics for evaluation included the cross-validated correlation coefficient (q^2), the non-cross-validated correlation coefficient (r^2), standard error of

estimate (SEE) and the F test value. The resulting model was further validated using the external test set compounds and gave a predictive r^2 (r_{pred}^2) indicating the difference between the predicted activities and the experimental activities of the test set.

For field calculations, we selected the steric and electrostatic fields for CoMFA, and the steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields for CoMSIA. Gasteiger-Hückel charges were assigned to all compounds. The grid resolution was set to 1.0 Å. Other settings were default.

3.2.4 HINT fields and HINT maps

The hydrophobic/polar field of HINT was combined with the steric and electrostatic fields of CoMFA for 3D-QSAR analysis. The HINT (Hydropathic INTeractions) scoring function[13] evaluates atom-atom interactions using a set of parameters derived from the solvation partition coefficients, LogPs, measured in a 1-octanol/water system (see Chapter 1).

The HINT program calculates a hydrophobic/polar field and an acid/base field. The test atom has a hydrophobic atom constant (a_t) and solvent accessible surface area (S_t) both equal to one. The field value A of each grid point is given by

$$A_t = \sum a_i S_i R_{it}$$

where a_i and S_i are for the atom i, and R_{it} is a function of the distance between the atom i and the test atom t. We used a resolution of 0.5 Å to calculate all the HINT fields in this study.

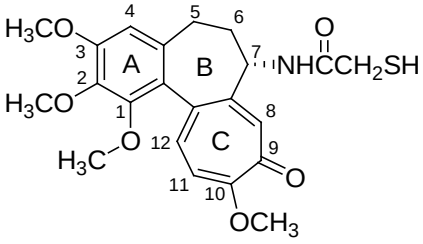
The HINT map for a molecule is the grid map containing the HINT field values described above. Two types of HINT maps, the hydrophobic/polar map and the acid/base map, were generated for all the compounds. To generate an overall map to represent all the compounds, the maps were linearly combined with weights related to their activity values (IC_{50}). In this study, we applied 0.1 (μM)/IC_{50} (μM) as the map weight for each compound.

3.3 Results and Discussion

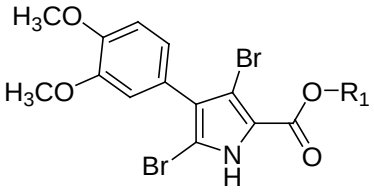
3.3.1 Dataset for modeling

Of the final selected 62 compounds, three different scaffolds could be categorized, namely pyrrole analogues, combretastatin analogues and pyrimidine analogues (**Table 3.1**). Fifty-three compounds covering all three scaffolds were used for the training set and the remaining 9 compounds were used for the test set based on random selection. Colchicine, whose comparatively large structure occupies most of the binding space (**Figure 3.1**), was tested as a reference compound in the experimental assays and therefore it was included into the test set. The pIC_{50} values for both the training set and the test set compounds covered a range of more than 3 log units. The final structures and pIC_{50} values for modeling are shown in **Table 3.1**.

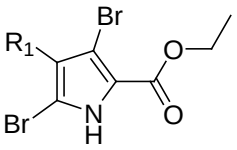
Table 3.1. Structures and activities of compounds in the training set and the test set.



1 (DAMA-colchicine)



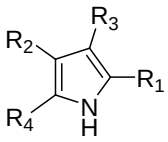
2-14



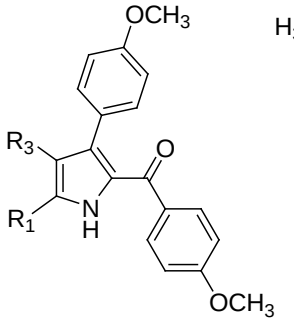
15-30

Cmpd	R1	IC_{50} (μM)	pIC_{50}
3.1*	—	0.014	7.70
3.2	Ethyl	0.036	7.44
3.3	Methyl	0.168	6.21
3.4	*n*-Propyl	0.050	7.17
3.5	*i*-Propyl	0.108	6.96
3.6	*t*-Butyl	1.82	5.74
3.7	*n*-Butyl	1.3	5.89
3.8	*n*-Hexyl	3.5	5.48
3.9	Benzyl	5.3	5.28
3.10	$-(CH_2)_3NMe_2$	4.6	5.34
3.11	$-(CH_2)_2NMe_2$	5.2	5.28
3.12	$-(CH_2)_3NMe_2H^+Cl^-$	8.0	5.10
3.13*	$-(CH_2)_2NMe_2H^+Cl^-$	10.7	4.97
3.14	4-Methoxyphenyl	18.3	4.74
3.15	Phenyl	10.3	4.99
3.16*	4-Methyphenyl	2.24	5.65
3.17	4-Chlorophenyl	0.919	6.04
3.18	4-Bromophenyl	0.312	6.51
3.19	4-Methoxyphenyl	0.843	6.07
3.20*	3-Methoxyphenyl	0.633	6.20
3.21	2-Bromo-4,5-dimethoxyphenyl	2.64	5.58
3.22	1-Napthyl	3.24	5.49
3.23*	3-Indolyl	1.98	5.70
3.24	4-Trifluoromethoxyphenyl	1.70	5.77
3.25	4-Thiomethyphenyl	0.626	6.20
3.26	3,4-Dichlorophenyl	0.806	6.09
3.27	3-Fluoro-4-methoxylphenyl	0.539	6.27
3.28	6-Ethoxyl-2-napthyl	1.99	5.70

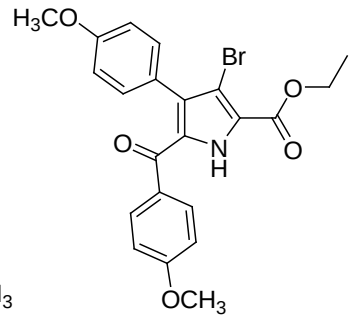
3.29	1,3-Benzodioxol-6-yl	1.80	5.74
3.30*	1,4-Benzodioxan-6-yl	4.36	5.36



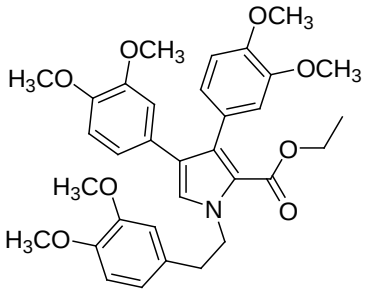
31-33



34-36

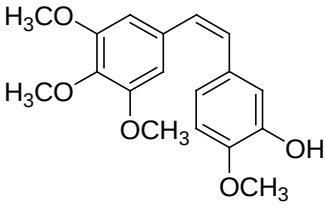


37

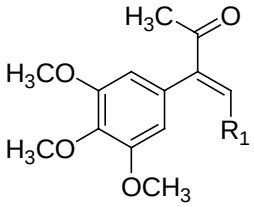


38

Cmpd	R1	R2	R3	R4	IC_{50} (μM)	pIC_{50}
3.31*	Ethoxycarbonyl	4-Methoxyphenyl	H	Cl	5	5.3
3.32	4-Methoxyphenyl	Ethoxycarbonyl	4-Methoxyphenyl	H	100	4.00
3.33	Ethoxycarbonyl	4-Methoxyphenyl	4-Methoxyphenyl	H	100	4.00
3.34	Ethoxycarbonyl	—	H	—	2.6	5.52
3.35	Ethoxycarbonyl	—	4-Methoxyphenyl	—	100	4.00
3.36	4-Methoxyphenylcarbonyl	—	4-Methoxyphenyl	—	100	4.00
3.37*	—	—	—	—	100	4.00
3.38	—	—	—	—	100	4.00

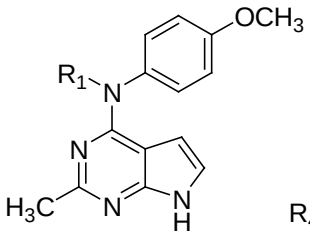


39 (Combretastatin A4)

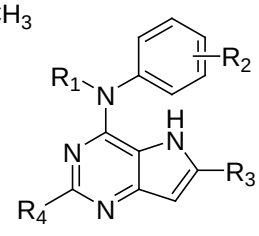


40-42

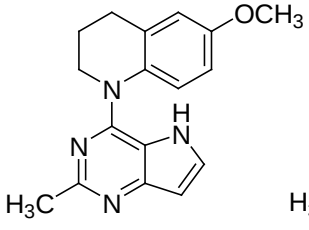
Cmpd	R1	IC_{50} (μM)	pIC_{50}
3.39	—	0.003	8.52
3.40	4-Methoxyphenyl	0.35	6.46
3.41	4-Methylphenyl	0.095	7.02
3.42	3-Hydroxy-4-methoxyphenyl	0.182	6.74



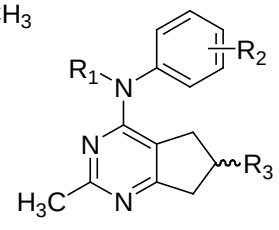
43-44



45-52



53



54-58 (S) **59-63** (R)

Cmpd	R1	R2	R3	R4	IC_{50} (μM)	pIC_{50}
3.43*	Methyl	—	—	—	0.183	6.74
3.44	H	—	—	—	100	4.00
3.45	Methyl	4-Methoxy	H	Methyl	0.096	7.02
3.46	Methyl	4-Methoxy	H	H	0.193	6.71
3.47	Methyl	H	H	Methyl	100	4.00
3.48	H	4-Methoxy	H	Methyl	100	4.00
3.49*	H	2,4-Dimethoxy	H	Methyl	100	4.00
3.50	Methyl	4-Methoxy	Methyl	Methyl	0.030	7.52
3.51	Methyl	4-Methoxy	Methyl	NH2	0.298	6.53
3.52	H	4-Methoxy	Methyl	NH2	100	4.00
3.53	—	—	—	—	0.043	7.37
3.54	Methyl	4-Methoxy	Methyl	—	0.012	7.92
3.55	H	4-Methoxy	Methyl	—	100	4.00
3.56	Methyl	3-Methoxy	Methyl	—	0.095	7.02
3.57	Methyl	2-Methoxy	Methyl	—	100	4.00
3.58	Methyl	H	Methyl	—	100	4.00
3.59	Methyl	4-Methoxy	Methyl	—	0.051	7.29
3.60	H	4-Methoxy	Methyl	—	100	4.00
3.61	Methyl	3-Methoxy	Methyl	—	0.402	6.40
3.62	Methyl	2-Methoxy	Methyl	—	100	4.00
3.63	Methyl	H	Methyl	—	100	4.00

The test set are the 10 compounds indicated by *. The other 53 compounds are in the training set. IC_{50}s are antiproliferative activities tested using human MDA-MB-435 cancer cells. pIC_{50}s are the negative log of the IC_{50}s.
The activities of compounds 55-58 and 60-63 were reported for racemic mixtures. For the compounds that did not have microtubule effect up to 40 μM, both enantiomers were assigned 100 μM. The ratio of the IC_{50} of compound 61 (the R enantiomer) over the IC_{50} of compound 56 (the S enantiomer) was assumed to be the same as the ratio of compound 59 (R) over compound 54 (S), which was experimentally tested.

3.3.2 Overview of the colchicine site and prediction of binding modes from docking

The colchicine site is located at the interface of α-tubulin and β-tubulin and is mostly buried in the β-tubulin subunit as indicated by the crystal structures (**Figure 3.1**). It is

surrounded by helices H7 and H8, loop T7, and strands S8 and S9 of β-tubulin, and loop T5 of α-tubulin (the nomenclature of the secondary structures can be referred to literature [25]). DAMA-colchicine occupies the pocket with it's A ring fitting into the subpocket A close to H7, the C ring in the subpocket C close to T5 and the B ring in the center of the pocket. As different ligands bind, the T7 and T5 loops can move to adapt to the changes, as shown by the crystal structures (**Figure 3.1**). To take this flexibility into account, we performed ensemble docking using five available crystal structures (PDBID: 1SA0, 1SA1, 3HKC, 3HKD and 3HKE), while 3HKB is the unliganded.In the latter structure, the T7 loop closes the entrance and thus occupies most of the site. It was not used in the docking study.

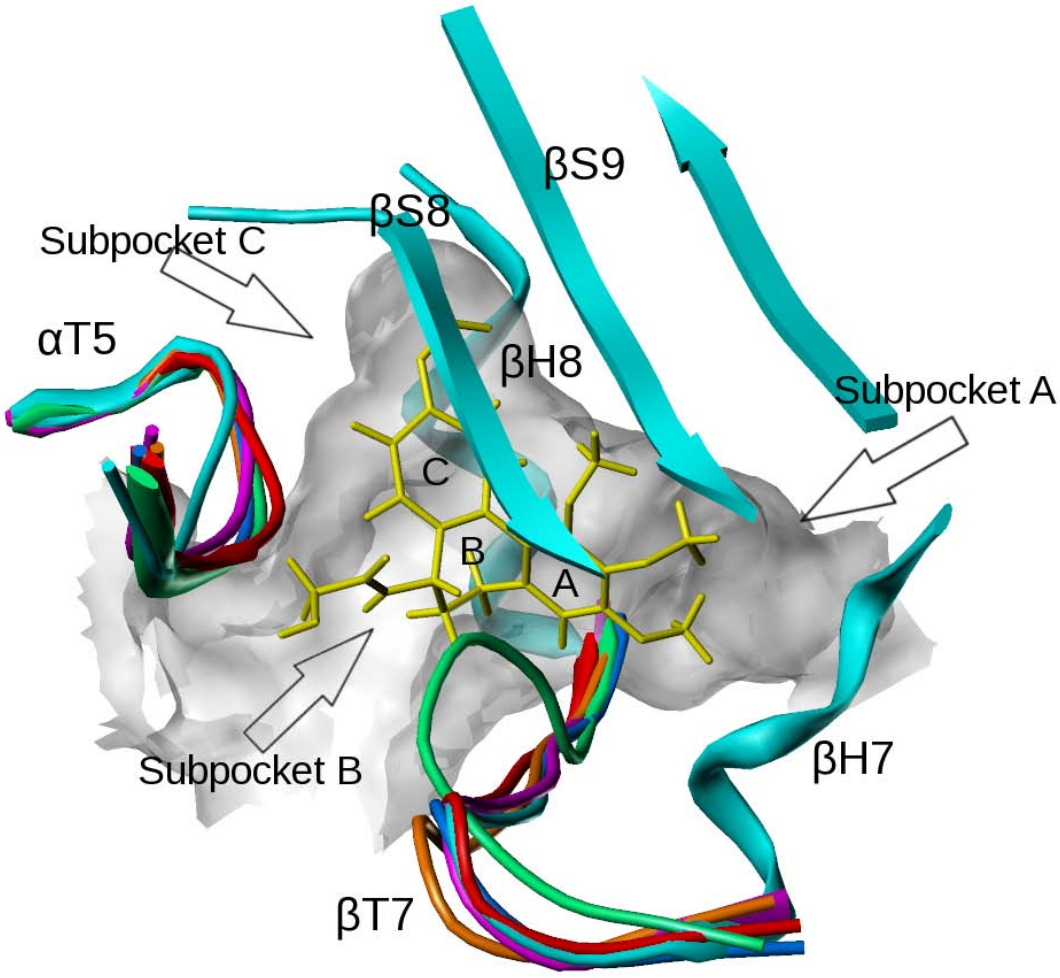


Figure 3.1. The colchicine site complexed with DAMA-colchicine (yellow). The loops of different colors from different crystal structures represent the flexibility of the pocket (1SA0: cyan; 3HKC: magenta; 3HKB: green; 3HKE: brown; 3HKD: red; 1SA1: blue).

The pyrrole compounds (**3.2**-**3.38**) adopted two distinct binding modes, which have been reported in our previous studies,[2] and in the previous chapter of this work. Mode I, represented by the most active compound, compound **3.2**, binds most favorably to the 3hkc model (**Figure 3.2**). The ester chain fits into the subpocket C with the carbonyl oxygen forming a hydrogen bond with the backbone nitrogen of Val181α, and the alkyl ends reaching the hydrophobic bottom. The pyrrole core is located in subpocket B, with

its NH forming hydrogen bonds with the side chain C=O of Asn258β and the backbone C=O of Thr179α. The dimethoxyphenyl group is clamped by the four-carbon side chains of Leu248β and Leu255β in subpocket A, and the two methoxys are locked by other residues that are deeper in the pocket. In addition, the polar residue Cys241β uses its SH group to form a weak hydrogen bond with either one of the two oxygens from the dimethoxyphenyls. Mode II, represented by compound **3.7**, tends to bind the 1SA0 structure more favorably (**Figure 3.2**). The dimethoxyphenyl group and the pyrrole core are located in the similar positions as they are located in the previous mode. The ester chain, however, shifts significantly away from the subpocket C due to its larger size. The chain exposes itself to the solvent and is anchored by a new hydrogen bond with Asn101α.

Mode I (compound **3.2**) and the complexed colchicine overlap well. The dimethoxyphenyl group of mode I is located near the trimethoxylphenyl A ring of colchicine and the ester chain of mode I mimics the combination of the methoxy and the carboxyl oxygen of colchicine's C ring. The pyrrole core coincides with half of the A ring and half of the B ring. Compared to mode I, the only portion where mode II (compound **3.7**) and colchicine overlap is **3.7**'s dimethoxyphenyl and colchicine's trimethoxylphenyl.

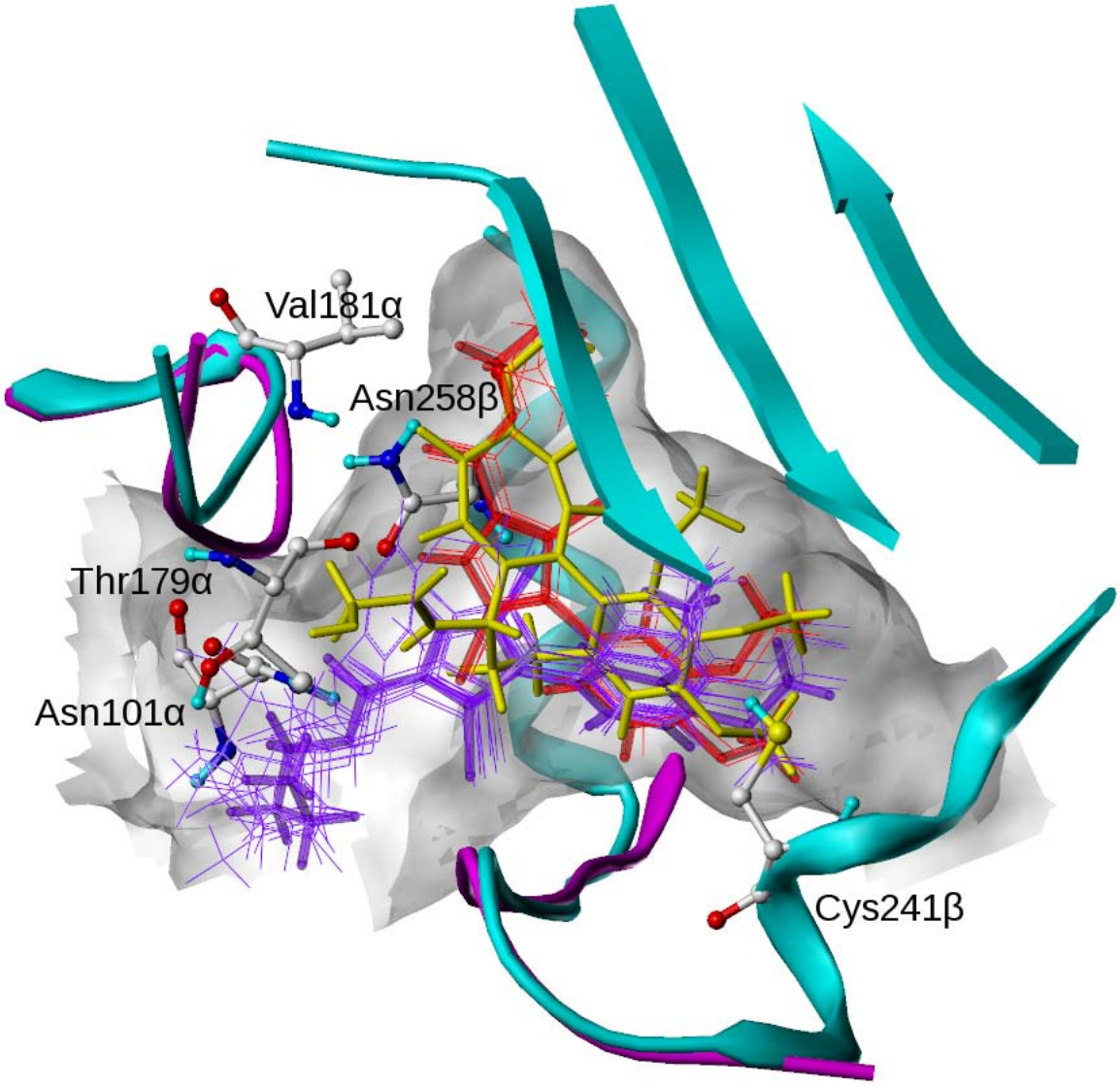


Figure 3.2. The two binding modes of the pyrrole compounds (Compound **3.2**: red; Compound **3.7**: purple) and colchicine (yellow) in the colchicine site (1SA0: cyan; 3HKC: magenta).

