COMPUTATIONAL CHEMISTRY STUDIES OF PRENYLTRANSFERASES AND SMALL LIGANDS

By

Li-Li Pan

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment to the requirements
for the degree of

Chemistry – Doctor of Philosophy

2015

ABSTRACT

COMPUTATIONAL CHEMISTRY STUDIES OF
PRENYLTRANSFERASES AND SMALL LIGANDS

By

Li-Li Pan

In this dissertation computational chemistry methods were applied to study the properties of prenyltransferase FtmPT1 and small ligand conformational changes. First, I have done thorough investigations on FtmPT1 enzyme catalysis by the quantum mechanics and molecular mechanics method. FtmPT1 is a fungal indole prenyltransferase that catalyzes the reaction of tryptophan derivatives with dimethylallyl pyrophosphate to form various biologically active compounds. Herein I describe detailed studies of FtmPT1 catalysis involving dimethylallyl pyrophosphate and Brevianamide F following the native pathway (yielding Tryprostatin B) and an alternate pathway observed in the Gly115Thr mutant of FtmPT1, yielding a novel cyclized product. Importantly, these two products arise from the same intermediate state, meaning that a step other than the cleavage of the dimethylallyl pyrophosphate (DMAPP; C-O) bond is differentiating between the two product reaction channels. From detailed potential of mean force (PMF) and 2-D PMF analyses I conclude that the rate-limiting step is the cleavage of the C-O bond in DMAPP, while the deprotonation/cyclization step determines the final product distribution. Hence, in the case of FtmPT1, the optimization of the necessary catalytic machinery guides the generation of the final product after the intermediate carbocation formation.

I also describe a conformational search algorithm using the “Movable Type” (MT) sampling method. Differing from traditional systematic and stochastic searching algorithms, this method uses probability information to facilitate the selection of the best conformations. The generated

ensembles provided good coverage of the available conformation space including available crystal structures. Furthermore, my approach directly provides the solvation free energies and the relative gas and aqueous phase free energies for all generated conformers. The method was validated against thorough analyses on thrombin ligands as well as against structures extracted from both the Protein Data Bank (PDB) and the Cambridge Structural Database (CSD). These studies demonstrate that this MT-based ligand conformational search algorithm is a powerful approach for delineating the conformer ensembles of molecular species.

This dissertation is dedicated to my parents, for their supports, encouragements, and constant loves that have sustained me throughout my life.

ACKNOWLEDGMENTS

First of all I would like to thank my advisor Professor Kenneth Merz *Jr*. for supervising me during my graduate studies in both University of Florida and Michigan State University. Professor Merz generously opened a door for me to the computational biochemistry world. I really appreciate his supervision, and always being so supportive to my research and career since I joined his group in January 2011.

I would like to thank my committee members here in MSU, Professor James Jackson, Professor Robert Cukier, and Professor Benjamin Levine for their time and kindness. I would also like to thank my former committee members in UF, Professor Rodney Bartlett, Professor Hai-Ping Cheng, Professor Gail Fanucci, and Dr. Erik Deumens for their time and suggestions for my qualifying exam. I also want to thank Professor Andy Zhong from University of Nebraska Omaha for his suggestions to my dissertation writing.

I would like to thank Professor Jun Li, my master advisor in Tsinghua University, for his supervision and supports for over ten years. I started my chemistry journey since I entered Tsinghua University in 2002, yet I did not find my way to theory until I met Professor Li in 2005. Professor Li has taught me not only how to start theoretical chemistry research, but also how to have a big heart on everything. He is the best person I have ever met in my life.

I would like to thank Professor So Hirata for his supervision for over two years. I appreciate his trainings to help me obtain a solid background on quantum chemistry. Studying in Professor

Hirata's group has motivated me to start learning how to do coding and derive equations. Although I no longer do research on that field, I still benefit a lot from the first two years' study in his group.

I would like to thank all current and former Merz group members especially Dr. Yue Yang, Dr. Dhruva Chakrovorty, Dr. Zheng Zheng, and Dr. Yipu Miao for their collaboration and helps during my Ph.D studies. I also want to thank my best friends Dr. Xiaochang Miao, Dr. Di Zhou and To-be Dr. Li Zhao for their constant encouragements. I really appreciate all supports and helps I have received from these wonderful people I have met since 2006.

I would like to thank my parents Mr. Lingmei Pan and Ms. Xiaojuan Lin for their supports. My father decided to give his daughters a better future so at age of seven I moved out from a small beach town in Southeast part of China and my life has been changed ever since. Although my parents do not know any science, nor did they have the chance to go to college, they have taught me a lot on how to deal with difficulties. I really appreciate their efforts for their daughters' future and I love them with all my heart. I also want to thank my sister Ms. Dandan Pan and her husband Dr. Zimin Wang for taking good care of my parents these years during my absence.

Finally I want to thank myself for persevering in computational chemistry for over ten years. I had lots of dreams during my undergraduate years, and had a number of choices upon graduation, but finally I chose chemistry as my long time career. Science is always one of the most difficult goals to achieve, and it is never simple to get a Ph.D in science. No matter what

the future may hold, I will still be proud of my research and every details in science, which I hope will become a source of delight when I look back in my old age.

PREFACE

I started doing computational chemistry research on biomolecular systems with Professor Kenneth Merz in 2011. I was attracted by biomolecular systems and determined to do a thorough research on them. I realized that the free energy term is a critical term that directly reflects the properties of a biosystem. For one thing, free energy surface is a key point to estimate the mechanisms for enzyme catalysis reactions. The accurate but expensive potential of mean force (PMF) is commonly used to determine free energy surfaces for reaction barriers. Therefore, finding a way to directly calculate the free energy surface is essential to shorten the mechanism searching time for reaction catalysis studies. For another thing, free energy could also be a judgment of a protein-ligand binding evaluation. Whether drug design and discovery will be accelerated by current sampling methods by quickly carrying out free energies as accurately as the PMF or other free energy calculation techniques, is another important research topic in the current computational chemistry field.

Hence, my dissertation focuses on two important computational chemistry problems: enzyme reaction catalyses and small ligand conformational search. These are referred to as the prenyl transfer reaction investigation for prenyltransferase FtmPT1, and the ligand conformational search algorithm development. Quantum mechanics / molecular mechanics (QM/MM), and classical molecular dynamics (MD) are the main techniques to approach this problem. The potential of mean force (PMF) and "Movable Type" (MT) energy sampling methods are the main tools I use in my research to obtain the free energy data for these systems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

AMBER – Assisted Model Building with Energy Refinement

DMA^+ – Dimethylallyl carbocation

DMAPP – Dimethylallyl Pyrophosphate

GAFF – General Amber Force Field

IPP – Isoprenyl Pyrophosphate

MD – Molecular Dynamics

PME – Particle Mesh Ewald

PT – PrenylTransferase

QM/MM – Quantum Mechanics/Molecular Mechanics

RESP – Restrained Electrostatic Potential

RMSD – Root Mean Squared Deviation

RMSF – Root Mean Square Fluctuation

SANDER – Simulated Annealing with NMR-Derived Energy Restraints

TS – Transition State

WHAM – Weighted Histogram Analysis Method

CHAPTER 1. INTRODUCTION TO PRENYLTRANSFERASE, THROMBIN, AND SMALL LIGAND CONFORMATIONS

1.1 PRENYLTRANSFEREASES TO CATALYZE PRENYL TRANSFER REACTIONS INSIDE THE ENZYME SYSTEMS

Prenyltransferases are a class of enzymes that can catalyze the prenyl transfer reactions at their active binding sites.[1-3] Prenyl transfer, as the name implies, describes the chemical reaction that transfers an allylic prenyl group to acceptor molecules from prenyl diphosphate compounds, as shown in Figure 1.1. There are three main classes of prenyltransferases (PTs): "(a) isoprenyl pyrophosphate synthases for linear isoprenoid compounds; (b) protein prenyltransferases for isoprenyl pyrophosphate transfer to a protein or a peptide; and (c) prenyltransferases to catalyze the cyclization of isoprenyl pyrophosphates."[4] Here the isoprenyl pyrophosphate (IPP) synthases of linear terpenoid compounds are a standard prenyl transfer reaction class. Vitamin K, for instance, is one class of the linear prenyl transfer products. (Figure 1.2) Meanwhile, the protein prenyltransferases that transfer the IPPs are also important in studying the proteins like FTase, a protein in the signal transduction pathway.. The Merz group has studied the binding and catalytic mechanism of Orf2 and FTase, two proteins important in the prenyl transfer reactions.[5-12] These two complexes contain a Mg^{2+} cation in the center of the substrate reactions. It is believed that this magnesium ion plays a critical role in regulating catalyses, ligand binding, and stabilization of the intermediate states. What is more, these two enzymes follow a (N/D)DxxD motif and '$\alpha\beta\beta\alpha$' duplicate rule, which are important in determining the formation of the final product.[13, 14] A representative of the aromatic prenyl transfer synthases is Orf2 in Figure 1.3, featuring a substrate bound at the active binding site.

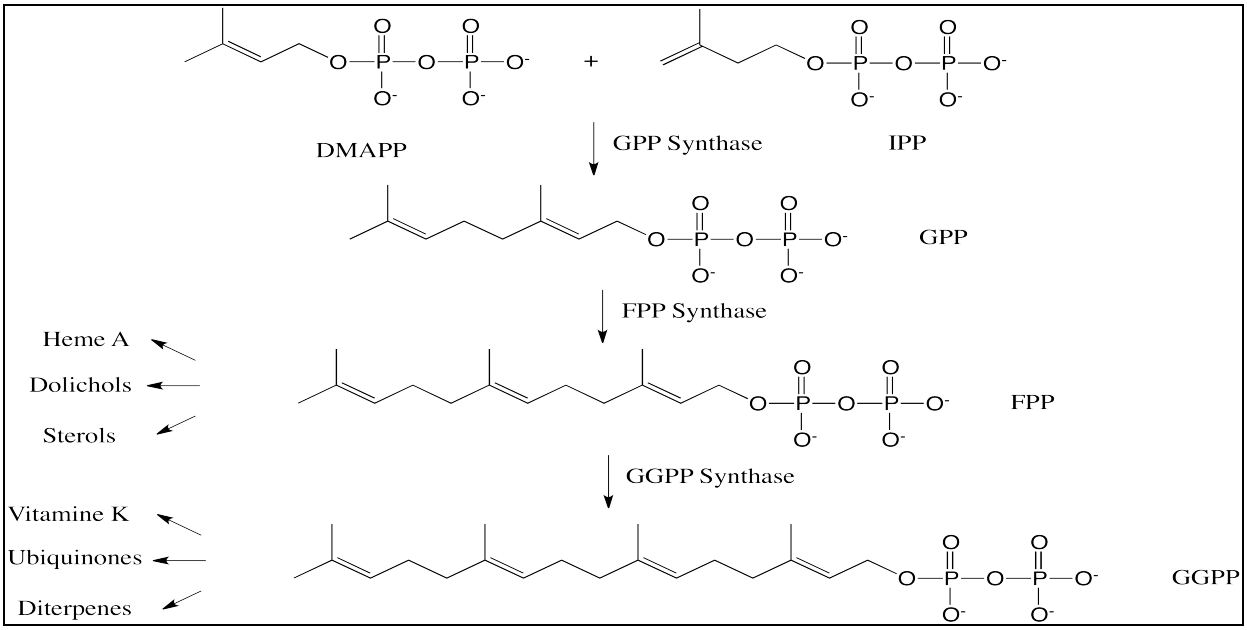


Figure 1.1 Examples for prenyltransferase catalyses[15]