The combretastatin analogues (compounds **3.39-3.42**) and the pyrimidine analogues (compound **3.43-3.68**) were predicted to bind most favorably to 3HKC. They adopt a similar binding mode as compound **3.2** (mode I) and colchicine do, as shown in **Figure 3.3** using the most active compound in each category. Subpocket A is occupied by a

hydrophobic moiety with hydrogen bond acceptors interacting with Cys241β. This moiety for compound **3.39** (combretastatin A-4) is its trimethoxyphenyl group with the oxygens as hydrogen bond acceptors, and for compound **3.54** (representing the pyrimidines) is its cyclopentapyrimidine group with one aromatic nitrogen as the acceptor. Subpocket C is occupied by the other end of these ligands. Compound **3.39** uses the methyl from the methoxy group to interact with the hydrophobic bottom of subpocket C, and the oxygen from the hydroxyl group to form a hydrogen bond with Val181α. Compound **3.54**'s methoxy group is also located inside subpocket C, but not as deep as is the methoxy of compound **3.39**. It can interact with the hydrophobic bottom (using the methyl) and Val181α (using the oxygen) at the same time. Subpocket B is occupied by the phenyl rings of both compound **3.54** and **3.39**.

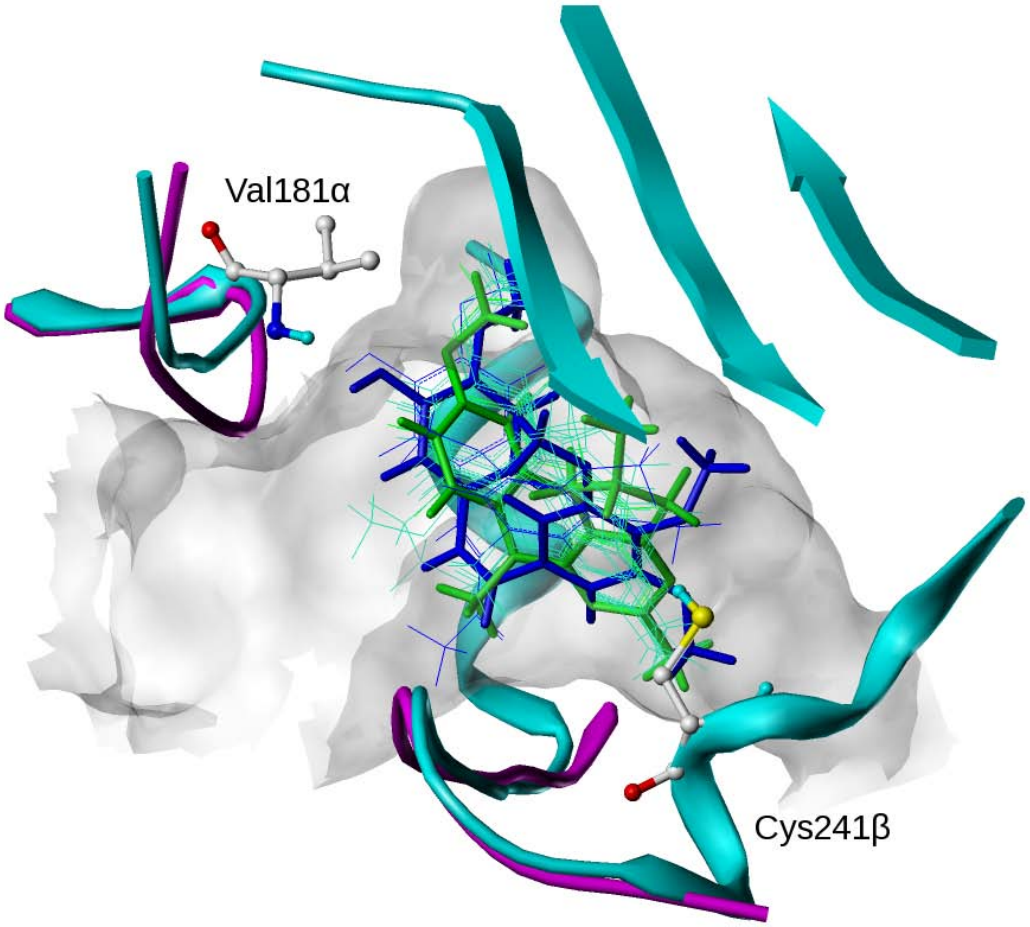


Figure 3.3 The binding modes of the combretastatin A-4 analogues and the pyrimidine analogues (compound **3.39** (combretastatin A-4): blue; compound **3.54**: green) in the colchicine site (1SA0: cyan; 3HKC: magenta).

The pharmacophore model (**Figure 3.4**) generalized from the binding modes identified above contains one hydrogen bond acceptor interacting with Cys241β, another hydrogen bond acceptor interacting with Val181α, and three hydrophobic centers in the subpockets A, B and C, respectively. Colchicine, compound **3.2** (mode I for the pyrroles), **3.39** and **3.54** are good examples that contain all the pharmacophore

features. Compound **3.7** (mode II for the pyrroles), however, only partially matches the model.

The logic of building a 3D-QSAR model relies on the assumption that the variations of activity can be fully explained by the variations in structure. Therefore, the identification of each compound's conformation and how to align the conformations with each other to generate the structural differences are two of the most important aspects. Here, we adopted two different approaches: 1) a docking-based approach, meaning that we used the docked poses (as described above) for 3D-QSAR modeling, and 2) a semi-ligand-based approach, meaning that we simply aligned structures of the different scaffolds based on the common features (pharmacophores) they adopted when interacting with the receptor. A traditional ligand-based 3D-QSAR approach was not used in this study because the substructure that was common to all the scaffolds was only a small portion of the chemical space, so that not enough information was available for a full alignment description.

The conformations of the compounds aligned according to the pharmacophore description are shown in **Figure 3.4**. The docked conformations are shown, as comparison, in **Figure 3.5**. The major difference between the two sets of conformations originates from the basis of the two approaches. The semi-ligand-based approach aligned all ligands of the same scaffold according to their common structures but aligned ligands from different scaffolds according to the defined pharmacophores. These two steps of alignment superimposed each subset of similar structures well, while structural differences between templates were comparatively maximized. The docking-based approach, however, allowed the ligands to move around according to

their *individual* interactions and complementarities with respect to the receptor. Therefore, the docking-based approach spreads the ligands evenly in the binding pocket and the ligand placement and conformations from the semi-ligand-based approach are concentrated in a more step-like manner.

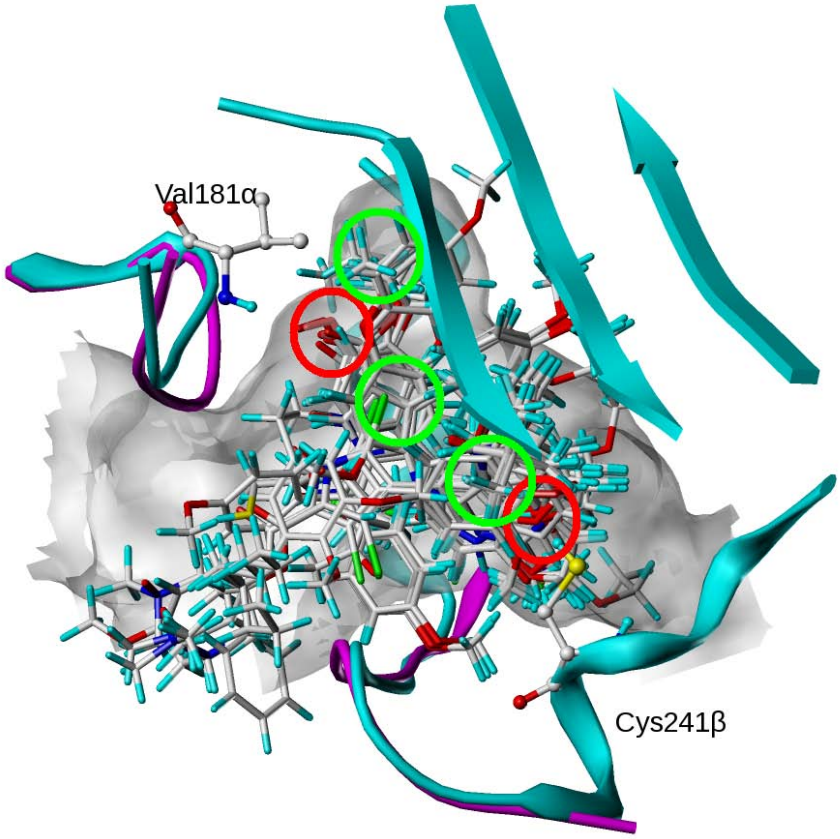


Figure 3.4. The pharmacophores and the aligned poses shown in the colchicine-site. The hydrogen bond acceptors are in red circles and the hydrophobic centers are in green circles.

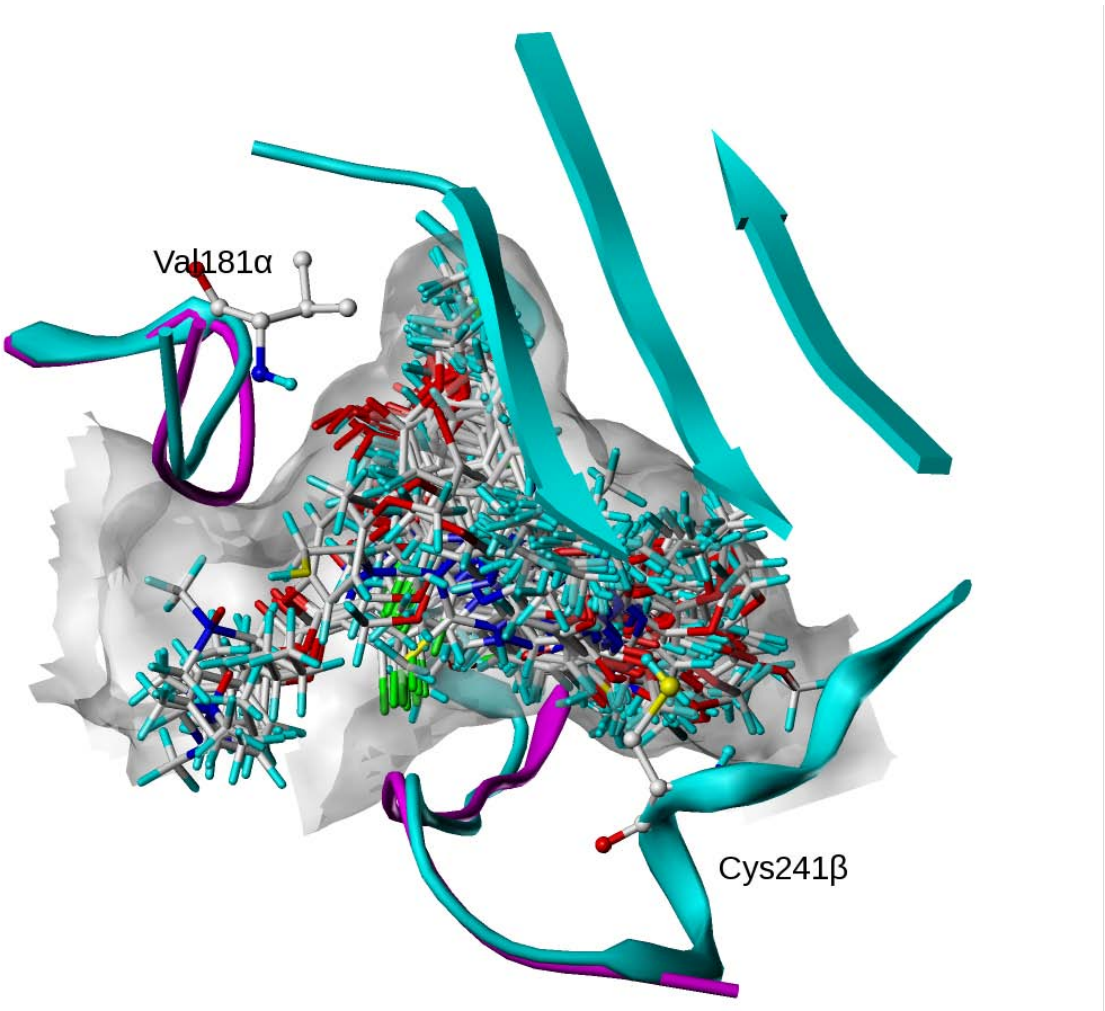


Figure 3.5. The docked poses of the colchicine site agents. The ligands predicted to not bind to the pocket due to steric clashes are not shown.

3.3.3 Analysis of QSAR statistics

Both docking-based poses and semi-ligand-based poses were used for QSAR modeling. Note that, in the cases where the ligand was unable to fit into the binding pocket due to strong steric clashes (compounds **3.32, 3.33, 3.35, 3.36, 3.38**), the docked pose was replaced by the aligned pose in the docking-based approach. In addition to the traditional CoMFA and CoMSIA methods, the hydrophobic/polar HINT

fields were added to the CoMFA fields which we call HINT-CoMFA. These field combinations have been used with previous success for hydrophobic datasets.[26] The QSAR models were evaluated based on the cross-validated correlation coefficient (q^2), the non-cross-validated correlation coefficient (r^2), standard error of estimate (SEE) and the F test value.

The statistical results listed in **Table 3.2** and the experimental and predicted activity plots in **Figure 3.6** indicate that the models built upon diverse scaffolds are reliable. The q^2 values were in the range of 0.500 to 0.621 and the r^2 values were all above 0.900. The predictability for the external test set was indicated by r_{pred}^2 and these values were in the range of 0.481 to 0.679. Compared to the CoMFA models, the combined HINT-CoMFA models gave better statistics. For the docking-based approach, HINT-CoMFA had an r^2 of 0.961 and a q^2 of 0.621, while CoMFA was somewhat less robust with an r^2 of 0.951 and a q^2 of 0.525. For the semi-ligand-based approach, HINT-CoMFA had an r^2 of 0.934 and a q^2 of 0.515, and CoMFA had an r^2 of 0.912 and a q^2 of 0.500. The HINT-CoMFA models did use more components compared to the CoMFA models (7 vs. 6 for docking-based and 7 vs. 5 for semi-ligand-based). As indicated by the higher q^2 values (0.621 vs. 0.525 for docking-based, 0.515 vs. 0.500 for semi-ligand-based), the additional components did not over-fit the model but picked up “real variance” generated by the HINT hydrophobic/polar field. The result from the external test set also demonstrated the improvement of HINT-CoMFA over CoMFA alone. HINT-CoMFA had r_{pred}^2 values of 0.638 for docking-based and 0.679 for semi-ligand-based. Standard CoMFA had r_{pred}^2 values of 0.566 and 0.481 correspondingly. The CoMSIA models were not significantly different from the HINT-CoMFA models in terms of statistics. But

their q^2, r^2 and r_{pred}^2 values were generally more or less higher than the corresponding values of the CoMFA models, indicating a slightly enhanced performance due to the additional fields provided in CoMSIA.

Table 3.2. Summary of the statistics of the 3D-QSAR models.

	Docking-Based			Semi-Ligand-Based		
	CoMFA	HINT-CoMFA	CoMSIA	CoMFA	HINT-CoMFA	CoMSIA
NOC	6	7	6	5	7	8
q^2	0.525	0.621	0.566	0.500	0.515	0.513
r^2	0.951	0.961	0.935	0.912	0.934	0.949
SEE	0.300	0.272	0.346	0.399	0.353	0.313
F value	150.020	157.000	110.199	97.344	91.005	103.006
r_{pred}^2	0.566	0.638	0.637	0.481	0.679	0.652

NOC: number of components. q^2: cross-validated correlation coefficient from leave-one-out. r^2: non-cross-validated correlation coefficient. SEE: standard error of estimate. F-value: from the F-test. r_{pred}^2: the predictive r^2 for the external test set.

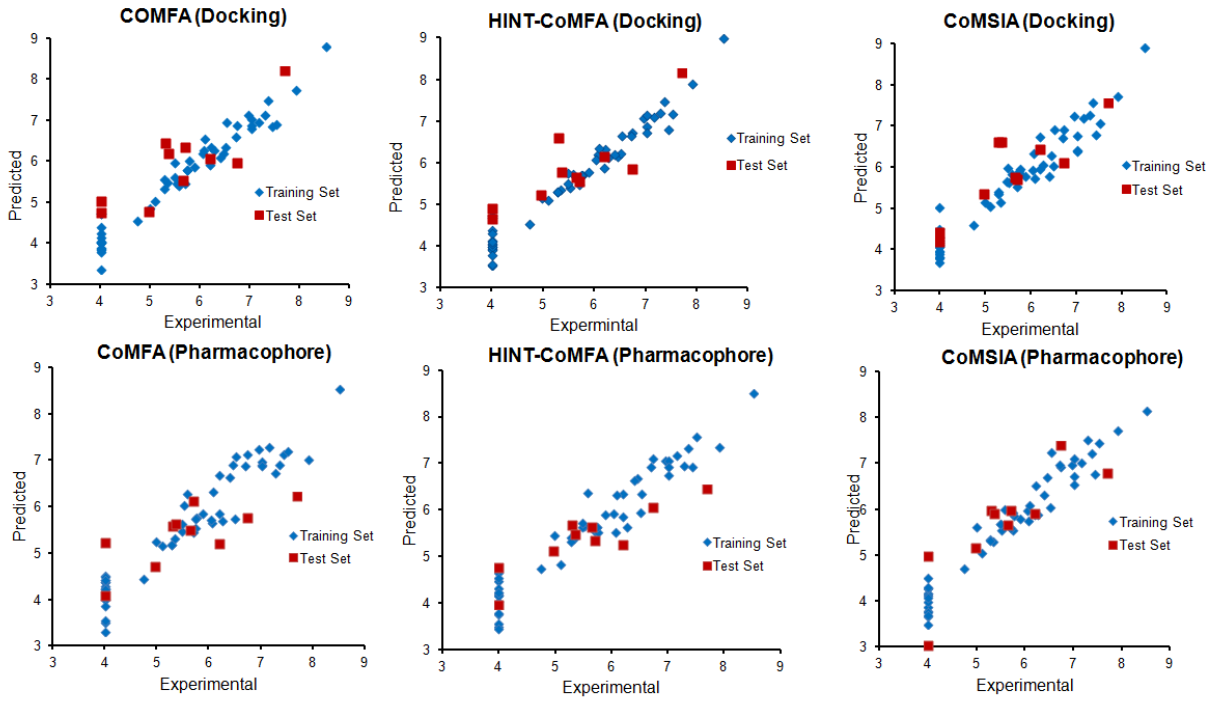


Figure 3.6. The scatter plots of the predicted pIC_{50} values verses the experimental pIC_{50} values. The CoMFA, HINT-CoMFA and CoMISA models based on either the docking-based approach or the semi-ligand-based (pharmacophore) approach are shown. The training set contains 53 compounds (blue) and the test set contains 10 compounds (red).

The docking-based and ligand-based approaches represent two scenarios in 3D-QSAR modeling. One is used when the structure of the receptor is known from experiments or homology modeling, so that the putative bioactive conformations of the ligands can be obtained from docking. The other scenario is when the structure of the receptor is unavailable or unsuitable for docking, so that the bioactive conformations of ligands have to be obtained from conformational search combined with energy minimization; the resulting conformations are overlaid based on substructure similarity and "experience". The ligand-based approach is generally regarded as better detecting the real "signals" from the real variance in the structures of the different ligands, while docking-based approach may confuse the regression method with "noise", i.e., the variation in the coordinates of a common substructure. See, for example, **Figure 3.5**, where the same functional groups seem to be "shaking" in the pocket. However, for structurally diverse ligands, the docking-based approach is more practical because the comparisons between different scaffolds are clearly indicated by their docked poses, while simple substructure similarity may not be enough for alignment using ligand-based approaches, as clearly shown in this study.

Here, in this particular study, we adopted a semi-ligand-based approach, wherein we took information from the docking to align the different scaffolds, and thus treated all ligands of the same scaffold with alignment rules based on common substructures to remove "noise". See **Figure 3.4** where the variance in structure is clear. The resulting statistics from both approaches are comparable (**Table 3.2**). No significant preference can be made towards either one approach although the docking-based alignment slightly outperformed the semi-ligand-based alignment in q^2, which was probably due to

the fact that docking placed individual functional groups more precisely in the pocket, especially those that were not identified as part of the pharmacophore. In addition, the "noise" from the docking-based approach did not seem to affect the detection of "signal" as seen with the satisfactory q^2 and r_{pred}^2 values.

3.3.4 Analysis of QSAR contour maps

The statistics from **Table 3.2** suggest the reliability of all 3D-QSAR models based on CoMFA, HINT-CoMFA and CoMSIA using either the docking-based approach or the semi-ligand-based approach. The docked poses of the ligands already revealed several residues that interact with the ligands and were expected to significantly affect activity. We analyzed the resulting contour maps of the 3D-QSAR models to see if the same features were detected and if other features were important.

The contour maps of the CoMFA, HINT-CoMFA and CoMSIA models based on the docking approach are shown in **Figure 3.7**-**3.11**. These maps identified regions that had significant impact on the activities according to the scalar products of standard deviations and coefficients used in the regression models. DAMA-colchicine is shown with the maps to facilitate spatial analysis. The evident fragmentation in contours is a result of complications in modeling due to the different scaffolds.