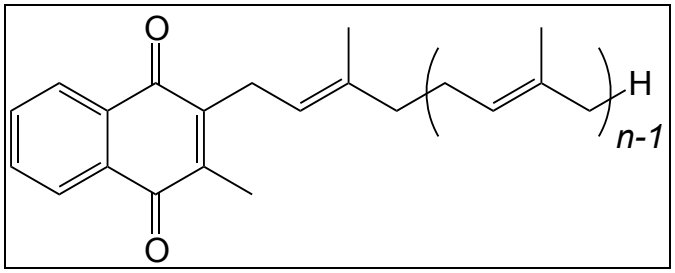


Figure 1.2 Structures of Vitamin K family

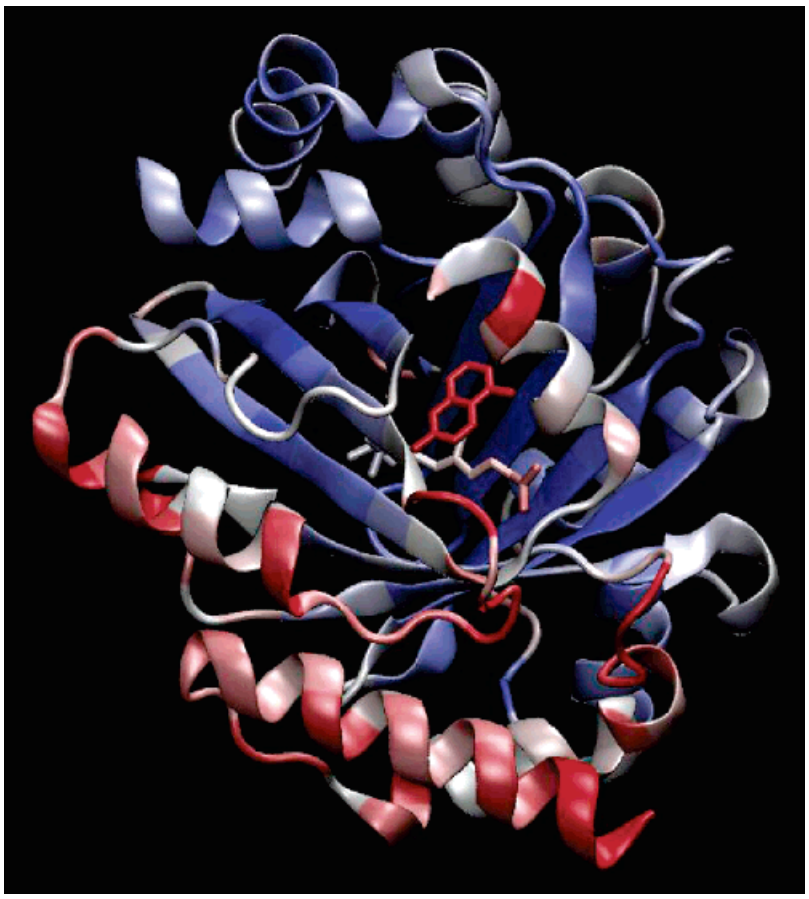


Figure 1.3 Orf2 conformation with ligand bound at the binding site[6]

In particular, aromatic prenyltranferases are currently of great research interest in fields as diverse as nutritional science and drug development because the products are natural products with low toxicity.[14-23] The aromatic prenyltransferases' catalysts are different from the protein prenyltransferases, which catalyze the large prenyl groups to cysteine protein residues in special amino acid sequences, *e.g.* farnesyl pyrophosphate[24] (FPP) or geranyl pyrophosphate[25] (GPP). Instead, the aromatic prenyltransferase exists to catalyze reactions between small isoprenoids such as isoprenyl pyrophosphate (IPP) or dimethylallyl pyrophosphate (DMAPP) and aromatic substrates. Its biosynthesis procedure happens during the primary and secondary metabolites for fungi, bacteria and plants, and hence attracts the attention of both theoretical and experimental

experts.[6, 11, 14, 17-19, 21, 26] A number of experimental studies have been done to examine the structures of these isoprenoids and related biosynthesis in the past 20 years, as well as computational studies of their conformational changes and reaction mechanisms in the past 10 years.[5-40] In terms of experimental research, S.-M. Li *et al*. have done pioneering work in prenyl transfer to fungi indole and corresponding aromatic compounds such as Tryprostatin A, B, Fumitremorgin B, C, Roquefortine C, etc.[18, 22, 37, 38, 41] These fungi's prenylated indole derivatives were successfully cloned and their crystal structures have been studied by the experimental experts as well.

The (N/D)DxxD motifs are essential to the aromatic prenyltransferases, as well as for prenyl diphosphate bindings and catalyses. Besides, metal cations such as Mg^{2+} and Zn^{2+} are also important for aromatic PTs since their existence can enhance or prohibit the prenyl transfer reactions. However, (N/D)DxxD and metal cations do not necessarily have to coexist for the aromatic prenyl transfers but one of them must be present so that the reaction and bindings in the enzyme can take place.[19, 23, 32] The fungal indole prenyltransferase synthase, for instance, does not include any metal cation inside. Instead, a (N/D)DxxD motif plays a critical role in the synthesis procedure. My targeted aromatic prenyltransferase system, the FtmPT1, includes two chains in the protein part: the α helices and β strands, which follow αααββααααββααα-ββααββααββ order. This order is consistent with the 'αββα' duplicate rule in aromatic prenyl transfer motifs. The detailed structures will be discussed in the following chapters.

As an aromatic prenyltransferase, the FtmPT1 could catalyze reactions between dimethyl allyl pyrophosphate (DMAPP) and a tryptophan derivative compound Brevianamide F to synthesize

another tryptophan derivative Tryprostatin B. As a cell growth inhibitor, Tryprostatin B is considered a promising cancer treatment drug. Furthermore, this enzyme does not only have the native Tryprostatin B product for the wild type system, but also could produce a novel five-member ring product for the Gly115Thr mutated system.[22] These different product formations by FtmPT1 catalysis are brought about by its special property of "regioselectivity".[37] Brevianamide F, one of the substrates, has C8 and C9 reactive atoms on its indole rings, while DMAPP, the other substrate, has C2 and C5 reactive atoms. The catalyzing process includes a prenylation reaction and a subsequent deprotonation reaction. In the prenylation reaction, the DMAPP first breaks its O-C bond and forms a dimethylallyl (DMA^+) carbocation intermediate compound, then the intermediate carbocation is attacked by a reactive carbon atom on the Brevianamide F indole ring and forms a C-C bond. The competitions happen when the reactants are trying to form a C-C bond between the indole ring and the DMA^+ cation and can result in different products respectively. The wild type FtmPT1 catalyzes the prenylation reaction between C5@DMAPP and C8@Brevianamide F, while the G115T mutated system catalyzes the reaction between C2@DMAPP and C9@Brevianamide F. Both of the reactions are finally completed with the extra proton transferred to Glu102, which forms a stable hydrogen bond with the nitrogen atom on the indole ring of Brevianamide F. The experiment also shows that Glu102 plays a significant role in the reaction, since no product would be yielded if this glutamic acid residue is replaced by any other residues. The final product forms and binds stably in the binding pocket by forming two hydrogen bonds, which finishes the whole biosynthesis process. From experiment results for the wild type, the winner is the C5-C8 combination; and for G115T, the winner is the C2-C9 combination with a reduction of product yield by 80%. The reactants and products are shown in Figure 1.4.

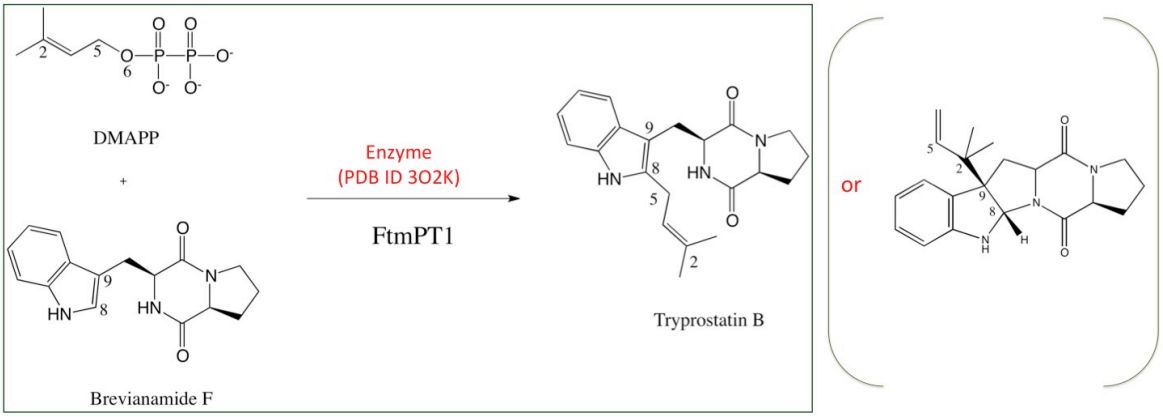


Figure 1.4 Reactions catalyzed by aromatic prenyltransferase FtmPT1 to produce different products via regioselectivity

Because of this interesting regioselectivity to produce different products, in my studies on the FtmPT1 system, the FtmPT1 catalysis was calculated to investigate the insights, not only for the wild type enzyme but also for the Gly115Thr mutated system. The reactions were reproduced by implementation of the potential of mean force (PMF) to evaluate the reaction free energy barriers for both C5-C8 and C2-C9 bond formations for comparison, and the assumptions of the reaction mechanisms were also validated. The proton transfer acceptors as well as the final products' binding conformations were identified. The quantum mechanical / molecular mechanics (QM/MM) technique was used to realize the processes.

1.2 CONFORMATIONAL SEARCH TO OBTAIN SMALL LIGAND CONFORMER ENSEMBLES

Conformational search for the small ligands and bioactive compounds is an essential computational chemistry process to approach the following cheminformatics procedures such as virtual screening, similarity searching, pharmacophore modeling, protein-ligand binding, etc..[42-48] Hence conformational search studies have been hot topics for decades.[49-53] Up to now, a number of conformational search algorithms have been developed to investigate the low-lying conformers for a given ligand or small compound.[54-85] These programs could generate various poses of conformers for a proposed molecule according to their specific criteria.

Conformational search programs have both systematic and random searching methods.[49] Systematic searching methods provide a searching logic of regular variation of the flexible geometrical parameters (Figure 1.5), while random searching methods include a structural random growth procedure (Figure 1.6). The former searching method then evaluates the conformer qualities using its relevant force field, while the latter utilizes its force field to minimize the conformers. The systematic method can have a broad coverage of the conformers, however the systematically generated variables may bring an exponential increase in the conformer generating time. Meanwhile, the random method can have a dramatic improvement of the generating speed, yet the coverage of the conformer energy surface is unknown due to its random generating algorithm. There are three major criteria that must be satisfied for the conformational search algorithm to represent the conformations of the molecule systems:[86] (1) a well defined molecular potential energy surface; (2) a highly efficient conformation sampling

method; (3) good convergence criteria. Therefore, finding the best solution for getting a good coverage of the energy surface without losing speed is the major concern in developing a new ligand conformational search algorithm. Figure 1.7 is a random conformer energy surface with ideal coordinates and various low energy wells.

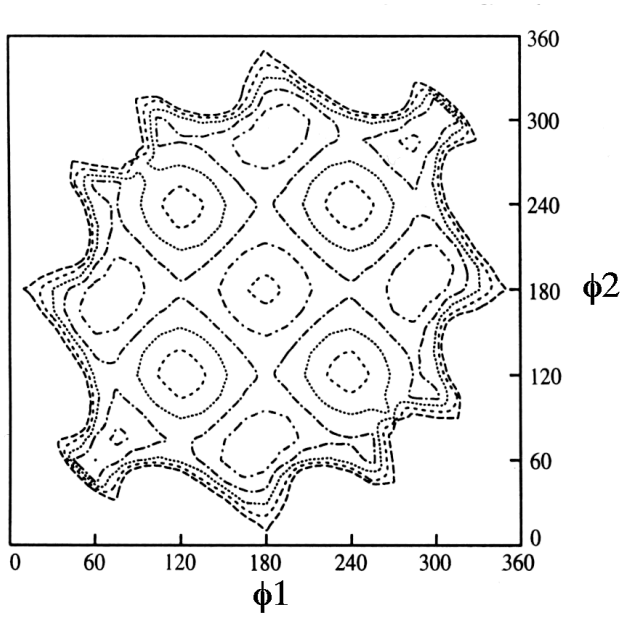


Figure 1.5 A general systematic variation of two dihedrals inside a molecule for the system search method[87]

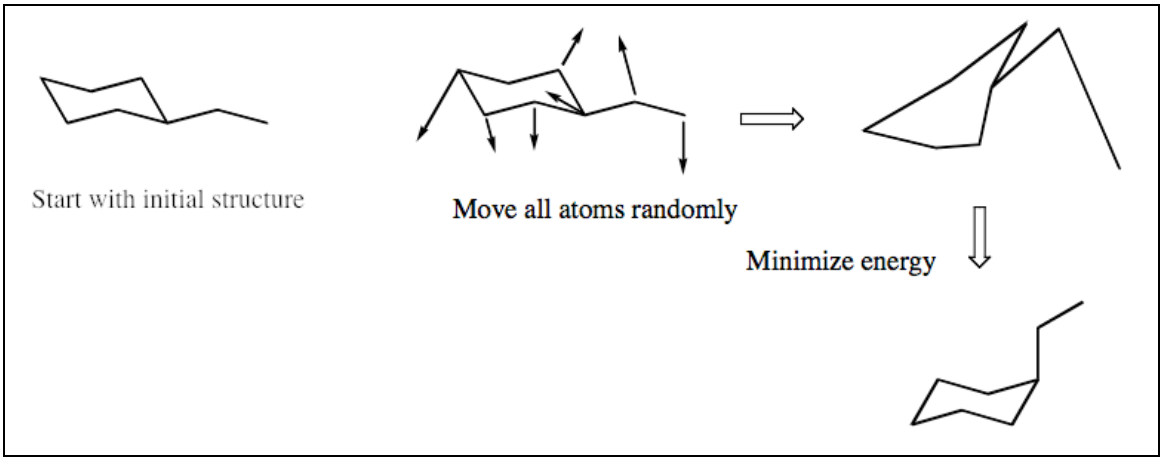


Figure 1.6 A general conformer generating process for a random search method[88]

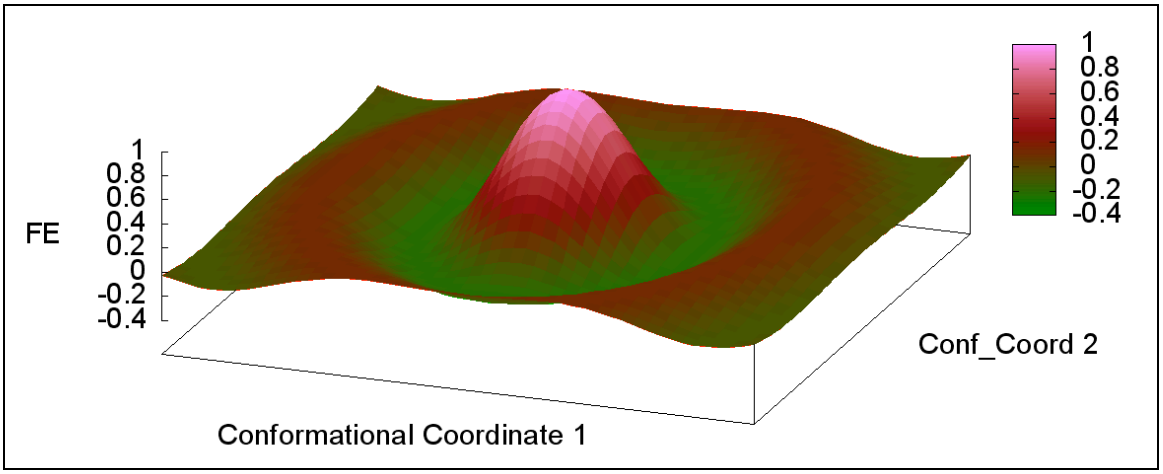


Figure 1.7 A sample free energy surface for a random molecule with respect to its two conformer coordinates

Nevertheless, the random searching algorithm is the major method that is widely implemented in modern conformational search programs. Genetic Algorithm,[67, 68, 89] Monte Carlo Simulated Annealing[75, 86, 90, 91] and distance geometry[73, 92-94] are the three major methods used among all random search algorithms. These algorithms have a big advantage on the generating speed, and

hence the efficiency of the programs. However, the random generating properties' determination of the coverage of low energy conformers is unknown. People nowadays are trying to use *ab initio* optimizations based on the generated conformers to determine the global minimum conformer of a given molecule.[95, 96] However, the cost of calculation time is much higher than a regular conformation search method. Therefore, finding a method to develop a program that can locate the global minimum quickly and efficiently is a major challenge in the modern conformational search field.

In my research on conformational search algorithm development, I implemented my group's newly-developed energy sampling method, the "Movable Type" method.[97] This "Movable Type" method, which has a similar logic to the wood-block printing method invented in ancient Chna, as shown in Figure 1.8, includes two independently-prepared databases that are ready for conformational search: (1) probabilities of bonds, angles, dihedrals, and nonbonded interactions with respect to pairwise atomic types in terms of distance; and (2) corresponded pairwise distant dependent energies.[97, 98] The free energy term for each atom pair hence can be directly evaluated without sampling by directly multiplying the two databases. By picking the largest numbers, I no longer need to do a systematic variation search to cover the energy surface while sacrificing the speed, nor do I need to do a random search to maintain a fast speed while sacrificing the energy surface coverage. In this way, I can get a fairly good coverage of the energy surface with a good speed. What is more, I could even carry out the relative gas phase free energies simultaneously through generating the conformers, and a following calculation of the solvation free energy could support the algorithm with solvation free energies[99] and relative aqueous phase free energies. The implementation of the "Movable Type" method could thus make my ligand conformational

search program more practical for protein-ligand binding usage. As a benchmark, I tested the program with approximately two hundred protein-bound ligand structures from the Protein Database Bank (PDB)[100, 101], as well as hundreds of small molecules from the Cambridge Structural Database (CSD)[102, 103], and further proved that this program is efficient and extensive for Computational Aided Drug Design (CADD) studies.

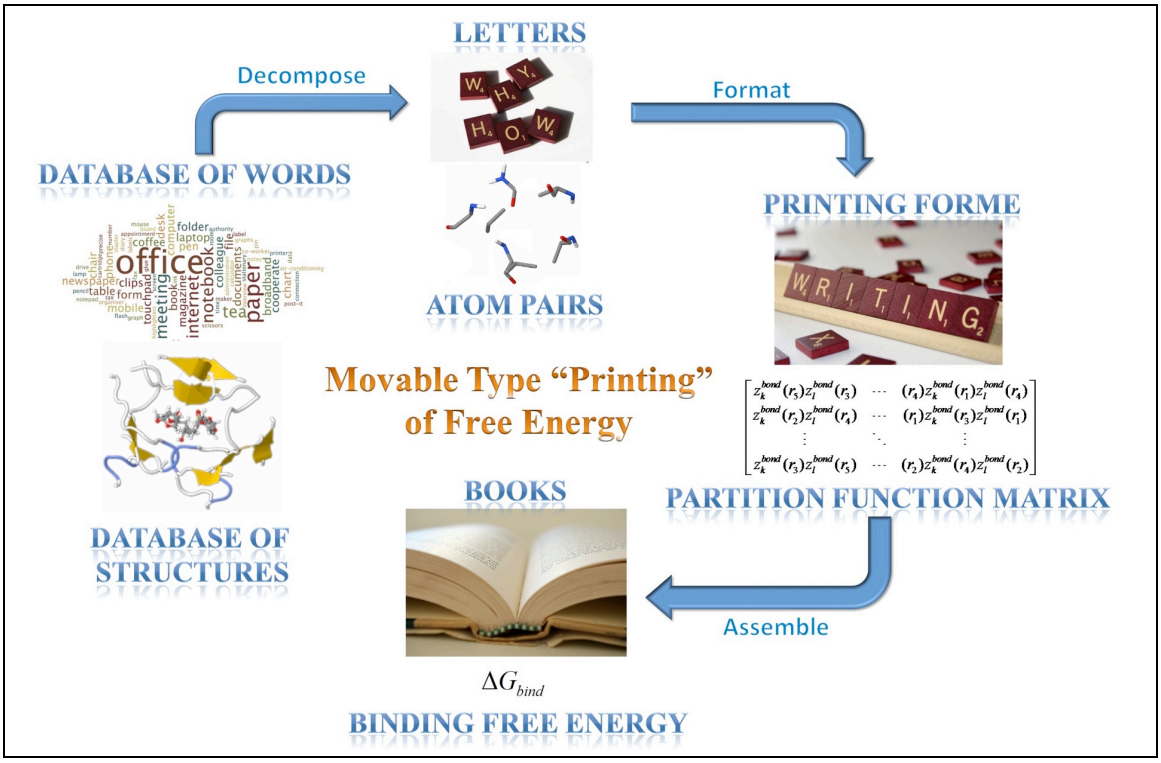


Figure 1.8 Introduction of "Movable Type" energy sampling method compared to movable type printing method[97]

REFERENCES

REFERENCES

1. Poulter, C. D.; Rilling, H. C., Prenyltransferase: the mechanism of the reaction. *Biochemistry-Us* **1976**, 15, 1079-83.