The CoMFA map for the docking-based models (**Figure 3.7**) suggests a favorable steric interaction (green) exists around the substitution at the ether oxygen on top of the ring C of colchicine, and beside it, a negatively charged group (red), such as the carbonyl oxygen of colchicine, would be favorable as well. Good examples from the ligand set

are the pyrrole compounds with the two distinct binding poses. The mode with higher activity (compound **3.2**) places the ester chain into the two regions with the carbonyl oxygen in the red region and the alkyl group in the green region. The mode with lower activity (compound **3.7**) places the ester chain away from the two regions because a too bulky chain could not fit as indicated by docking. From the view of the receptor, these two features correlate with the two regions. The small subpocket C correlates with the green region, so that a properly sized group would be favorable for activity but a too large group would not fit into the pocket. And the backbone NH of the residue Val181α points to the negatively charged red region to form hydrogen bonds.

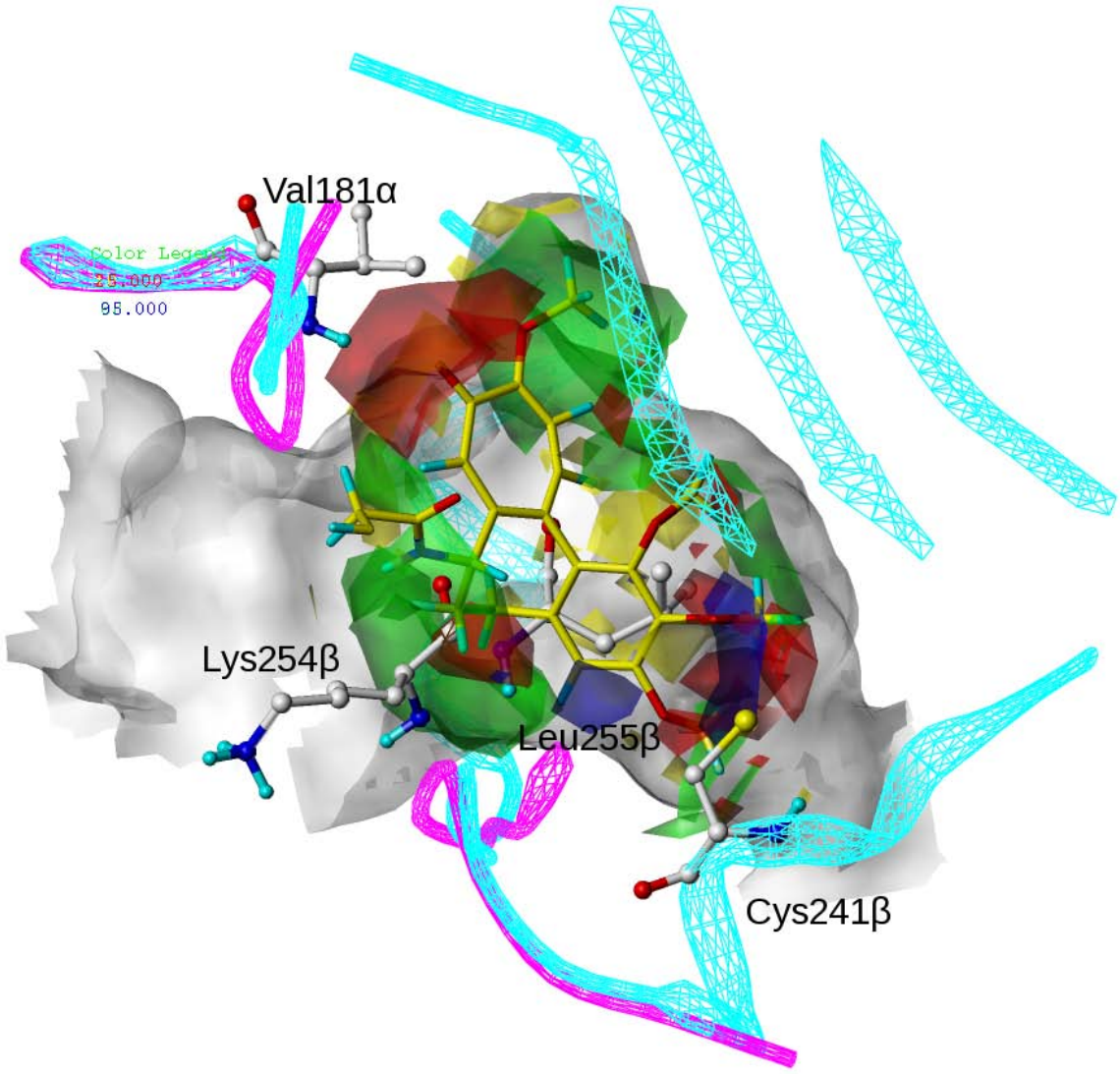


Figure 3.7. The contour map of the CoMFA model based on docked poses. Colchicine is shown in yellow. Green and yellow indicates favorable and unfavorable steric interactions, respectively. Blue regions favor electropositive groups and red regions favor electronegative groups.

The large green region around the B ring of colchicine suggests favorable steric interactions with the receptor. The pyrimidine analogues **3.43**, **3.45**, **3.51** and **3.54** that possess a methyl substitution on the amino nitrogen showed significantly higher activity than the corresponding analogues **3.44**, **3.48**, **3.52** and **3.55** that do not have the methyl

group. The methyl group lies over the ring B of colchicine and is surrounded by the four-carbon-atom-long side chain of Lys254β, which explains the favorable impact on activity. The red region right next to the green region favors negatively charged groups. The carbonyl oxygens of compounds **3.40**, **3.41** and **3.42** found in this region and they were predicted to interact with the backbone NH of Leu255β of the protein.

The contour map around the trimethoxylphenyl group (the A ring) of colchicine are mainly composed of two major green regions favorable for steric interactions, and several red regions favoring negatively charged groups and blue regions favoring positively charged groups. The two green regions represent the higher activity of combretastatin analogues that have a third methoxyl group (such as compound **3.39**) and the pyrimidine analogues that have a five-membered ring (such as compound **3.54**) beyond the third methoxyl. The available crystal structures confirm the flexibility of this portion of the pocket.[10] Ligands with distinct shapes can go further into the pocket, providing room for exploration and development of better binders. The red regions represent the importance of hydrogen bond acceptors commonly observed for good colchicine-site binders. The blue regions are actually complementary to the red regions because they can constrain the hydrogen bond acceptors in a hydrophobic environment. We explored this hydrogen bonding functionality by testing compounds with strong and weak hydrogen bond accepting ability in our previous study (Chapter 2) and the hydrogen bond donating ability of Cys241β has been related to it.[14] βIII-tubulin possess a serine mutation to the Cys241β residue.[27] The stronger hydrogen bond donating ability of serine provides an opportunity for the colchicine site agents to be selective for βIII-tubulin, an isoform overexpressed in cancer.

The few yellow regions around the ligand indicate unfavorable steric interaction. One notable yellow region is the space between ring A and ring C of colchicine. Residues from the receptor push against the ligand, leaving no room for a bulky substitution. The methoxyphenyl group at the C-3 position of compound **3.33** is predicted by docking to intrude into the protein. No antiproliferative activity or microtubule effect was observed for this compound. While Compound **3.34**, which moves the bulky methoxylphenyl to the C-5 position, remained active.

The HINT-CoMFA model generated a similar map showing the steric and electrostatic fields compared to the map generated by the CoMFA model (**Figure 3.8**). In general, the similar regions that could impact activity were identified, although the sizes and shapes of contours were somewhat different. We focus here on the map showing the HINT field (**Figure 3.9**), which provides additional hydrophobic/polar information about the model. Green contours suggest favorable hydrophobic interaction. The regions are related to the partially hydrophobic subpocket C (the alkyl side chain of Val181α), the hydrophobic alkyl side chain of Lys254β around the ring B of colchicine and the hydrophobic subpocket A surrounded by Leu248β and Leu255β. The yellow regions indicate favorable polar interactions. They are related to the hydrogen bond donors from the receptor: the backbone NH group of Val181 and the SH group of Cys241β.

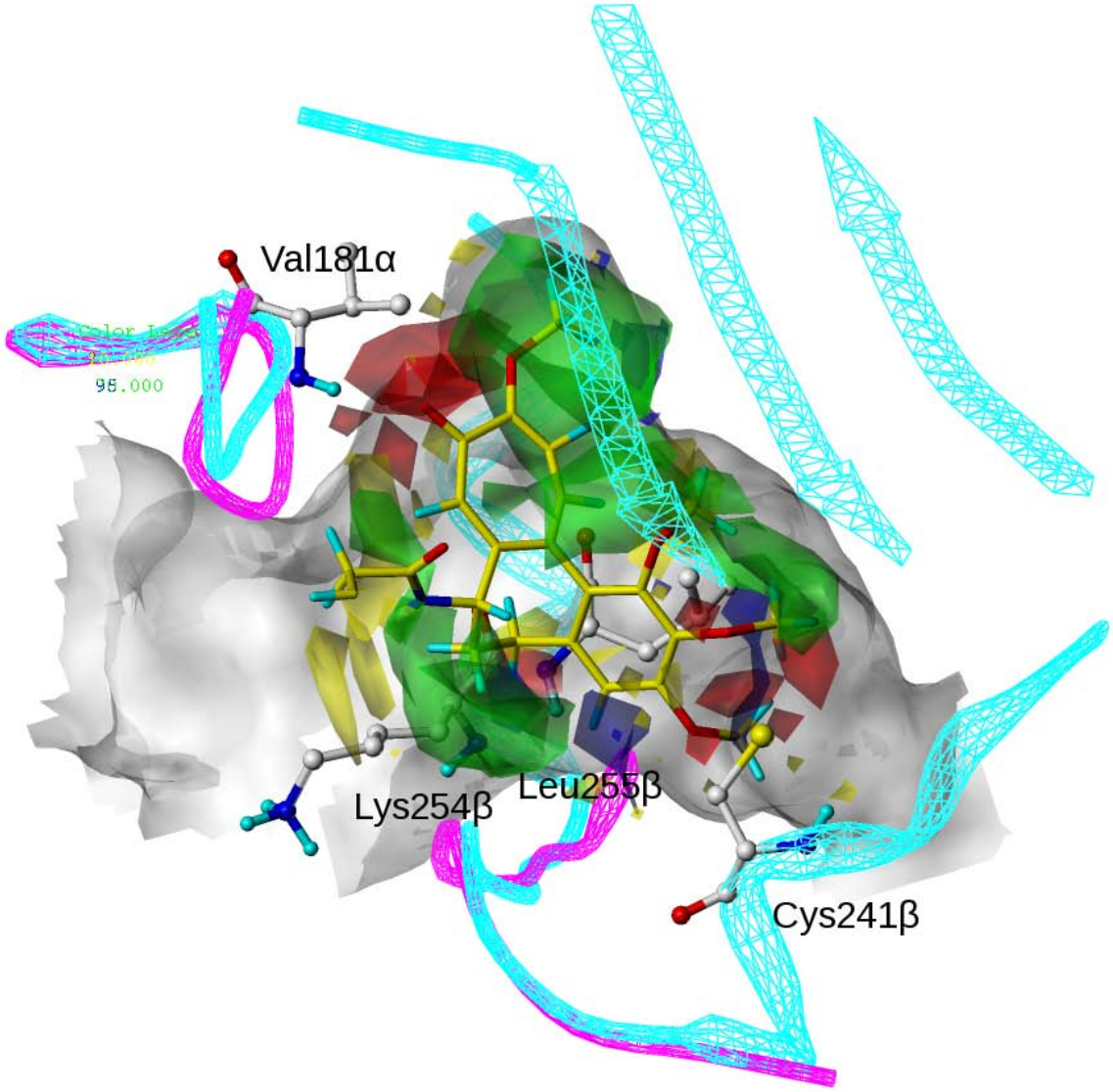


Figure 3.8. The CoMFA contour map of the HINT-CoMFA model based on docked poses. The contours represent the steric and electrostatic fields are shown. Colchicine is shown in yellow. Green and yellow indicates favorable and unfavorable steric interactions respectively. Blue regions favor electropositive groups and red regions favor electronegative groups.

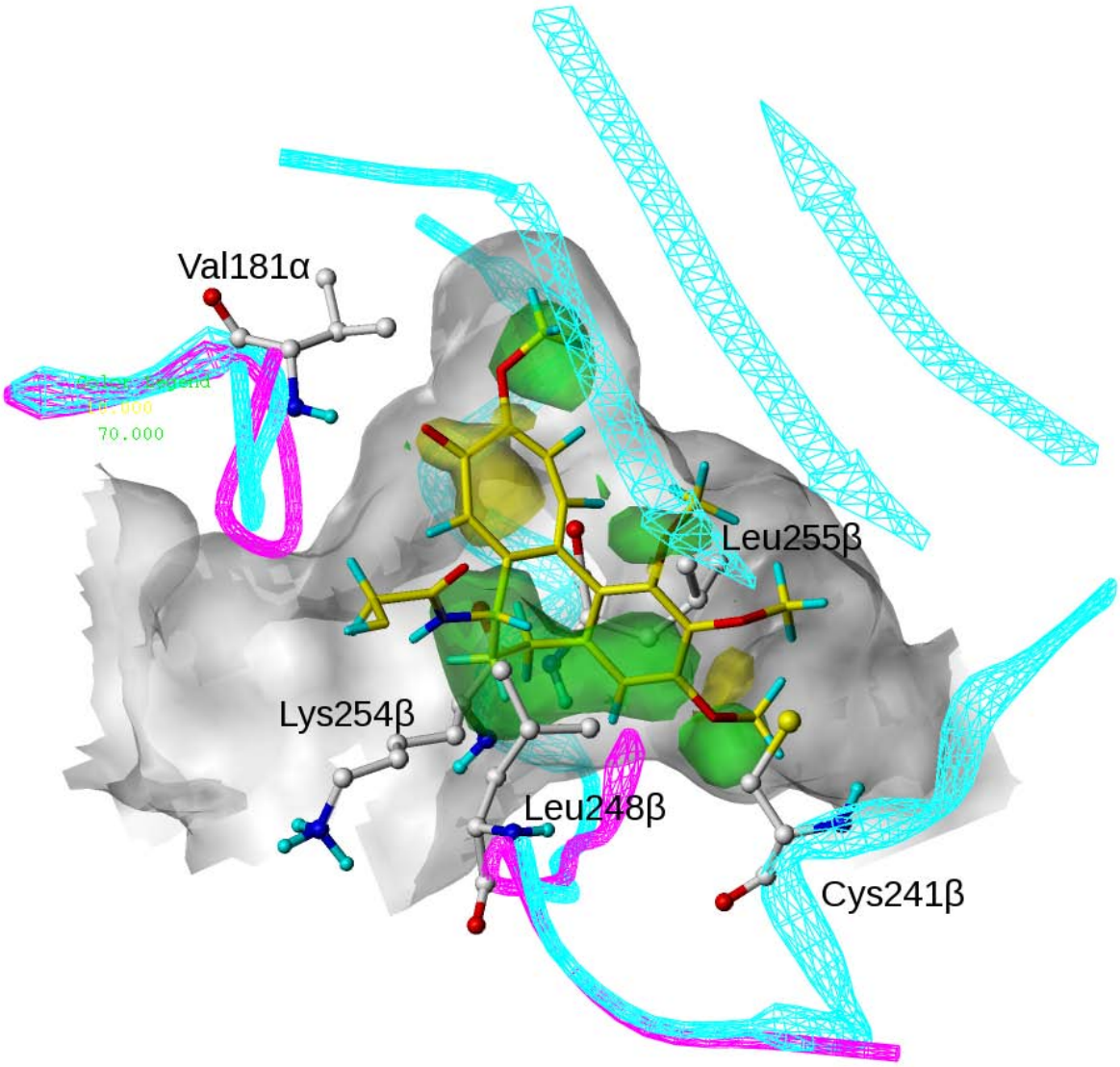


Figure 3.9. The HINT contour map of the HINT-CoMFA model based on docked poses with the contours representing the HINT hydrophobic/polar field. Colchicine is shown in yellow. Green indicates favorable hydrophobic interactions and yellow indicates favorable polar interactions.

Although differences exist, the contour map (**Figure 3.10**) of the CoMSIA model shows similar features as the maps of the CoMFA and HINT-CoMFA do: the favorable steric region (green) on top of ring C of colchicine, unfavorable steric region (yellow) in the space between ring C and ring A, favorable electronegative regions (red) around the

carbonyl oxygen and the trimethoxylphenyl group of ring A, favorable polar regions (black) around the same carbonyl oxygen and trimethoxyl group, favorable hydrophobic regions (purple) around ring B and the trimethoxyphenyl group of ring A.

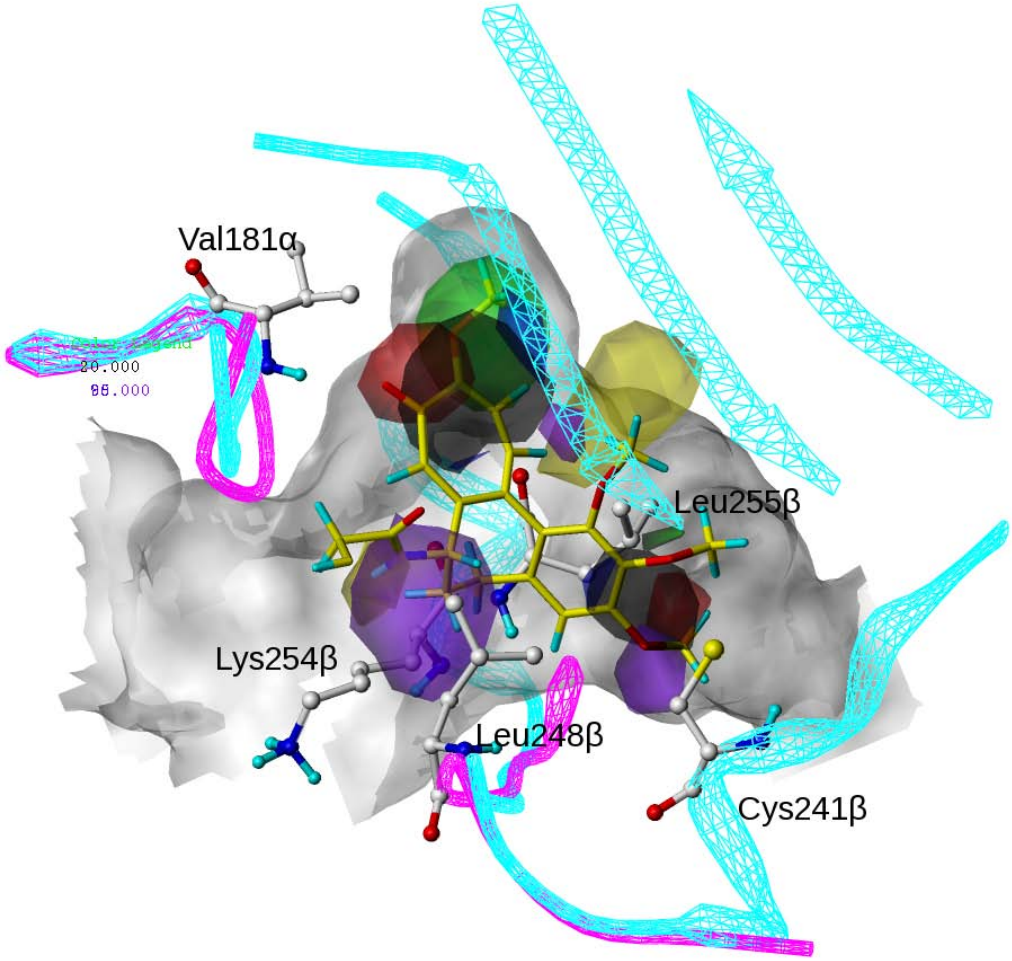


Figure 3.10. The contour map of the CoMSIA model based on docked poses. The contours represent the steric, electrostatic and hydrophobic fields from CoMSIA. Colchicine is shown in yellow. Green and yellow indicates favorable and unfavorable steric interactions respectively. Blue regions favor electropositive groups and red regions favor electronegative groups. Purple and black indicate favorable hydrophobic and polar interactions respectively.

The CoMSIA hydrogen bond donor and acceptor field types that CoMFA and HINT-CoMFA lack provide additional information (**Figure 3.11**). The blue region around the

backbone carboxyl oxygen of Thr179α represent a favorable hydrogen bond donating feature of the pyrrole compounds (such as compound **3.2**) which contain a NH group on the pyrrole core. The red region next to it can be related to the favorable hydrogen bond accepting feature of the other binding mode (compound **3.7**), whose carbonyl oxygen interacts with the side chain NH_2 of Asn101α. The other regions provide the same information as the other models do.