2. Poulter, C. D.; Satterwhite, D. M., Mechanism of Prenyl-Transfer Reaction - Studies with (E)-3-Trifluoromethyl-2-Buten-1-Yl and (Z)-3-Trifluoromethyl-2-Buten-1-Yl Pyrophosphate. *Biochemistry-Us* **1977**, 16, 5470-5478.

3. Poulter, C. D.; Rilling, H. C., Prenyl Transfer-Reaction - Enzymatic and Mechanistic Studies of 1'-4 Coupling Reaction in Terpene Biosynthetic-Pathway. *Accounts Chem Res* **1978**, 11, 307-313.

4. Liang, P. H.; Ko, T. P.; Wang, A. H., Structure, mechanism and function of prenyltransferases. *Eur J Biochem* **2002**, 269, 3339-54.

5. Cui, G. L.; Wang, B.; Merz, K. M., Computational studies of the farnesyltransferase ternary complex - Part I: Substrate binding. *Biochemistry-Us* **2005**, 44, 16513-16523.

6. Cui, G.; Li, X.; Merz, K. M., Jr., Understanding the substrate selectivity and the product regioselectivity of Orf2-catalyzed aromatic prenylations. *Biochemistry-Us* **2007**, 46, 1303-11.

7. Cui, G.; Merz, K. M., . Computational studies of the farnesyltransferase ternary complex part II: The conformational activation of farnesyldiphosphate. *Biochemistry-Us* **2007**, 46, 12375-12381.

8. Cui, G. L.; Li, X.; Merz, K. M., Alternative binding modes in Orf2 catalyzed prenylation: Identification and validation. *Abstr Pap Am Chem S* **2007**, 233.

9. Cui, G. L.; Merz, K. M., The intrinsic dynamics and function of nickel-binding regulatory protein: Insights from elastic network analysis (vol 94, pg 3769, 2008). *Biophys J* **2008**, 94, 5081-5081.

10. Yang, Y.; Chakravorty, D. K.; Merz, K. M., Jr., Finding a needle in the haystack: computational modeling of Mg2+ binding in the active site of protein farnesyltransferase. *Biochemistry-Us* **2010**, 49, 9658-66.

11. Yang, Y.; Miao, Y. P.; Wang, B.; Cui, G. L.; Merz, K. M., Catalytic Mechanism of Aromatic Prenylation by NphB. *Biochemistry-Us* **2012**, 51, 2606-2618.

12. Yang, Y.; Wang, B.; Ucisik, M. N.; Cui, G. L.; Fierke, C. A.; Merz, K. M., Insights into the Mechanistic Dichotomy of the Protein Farnesyltransferase Peptide Substrates CVIM and CVLS. *J Am Chem Soc* **2012**, 134, 820-823.

13. Tello, M.; Kuzuyama, T.; Heide, L.; Noel, J. P.; Richard, S. B., The ABBA family of aromatic prenyltransferases: broadening natural product diversity. *Cell Mol Life Sci* **2008**, 65, 1459-1463.

14. Saleh, O.; Haagen, Y.; Seeger, K.; Heide, L., Prenyl transfer to aromatic substrates in the biosynthesis of aminocoumarins, meroterpenoids and phenazines: The ABBA prenyltransferase family. *Phytochemistry* **2009**, 70, 1728-1738.

15. Nurenberg, G.; Volmer, D. A., The analytical determination of isoprenoid intermediates from the mevalonate pathway. *Anal Bioanal Chem* **2012**, 402, 671-685.

16. Chen, A. J.; Kroon, P. A.; Poulter, C. D., Isoprenyl Diphosphate Synthases - Protein-Sequence Comparisons, a Phylogenetic Tree, and Predictions of Secondary Structure. *Protein Sci* **1994**, 3, 600-607.

17. Kuzuyama, T.; Noel, J. P.; Richard, S. B., Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. *Nature* **2005**, 435, 983-987.

18. Li, S. M., Evolution of aromatic prenyltransferases in the biosynthesis of indole derivatives. *Phytochemistry* **2009**, 70, 1746-1757.

19. Bonitz, T.; Alva, V.; Saleh, O.; Lupas, A. N.; Heide, L., Evolutionary Relationships of Microbial Aromatic Prenyltransferases. *Plos One* **2011**, 6.

20. Sinko, W.; de Oliveira, C.; Williams, S.; Van Wynsberghe, A.; Durrant, J. D.; Cao, R.; Oldfield, E.; McCammon, J. A., Applying Molecular Dynamics Simulations to Identify Rarely Sampled Ligand-bound Conformational States of Undecaprenyl Pyrophosphate Synthase, an Antibacterial Target. *Chem Biol Drug Des* **2011**, 77, 412-420.

21. Heide, L., Prenyl transfer to aromatic substrates: genetics and enzymology. *Curr Opin Chem Biol* **2009**, 13, 171-179.

22. Jost, M.; Zocher, G.; Tarcz, S.; Matuschek, M.; Xie, X. L.; Li, S. M.; Stehle, T., Structure-Function Analysis of an Enzymatic Prenyl Transfer Reaction Identifies a Reaction Chamber with Modifiable Specificity. *J Am Chem Soc* **2010**, 132, 17849-17858.

23. Yu, X.; Xie, X. L.; Li, S. M., Substrate promiscuity of secondary metabolite enzymes: prenylation of hydroxynaphthalenes by fungal indole prenyltransferases. *Appl Microbiol Biot* **2011**, 92, 737-748.

24. Davisson, V. J.; Neal, T. R.; Poulter, C. D., Farnesyl-Diphosphate Synthase - Catalysis of an Intramolecular Prenyl Transfer with Bisubstrate Analogs. *J Am Chem Soc* **1993**, 115, 1235-1245.

25. Brajeul, S.; Delpech, B.; Marazano, C., Sulfonium salts as prenyl, geranyl, and isolavandulyl transfer agents towards benzoylphloroglucinol derivatives. *Tetrahedron Lett* **2007**, 48, 5597-5600.

26. Chen, J.; Morita, H.; Wakimoto, T.; Mori, T.; Noguchi, H.; Abe, I., Prenylation of a Nonaromatic Carbon of Indolylbutenone by a Fungal Indole Prenyltransferase. *Org Lett* **2012**, 14, 3080-3083.

27. Amatore, C.; Thiebault, A.; Verpeaux, J. N., Unexpected Single Electron-Transfer Catalyzed Cyclization of Prenyl Sulfone Dimer - Evidence for Radical-Anion Coupling in the Outer-Sphere Oxidation of Prenyl Sulfone Carbanion. *J Chem Soc Chem Comm* **1989**, 1543-1545.

28. Zhang, D. L.; Daniels, L.; Poulter, C. D., Biosynthesis of Archaebacterial Membranes - Formation of Isoprene Ethers by a Prenyl Transfer-Reaction. *J Am Chem Soc* **1990**, 112, 1264-1265.

29. Casey, P. J.; Seabra, M. C., Protein prenyltransferases. *J Biol Chem* **1996**, 271, 5289-5292.

30. Poulter, C. D., Mechanistic studies of the prenyl transfer reaction with fluorinated substrate analogs. *Biomedical Frontiers of Fluorine Chemistry* **1996**, 639, 158-168.

31. Hong, B. C.; Hong, J. H.; Tsai, Y. C., Regio- and enantioselective prenyl anion transfer: Application to the total synthesis of (-)-rosiridol. *Angew Chem Int Edit* **1998**, 37, 468-470.

32. Liang, P. H.; Ko, T. P.; Wang, A. H. J., Structure, mechanism and function of prenyltransferases. *Eur J Biochem* **2002**, 269, 3339-3354.

33. Tantillo, D. J., The carbocation continuum in terpene biosynthesis--where are the secondary cations? *Chem Soc Rev* **2010**, 39, 2847-54.

34. Luk, L. Y. P.; Qian, Q.; Tanner, M. E., A Cope Rearrangement in the Reaction Catalyzed by Dimethylallyltryptophan Synthase? *J Am Chem Soc* **2011**, 133, 12342-12345.

35. Gordon, L. B.; Kleinman, M. E.; Miller, D. T.; Neuberg, D. S.; Giobbie-Hurder, A.; Gerhard-Herman, M.; Smoot, L. B.; Gordon, C. M.; Cleveland, R.; Snyder, B. D.; Fligor, B.; Bishop, W. R.; Statkevich, P.; Regen, A.; Sonis, A.; Riley, S.; Ploski, C.; Correia, A.; Quinn, N.; Ullrich, N. J.; Nazarian, A.; Liang, M. G.; Huh, S. Y.; Schwartzman, A.; Kieran, M. W., Clinical trial of a farnesyltransferase inhibitor in children with Hutchinson-Gilford progeria syndrome. *P Natl Acad Sci USA* **2012**, 109, 16666-16671.

36. Spindler, S. R.; Li, R.; Dhahbi, J. M.; Yamakawa, A.; Mote, P.; Bodmer, R.; Ocorr, K.; Williams, R. T.; Wang, Y. S.; Ablao, K. P., Statin Treatment Increases Lifespan and Improves Cardiac Health in Drosophila by Decreasing Specific Protein Prenylation. *Plos One* **2012**, 7.

37. Wollinsky, B.; Ludwig, L.; Xie, X. L.; Li, S. M., Breaking the regioselectivity of indole prenyltransferases: identification of regular C3-prenylated hexahydropyrrolo[2,3-b]indoles as side products of the regular C2-prenyltransferase FtmPT1. *Organic & biomolecular chemistry* **2012**, 10, 9262-9270.

38. Yu, X.; Liu, Y.; Xie, X. L.; Zheng, X. D.; Li, S. M., Biochemical Characterization of Indole Prenyltransferases Filling the Last Gap of Prenylation Positions by a 5-DimethylAllyTryptophan Synthase from Aspergillus Clavatus. *J Biol Chem* **2012**, 287, 1371-1380.

39. Mahmoodi, N.; Tanner, M. E., Potential Rearrangements in the Reaction Catalyzed by the Indole Prenyltransferase FtmPT1. *Chembiochem* **2013**, 14, 2029-2037.