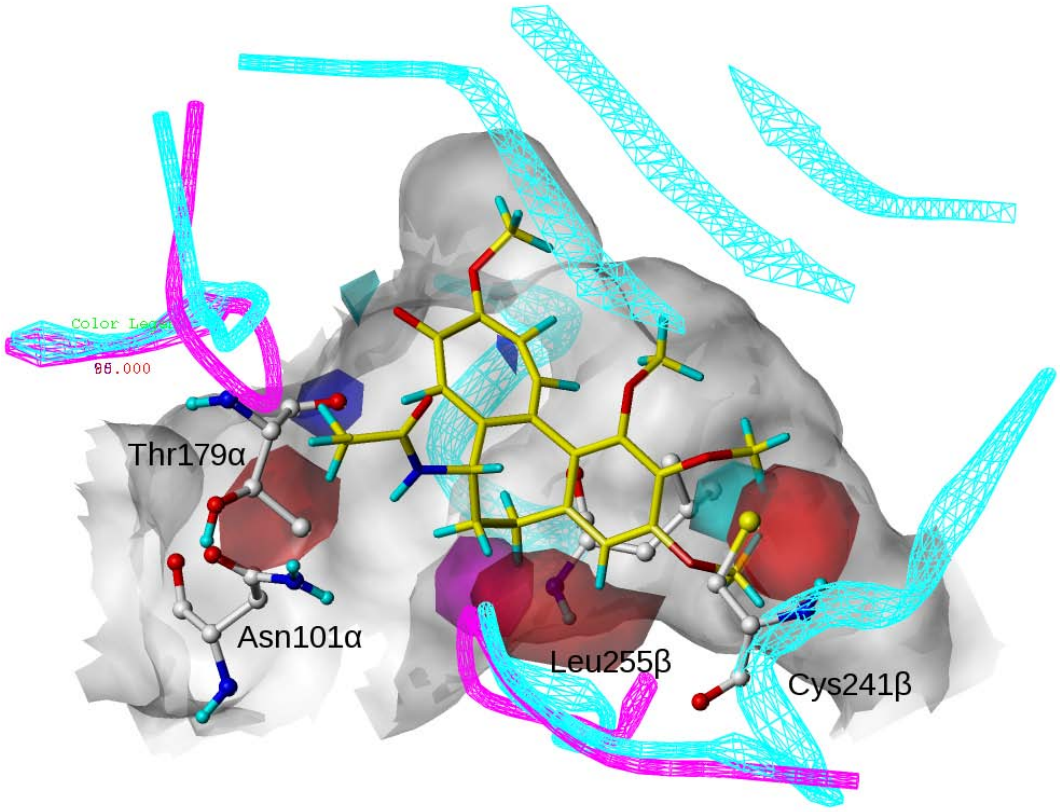


Figure 3.11. The contour map of the CoMSIA model based on docked poses. The contours represent the hydrogen bond donor and acceptor fields from CoMSIA. Colchicine is shown in yellow. Blue and cyan represent regions that favor and not favor hydrogen bond donors. Red and magenta represent regions that favor and not favor hydrogen bond acceptors.

The maps of the CoMFA, HINT-CoMFA and CoMSIA models based on the semi-ligand approach (**Figure 3.12**-**3.16**) agreed with maps based on the docking-based approach and did not show significant differences. Most of the features that were identified by the docking-based approach appeared on the maps based on the semi-ligand approach as well.

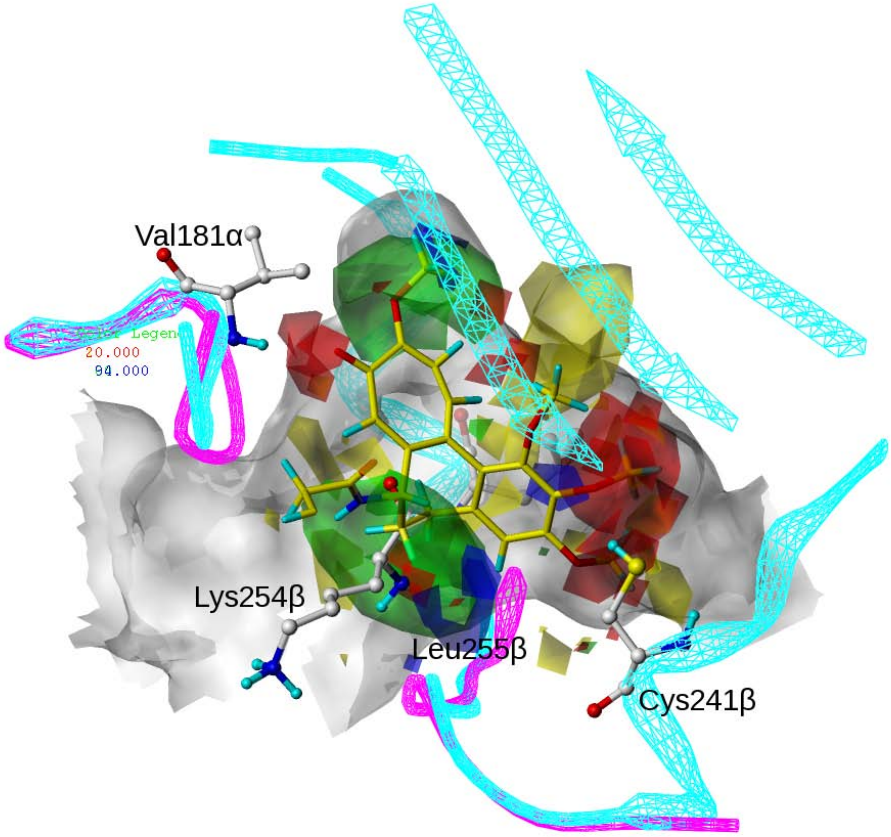


Figure 3.12. The contour map from CoMFA based on the semi-ligand approach. Colchicine is shown in yellow. Green and yellow indicates favorable and unfavorable steric interactions respectively. Blue regions favor electropositive groups and red regions favor electronegative groups.

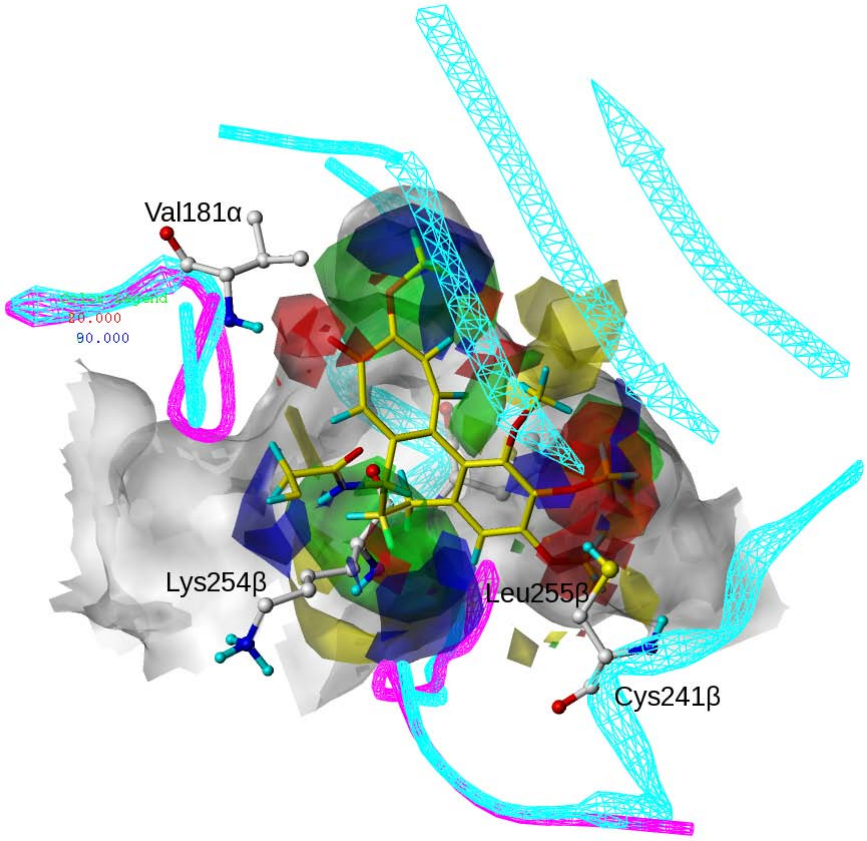


Figure 3.13. The contour map from the HINT-CoMFA model based on the semi-ligand approach. The contours represent the steric and electrostatic fields from CoMFA are shown. Colchicine is shown in yellow. Green and yellow indicates favorable and unfavorable steric interactions respectively. Blue regions favor electropositive groups and red regions favor electronegative groups.

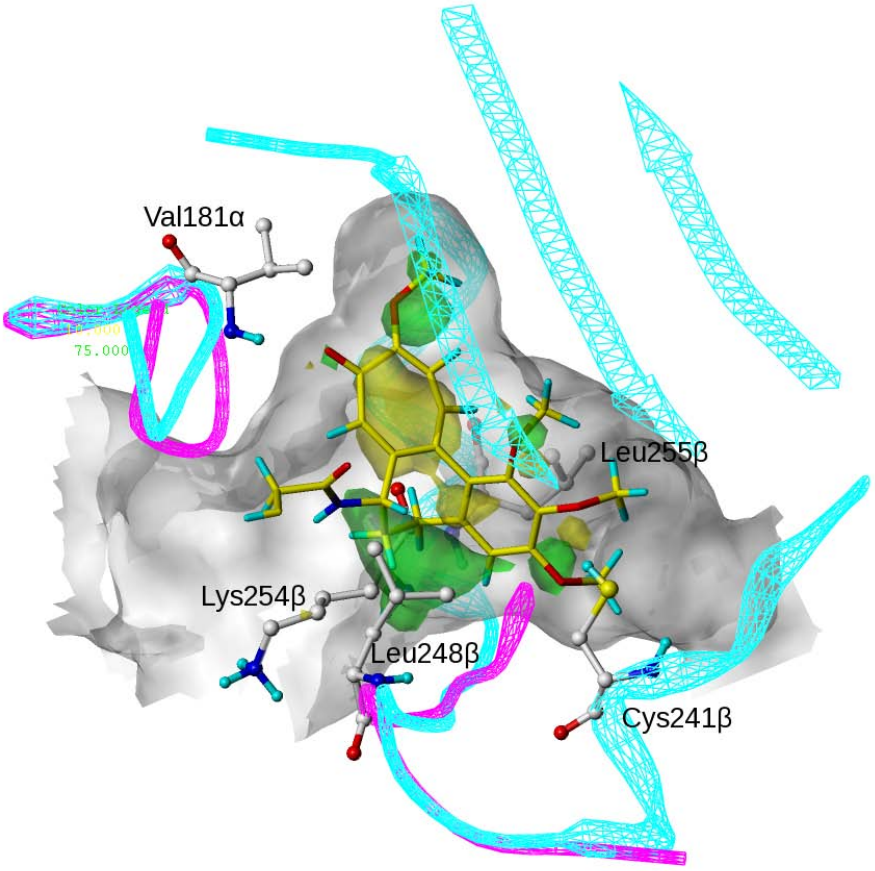


Figure 3.14. The contour map from the HINT-CoMFA model based on the semi-ligand approach with the contours representing the HINT hydrophobic/polar field. Colchicine is shown in yellow. Green indicates favorable hydrophobic interactions and yellow indicates favorable polar interactions.

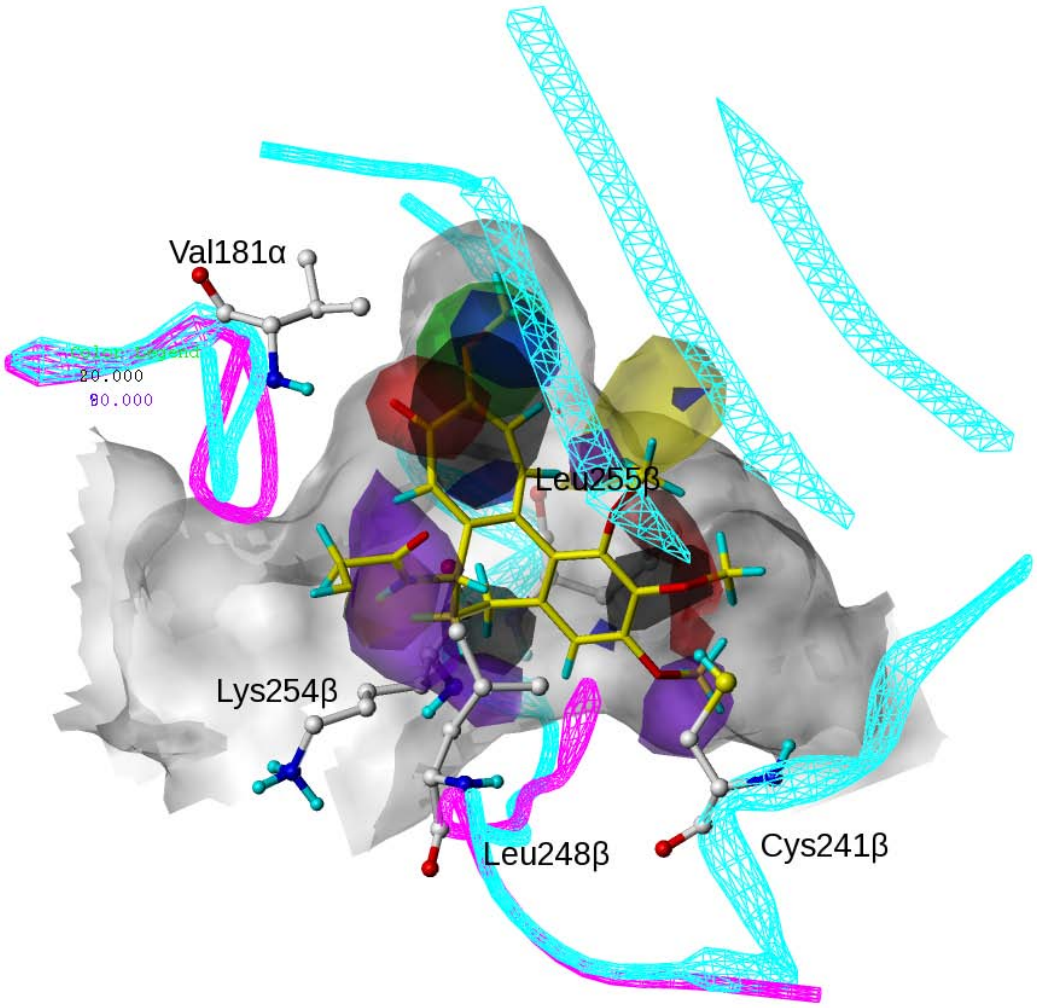


Figure 3.15. The contour map from the CoMSIA model based on the semi-ligand approach. The contours represent the steric, electrostatic and hydrophobic fields from CoMSIA. Colchicine is shown in yellow. Green and yellow indicates favorable and unfavorable steric interactions respectively. Blue regions favor electropositive groups and red regions favor electronegative groups. Purple and black indicate favorable hydrophobic and polar interactions respectively.

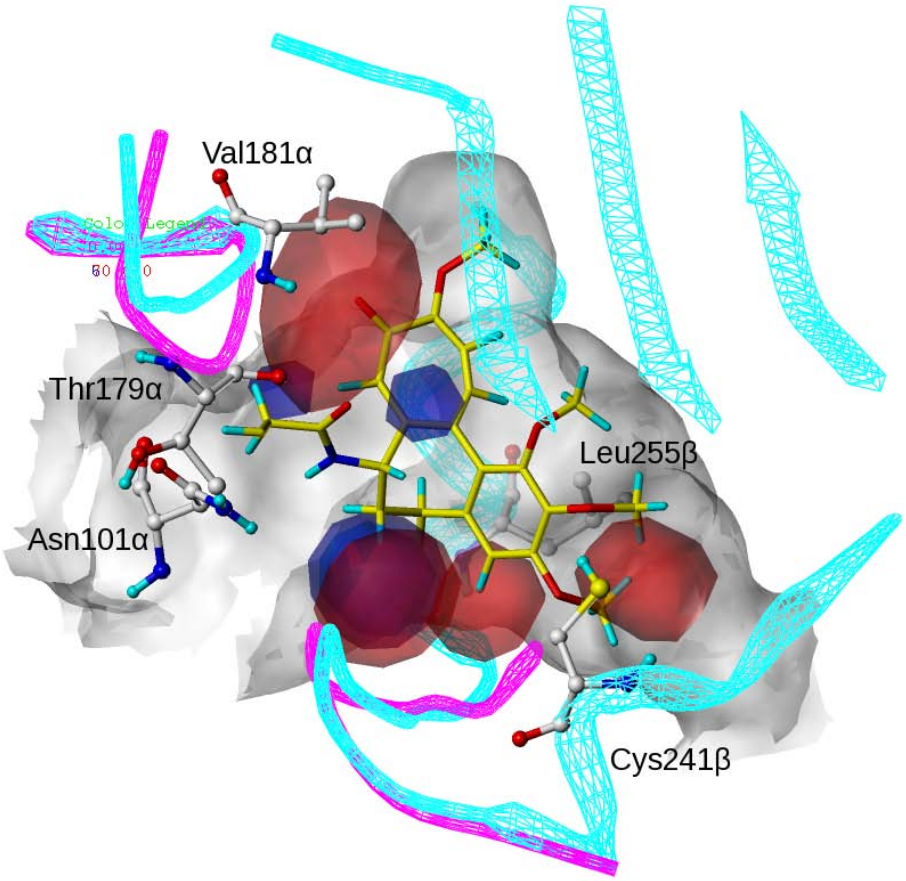


Figure 3.16. The contour map from the CoMSIA model based on the semi-ligand approach. The contours represent the hydrogen bond donor and acceptor fields from CoMSIA. Colchicine is shown in yellow. Blue and cyan represent regions that favor and not favor hydrogen bond donors. Red and magenta represent regions that favor and not favor hydrogen bond acceptors.

The contour maps of all 3D-QSAR models correlate very well with the structure of the colchicine site. Most identified regions that had significant impact on activity can be explained by the functions of the neighboring residues, which indicates the binding poses of the compounds were most likely correctly predicted. Another important observation is that the phamacophores of the colchicine-site binders were validated by the 3D-QSAR models. The hydrogen bond acceptors related to Cys241β and Val181α,

and the three hydrophobic centers in subpockets A, B and C, are all shown on the contour maps. 3D-QSAR analysis is a technique based on statistical correlation. It means that these pharmacophores are not only present in the highly active colchicine-site binders, as we see them in the docking study; removing them can also decrease activity, as indicated in 3D-QSAR by analyzing the activity change for compounds with and without these pharmacophores in the dataset. The other features on the maps were also identified to have significant impact on activity. They are particularly important for improving the activities of the highly active compounds because it is possible that these features have not been consolidated in one or more to these compounds as of yet. One such feature will be discussed later.

The docking-based and the semi-ligand-based approaches generated reliable models and the two types of models were comparable in terms of statistics and contour maps. The similarity was due to the fact that the semi-ligand-based approach adopted the information from the docked poses to align different scaffolds, which proved to be effective because both model generated satisfying statistics and explainable contour maps. The “noise” present in the docked poses did not affect the statistics or generate significant differences in contour maps compared to the semi-ligand-based approach. It can be explained by: 1) most compounds followed their representatives, the most active compounds in different categories, to adopt similar docked poses (Figure **3.2** and Figure **3.3**). Their shapes were complementary to the binding pocket so little room was left for them to move around to create “noise”; 2) the docked poses that were significantly different from the others due to serious steric clashes were replaced by the aligned poses (Compound **3.32, 3.33, 3.35, 3.36, 3.38**); and 3) the abundant structural

changes across different scaffolds significantly outweighed the “noise”. In fact, as more residues were involved in binding than indicated by the pharmacophores identified for the semi-ligand-based approach, docking was able to place the ligands more delicately in the colchicine site, which might contribute to the better performance (higher q^2, r^2, etc.) in the model validations.

3.3.5 Highlight important features of compounds using an overall HINT map

The previously mentioned CoMFA, HINT-CoMFA and CoMSIA models and corresponding contour maps were based on statistical analysis. The success of a statistical model depends on many factors including accuracy of the input data and variance in the training set and test set compounds. In the case of a selection of unique compounds, where the features of any compound that affect activity significantly are not covered by any other compounds in the dataset, a statistical analysis based on cross-validation would disqualify the model. The uniqueness of each compound, however, may be valuable for drug design ideas. Thus, we introduced a simple linear combination of HINT maps of compounds to highlight the combination of uniqueness and commonality.

The idea is to calculate the HINT maps that contain a hydrophobic/polar field and an acid/base field for each compound. Then a “weight” that corresponds to the activity of individual compound is applied to each of these maps. An overall HINT map that represents the whole set is constructed by a linear combination of the individual weighted maps by addition of field values. Two types of information are highlighted by the method. First, the highly active compounds would have higher weights against

weaker compounds, so that the unique features of the highly active compounds would be shown to be more significant. Second, the common features of the maps would also be significant due to this addition. Two factors are critical for constructing a useful overall HINT map. The weight should be able to distinguish active compounds from weak compounds. As the number of weak compounds increases, the field values from the addition of weak compounds can exceed the field values of the highly active compounds. Mathematical operations such as exponentiation can be used to expand the distance between the weight of a highly active compound and the weight of a weaker compound. In this study, we simply applied a one order operation, using 0.1 (μM)/IC50 (μM) as the weight for each compound. Another important factor is the choice of contour values for the display of the final overall HINT map. The selection of contour values can be iteratively obtained, with the aim of obtaining desired features from the map.

The overall HINT map based on all the colchicine site agents from the entire dataset is shown in **Figure 3.17**. The map agrees with the contour maps from the 3D-QSAR analysis. The features shown on the HINT overall map include the hydrogen bond acceptors related to Cys241β and Val181α, and the large hydrophbobic area covering subpocket A, B and C of the site. These features agree with the pharmacophores identified from docking, representing the commonality of the highly active compounds and the whole dataset. One unique feature shown on the map belongs to the most active pyrrole analogue compound **3.2** (also common to most of the pyrrole analogues). This feature is the hydrogen donating ability of the NH on the pyrrole core interacting with Thr179α. Another unique feature is the large green region next to the beta strands,

which is extended from the boundary of the trimethoxyphenyl ring A of colchicine (**Figure 3.18**). The extension was generated partially by the cyclopentapyridine groups of the highly active pyrimidine analogues such as compound **3.54** (**Figure 3.18**), and was also identified by the statistical 3D-QSAR models (as favorable steric green regions surrounding the trimethoxyphenyl group of colchicine in **Figure 3.7** and **3.8,** and in a clearer view in **Figure 3.19**), the same information is directly shown in the overall HINT map.

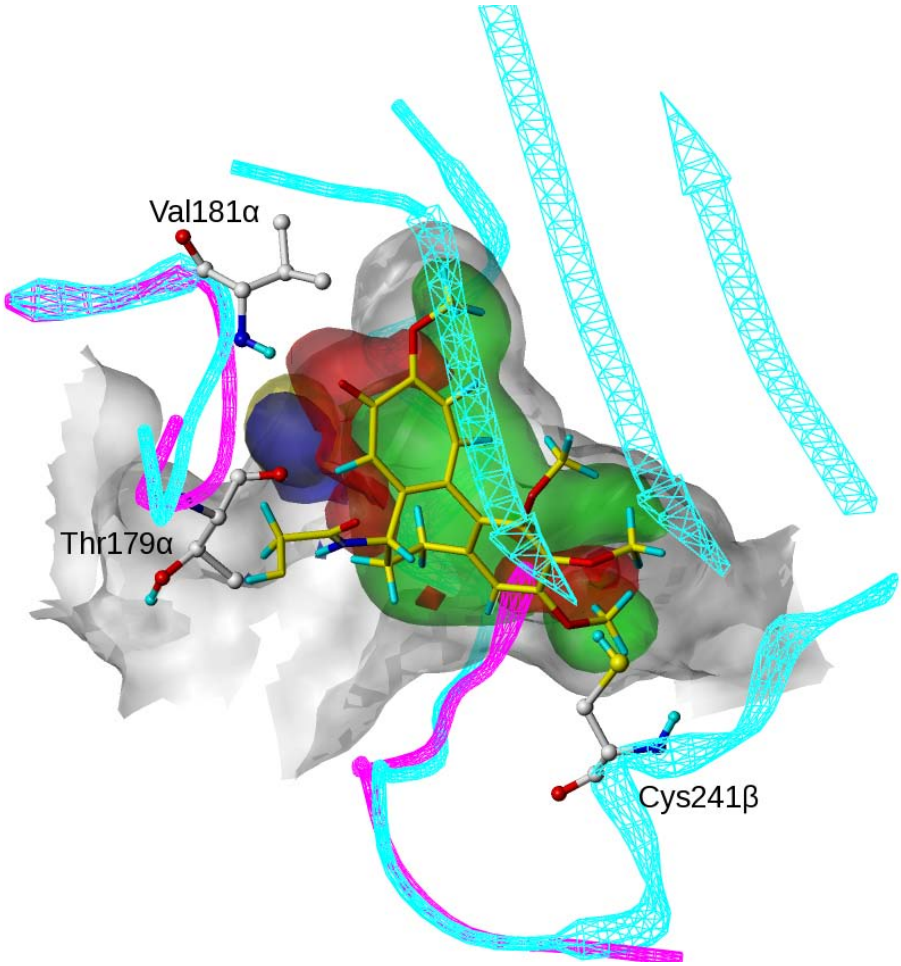


Figure 3.17. The overall HINT map based on the whole set of colchicine-site agents. Colchicine is shown in yellow. Green and yellow represent hydrophobic and polar regions of the compound set. Blue and red represent acidic and basic regions of the compound set.

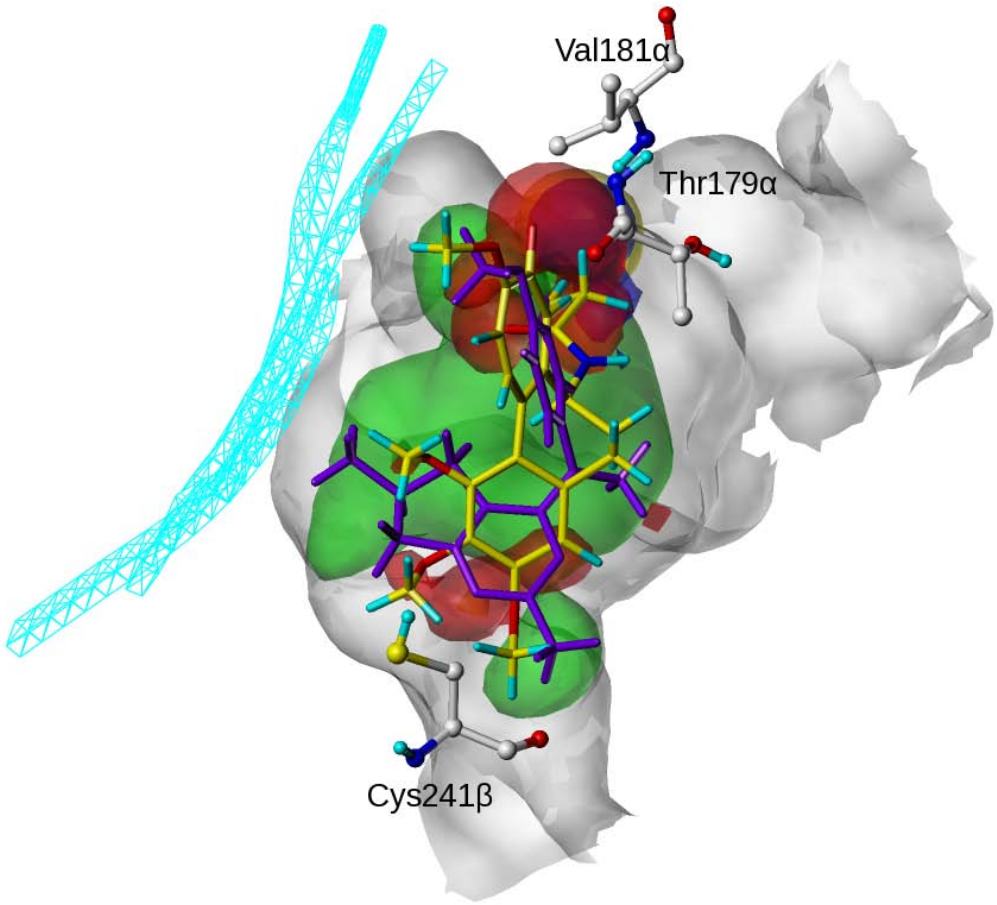


Figure 3.18. The overall HINT map based on the whole set of colchicine-site agents (a different view compared to **Figure 3.17**). Colchicine is shown in yellow and compound **3.54** in purple. The green region not covered by the trimethoxylphenyl group of colchicine indicates room for exploration.

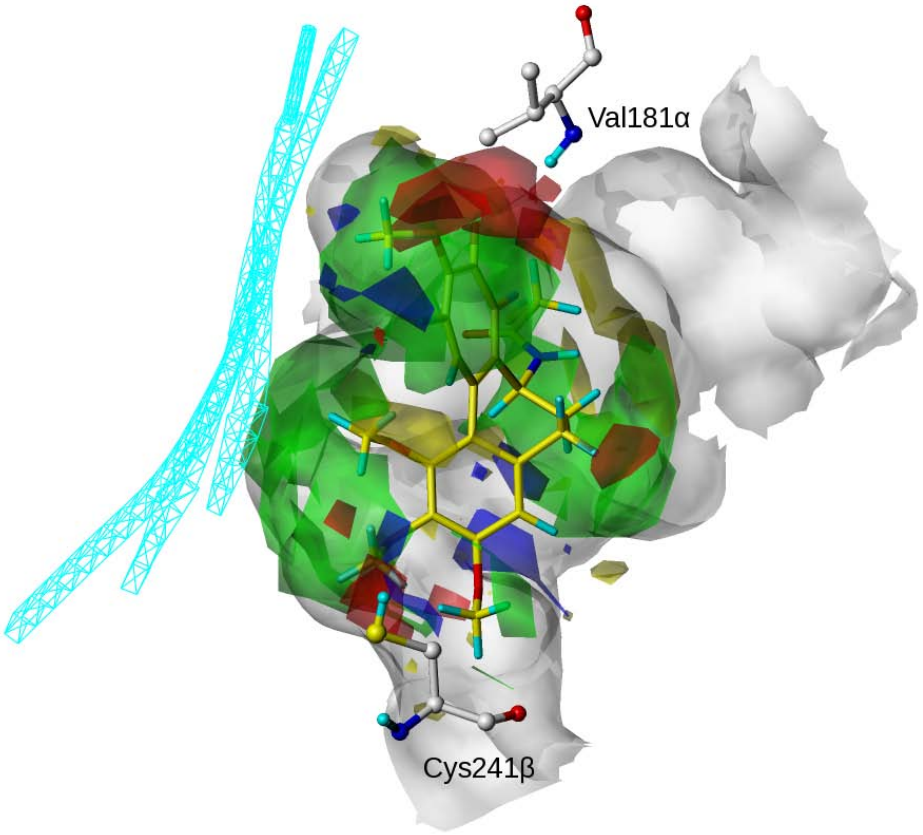


Figure 3.19. A different view of the contour map of the HINT-CoMFA model based on docked poses (**Figure 3.8**). Colchicine is shown in yellow. Green and yellow indicates favorable and unfavorable steric interactions respectively. Blue regions favor electropositive groups and red regions favor electronegative groups.

Compared to the overall HINT map approach, a statistical 3D-QSAR model possesses the following advantages. First, the important features of compounds identified by statistics have a higher chance being the ones that affect activity. In the case of the non-statistical HINT map, however, the features identified are dominated by those present in the highly active compounds. Whether removing them would affect activity is not tested. As shown in the contour maps of 3D-QSAR models, identified contours are usually fragmented, indicating how specific features in specific areas affect activity

(**Figure 3.7**). In the overall HINT map, features that may or may not affect activity are gathered together, usually giving smooth contours (**Figure 3.17**). Second, due to the same reason, the features that can be detected by comparing a modestly-active compound and a weaker compound may not be detected by the HINT map approach. However, as stated before, the success of 3D-QSAR modeling depends on many factors. In the cases of compounds having multiple scaffolds, especially when different compounds are not tested in the same lab, a slight disagreement among scaffolds on how one functional group affects activity might generate unattractive statistics and therefore confuse users of whether to use the results such as the contour maps. The overall HINT map approach is more relaxed on statistical validation but more focused on hypothesis generation. Its advantages are mostly shown when dealing with an unstructured dataset with multiple scaffolds and inconsistent activity measurements.

The overall HINT map approach represents an alternative to the statistical 3D-QSAR approach whose success relies heavily on the quality of the data set. Combining the two approaches would provide valuable information in both the early-stage and late-stage of drug design.

3.3.6 Summary of features identified by 3D-QSAR analysis for colchicin-site agents

We summarize here in **Figure 3.20** the features identified by 3D-QSAR analysis for the colchicine-site agents. Colchicine is shown as the binding ligand in the site. A favorable ligand would have four hydrogen bond acceptors interacting with Val181α, Asn101α, Leu255 β and Cys241β, a hydrogen bond donor interacting with Thr179α, three

hydrophobic centers interacting with three hydrophobic subpockets and an extra hydrophobic group for the empty volume next to strands S8 and S9 as indicated in the figure. Colchicine lacks the hydrogen bond acceptors for Asn101α and Leu255 β, the hydrogen bond donor for Thr179α and the extra hydrophobic group for the room, which could be considered for optimization.

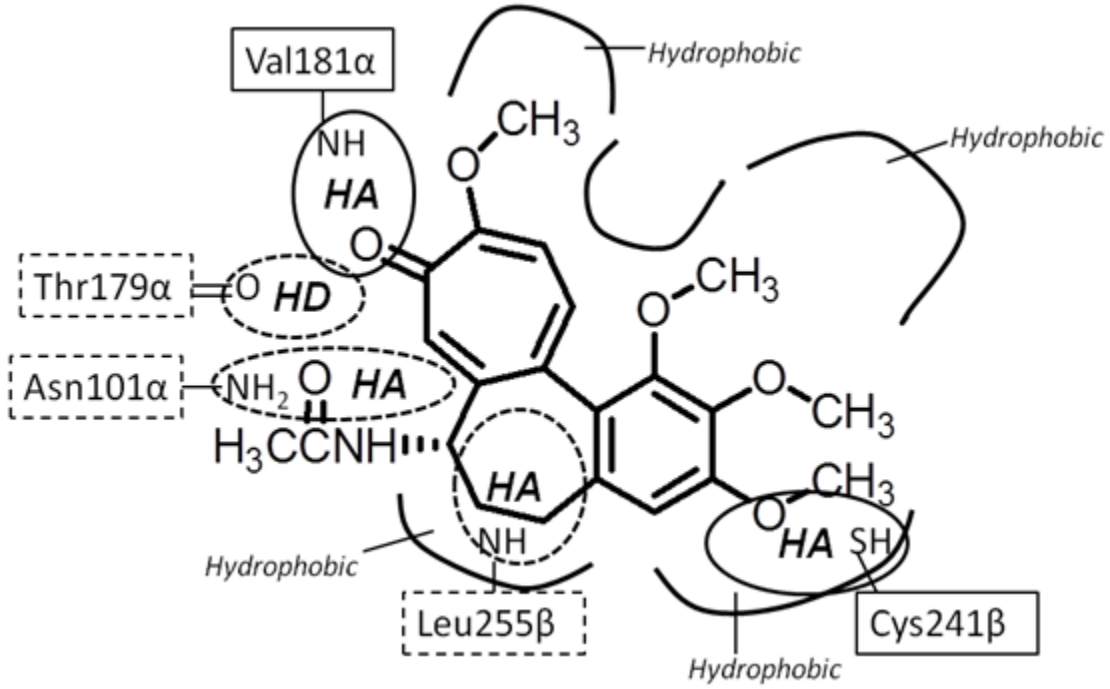


Figure 3.20. The summary of features identified by 3D-QSAR analysis for colchicine-site agents. HA indicates a favorable hydrogen bond acceptor and HD indicates a favorable hydrogen bond donor. The oval circles the favorable functional group and the interacting residue. Solid lines indicate that the interaction is present for colchicine and dashed lines indicate that the interaction is not present. The curves indicate the shape of the colchicine site.

3.4 Conclusions

A number of selected colchicine-site agents tested in the same laboratory were studied in order to understand their interactions with the colchicine site on microtubules. By

applying ensemble docking using the HINT scoring function, the binding conformations of the compounds and the related receptor structures were identified. The pharmacophore model for the ligands contains a hydrogen bond acceptor interacting with Cys241β, another hydrogen bond acceptor interacting with Val181α, and three hydrophobic centers in the subpockets A, B and C, respectively.

The docked conformations as well as the aligned conformations based on the pharmacophores were then used to construct 3D-QSAR models. In addition to the traditional CoMFA and CoMSIA methods, a hydrophobic/polar HINT field was combined with CoMFA to form the HINT-CoMFA method, and the HINT field proved to be a good supplement to the CoMFA fields. The cross-validated correlation coefficients (q^2) using the leave-one-out (LOO) method, the non-cross-validated regression correlation coefficients (r^2) and the predictive r^2 (r_{pred}^2) using an external test set were 0.525, 0.951 and 0.566 for the CoMFA model, 0.621, 0.961 and 0.638 for the HINT-CoMFA model, and 0.566, 0.935 and 0.637 for the CoMSIA model, all based on the docked conformations. The corresponding statistics based on the aligned conformations were 0.500, 0.912 and 0.481 for the CoMFA model, 0.515, 0.934 and 0.679 for the HNT-CoMFA model, and 0.513, 0.949 and 0.652 for the CoMSIA model. The contour maps of the 3D-QSAR models were analyzed and compared with the binding site of the protein. The regions indicating favorable and unfavorable interactions can be directly related to the specific residues. The statistics and the agreement between the contour maps and the structure of the receptor demonstrated the robustness and the reliability of the models.

In addition to the statistical 3D-QSAR approach, we introduced the overall HINT map approach, which is a linear combination of weighted HINT maps of individual compounds. The overall HINT map highlights the uniqueness of the highly active compounds and the commonality of all the compounds in the data set. In this study, the HINT overall map agrees well with the receptor and the 3D-QSAR models.

By combining different approaches, including statistical 3D-QSAR methods and constructing a non-statistical HINT overall map, detailed insights for how the ligand structure affects activity and interactions with the colchicine site were understood. The models will be a help for optimization and design of colchicine-site agents.

References

1. Jordan, M. A.; Wilson, L. Microtubules as a Target for Anticancer Drugs. *Nat. Rev. Cancer.* **2004**, *4*, 253-265.
2. Dumontet, C.; Jordan, M. A. Microtubule-Binding Agents: A Dynamic Field of Cancer Therapeutics. *Nat. Rev. Drug Discovery.* **2010**, *9*, 790-803.
3. Bennett, M. J.; Barakat, K.; Huzil, J. T.; Tuszynski, J.; Schriemer, D. C. Discovery and Characterization of the Laulimalide-Microtubule Binding Mode by Mass Shift Perturbation Mapping. *Chem. Biol.* **2010**, *17*, 725-734.
4. Stanton, R. A.; Gernert K. M.; Nettles, J. H.; Aneja, R. Drugs That Target Dynamic Microtubules: A New Molecular Perspective. *Med. Res. Rev*. **2011**, *31*, 443-481.
5. Lu, Y.; Chen, J.; Xiao, M.; Li, W.; Miller, D. D. An Overview of Tubulin Inhibitors that Interact with the Colchicine Binding Site. *Pharm. Res.* **2012**, *29*, 2943-2971.
6. Seve, P.; Dumontet C. Is Class III Beta-tubulin a Predictive Factor in Patients Receiving Tubulin-binding Agents? *Lancet Oncol.* **2008**, *9*, 168-175.
7. Stengel, C.; Newman, S. P.; Leese, M. P.; Potter, B. V.; Reed, M. J.; Purohit, A. Class III Beta-tubulin Expression and in vitro Resistance to Microtubule Targeting Agents. *Br. J. Cancer.* **2010**, *102*, 316-324.
8. Chen, J.; Liu, T.; Dong, X.; Hu, Y. Recent Development and SAR Analysis of Colchicine Binding Site Inhibitors. *Mini-Rev. Med. Chem.* **2009**, *9*, 1174-1190.

9. Ravelli, R. B.; Gigant, B.; Curmi, P. A.; Jourdain, I.; Lachkar, S.; Sobel, A.; Knossow, M. Insight into Tubulin Regulation from a Complex with Colchicine and a Stathmin-like Domain. *Nature*. **2004**, *428*, 198-202.
10. Dorleans, A.; Gigant, B.; Ravelli, R. B.; Mailliet, P.; Mikol, V.; Knossow, M. Variations in the Colchicine-binding Domain Provide Insight into Structural Switch of Tubulin. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 13775-13779.
11. Barbier, P.; Dorleans, A.; Devred, F.; Sanz, L.; Allegro, D.; Alfonso, C.; Knossow, M.; Peyrot, V.; Andreu, J. M. Stathmin and Interfacial Microtubule Inhibitors Recognize a Naturally Curved Conformation of Tubulin Dimers. *J. Biol. Chem.* **2010**, *285*, 31672-31681.
12. Chakraborti, S.; Chakravarty, D.; Gupta, S.; Chatterji, B. P.; Dhar, G.; Poddar, A.; Panda, D.; Chakrabarti, P.; Ghosh, D. S.; Bhattacharyya, B. Discrimination of Ligands with Different Flexibilities Resulting from the Plasticity of the Binding Site in Tubulin. *Biochemistry*. **2012**, *51*, 7138-7148.
13. Kellogg, E. G.; Abraham, D. J. Hydrophobicity: Is LogPo/w More than the Sum of Its Parts? *Eur. J. Med. Chem.* **2000**, *35*, 651-661.
14. Da, C.; Telang, N.; Gupton, J. T.; Mooberry, S. L.; Kellogg, G. E. Developing Novel C-4 Analogues of Pyrrole-Based Antitubulin Agents: Weak but Critical Hydrogen Bonding in the Colchicine Site. *Med. Chem. Commun.* **2013**, *4*, 417-421.
15. Da, C.; Telang, N.; Barelli, P.; Jia, X.; Gupton, J. T.; Mooberry, S. L.; Kellogg, G. E. Pyrrole-Based Antitubulin Agents: Two Distinct Binding Modalities are Predicted for C-2 Analogues in the Colchicine Site. *ACS Med. Chem. Lett.* **2012**, *3*, 53–57.
16. Tripathi, A.; Fornabaio, M.; Kellogg, G. E.; Gupton, J. T.; Gewirtz, D. A.; Yeudall, W. A.; Vega, N. E.; Mooberry, S. L.; Docking and Hydrophobic Scoring of Polysubstituted Pyrrole Compounds with Antitubulin Activity. *Bioorg. Med. Chem*. **2008**, *16*, 2235-2242.
17. Mooberry, S. L.; Weiderhold, K. N.; Dakshanamurthy, S.; Hamel, E.; Banner, E. J.; Kharlamova, A.; Hempel, J.; Gupton, J. T. Brown, M. L. Identification and Characterization of a New Tubulin-Binding Tetrasubstituted Brominated Pyrrole. *Mol. Pharmacol.* **2007**, *72*, 132-140.
18. Lee, M; Brockway, O.; Dandavati, A.; Tzou, S.; Sjoholm, R.; Satam, V.; Westbrook, C.; Mooberry, S. L.; Zeller, M.; Babu, B.; Lee, M. A Novel Class of trans-Methylpyrazoline Analogues of Combretastatins: Synthesis and in-vitro Biological Testing. *Eur. J. Med. Chem.* **2011**, *46*, 3099-3104.
19. Lee, L.; Robb, L. M.; Lee, M.; Davis, R.; Mackay, H.; Chavda, S.; Babu, B.; O'Brien, E. L.; Risinger, A. L.; Mooberry, S. L.; Lee, M. Design, Synthesis, and Biological Evaluations of 2,5-Diaryl-2,3-Dihydro-1,3,4-Oxadiazoline Analogues of Combretastatin-A4. *J. Med. Chem.* **2010**, *53*, 325-334.