40. Pan, L. L.; Yang, Y.; Merz, K. M., Jr., Origin of product selectivity in a prenyl transfer reaction from the same intermediate: exploration of multiple FtmPT1-catalyzed prenyl transfer pathways. *Biochemistry-Us* **2014**, 53, 6126-38.

41. Steffan, N.; Grundmann, A.; Yin, W. B.; Kremer, A.; Li, S. M., Indole prenyltransferases from fungi: a new enzyme group with high potential for the production of prenylated indole derivatives. *Curr Med Chem* **2009**, 16, 218-31.

42. Lorber, D. M.; Shoichet, B. K., Flexible ligand docking using conformational ensembles. *Protein Sci* **1998**, 7, 938-950.

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

44. Kristam, R.; Gillet, V. J.; Lewis, R. A.; Thorner, D., Comparison of conformational analysis techniques to generate pharmacophore hypotheses using catalyst. *J Chem Inf Model* **2005**, 45, 461-476.

45. Stahura, F. L.; Bajorath, M., New methodologies for ligand-based virtual screening. *Curr Pharm Design* **2005**, 11, 1189-1202.

46. Renner, S.; Schwab, C. H.; Gasteiger, J.; Schneider, G., Impact of conformational flexibility on three-dimensional similarity searching using correlation vectors. *J Chem Inf Model* **2006**, 46, 2324-32.

47. Conformations and 3D pharmacophore searching. *Drug discovery today. Technologies* **2010**, 7, e203-70.

48. Verma, J.; Khedkar, V. M.; Coutinho, E. C., 3D-QSAR in Drug Design - A Review. *Curr Top Med Chem* **2010**, 10, 95-115.

49. Leach, A. R., A survey of methods for searching the conformational space of small and medium-size molecules. *Reviews of Computational Chemistry, Wiley, New York* **1991**, 2, 1–55.

50. Perola, E.; Charifson, P. S., Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. *J Med Chem* **2004**, 47, 2499-510.

51. Agrafiotis, D. K.; Gibbs, A. C.; Zhu, F.; Izrailev, S.; Martin, E., Conformational sampling of bioactive molecules: a comparative study. *J Chem Inf Model* **2007**, 47, 1067-86.

52. Bonnet, P.; Agrafiotis, D. K.; Zhu, F. Q.; Martin, E., Conformational Analysis of Macrocycles: Finding What Common Search Methods Miss. *J Chem Inf Model* **2009**, 49, 2242-2259.

53. Ebejer, J. P.; Morris, G. M.; Deane, C. M., Freely Available Conformer Generation Methods: How Good Are They? *J Chem Inf Model* **2012**, 52, 1146-1158.

54. The 3D structure generator CORINA is available from Molecular Networks GmbH. *Erlangen, Germany*, http://www.molecular-networks.com.

55. Rubicon, Version 4.7.1. *Daylight Chemical Information Systems, Inc. Aliso Viejo, CA*, http://www.daylight.com.

56. Hawkins, P. C. D. S., A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T, OMEGA 2.5.1.4: OpenEye Scientific Software, Santa Fe, NM. .

57. Landrum, G., RDKit: Open-source cheminformatics. http://www.rdkit.org.

58. Klebe, G.; Mietzner, T., A Fast and Efficient Method to Generate Biologically Relevant Conformations. *J Comput Aid Mol Des* **1994**, 8, 583-606.

59. Sadowski, J.; Gasteiger, J.; Klebe, G., Comparison of Automatic 3-Dimensional Model Builders Using 639 X-Ray Structures. *J Chem Inf Comp Sci* **1994**, 34, 1000-1008.

60. Makino, S.; Kuntz, I. D., Automated flexible ligand docking method and its application for database search. *J Comput Chem* **1997**, 18, 1812-1825.

61. Balducci, R.; Pearlman, R. S., CONFORT: A rational conformation analysis tool. *Abstr Pap Am Chem S* **1999**, 217, U644-U645.

62. Hare, B. J.; Walters, W. P.; Caron, P. R.; Bemis, G. W., CORES: An automated method for generating three-dimensional models of protein/ligand complexes. *J Med Chem* **2004**, 47, 4731-4740.

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

64. Sadowski, J.; Bostrom, J., MIMUMBA revisited: Torsion angle rules for conformer generation derived from X-ray structures. *J Chem Inf Model* **2006**, 46, 2305-2309.

65. Leite, T. B.; Gomes, D.; Miteva, M. A.; Chomilier, J.; Villoutreix, B. O.; Tuffery, P., Frog: a FRee Online druG 3D conformation generator. *Nucleic Acids Res* **2007**, 35, W568-W572.

66. Li, J.; Ehlers, T.; Sutter, J.; Varma-O'Brien, S.; Kirchmair, J., CAESAR: A new conformer generation algorithm based on recursive buildup and local rotational symmetry consideration. *J Chem Inf Model* **2007**, 47, 1923-1932.

67. Pavlov, T.; Todorov, M.; Stoyanova, G.; Schmieder, P.; Aladjov, H.; Serafimova, R.; Mekenyan, O., Conformational coverage by a genetic algorithm: Saturation of conformational space. *J Chem Inf Model* **2007**, 47, 851-863.

68. Vainio, M. J.; Johnson, M. S., Generating conformer ensembles using a multiobjective genetic algorithm. *J Chem Inf Model* **2007**, 47, 2462-2474.

69. Brameld, K. A.; Kuhn, B.; Reuter, D. C.; Stahl, M., Small molecule conformational preferences derived from crystal structure data. A medicinal chemistry focused analysis. *J Chem Inf Model* **2008**, 48, 1-24.

70. Chen, I. J.; Foloppe, N., Conformational sampling of druglike molecules with MOE and catalyst: Implications for pharmacophore modeling and virtual screening. *J Chem Inf Model* **2008**, 48, 1773-1791.

71. Dorfman, R. J.; Smith, K. M.; Masek, B. B.; Clark, R. D., A knowledge-based approach to generating diverse but energetically representative ensembles of ligand conformers. *J Comput Aid Mol Des* **2008**, 22, 681-691.

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

73. Lagorce, D.; Pencheva, T.; Villoutreix, B. O.; Miteva, M. A., DG-AMMOS: a new tool to generate 3d conformation of small molecules using distance geometry and automated molecular mechanics optimization for in silico screening. *BMC chemical biology* **2009**, 9, 6.

74. Liu, X. F.; Bai, F.; Ouyang, S. S.; Wang, X. C.; Li, H. L.; Jiang, H. L., Cyndi: a multi-objective evolution algorithm based method for bioactive molecular conformational generation. *BMC bioinformatics* **2009**, 10.

75. Sperandio, O.; Souaille, M.; Delfaud, F.; Miteva, M. A.; Villoutreix, B. O., MED-3DMC: a new tool to generate 3D conformation ensembles of small molecules with a Monte Carlo sampling of the conformational space. *Eur J Med Chem* **2009**, 44, 1405-9.

76. Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T., Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J Chem Inf Model* **2010**, 50, 572-584.

77. Miteva, M. A.; Guyon, F.; Tuffery, P., Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Res* **2010**, 38, W622-7.

78. Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C., ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers. *J Chem Inf Model* **2010**, 50, 534-546.

79. Grebner, C.; Becker, J.; Stepanenko, S.; Engels, B., Efficiency of Tabu-Search-Based Conformational Search Algorithms. *J Comput Chem* **2011**, 32, 2245-2253.

80. O'Boyle, N. M.; Vandermeersch, T.; Flynn, C. J.; Maguire, A. R.; Hutchison, G. R., Confab - Systematic generation of diverse low-energy conformers. *Journal of cheminformatics* **2011**, 3.

81. Hawkins, P. C. D.; Nicholls, A., Conformer Generation with OMEGA: Learning from the Data Set and the Analysis of Failures. *J Chem Inf Model* **2012**, 52, 2919-2936.

82. Molecular Operating Environment (MOE), Version 2013.08. *Chemical Computing Group Inc. , 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7* **2013**.

83. Oda, A.; Yamaotsu, N.; Hirono, S.; Takano, Y.; Fukuyoshi, S.; Nakagaki, R.; Takahashi, O., Evaluations of the conformational search accuracy of CAMDAS using experimental three-dimensional structures of protein-ligand complexes. *Jounal of Physics: COnference Series* **2013**, 454.

84. Macromodel, Version 10.4. *Schrödinger, Inc., New York, NY* **2014**.

85. Inc., A. S., Catalyst, Release 4.7. *San Diego: Accelrys Software Inc*. **2014**.

86. Parish, C.; Lombardi, R.; Sinclair, K.; Smith, E.; Goldberg, A.; Rappleye, M.; Dure, M., A comparison of the Low Mode and Monte Carlo conformational search methods. *J Mol Graph Model* **2002**, 21, 129-50.

87. Leach, A. R., Molecular Modeling - Principles and Applications, 2nd ed. *Prentice Hall* **2001**, Chaoter 9, p 457.

88. Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C., Conformations of Cycloheptadecane - a Comparison of Methods for Conformational Searching. *J Am Chem Soc* **1990**, 112, 1419-1427.

89. Mekenyan, O.; Dimitrov, D.; Nikolova, N.; Karabunarliev, S., Conformational coverage by a genetic algorithm. *J Chem Inf Comp Sci* **1999**, 39, 997-1016.

90. Wilson, S. R.; Cui, W.; Moskowitz, J. W.; Schmidt, K. E., Applications of Simulated Annealing to the Conformational-Analysis of Flexible Molecules. *J Comput Chem* **1991**, 12, 342-349.

91. Chang, G.; Guida, W. C.; Still, W. C., An Internal Coordinate Monte-Carlo Method for Searching Conformational Space. *J Am Chem Soc* **1989**, 111, 4379-4386.

92. Havel, T. F.; Kuntz, I. D.; Crippen, G. M., The Theory and Practice of Distance Geometry. *B Math Biol* **1983**, 45, 665-720.

93. Spellmeyer, D. C.; Wong, A. K.; Bower, M. J.; Blaney, J. M., Conformational analysis using distance geometry methods. *J Mol Graph Model* **1997**, 15, 18-36.

94. Feuston, B. P.; Miller, M. D.; Culberson, J. C.; Nachbar, R. B.; Kearsley, S. K., Comparison of knowledge-based and distance geometry approaches for generation of molecular conformations. *J Chem Inf Comp Sci* **2001**, 41, 754-763.

95. Fu, Z.; Li, X.; Merz, K. M., Accurate Assessment of the Strain Energy in a Protein-Bound Drug Using QM/MM X-ray Refinement and Converged Quantum Chemistry. *J Comput Chem* **2011**, 32, 2587-2597.

96. Fu, Z.; Li, X.; Merz, K. M., Conformational Analysis of Free and Bound Retinoic Acid. *J Chem Theory Comput* **2012**, 8, 1436-1448.