20. Gangjee, A.; Pavana, R. K.; Li, W.; Hamel, E.; Westbrook, C.; Mooberry, S. L. Novel Water-Soluble Substituted Pyrrolo[3,2-d]pyrimidines: Design, Synthesis, and Biological Evaluation as Antitubulin Antitumor Agents. *Pharm. Res.* **2012**, *29*, 3033-3039.
21. Gangjee, A.; Zhao, Y.; Hamel, E.; Westbrook, C.; Mooberry, S. L. Synthesis and Biological Activities of (R)- and (S)-N-(4-Methoxyphenyl)-N,2,6-Trimethyl-6,7-Dihydro-5H-Cyclopenta[d]pyrimidin-4-Aminium Chloride as Potent Cytotoxic Antitubulin Agents. *J. Med. Chem.* **2011**, *54*, 6151-6155.
22. Gangjee, A.; Zhao, Y.; Lin, L.; Raghavan, S.; Roberts, E. G.; Risinger, A. L.; Hamel, E.; Mooberry, S. L. Synthesis and Discovery of Water-Soluble Microtubule Targeting Agents that Bind to the Colchicine Site on Tubulin and Circumvent Pgp Mediated Resistance. *J. Med. Chem.* **2010**, *53*, 8116-8128. SYBYL 8.1, Tripos Interactional, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
23. SYBYL 8.1, Tripos Interactional, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
24. Jones, G.; Willett, P.; Glen, R. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *J. Mol. Biol.* **1995**, *245*, 43-53
25. Löwe, J.; Li, H.; Downing, K. H.; Nogales, E. Refined structure of alpha beta-Tubulin at 3.5 A Resolution. *J. Mol. Biol.* **2001**, *313*, 1045-1057.
26. Sarkar, A.; Kellogg, G. E. Hydrophobicity – Shake Flasks, Protein Folding and Drug Discovery. *Curr. Top. Med. Chem.* **2010**, *10*, 67-83.
27. Massarotti, A.; Coluccia, A.; Silvestri, R.; Sorba, G.; Brancale, A. The Tubulin Colchicine Domain: a Molecular Modeling Perspective. *ChemMedChem.* **2012**, *7*, 33-42.

CHAPTER 4

Incorporation of Tautomerism within HINT

4.1 Introduction

4.1.1 Background of tautomers

Tautomerism is defined as the transfer of a chemical group and the rearrangement of single and double bonds. For the most common and simple cases of prototropic tautomers that belong to the same compound, the differences are the positions of certain hydrogen atoms that can shift between carbons and heteroatoms, and the positions of related single and double bonds. Frequent examples are keto-enol, imine-enamine forms and nitrogen-containing aromatic heterocycles. In this chapter, tautomers only refer to prototropic tautomers unless stated otherwise, because these isomers are most relevant to our goal of improving virtual screening.

4.1.2 Tautomers in drug discovery

For a given compound that can tautomerize, different tautomeric forms can exist. The most obvious difference that medicinal chemists interested in drug discovery note is probably the change of positions of hydrogen bond donors and acceptors, which may have significant impact on the interactions with a receptor. This, in turn has an important impact on molecular modeling of protein-drug interactions. Docking is a commonly used and very effective modeling technique to evaluate receptor-ligand interactions for binding mode prediction of specific compounds or for virtual screening of hits targeting

the receptor. Traditionally, one compound is treated as one structure in docking and this simplification has not seemed to impede the success of the technique in practical drug discovery and development projects. However, as more accurate predictions of ligand binding are needed, more detailed descriptions of the complex structures in modeling seem to be emergent as critical issues, and medicinal chemists would not want to miss a hit compound whose tautomeric forms are predicted to be good binders in virtual screening, while its “database” encoded structure is not.

When dealing with a compound that can tautomerize, modeling its multiple tautomeric structures should be more comprehensive and computationally expensive than considering just the one structure that represents the most energetically favorable form in aqueous solution. Moreover, as reported by Milletti et al., 29% of the compounds in commercial databases are potentially tautomeric and 7.8% of them are not even represented by the most stable form predicted in water.[1] In addition to docking, pharmacophores, chemical descriptors and structure searches are all affected by tautomerization. Recently, the issues related to incorporating tautomerisim into molecular modeling have been highlighted in a number of review articles.[2-4]

4.1.3 Existing approaches that deal with tautomers

A number of software tools have been developed to deal with tautomers, both open-source and commercial. Regardless of the additional functions these programs have, the first step is identification and enumeration. TauTGen[5] is an application that enumerates tautomeric forms from a given structure. The user has to manually provide input information such as which heavy atoms the hydrogens should be attached to and

the minimum and maximum number of hydrogens allowed. The application is written in the C programming language and the source code is available online.[6] CACTVS[7] uses a more automated approach, and contains a set of 21 predefined SMIRKS-based (a language based on SMILES) transformations and an engine to generate corresponding tautomers without user manipulation. ChemProp[8] takes information from the InChI code of a structure to generate tautomers. The InChI code recognizes the hydrogen atoms that can shift among heteroatoms as mobile H atoms. ChemProp uses such information as the key input for its algorithm. TauThor[1] generates tautomers recursively according to the general scheme of tautomerization, from HX-Y=Z to X=Y-ZH, where X, Y and Z represent C, N, O or S. TauThor is expected to be more comprehensive than the approaches using predefined transformations, but is also more computationally intensive. Other commercially available applications include Pipeline Pilot (Accelrys), LigPrep (Schrödinger) and Marvin (ChemAxon).

For the same compound, different tautomeric forms possess different internal energies, thus affecting their populations in solution and thereby their contributions to the binding of the compound. While it is a challenge to predict their energies and populations accurately,[2-4] some form of penalty needs to be applied to the high-energy tautomers, as false positives have been related to energetically unreasonable tautomeric forms being recognized as hits in virtual screening studies with tautomer-enriched databases.[9-11] To compensate and correct, different levels of quantum chemistry calculations,[5,8] empirical rules (such as tautomers with more aromatic structures are favorable),[7] and penalties based on the predicted pKa values of the moveable

hydrogens[1] are combined with identification and enumeration tools in various different application packages.

4.1.4 The HINT approach

The HINT (Hydropathic INTeraction)[12] software package contains multiple applications for cheminformatics and molecular modeling research.[13] It has solid software infrastructure being built from a toolkit of functions, includes abundant atom types and bond types representing different chemical structures and useful built-in functions that can handle structural changes. More importantly, the HINT scoring function has been successfully applied to evaluate protein-ligand interactions for docking in many systems.[13] With a tautomer function written as part of HINT, our expectation was that more accurate prediction of binding would be possible. In addition, incorporation of tautomerism within other HINT tools would be more convenient; e.g., as part of the Computational Titration suite, an application that considers multiple ionization states in binding.[14] The aim of this present study was to construct a workflow and related applications to incorporate tautomers into the HINT infrastructure. The potential uses of such a tool range from docking/virtual screening, QSAR, and into other modeling areas.

4.2 Materials and Methods

4.2.1 HINT infrastructure

HINT (Hydropathic INTeractions) is a compilation of applications[13] designed for quantifying and visualizing molecular interactions. It contains a powerful system that handles input, output and representation of small molecule and macromolecule structures. The representation of a structure in HINT is based on the descriptions of

atoms. Such information includes atom types, connection tables indicating how the atoms are connected, and the corresponding bond types. The information provides the key input for the associated algorithms to work. The workflow can be described as: 1) the HINT infrastructure reads the original molecular structure files; 2) builds a molecule object from these data; 3) passes the molecule information to our algorithms; 4) our algorithms identify the tautomers and generate the respective tautomeric structures; and 5) passes the new information back to HINT as new molecule objects for further processing; i.e., partitioning, scoring, etc.

4.2.2 HINT binding score

The HINT scoring function evaluates atom-atom interactions using a set of parameters derived from the solvation partition coefficients, LogPs, measured in a 1-octanol/water system, as described in Chapter 1. HINT calculated hydropathic interactions include hydrophobic interactions, hydrogen bonding, acid-base interactions and Coulombic interactions. In this study, the crystal structures were taken from the PDB database and prepared in Sybyl 8.1.[15] The hydrogen atoms were added and optimized using the Tripos force field with Gasteiger-Hückel charges to a gradient of 0.005 kcal mol^{-1} $Å^{-1}$. The ligand was then extracted, leaving only the protein structure. To get the binding poses of the tautomeric forms, the tautomeric structures were placed back into the protein according to the coordinates of the original ligand, and a minimization was then conducted for each pose using the Tripos force field with the atoms from the protein constrained. A HINT binding score was calculated for each tautomeric form binding to the protein structure. Although we ultimately plan to structurally incorporate water into

these calculations, in the present, proof-of-concept, study, all water molecules were also deleted from the complex model.

4.2.3 Tautomer energy prediction

The energies of the small molecule tautomers were predicted by using the calculated heat of formation ΔH_f from the semiempirical quantum chemistry method, PM3, implemented in the MOPAC module in SYBYL 8.1.[15] Before PM3 optimization and calculation, the molecular structures were optimized by the Tripos force field with Gasteiger-Hückel charges to a gradient of 0.005 kcal mol^{-1} $Å^{-1}$.

4.3 Results and Discussion

4.3.1 General workflow

The general workflow of our incorporation of tautomerism within HINT is shown in **Figure 4.1**. First, the molecular structure files that HINT recognizes (such as Sybyl .mol2 files) are read by HINT and converted to an internal molecule structure objects. Then, the tautomer module analyzes these molecule structure objects and recognizes structures that can tautomerize. The prospective tautomeric forms are then generated by the module as new molecule objects, while the corresponding penalty scores are assigned to these new forms as well as the original one. The penalty scores can be derived from experimental measurements stored in a database or obtained through quantum chemistry calculations. They are next converted to HINT score units to facilitate HINT scoring for simple docking and large-scale virtual screening.

The identification, enumeration and penalty assignment are the core steps of the tautomer module. We developed two separate approaches to handle these functions. The general search tool uses an algorithm based on intuitive hydrogen shifts to recognize and enumerate tautomers. The penalty-related energy prediction is performed by an outside quantum chemistry application. Ultimately, these penalty calculations will be incorporated within the module. The second approach uses a database that contains commonly observed tautomeric structures, with the corresponding penalty scores having been pre-calculated. The tautomer database method runs notably faster than the general search tool as it was designed for large-scale computing tasks such as virtual screening. The general search tool is able to find tautomeric patterns that are not stored in the tautomer database and is best applied to small-scale computing tasks such as the simple docking of analogues.

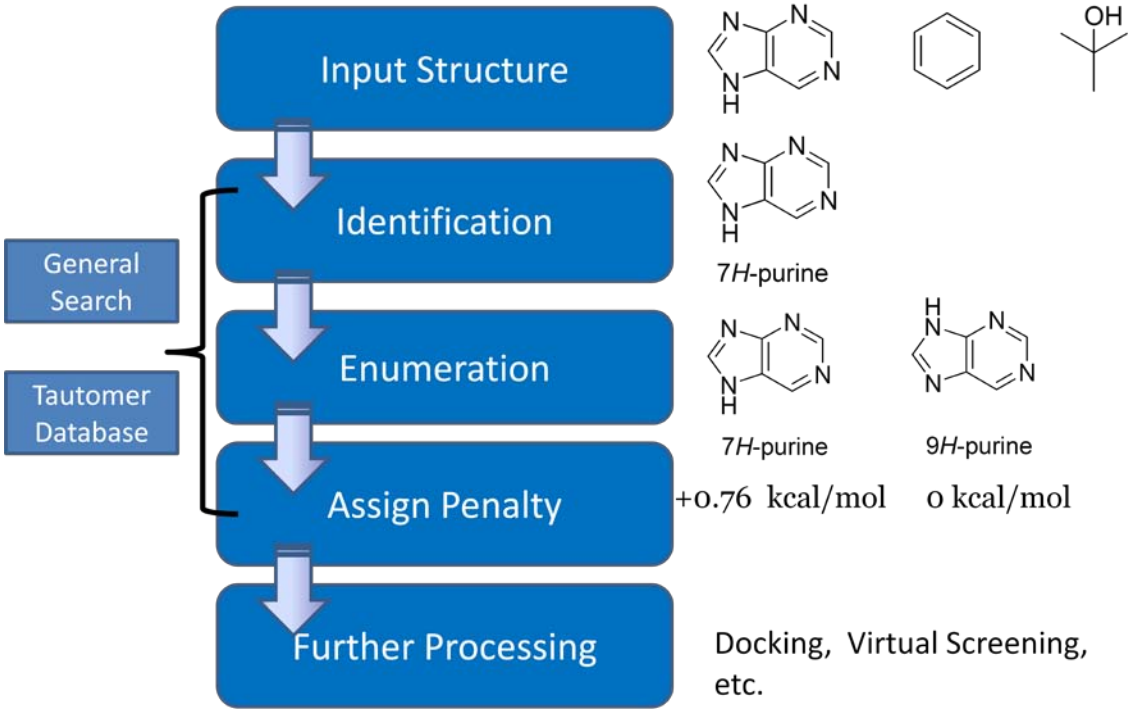


Figure 4.1. The general workflow of tautomer processing in HINT.

4.3.2 Tautomer identification and enumeration for the general search tool

The algorithm to identify tautomers is based on the recognition of hydrogen shifts between heavy atoms. The process is simple and intuitive as shown in **Figure 4.2**. We only consider the hydrogen shifts between heteroatoms. The shifts can cross fused rings. Aromatic bonds are considered as alternating single and double bonds; i.e., as in the Kekule formalism. Compound **4.1** contains two moveable hydrogen atoms. The hydrogen attached to the oxygen can go through a 1,3-shift to form compound **4.2** and a 1,9-shift to form compound **4.5**. The hydrogen attached to the nitrogen can go through two different 1,5-shifts to form compounds **4.3** and **4.4**. If two hydrogen atoms shift together, compound **4.6** can be formed.

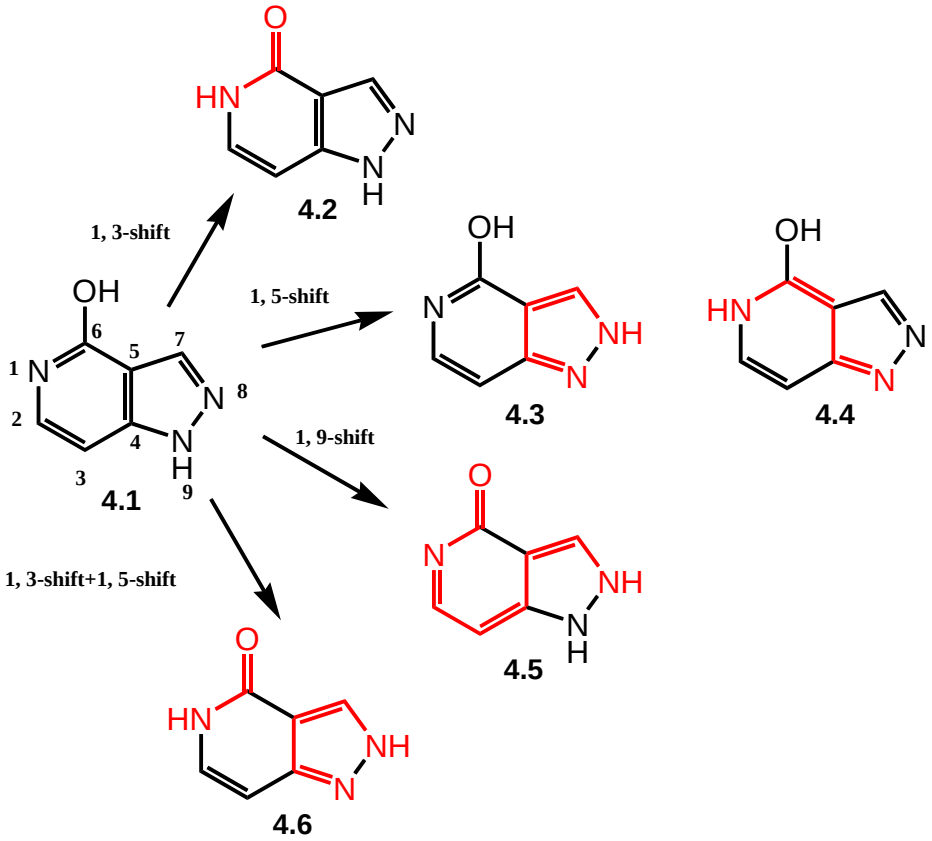


Figure 4.2. The shifts of the movable hydrogen atoms of compound **4.1** and corresponding tautomeric forms. The paths of the shifts are shown in red.

The algorithm contains 4 steps:

Step 1: Identify the patterns of shifts from the original structure. The general scheme is described as:

$$HA-(B=C)_n-D=E$$

where *A* is an sp^3 N, O or S; *B, C* and *D* are sp^2 N or C; E is sp^2 N, O or S; n is equal to 0 to 3 representing (1,3), (1,5), (1,7) and (1,9)-shifts respectively; and the bond types are alternating single and double. The algorithm processes all atoms of the structure to identify the atoms that have the same atom types and connections as described in the shift patterns. In the case of compound **4.1**, we found 4 different shifts (one 1,3-shift, two 1,5-shifts and one 1,9-shift). It is worth mentioning that although longer-range shifts exist in large rings and very long chain systems, we only considered the most commonly seen shifts, which are those up to (1,9)-shifts.

Step 2: Create a binary matrix to represent the combinations of shifts. We applied a brute-force attack method to exhaustively list all the possible combinations using an *m*×*n* binary matrix (**Figure 4.3**). In the matrix, 0 is used to indicate a specific shift is not performed and 1 indicates it is performed. The columns represent different shifts and the rows represent the tautomeric forms that are composed of the shifts. The number of all possible tautomeric forms, m, is equal to 2^n, where n is the number of shifts.

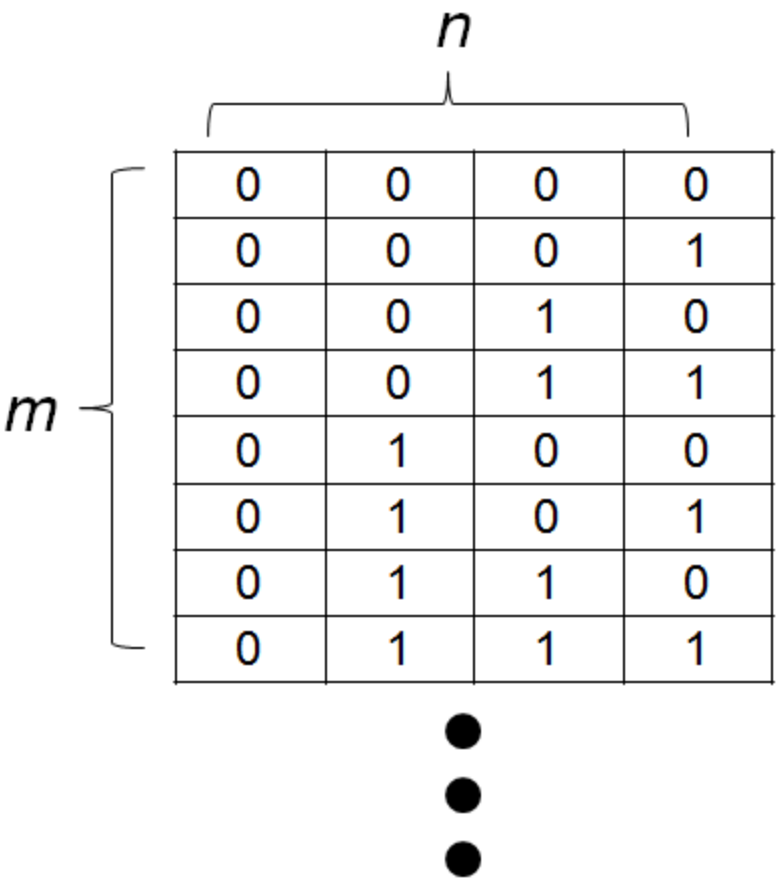


Figure 4.3. The *m*×*n* (m=2^n) binary matrix representing the combinations of the shifts. Each column indicates one shift. 0 indicates the shift is not performed and 1 indicates it is performed. Each row represents one tautomeric form with shifts that do or do not occur.

Step 3: The binary matrix is verified to ensure the prospective tautomeric forms can actually exist. The potential problems we consider are purely based on the scheme we use in the algorithm and not at this point based on energy. Errors will occur in the cases (see **Figure 4.4**) where a) two or more shifts that share the same moveable hydrogen occur at the same time; b) two or more shifts that share the same end hetero atom occur at the same time; c) two or more shifts that share the same double bond in the

path occur at the same time. Thus, the algorithm checks for the generalized situation. If one prospective tautomeric form contains two or more potential shifts that share any atom, the entry (row) corresponding to that tautomeric form will be removed from the matrix. In the case of compound **4.1**, the initial matrix contains 2^4=16 rows. After the filter of step 3, the matrix is reduced to 6 rows representing 6 tautomeric forms including the original one.