97. Zheng, Z.; Ucisik, M. N.; Merz, K. M., The Movable Type Method Applied to Protein-Ligand Binding. *J Chem Theory Comput* **2013**, 9, 5526-5538.

98. Zheng, Z.; Merz, K. M., Development of the Knowledge-Based and Empirical Combined Scoring Algorithm (KECSA) To Score Protein-Ligand Interactions. *J Chem Inf Model* **2013**, 53, 1073-1083.

99. Zheng, Z.; Wang, T.; Li, P. F.; Merz, K. M., KECSA-Movable Type Implicit Solvation Model (KMTISM). *J Chem Theory Comput* **2015**, 11, 667-682.

100. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000**, 28, 235-42.

101. Battle, G.; Kleywegt, G.; Velankar, S.; Oldfield, T.; Gore, S.; Mir, S.; Dana, J., Chemistry-related resources at the Protein Data Bank in Europe. *Abstr Pap Am Chem S* **2013**, 245.

102. Allen, F. H., The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr B* **2002**, 58, 380-388.

103. Henderson, S.; Bruno, I. J., Cambridge Structural Database: Moving with the times. *Abstr Pap Am Chem S* **2013**, 245.

CHAPTER 2. PROJECT OBJECTIVES

The main objectives of this work were:

A. Find the reaction mechanism for FtmPT1 reaction catalyses, including reactions in both wild type and Gly115Thr mutated systems. There are two competitive reactions happening in both systems, so I determined the reaction pathway by evaluating the reaction free energies. The experimental experts estimated the residue Glu102 as the proton acceptor for the proton transfer reaction, so I also watched the reactions to value the importance of this residue. Furthermore, the mutated system catalysis brings up a different regioselectivity compared to the wild type catalysis, so I also tried to find out the potential dynamics effects on the reaction pathway preference besides the thermodynamics ways. I also implemented quantum mechanics calculations on the key states of the mutated system reaction catalysis to evaluate the energy difference I carried out by using QM/MM method.

B. For the conformational search algorithm development, I am trying to find out an efficient way for the program to maintain a fast generating speed as well as the high-quality coverage of the conformer energy surface. Hence, I targeted the development of the program itself, and used hundreds of PDB and CSD structures to evaluate the quality of my program.

CHAPTER 3. ORIGIN OF PRODUCT SELECTIVITY IN A PRENYL TRANSFER REACTION FROM THE SAME INTERMEDIATE: EXPLORATION OF MULTIPLE FTMPT1 CATALYZED PRENYL TRANSFER PATHWAYS

3.1 INTRODUCTION

Prenyl transfer reactions are ubiquitous in nature and are utilized by organisms to generate a wide range of biologically active small molecules.[1-4] The fungal indole prenyltransferases (indole PTs) are a unique example of this class of enzyme. They differentiate themselves from other PTs, by not requiring metal cofactors such as Mg^{2+} and Zn^{2+} to affect the prenyl transfer reaction, nor do they require the (N/D)DxxD motif seen in other PTs for binding or catalysis.[5-22] However, the 'αββα' duplicate chain order rule for the α helices and β strands is satisfied in this class of PTs.[20] Recently a number of experimental papers have been published on fungal indole PTs exploring their regioselectivity.[23-26] Multiple reaction channels are possible, so the fundamental question is how one reaction channel is favored over another by the characteristics of the active site of the protein catalyst involved. Importantly, once the C-O bond in DMAPP is cleaved the dimethyl allyl carbocation can react at one of two ends of the molecule: the less sterically crowded C5 position or the more crowded C2 site (see Figure 3.1). Indeed, both product types are observed and the fundamental question arises how can the formation of a sterically more crowded product be favored over a less crowded reaction channel by active site characteristics. Both pathways yield interesting molecules, so the ultimate goal is to tune one pathway over another in order to synthesize the more desirable product. Figure 3.1 shows, as an example, the possible reaction products for the reaction of Brevianamide F and DMAPP as catalyzed by FtmPT1.

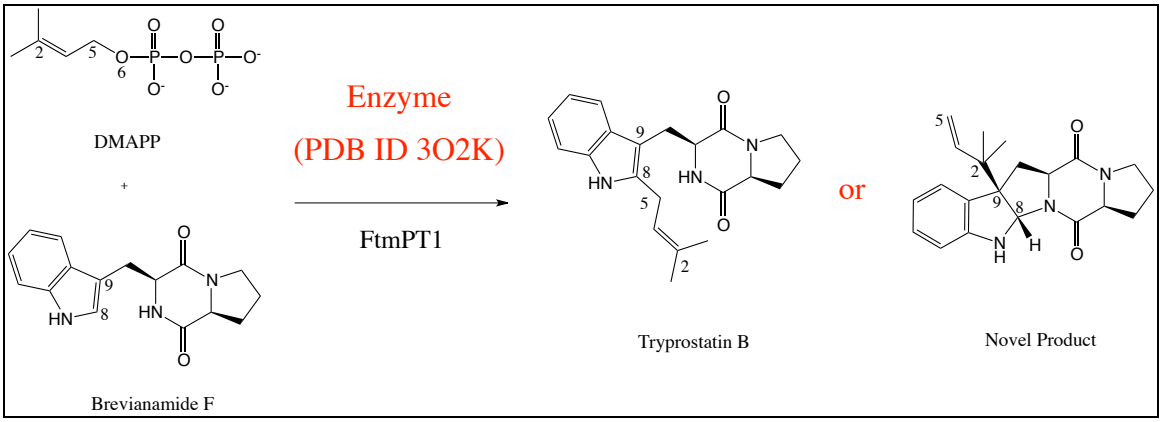


Figure 3.1 Catalytic reactions between DMAPP and Brevianamide F catalyzed by FtmPT1: (a) C5-C8 and (b) C2-C9 reaction channels.

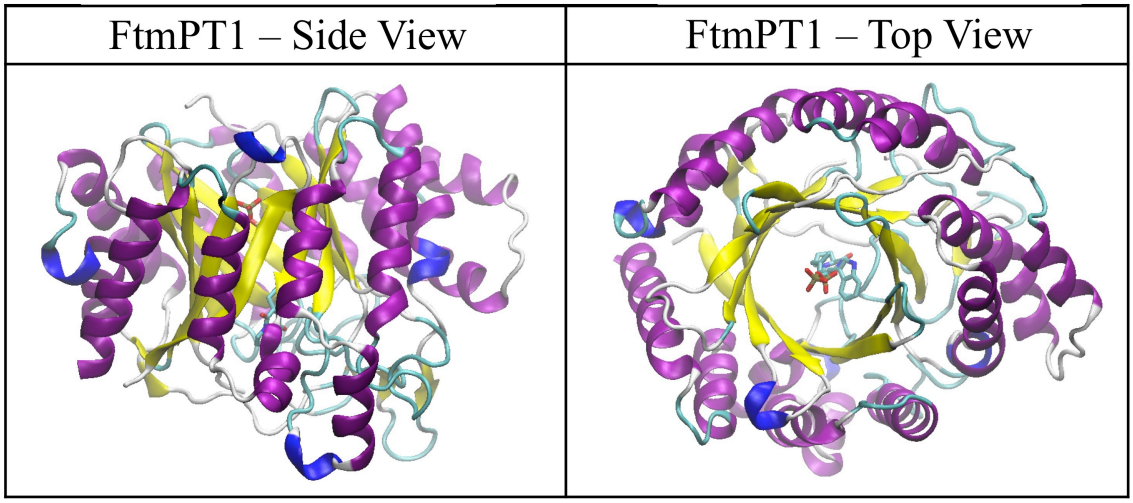


Figure 3.2 A general view of FtmPT1 complex system. Purple ribbons represent α-helices, yellow bands represent β-strands, and green and white tubes represent the loops. The licorice structures in the center of the binding site represent the substrates DMAPP and Brevianamide F.

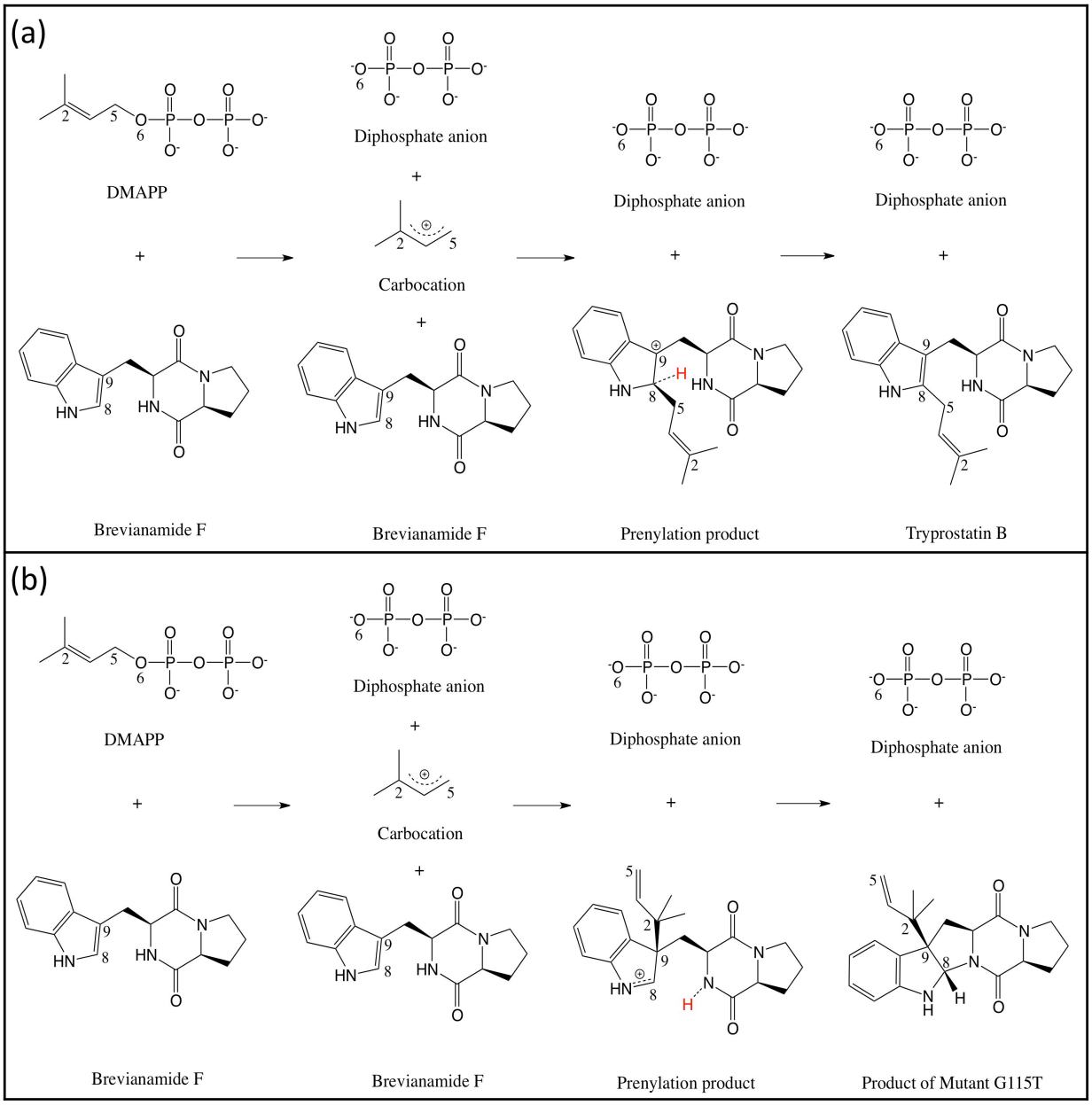


Figure 3.3 Hypothesized reaction mechanisms for complete reactions in FtmPT1 between DMAPP and Breivianamide F of (a) C5-C8 and (b) C2-C9 selections.

As shown in Figure 3.2, FtmPT1 is a PT whose polypeptide chain sequence is αααββαααββααα – ββαααββααββ, which satisfies the indole PT's 'αββα' motif. Produced by *Aspergillus fumigatus*, this fungal indole prenyltransferase catalyzes the reaction between dimethylallyl pyrophosphate (DMAPP) and L-Trp derivatives. Specifically, in this enzyme, the reaction of DMAPP with Brevianamide F (*cyclo*-$_{L}$-Trp-$_{L}$-Pro) to form the cell growth-inhibiting compound, Tryprostatin B has been investigated.[12] Via biochemical and X-ray crystallographic techniques Li et.al in 2010,[12] identified two reaction channels for the prenyl transfer reaction between DMAPP and Brevianamide F: A C5-C8 reaction between the C5@DMAPP and the C8@Indole for the wild type enzyme and C2-C9 between the C2@DMAPP and the C9@Indole for the Gly115Thr mutant (the C5-C8 is seen in the mutant, but as a side reaction). The native reaction channel produces Tryprostatin B, while the Gly115Thr mutant produces a product not seen in the native enzyme (see Figure 3.1). Further studies, regarding the regioselectivity of FtmPT1 have been reported.[25, 26] Furthermore, nonaromatic carbon atoms can be prenylated along with aromatic carbon atoms on indole rings.[24] All these experimental papers demonstrate various regioselectivities observed in FtmPT1 making it a promising enzyme to form novel products of value to the pharmaceutical industry.

Given the interesting behavior of this reaction and its sensitivity to point mutations I decided to gain further insight into the catalytic mechanism of native FtmPT1 by studying the reaction between Brevianamide F and DMAPP. The reaction pathway consists of the prenyl transfer step, which is then followed by a proton transfer step and substrate loss to complete the reaction process. According to Li *et al*., besides the C5-C8 regioselectivity for the wild type enzyme, and C2-C9 for the G115T mutant (with a product yield reduction of 80%), they hypothesize that

Glu102 is the proton acceptor for the deprotonation step in the reaction. Given the bifurcation in the preferred reaction channel between the native and mutant proteins I decided to determine what was blocking the C2-C9 reaction channel in the native protein (see Figure 3.1) using modern theoretical tools. To reach this goal I hypothesized an initial reaction mechanism for both the C2-C9 and C5-C8 reaction pathways respectively, as shown in Figure 3.3. My goal was to elucidate three important aspects of FtmPT1 catalysis: (1) elucidate the energetics of the C5-C8 and C2-C9 reaction channels in wild type enzyme (2) identify the proton acceptors in the active site – whether it is Glu102, surrounding water molecules or another residue in the active site region. (3) identify the prenylation reaction as being either an S_N1 or an S_N2 process, with the former involving a free carbocation. Herein, I describe the results of detailed quantum mechanical/molecular mechanical (QM/MM) on the catalytic mechanism of FtmPT1.

3.2 METHODS

3.2.1 General Features

All molecular dynamics simulations were performed in the Assisted Model Building with Energy Refinement (AMBER, version 11) molecular dynamics package.[27] The crystal structure of FtmPT1 in complex with dimethylallyl S-thiolodiphosphate ($C_5H_{14}O_6P_2S$) and Brevianamide F ($C_{16}H_{17}N_3O_2$) were used as the starting geometry (PDBID 3O2K).[12] The PDB structure was modified by replacing sulfur with oxygen in dimethylallyl S-thiolodiphosphate to form dimethylallyl-pyrophosphate. Using the LEaP module of AMBER hydrogen atoms were added to the structure.

The complex system was then solvated in a truncated octahedral cell with TIP3P[28] explicit water molecules. The ff99sb force field[29] was employed for the protein system and the generalized AMBER force field (GAFF)[30] was used for the substrate molecules to construct the parameter topology files. In all QM/MM molecular dynamics (MD) simulations, unless otherwise noted, I used the self-consistent charge density functional tight binding (SCC-DFTB) method,[31] as implemented in AMBER11. This approach has worked well in the past for prenylation type reactions[32-35] among other reaction types and allows us to carry out extensive sampling along the reaction pathways. The charges for the substrate molecules were determined according to the restrained electrostatic potential (RESP) methodology[36] by a prior structure optimization at the M06L/6-31+G(d,p) level of theory with a following electrostatic potential calculation at the HF/6-31+G(d,p) levels of theory using the Gaussian09 electronic structure program[37], and the Antechamber module in the AMBERTools suite of programs. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) method,[38-41] and 8.0 Å cutoffs were used for the non-bonded interactions. All bonds with hydrogen atoms were constrained using SHAKE[42, 43] except those included in the QM region for the calculations on the proton transfer step.

The entire system was initially minimized to remove close contacts using weak restraints. The system temperature was gradually increased to 300K over 50ps using a restrained MD simulation with a 1fs time step. Then the restraints were slowly removed prior to a constant NPT 450ps MD run. The SANDER module[44, 45] was used for the minimization step and the 500ps MD runs. The following 22ns constant NVT MD simulation was subsequently finished using the PMEMD program. During the 22ns production run, a 2fs time step was used and snapshots were saved

every 2ps. The 1ns product equilibration was completed using a similar procedure of initial minimization using the QM/MM MD capabilities in SANDER. The corresponding reactive substrate molecules, residues and relevant water molecules were included in the QM region.

The Visual Molecular Dynamics (VMD)[46, 47] program was used for structure analysis. GNUPlot[48] was used for all free energy profile (FEP) plots.

3.2.2 Prenyl Transfer Reaction Calculations

3.2.2.1 C5-C8 Bond Formation

QM/MM MD studies were used to obtain 2D free energy profiles where DMAPP and Brevianamide F were in the QM region. In the C5-C8 case, the two reaction coordinates were bond cleavage along C5-O6 and bond formation along C5-C8. For C5-C8 bond formation Steered Molecular Dynamics (SMD)[49-53] was first applied to propagate the trajectory along the reaction coordinate, covering a bond distance range of 5.3Å to 1.4Å at an interval of 0.1Å, using a force constant of 5000 kcal/(molÅ^2), and a pulling speed of 0.02 Å/ps. A 2-D umbrella sampling was performed using as starting structures the C5-C8 SMD snapshots at the corresponding C5-C8 distances. I next propagated along C5-O6 at an interval of 0.1Å, resulting in 2040 windows within a C5-C8 region bound by [1.4 Å, 5.3 Å] and C5-O6 region bounded by [1.4 Å, 6.4 Å]. An initial 50 ps constant NVT equilibration was followed by another 50 ps constant NVT production run, during which the data were collected. The force constants for C5-C8 and C5-O6 were 500-600 kcal/(molÅ^2) and 300-500 kcal/(molÅ^2), where the larger force constants were used for the higher energy regions to assure thorough sampling at the desired location of the 2-D surface. The two-dimensional weighted histogram analysis method (WHAM-

2D)[54] was used to analyze the probability density and obtain the free energy profiles (FEP, 2D) for the unbiased system along the two reaction coordinates. All of the Steered MD and umbrella sampling calculations, unless otherwise indicated, used a time step of 1fs.

In addition to the 2D umbrella sampling simulations, 1D umbrella sampling simulations were also performed for both C5-C8 and C5-O6 as single reaction pathways. Starting structures of the corresponding umbrella sampling simulations were extracted from the corresponding Steered MD trajectories at an interval of 0.1 Å between adjacent windows for both the O5-C6 and C5-C8 reaction coordinates in the 1D PMF calculations. The C5-C8 distance decreased from 5.3 Å to 1.4 Å, while the O6-C5 increased from 1.4 Å to 4.6 Å, resulting in 40 windows for the C5-C8 reaction coordinate and 33 windows for the O6-C5 reaction coordinate. Each of the windows was first equilibrated for 100 ps under NVT simulation conditions and then followed by a 150 ps production NVT simulation, during which the data were collected. The 1-D weighted histogram analysis method (WHAM) was implemented to analyze the probability density and obtain the free energy profiles (FEP, 1D) for the unbiased system along the corresponding reaction coordinates.

3.2.2.2 C2-C9 Bond Formation

Carbon-carbon bond formation at C2-C9 was also examined using both 1D PMF and 2D PMF calculations. Similar to my C5-C8 bond formation studies outlined in Section 2.2.1 the C2-C9 reaction coordinate was mapped out using Steered MD simulations with an interval of 0.1Å between adjacent windows covering a distance interval of [4.5 Å, 1.4 Å]. The 1D PMF profile for C2-C9 formation was accomplished using 100 ps of equilibration and 150 ps of sampling

along the C2-C9 distance interval of [4.5 Å, 1.4 Å]. The 2D profile was generated using 50 ps of constant NVT equilibration and 50 ps of constant NVT sampling using a C2-C9 distance region of [4.5 Å, 1.4 Å] and for C5-O6 a distance interval of [1.4 Å, 6.4 Å] again resulting in 2040 windows. The force constant used for the Steered MD studies was 5000 kcal/molÅ^2 with a pulling speed of 0.02 Å/ps. For the 1D and 2D PMF simulations the force constants were 500-600 kcal/(molÅ^2) for C2-C9 and 300-500 kcal/(molÅ^2) for C5-O6, with higher value being used in higher energy regions of the profiles. The QM region employed included DMAPP and Breivianamide F.

3.2.3 Final Product Formation

2.3.1 Proton Transfer to Form the Final Product for the C5-C8 Reaction Pathway

The reaction system required a proton transfer step to remove the hydrogen atom connected to C8 on the indole ring of Brevianamide F after the C5-C8 prenylation step in order to obtain the final product (Tryprostatin B). First I needed to locate the minimum for the intermediate state by varying the distance between the diphosphate anion and the carbocation of the prenylation product using umbrella sampling. Starting from the protonated prenylation product, an umbrella sampling calculation was done for the C5-O6 distance, covering an interval of [6.0 Å, 15.0 Å] at 0.25 Å increments and a force constant of 100 kcal/(molÅ^2). Since this calculation was for the non-reactive pathway, the sampling was done using classical MD under NVT conditions using a time step of 1fs, with a 551 ns equilibration stage followed by a 1 ns data collection step. This simulation was carried out to energetically characterize the prenylation product minimum. This free energy profile is given in Figure 3.4.