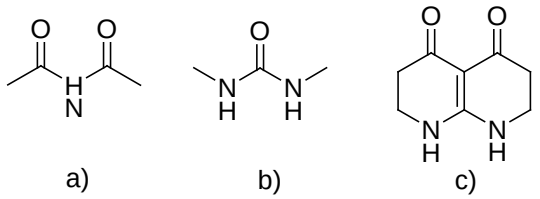


Figure 4.4. Cases where errors will occur if identified shifts are performed simultaneously.

Step 4: Enumerate the prospective tautomeric forms based on the matrix. The internal structural information of the resulting tautomeric forms is generated in this step. The original molecule structure objects are copied to give the new structure objects. Within these, only the substructures related to the shifts are changed (**Figure 4.5**), while the remainder of the structure is kept the same. The appropriate single bonds are changed to double bond types and double bonds to single bond types, accordingly. The two end atoms of the shift change their atom types from sp^3 to sp^2 and sp^2 to sp^3. The coordinates of the moved hydrogen are then regenerated and the connection tables associated with the move are reconstructed thourg operations on the molecule object for that tautomer.

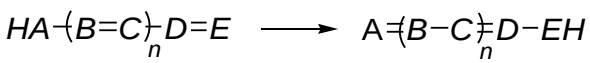


Figure 4.5. The changes occur during a hydrogen shift.

The general algorithm to identify and enumerate prospective tautomeric forms is simple and intuitive. It detects up to 1,9-hydrogen shifts between heteroatoms and uses the information for enumeration. If necessary, long-range shifts that exceed 1,9-shifts can be easily added to expand the algorithm to handle larger systems.

4.3.3 Energy prediction and HINT penalty scores

The tautomeric forms after the general search should then be converted to external mol2 files and subjected to Molecular Mechanics (MM) optimization and then PM3 calculations in Sybyl 8.1 (or another modeling suite). In our implementation of this tool, we assumed that the heat of formation ΔH_f from PM3 represents the energy of each tautomer. Then, the relative energies compared to the most stable form in the group were calculated. The most stable form would get no penalty and the rest of the forms would receive penalty scores directed correlated to the relative PM3 energies. In previous studies, around 515 HINT units were correlated to 1 kcal mol^{-1} of binding energy.[16] However, we found the penalty scores resulting from this relationship were far too high for most of the tautomeric forms, considering that HINT binding scores in the range of 500-2000 are commonly seen, and 5 and 28 kcal mol^{-1} are both suggested as cut-offs for high-energy tautomers.[17,18] For the remainder of this work, 100 HINT units/ kcal mol^{-1} has been arbitrarily used. Further studies are needed to find the more optimum relationship.

4.3.4 Kekule structure assignment for aromatic molecules

One required step before a molecular structure is processed by the general search algorithm is to assign a Kekule structure to the molecule if it is aromatic. This is because the general tautomer search algorithm is based on the recognition and manipulation of shift patterns, which are represented by alternating single and double bonds and sp^2 and sp^3 atom types. Aromatic molecules, however, usually use only one bond type (normally “ar” indicating “aromatic” or “1.5” indicating a bond between single and double) instead of single and double bond types to represent the delocalized nature of bonding, and, while formally sp^2, use aromatic atom types in molecular mechanics force fields instead of sp^2. Considering how common aromatic molecules are tautomers (such as compound **4.1** in **Figure 4.2**), we developed an algorithm to assign a Kekule structure with alternating single and double bonds to preprocess each aromatic molecule (**Figure 4.6**). The algorithm is similar to an algorithm developed to create molecular skeletons to meet valence rules.[8] Following are the steps of this algorithm:

Step 1: remove all atoms that are not aromatic from the structure and treat all the connected bonds as single bonds. Each atom is defined as a node as in graph theory.

Step 2: Find the atom that has the minimum number of connected neighbors (the node with the minimum degree). In the case of more than one atom meeting this criterion, any of them can be selected. Assign a double bond between the selected atom and one of its neighbors.

Step 3: Temporarily remove the two paired atoms and the bonds connected to them from the structure.

Step 2 and *3* are repeated until all the aromatic atoms are paired using double bonds, and no atom is left in the structure. Then the structure is reconstructed by tracing back how the double bonds are assigned.

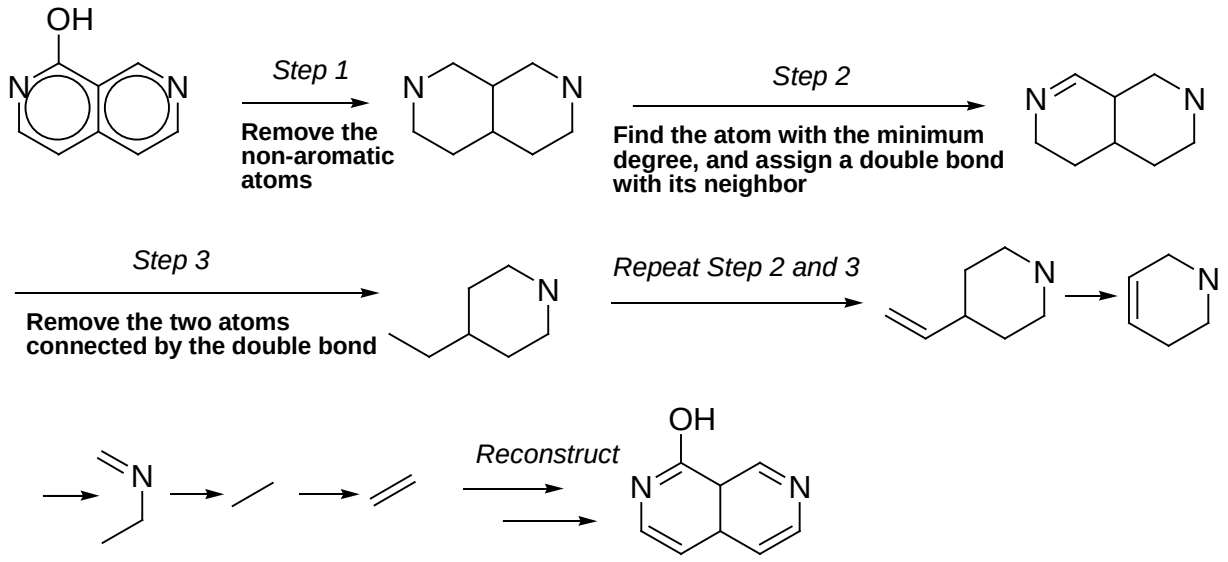


Figure 4.6. The algorithm for assigning a Kekule structure formalism to an arbitrary aromatic molecule.

4.3.5 Case study on pterin binding to the ricin toxin-A chain

Pterin (**Figure 4.7**) represents the type of structures that are able to tautomerize. Pterin-based compounds have been shown to bind to the ricin toxin-A chain (RTA).[19] What tautomeric form pterin uses to bind to the protein is of interest and has been previously studied by crystallographic analysis combined with molecular mechanics and quantum mechanics.[20,21] Here, we applied the workflow developed for tautomers described to answer the question of which tautomer binds best to the active site and rationalize the energetic thereof. The crystal structure of ricin toxin-A chain complexed with a pterin-based compound PTA (pteroic acid) was taken from PDB (pdbid: 1BR6). The ligand

was extracted and the pterin part of the structure was subjected to our workflow (**Figure 4.7**). The original pterin structure (**Pterin_aromatic**) was assigned a Kekule structure, processed by our general search algorithm and then each tautomer was analyzed by PM3 calculations to determine its internal energy. Then, all the tautomeric structures were merged back to the protein's pocket based on the original coordinates of the complex Their positions were optimized by the Tripos force field and their binding interactions were scored by HINT. The resulting 10 tautomeric structures including the original (**Pterin2** and **Pterin5** are the same molecule) and their corresponding energies and scores are shown in **Figure 4.7** and **Table 4.1**. HINT scores summarize different types of energy of binding, including hydrophobic interactions, hydrogen bonding, acid-base interactions and Coulombic interactions. We focus on the relative HINT scores in order to compare different tautomeric forms and answer the question of which forms should be bound.

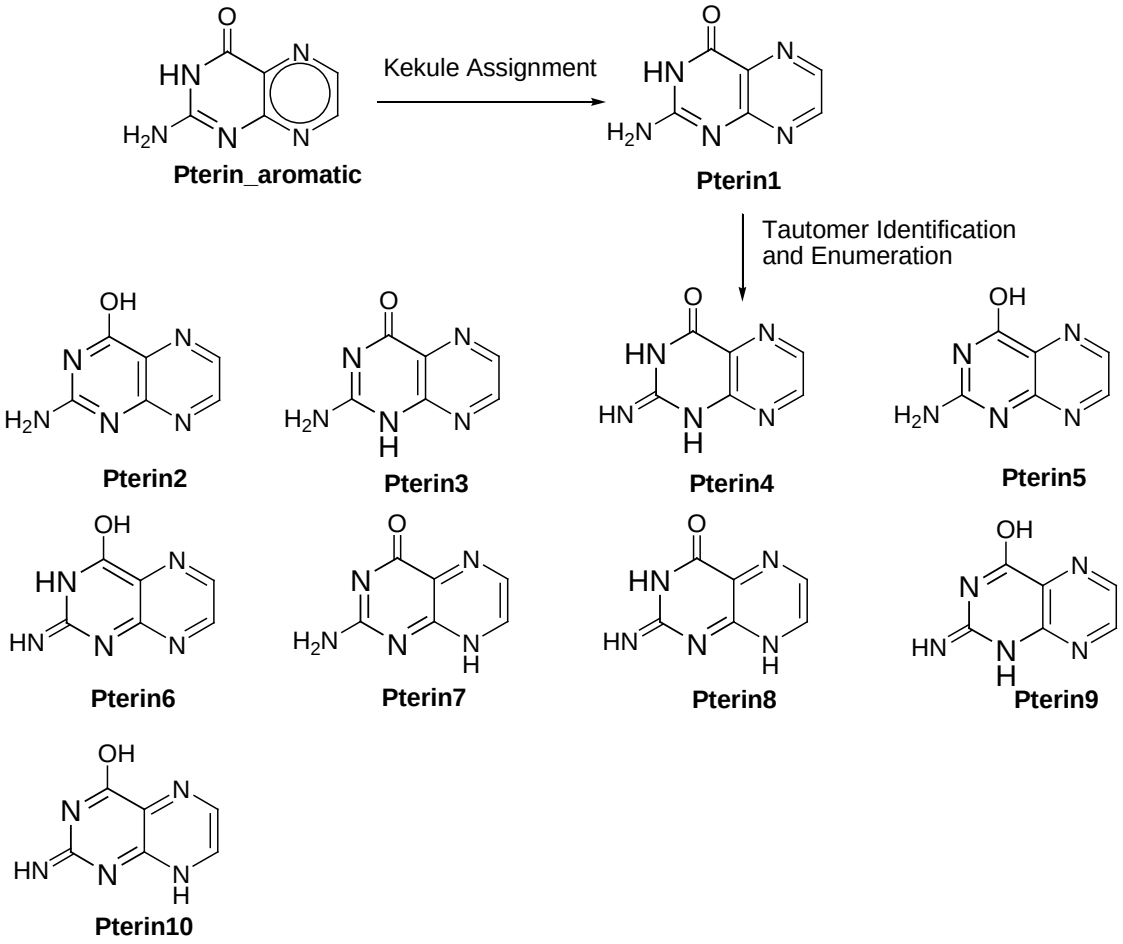


Figure 4.7. Tautomer identification and enumeration for pterin.

Table 4.1. Summary of all the tautomeric forms of pterin.

Structure	$\Delta(\Delta H_f)$[a]	Penalty[b]	Relative HINT Binding Score[c]	Corrected Relative HINT Score[d]
Pterin1	0	0	0	0
Pterin2	1.32	132	-703	-835
Pterin3	3.11	311	337	26
Pterin4	2.68	268	123	-145
Pterin5	1.32	132	-703	-835
Pterin6	26.73	2673	-800	-3473
Pterin7	12.99	1299	-1179	-2478
Pterin8	18.09	1809	-793	-2602
Pterin9	8.66	866	85	-781

Pterin10	26.15	2615	-872	-3487

a: The heat of formation ΔH_f (kcal mol^{-1}) was calculated by PM3. The relative energy $\Delta(\Delta Hf)$ (kcal mol^{-1}) was the energy compared to the most stable form of pterin.
b: HINT penalty score = $100 \times \Delta(\Delta Hf)$.
c: The HINT binding scores relative to the score of Pterin1.
d: Corrected Relative HINT Score = Relative HINT Binding Score – Penalty.

More stringent computational studies and crystallographic analysis have suggested that **Pterin3** is preferred over **Pterin1** in binding to the ricin toxin-A chain, although **Pterin1** is the most stable form in solution. Our data agrees with this suggestion, as **Pterin3** received a higher HINT binding score due to notably better hydrogen bonding network (**Figure 8**), which is compensated by a penalty score due to its higher internal energy compared to **Pterin1**. The combination of the binding scores and the penalty scores made **Pterin1** and **Pterin3** stand out as the most reasonable models as they had the highest corrected HINT scores. Other forms were either poorly bound or were predicted to have too high energies. This case study illustrates the idea that considering tautomerism in virtual screening, in which a compound with good tautomeric binding forms could be missed because the most energetically accessible form was not a good match for the site.

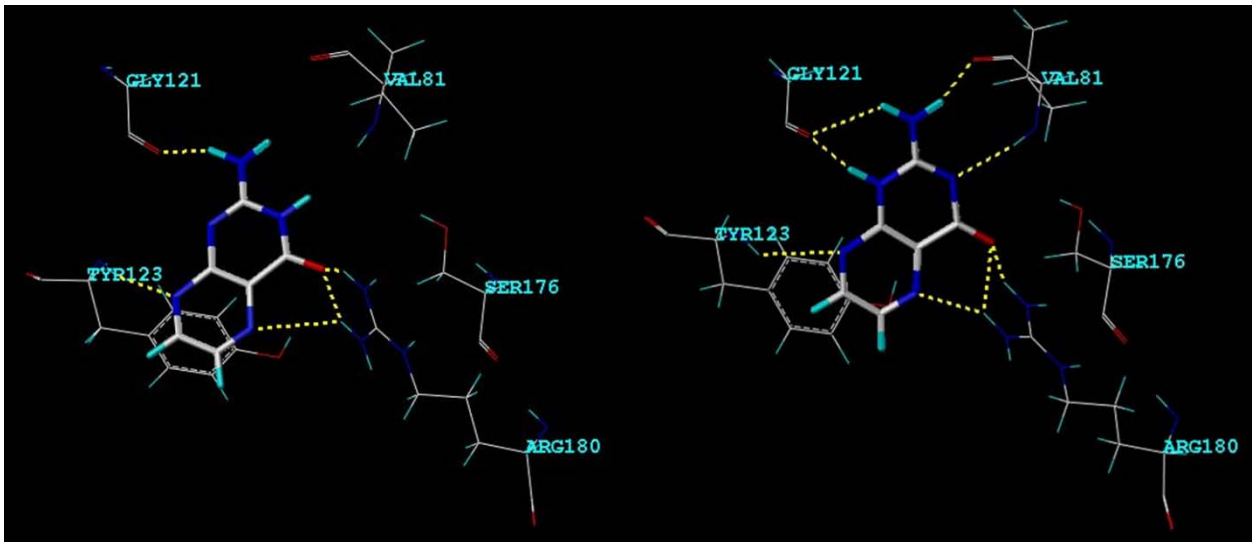


Figure 4.8. Two favorable tautomeric forms of pterin binding to the ricin toxin-A chain. Pterin3 shown on the right tend to satisfy more hydrogen bonds and thus binds better than Pterin1 does shown on the left.

4.3.6 Tautomer database

The tautomer database is designed to speed up the process of tautomer identification, enumeration by bypassing the quantum chemistry calculations for relative energies. The general search algorithm as described above identifies and enumerates all possible tautomeric forms based on hydrogen shifts with no concern of energy. Whether the forms have too high energy to exist, such that they are not contributing to binding is not considered until the quantum chemistry calculation is performed, which is by far the slowest part of the procedure. The basic idea of a tautomer database is that it is a collection of commonly observed tautomeric structures. These structures will be matched to the input molecular structure as substructures. Each tautomeric structure is related to only a handful of other structures that belong to the same category (**Figure 4.9**) and have tolerable energies, and/or are regularly considered by researchers. The

penalty scores corresponding to the forms are calculated in the same way as the general search tool does and they are stored in the database so that the time needed for quantum chemistry calculation can be saved.

Our proof-of-concept database contains 69 tautomeric forms belonging to 25 annular tautomeric systems (**Figure 4.9**). Most of the forms were collected from a published annular tautomerism study and have been identified as populated in public structural databases such as CSD and PDB.[22] Five-membered, six-membered and bicyclic rings with nitrogen form the majority of the database, including pyrazole(**4a-b**), imidazole (**5a-b**), triazoles (**1a-c and 2a-c**), tetrazoles (**3a-b**) and their fused rings such as indazole (**8a-b**). These heterocyclic rings are frequently used in organic synthesis as building blocks and therefore they are common in organic compounds.

The structures in the tautomer database are matched as exact substructures (atom types, bond types and connections) to the input molecule. It may seem that there are duplicate structures in the database. **1a** and **1b**, for example, appear to be the exact same structure. However, because they are matched to the input molecule as substructures, if the parent compound (the input) is not symmetrical with respect to the middle nitrogen (for example, two different substitutions on the two carbon atoms), **1a** and **1b** are actually two different forms.

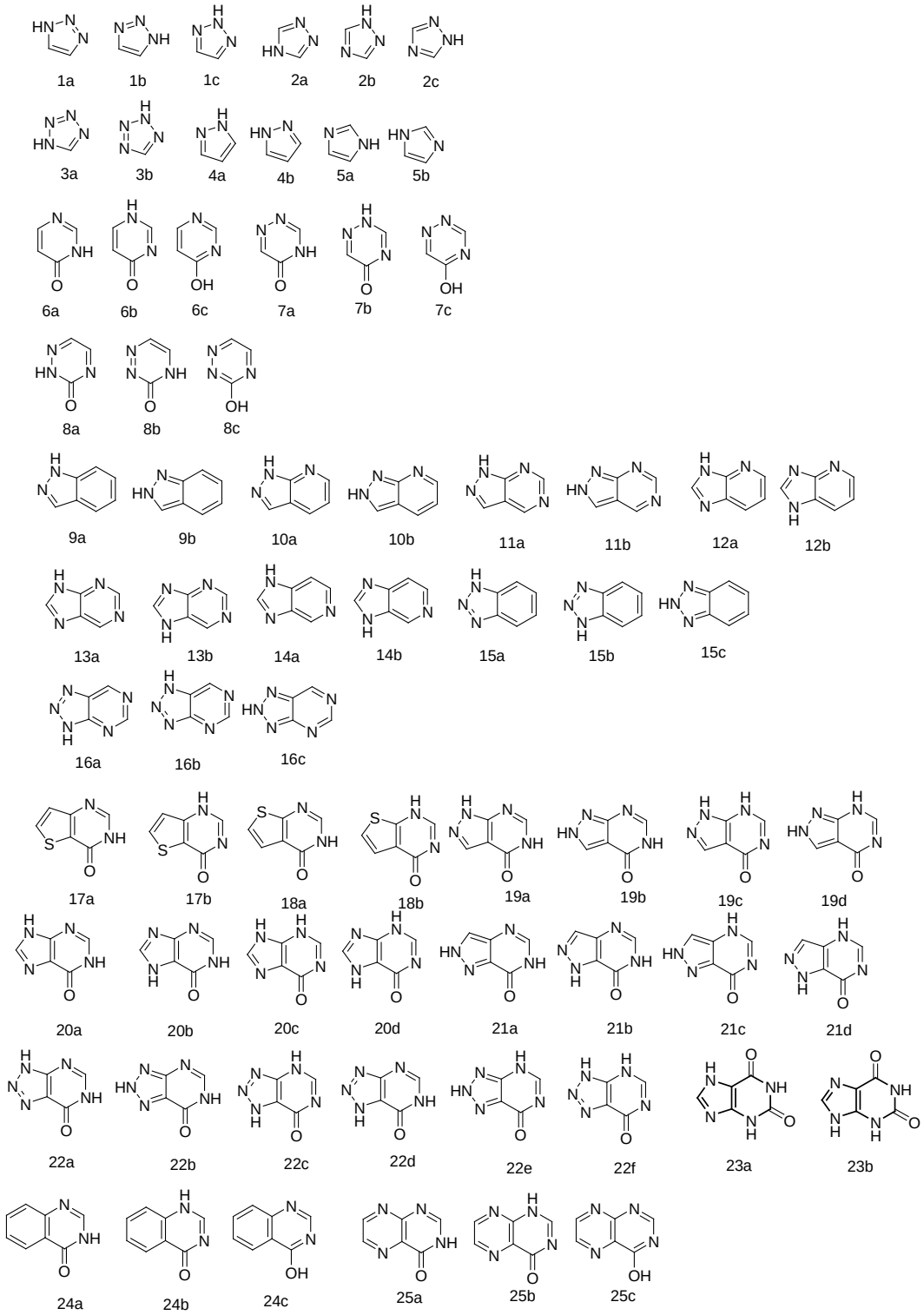


Figure 4.9. The annular tautomeric forms in the initial tautomer database.

The identification and enumeration using the tautomer database are performed as follows (**Figure 4.10**):

Step 1: the tautomeric structures in the database are matched as substructures to the input molecule (e.g., compound **4.7**).

Step 2: if a match is found (**5a** and **9a, Figure 4.9**), relate the substructures within the same family (i.e., **5b** and **9b**) to the prospective tautomers.

Step 3: the prospective tautomer list includes the combinations of all the substructures. A matrix with iteration numbers similar to that of the binary matrix (**Figure 4.3**) is constructed to facilitate enumeration. 4 tautomeric structures as new molecule objects (compound **4.7-4.10**) including the original **4.7** are the output from the database search algorithm.

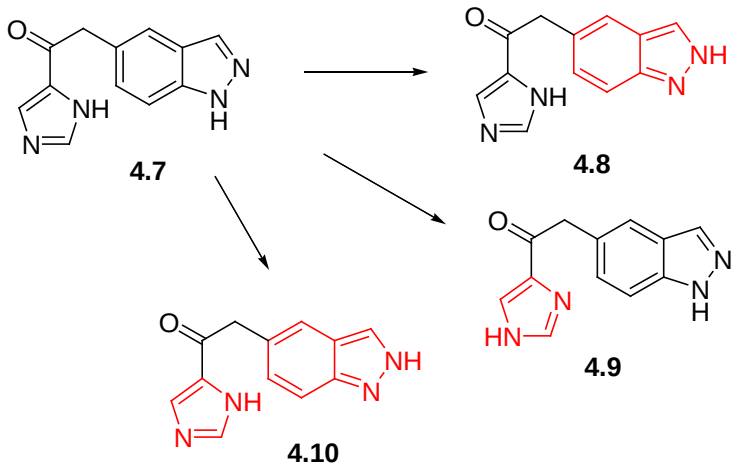


Figure 4.10. The prospective tautomeric forms of compound **4.7** generated by the tautomer database search algorithm. The identified tautomeric substructures are shown in red.

The penalty scores of the substructures in the database are pre-calculated by converting the heat of formation ΔH_f values to HINT scores (**Table 4.2**). The parent structures from the input are assigned with the summation of the penalty scores of their tautomeric substructures. For example, compound **4.8** is composed of the substructures **5a** and **9b**. Its penalty score is the penalty score of **5a** plus the penalty score of **9b**.

Table 4.2. Penalty scores and relative energies of the structures in the tautomer database.

Structure	$\Delta(\Delta H_f)$[a]	Penalty[b]	Structure	$\Delta(\Delta Hf)$[a]	Penalty[b]	Structure	$\Delta(\Delta H_f)$[a]	Penalty[b]
1a	0	0	10a	0	0	19d	8.76	876
1b	0	0	10b	6.98	698	20a	0	0
1c	2.81	281	11a	0	0	20b	0.67	67
2a	0.03	3	11b	6.8	680	20c	11.08	1108
2b	0	0	12a	0	0	20d	3.61	361
2c	0	0	12b	3.42	342	21a	3.56	356
3a	0	0	13a	0	0	21b	0	0
3b	2.16	216	13b	3.56	356	21c	10.04	1004
4a	0	0	14a	0	0	21d	5.5	550
4b	0	0	14b	0.12	12	22a	0	0
5a	0	0	15a	0	0	22b	8.06	806
5b	0	0	15b	0	0	22c	5.1	510
6a	0	0	15c	7.59	759	22d	2.15	215
6b	8.11	811	16a	0	0	22e	11.97	1197
6c	1.24	124	16b	3.83	383	22f	10.59	1059
7a	0	0	16c	9.88	988	23a	0	0
7b	7.64	764	17a	0	0	23b	6.95	695
7c	1.83	183	17b	6.46	646	24a	0	0
8a	1.88	188	18a	0	0	24b	6.63	663
8b	1.05	105	18b	8.43	843	24c	4.55	455
8c	0	0	19a	0	0	25a	0	0
9a	0	0	19b	4.64	464	25b	3.19	319
9b	4.97	497	19c	10.76	1076	25c	4.02	402

a: The heat of formation ΔH_f (kcal mol^{-1}) was calculated by PM3. The relative energy $\Delta(\Delta Hf)$ (kcal mol^{-1}) was the energy compared to the most stable form in the same category.

b: HINT penalty score = $100 \times \Delta(\Delta Hf)$.

4.3.7 Tautomer enrichment in a virtual screening benchmarking database

DUD (Directory of Useful Decoys) is a database of active compounds against specific protein targets and decoy compounds with similar physical properties to the actives but no activity.[23] The database is commonly used for benchmarking docking algorithms for virtual screening. Recently, it has also been used to test the effect of considering tautomerism in virtual screening because the active compounds in the database, as well as the large number of decoys, contain tautomers. In order to test our application and to demonstrate the enrichment of tautomeric structures in the commonly used database, we applied our tautomer database search to a total of around 3000 active compounds and 100,000 decoys against a total of 40 targets from the DUD database. If a structure in DUD contains a substructure that is the same as any tautomeric structure in our tautomer database, the structure is considered tautomeric. The results are listed in **Table 4.3**. The overall enrichment was low. 7% of the active compounds and 4% of the decoy compounds were determined to have substructures that matched to the tautomeric structures in our database. However, in some specific systems, the ratios were considerably larger. In fact, 7 out of 40 targets (ADA, CDK2, COX2, HIVPR, HSP90, P38, VEGFr2) had more than 15% of actives as potential tautomers. And 2 of the 40 had over 30% (HSP90: 54%; P38: 33%). The tautomer ratio for decoys was more consistent across different targets. The range was from 1% to 12%. In addition, the tautomeric structures identified in DUD were determined to be commonly observed or “significant” in our tautomer database. Considering their corresponding prospective tautomeric counterparts is certainly a reasonable step in order to achieve better interpretation of a binding event. Moreover, our tautomer database only partially covers

the types of tautomers existing in chemistry. We would expect higher enrichment if the general search tool was applied.

Table 4.3. Tautomer counts and ratios of active and decoy compounds from the DUD database.

Target	Actives	Tautomers	Ratio	Decoys	Tautomers	Ratio
ACE	49	0	0.00	1727	42	0.02
ACHE	105	0	0.00	3714	53	0.01
*ADA	23	6	0.26	821	21	0.03
ALR2	26	0	0.00	918	62	0.07
AmpC	21	0	0.00	732	24	0.03
AR	74	0	0.00	2628	75	0.03
*CDK2	50	9	0.18	1779	124	0.07
COMT	11	0	0.00	430	10	0.02
COX1	25	0	0.00	849	49	0.06
*COX2	348	62	0.18	12464	220	0.02
DHFR	201	0	0.00	7145	826	0.12
EGFr	444	12	0.03	14894	546	0.04
ER_agonist	67	0	0.00	2355	145	0.06
ER_antagonist	39	0	0.00	1395	13	0.01
FGFr1	118	1	0.01	4205	128	0.03
FXa	142	1	0.01	5095	260	0.05
GART	21	0	0.00	753	15	0.02
GPB	52	0	0.00	1850	53	0.03
GR	78	0	0.00	2797	21	0.01
*HIVPR	53	9	0.17	1885	36	0.02
HIVRT	40	0	0.00	1437	63	0.04
HMGR	35	0	0.00	1241	38	0.03
*HSP90	24	13	0.54	860	73	0.08
InhA	85	0	0.00	3035	99	0.03
MR	15	0	0.00	535	16	0.03
NA	49	0	0.00	1745	39	0.02
*P38	256	85	0.33	8387	352	0.04
PARP	33	0	0.00	1176	82	0.07
PDE5	51	6	0.12	1809	28	0.02
PDGFrb	157	0	0.00	5614	154	0.03
PNP	25	2	0.08	882	94	0.11
PPARg	81	2	0.02	2906	49	0.02
PR	27	0	0.00	967	12	0.01
RXRa	20	0	0.00	708	17	0.02
SAHH	33	0	0.00	1159	63	0.05
SRC	155	1	0.01	5793	259	0.04

Thrombin	65	0	0.00	2292	67	0.03
TK	22	0	0.00	784	30	0.04
Trpsin	44	0	0.00	1544	48	0.03
*VEGFr2	74	17	0.23	2641	191	0.07
Total	3238	226	0.07	113951	4497	0.04

Only the substructures that match to the tautomeric forms in our tautomer database were counted. When counting the totals, duplicate structures from different target sets were not removed.
* : the targets that have more than 15% tautomers in actives.

4.3.8 Potential use in virtual screening and in combination with Computational Titration

One of the main applications that we have planned for the tautomer module in HINT is docking for virtual screening. It has been shown in the pterin case that a tautomeric form with a higher energy may have better binding than the most stable form. Virtual screening workflows should be expected to provide higher accuracy if tautomerism is considered when such structures are present. During the docking of a compound in virtual screening, its tautomeric forms generated by our general search algorithm or the tautomer database are also docked to the same target in addition to the original form. The HINT score of each form is the HINT docking score minus the HINT penalty score. The final HINT score representing the compound is the best HINT score among the scores of all the forms. In this way, a potential good binder of the target would not be missed if one of its tautomeric forms binds well but the default form does not. Also, the tautomeric forms that bind well but have intolerable high energy would be penalized to prevent them from causing false positives, which several researcher have encountered when not using penalty scores.[9-11]

Computational titration is a module in HINT that iterates the ionization states of the residues on protein and the ligand in a binding pocket, to find the optimum combination of states.[14] When treating a pocket with multiple ionizable residues and a ligand with multiple tautomeric forms, combining the computational titration module and the tautomer module would provide more accurate description of binding. The combination of the two approaches would generate an ensemble of different complexes for HINT scoring. The complex ensemble with optimized ionization states possessing the highest HINT interaction score (corrected by penalty scores) would be considered to be ideal and subject to further analysis. Considering both ionization states and tautomerism is a further step towards accurate prediction of binding, which has been reported in several studies[9,10, 24-26] and is an on-going project of our group.

4.4 Conclusions

In this study we proposed and developed a workflow to incorporate tautomerism within HINT. A simple and intuitive algorithm is used to identify and enumerate tautomeric forms based on hydrogen shifts of the input structure. A penalty score is assigned to each tautomeric form based on the energy calculated by the PM3 method. The workflow was tested on pterin, a structure that can tautomerize and bind to ricin toxin-A chain. The favorable binding forms were identified and the results agree with more stringent computational studies and crystallographic analysis.

In the second part of the study, a database containing commonly observed tautomeric structures and their pre-calculated penalty scores was built to speed up the workflow and to facilitate large-scale computing such as virtual screening. The use of the

database was illustrated in the determination of the tautomer enrichment for a popular virtual screening benchmarking database, DUD. Although the overall tautomer ratio was low, specific protein targets that had considerable enrichment were identified and attention should be paid when conducting benchmarking virtual screening studies.

Incorporation of tautomerism represents an on-going process of improving the docking of small molecule ligands to proteins. Our workflow provides a computationally efficient way to achieve the proof-of-concept goal. Future studies will be focused on the virtual screening using the DUD database to evaluate the impact of considering tautomers on the hit rate and the false-positive rate, and the combination of the tautomerism algorithms with Computational Titration, which considers the ionization states during binding, should prove to be very beneficial for computer-aided drug design.

References

1. Milletti, F.; Storchi, L.; Sforna, G.; Cross, S.; Cruciani, G. Tautomer Enumeration and Stability Prediction for Virtual Screening on Large Chemical Databases. *J. Chem. Inf. Model.* **2009**, *49*, 68-75.
2. Martin, Y. C. Let's Not Forget Tautomers. *J. Comput. Aided Mol. Des.* **2009**, *23*, 693-704.
3. Katritzky, A. R.; Hall, C. D.; El-Gendy, Bel-D.; Draghici, B. Tautomerism in Drug Discovery. *J. Comput. Aided Mol. Des.* **2010**, *24*, 475-484.
4. Sayle, R. A. So You Think You Understand Tautomerism? *J. Comput. Aided Mol. Des.* **2010**, *24*, 485-496.
5. Haraniczyk, M.; Gutowski, M. Quantum Mechanical Energy-Based Screening of Combinatorially Generated Library of Tautomers. TauTGen: A Tautomer Generator Program. *J. Chem. Inf. Model.* **2007**, *47*, 686-694.
6. Haraniczyk, M.; Gutowski, M. TauTGen. http://tautgen.sourceforge.net/ (accessed Mar 20, 2013).

7. Oellien, F.; Cramer, J.; Beyer, C.; Ihlenfeldt, W. D.; Selzer, P. M. The Impact of Tautomer Forms on Pharmacophore-Based Virtual Screening. *J. Chem. Inf. Model.* **2006**, *46*, 2342-2354.
8. Thalheim, T.; Vollmer, A.; Ebert, R. U.; Kühne, R.; Schüürmann, G. Tautomer Identification and Tautomer Structure Generation Based on the InChI Code. *J. Chem. Inf. Model.* **2010**, *50*, 1223-1232.
9. ten Brink, T.; Exner, T. E. Influence of Protonation, Tautomeric, and Stereoisomeric States on Protein-Ligand Docking Results. *J. Chem. Inf. Model.* **2009**, *49*, 1535-1546.
10. Clark, R. D.; Shepphird, J. K.; Holliday, J. The Effect of Structural Redundancy in Validation Sets on Virtual Screening Performance. *J Chemometrics*. **2009**, *23*, 471-478.
11. Milletti, F.; Vulpetti, A. Tautomer Preference in PDB complexes and Its Impact on Structure-Based Drug Discovery. *J. Chem. Inf. Model.* **2010**, *50*, 1062-1074.
12. Kellogg, E. G.; Abraham, D. J. Hydrophobicity: Is LogPo/w More than the Sum of Its pArts? *Eur. J. Med. Chem*. **2000**, *35*, 651-661.
13. Sarkar, A.; Kellogg, G. E. Hydrophobicity – Shake Flasks, Protein Folding and Drug Discovery. *Curr. Top. Med. Chem*. **2010**, *10*, 67-83.
14. Spyrakis, F.; Fornabaio, M.; Cozzini, P.; Mozzarelli, A.; Abraham, D. J.; Kellogg, G. E. Computational titration analysis of a multiprotic HIV-1 protease-ligand complex. *J. Am. Chem. Soc.* **2004**, *126*, 11764-11765.
15. SYBYL 8.1, Tripos Interactional, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
16. Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. 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.
17. Minkin, V. I.; Garnovskii, A. D.; Elguero, J.; Katritzky, A. R.; Denisko, O. V. The Tautomerism of Heterocycles: five-membered rings with two or more heteroatoms. In *Adv. Heterocycl. Chem.*, Academic Press: New York, NY, 2000; Vol. 76, pp 157-323.
18. Elguero, J.; Katritzky, A. R. Tautomerism: Pyridone and Hydropyridines. In *Handbook of Heterocyclic Chemistry*; Pergamon: Amsterdam, The Netherlands, 2000; pp 47-51.
19. Yan, X.; Hollis, T.; Svinth, M.; Day, P.; Monzingo, A. F.; Milne, G. W.; Robertus, J. D. Structure-Based Identification of a Ricin Inhibitor. *J. Mol. Biol.* **1997**, *266*, 1043-1049.
20. Yan, X.; Day, P.; Hollis, T.; Monzingo, A. F.; Schelp, E.; Robertus, J. D.; Milne, G. W.; Wang, S. Recognition and Interaction of Small Rings with the Ricin A-Chain Binding Site. *Proteins*. **1998**, *31*, 33-41.

21. Jaramillo, P.; Coutinho, K.; Canuto, S. Solvent Effects in Chemical Processes. Water-Assisted Proton Transfer Reaction of Pterin in Aqueous Environment. *J. Phys. Chem. A.* **2009**, *113*, 12485-12495.
22. Cruz-Cabeza, A. J.; Schreyer, A.; Pitt, W. R. Annular Tautomerism: Experimental Observations and Quantum Mechanics Calculations. J*. Comput. Aided. Mol. Des.* **2010**, *24*, 575-586.
23. Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789-6801.
24. Kalliokoski, T.; Salo, H. S.; Lahtela-Kakkonen, M.; Poso, A. The Effect of Ligand-Based Tautomer and Protomer Prediction on Structure-Based Virtual Screening*. J. Chem. Inf. Model.* **2009**, *49*, 2742-2748.
25. Todorov, N. P.; Monthoux, P. H.; Alberts, I. L. The Influence of Variations of Ligand Protonation and Tautomerism on Protein-Ligand Recognition and Binding Energy Landscape. *J. Chem. Inf. Model.* **2006**, *46*, 1134-1142.
26. Park, M.; Gao, C.; Stern, H. A. Estimating Binding Affinities by Docking/Scoring Methods using Variable Protonation States. *Proteins.* **2011**, *79*, 304-314.

CHAPTER 5

Conclusions

The hydrophobic effect is a universal phenomenon that describes the tendency of nonpolar moieties to stay together. The concept has been successfully applied in the HINT (Hydropathic INTeractions) scoring function, which uses the hydrophobic atom constants derived from the experimentally measured partition coefficient ($LogP_{o/w}$) values to evaluate hydrophobic interactions, hydrogen bonding, acid-base interactions and Coulombic interactions for ligand-protein and protein-protein binding. We wanted to test the HINT scoring function and related applications to see if they could be used to model the binding of tubulin colchicine-site anticancer agents and provide guidance for design and optimization; and how we could incorporate tautomerism in order to improve prediction of binding in general.

Microtubules have been treated as a target for cancer therapies for a long period of time, due to the fact that they are one of the major cytoskeletal components in eukaryotic cells and their critical functions, such as maintenance of cell shape, protein trafficking, signaling and segregation of chromosomes during mitosis. Owing to the fact that microtubules are important regulators of endothelial cells, recently colchicine-site agents are being intensively developed as angiogenesis inhibitors (prevent new blood vessel formation) and vascular disrupting agents (destroy existing vasculature) for cancer treatment. Combretastatins, one family of colchicine-site agents, are progressing through clinical trials for this purpose. In addition, colchicine-site agents might be able to

circumvent βIII-tubulin overexpression, which compromises the clinical use of taxanes and vinca alkaloids.

The in-house developed pyrrole-based compounds targeting the colchicine site were studied first. Docking these compounds into the colchicine site with HINT predicted two distinct binding modes. The mode that overlapped very well with colchicine corresponded to the highly active compounds. The other mode only partially overlapped with colchicine and belonged to the weaker compounds. Of the residues that were identified to participate in binding, Cys241β was revealed to form a critical hydrogen bond with the ligand. Although this interaction was supposed to be weak, loss of it appeared to shift the antiproliferative mechanism of action away from microtubule inhibition.

We further analyzed a collection of colchicine-site agents with different scaffolds including the pyrrole-based compounds. These compounds were tested for antiproliferative activities and microtubule effects in the same laboratory. By applying the same docking procedure, their binding modes were predicted and a pharmacophore model was generalized to have a hydrogen bond acceptor interacting with Cys241β, another hydrogen bond acceptor interacting with Val181α, and three hydrophobic centers in the subpockets A, B and C respectively. 3D-QSAR (Quantitative Structure-Activity Relationship) models were constructed based on the binding modes to correlate structural changes with antiproliferative activities. The reliability of the models was indicated by good statistics and the observation that the contour maps showing the favorable and unfavorable functional groups for activity were directly related to the residues around them. We also introduced a new approach, a linear combination of

weighted HINT maps of individual compounds, to highlight the unique features (functional groups) of highly active compounds and the commonality of all compounds in the dataset. The new method was successfully applied to colchicine-site agents and the generated maps agreed with the contours maps of the 3D-QSAR models and may provide guidance for later drug design efforts.

In the process of improving binding prediction of HINT, considering tautomerism was also recently highlighted. We proposed and developed a workflow to incorporate tautomerism within HINT scoring. A simple and intuitive algorithm based on hydrogen shift patterns was developed for a general search tool to identify and enumerate tautomeric structures. The HINT penalty scores for the tautomeric structures were designed to be converted from energies predicted by the semi-empirical PM3 method for quantum chemistry calculations. The workflow was successfully tested on the prediction of pterin binding to ricin toxin-A chain, with the correct tautomeric forms of pterin identified. A database containing a number of commonly observed tautomeric structures and pre-calculated penalty scores was also built to facilitate large-scale computing tasks such as virtual screening. The database approach was tested on a virtual screening benchmarking database DUD, to identify DUD's tautomer enrichment.

Overall, we have answered the question of whether and how we can use molecular modeling techniques based on HINT to explain the activities of colchicine-site agents. We predicted their binding modes in the colchicine site and understood how structural changes would affect activity. Tautomerism has also been incorporated within HINT to consider more details about ligand binding and to improve the prediction results. The models and insights we obtained for the colchicine-site agents and the tautomer module

in HINT will provide valuable guidance and better modeling results in the design and development process.

Vita

Chenxiao Da was born on March 14, 1986, in Nanjing, Jiangsu, China. In 2008, He received his Bachelor of Engineering degree in Pharmaceutical Engineering from East China University of Science and Technology, Shanghai, China, where he was chosen to join the Department of Excellent Students in Science and Technology and was awarded the fellowship for top-tier students for three years. He subsequently joined the Department of Medicinal Chemistry at Virginia Commonwealth University for his PhD. In 2012, he worked at Merck as an intern through the Future Talent Program. He has also been active in extracurricular activities. He served as the President of the Medicinal Chemistry Graduate Student Association and the Alpha Student Chapter of the American Chemical Society's Medicinal Chemistry Division during the year 2012. He also founded the VCU MedChem Soccer Club in 2011.