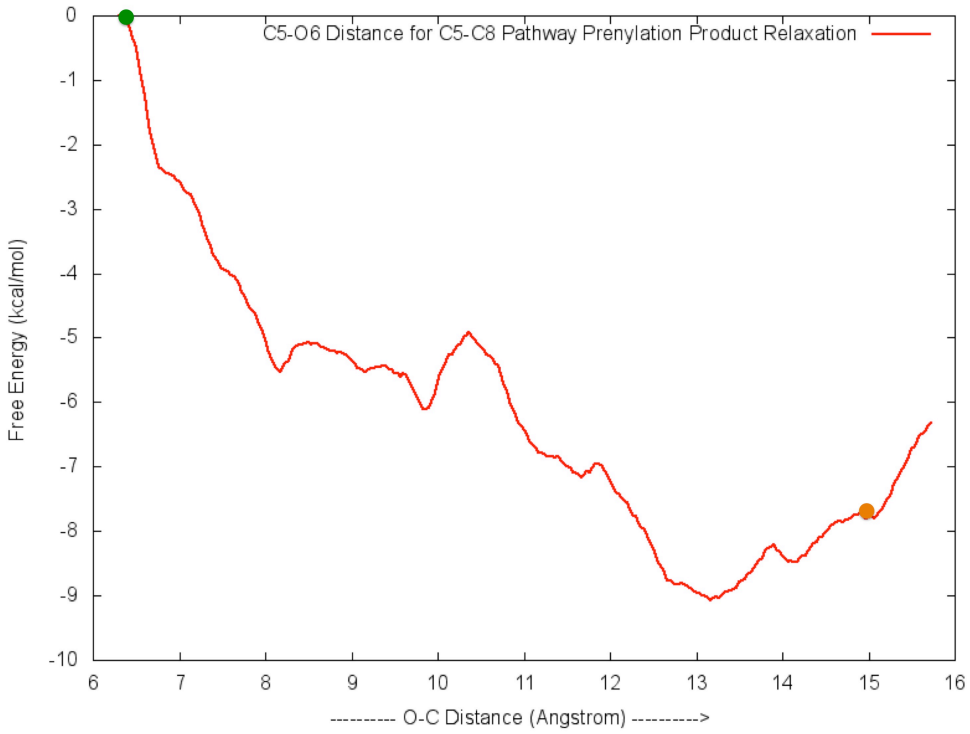


Figure 3.4 Free energy profile (in kcal/mol) for the relaxation of the prenylation product along the C5-C8 pathway is shown as a red line. The green dot at (6.35 Å, 0.0 kcal/mol) represents the starting point of the relaxation process, *i.e*. the ending point for prenyl transfer reaction; the orange dot at (15.00 Å, -7.8 kcal/mol) represents ending point of the relaxation, which served as the starting point for the QM/MM study of the proton transfer reaction simulations. This point arose from a 20ns classical simulation followed by a 1ns QM/MM simulation.

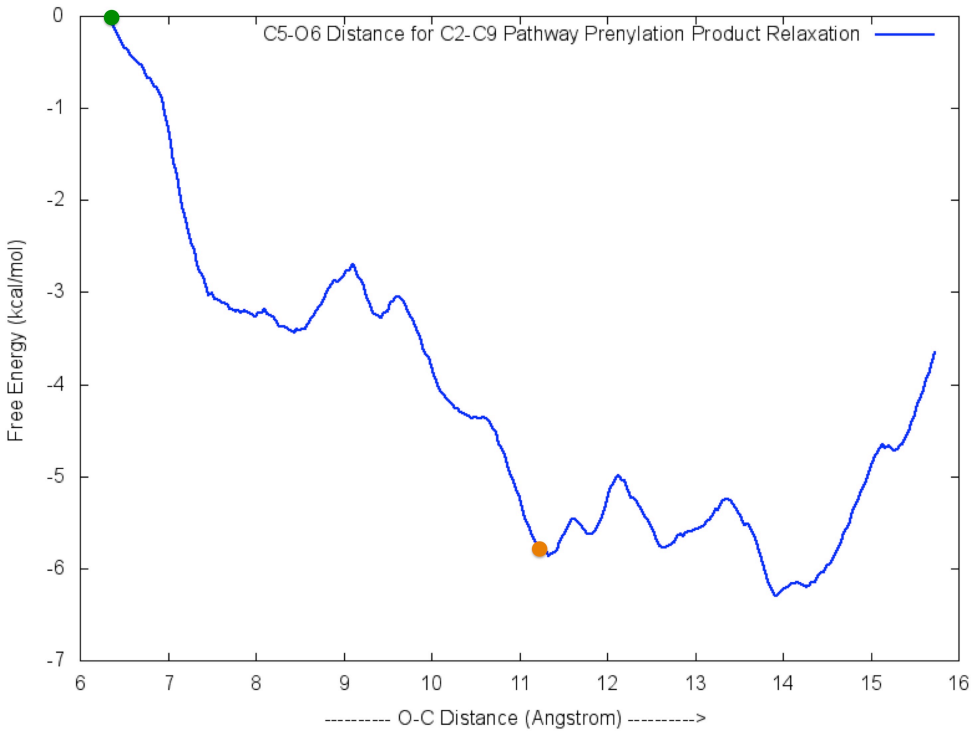


Figure 3.5. Free energy profile (in kcal/mol) for the relaxation of the prenylation product along the C2-C9 pathway is shown as a red line. The green dot at (6.35 Å, 0.0 kcal/mol) represents the starting point of the relaxation process, *i.e*. the end point of the prenyl transfer reaction; the orange dot at (11.15 Å, -5.7 kcal/mol) represents ending point of relaxation, which served as the starting point for the QM/MM study of the proton transfer reaction simulations. This point arose from a 20ns classical simulation followed by a 1ns QM/MM simulation.

From the end of the prenylation reaction I carried out a 20 ns constant NTP MD equilibration of the protonated prenylation product complex followed by a 1ns constant NVT QM/MM MD equilibration prior to the subsequent QM/MM Steered MD and Umbrella Sampling steps. The position of the minima arising from these simulations versus the classical PMF simulations described above is given in Figure 3.4.

Analysis of the system after MD equilibration, the only two possible receptors of the proton were Glu102 and the water molecule adjacent to this residue. Hence, I included the side chain of Glu102 as well as the water molecule into the QM region, together with protonated C5-C8 prenylation product. The diphosphate anion was not included in the QM region for this step. The Steered MD and Umbrella Sampling calculations for the H(@C8)-C8 bond cleavage covered a distance interval of [1.1 Å, 3.5 Å]. The force constant for the Steered MD study was 5000 kcal/molÅ^2 with a pulling speed of 0.02 Å/ps. For the Umbrella Sampling calculations, the force constant was 300 kcal/(molÅ^2) for regions the before and after windows in the transition state region, 600 kcal/(molÅ^2) for windows close to the transition state, and 100 kcal/(molÅ^2) for windows far away from the transition state. WHAM-1D was used afterwards to obtain the 1D-FEP curve for this deprotonation step.

3.2.3.2 Cyclization and Deprotonation to Form the Final Product for the C2-C9 Reaction Pathway

An extra cyclization step via C8-N7 bond formation coupled with a proton transfer step via H(@N7)-N7 bond breakage were required to obtain the final product for the C2-C9 prenylation reaction pathway. Similar to the C5-C8 deprotonation step study, only a neighboring water molecule and the side chain of Glu102 as well as the C2-C9 prenylation product were included in the QM region. Starting from the protonated prenylation product, I carried out another 1D-FEP curve using NVT MD umbrella sampling along the C5-O6 pathway over a distance increment of [6.0 Å, 15.0 Å] at intervals of 0.25 Å and a force constant of 100 kcal/(molÅ^2). This simulation fully classical was used to energetically characterize the minimum for this intermediate state.

I also simulated the prenylated C2-C9 system for 20 ns of MD simulation and 1ns of QM/MM MD simulation (to generate the starting minimum for the subsequent QM/MM simulations) followed by a 1D Steered MD study along the C8-N7 cyclization reaction pathway. The Steered MD force constant was 5000 kcal/molÅ^2 with a pulling speed of 0.02 Å/ps. A 2D umbrella sampling study was carried out afterwards using snapshots from the Steered MD simulations for the C8-N7 reaction pathway as starting points. The H(@N7)-N7 distance was then varied as well, resulting in 456 windows with a C8-N7 distance interval of [3.2 Å, 1.4 Å] and a N-H distance interval of [0.9 Å, 3.2 Å]. The force constants for C8-N7 bond formation were 300-500 kcal/(molÅ^2) and for the N-H bond cleavage values between 500-1000 kcal/(molÅ^2) were used. The higher force constants were used at windows close to the saddle points. WHAM-2D was used to obtain the final 2D profile.

3.2.4 Interaction Energy Decomposition

For both the C5-C8 and C2-C9 prenylation reaction pathways, "π-chambers"[35, 55-58] were used to encapsulate the carbocation intermediate states. The intermediate carbocations formed a cation-π interaction with the π-chamber, which was formed by phenol rings from three tyrosine residues in the active site region - TrY382, Tyr435 and Tyr450. To quantitatively analyze the interaction energies between each phenol ring and the carbocation, I used high-level QM calculations for the energy analysis. The calculations included only the formed carbocation and the three tyrosine residues. The backbones of the tyrosine residues were removed with only the phenol ring and the hydrogen capped β-carbon being used in the subsequent calculations.

Five snapshots were picked from the umbrella sampling QM/MM MD simulation results at the intermediates states for both the C5-C8 and C2-C9 calculations. The reduced system of four aromatic fragments and the carbocation was first optimized at the M06L/6-31+G(d,p) level of theory, with all heavy atoms fixed. The optimization was then followed by a set of single point calculations with Basis Set Superposition Error (BSSE) corrections to more quantitatively study the interactions between the carbocation and the π-chamber.

In addition, in the FtmPT1 active site, there are more residues that could affect the catalytic mechanism and the resulting product distribution. Hence, I used the energy decomposition method within the MM-GBSA module[59-62] in AMBER11 to analyze how the active site residues stabilize/destabilize the reactive species. About 100 residues within a 10Ångstoms distance of the diphosphate anion, the dimethylallyl carbocation and Brevianamide F were studied and analyzed via pairwise free energy decomposition. In order to facilitate this analysis, 42 ns NVT MD simulations of the starting, intermediate and product states for the C5-C8 prenylation reaction were carried out to save snapshots for the analysis. Snapshots over the 42ns MD were obtained every 50 ps for the MM-GBSA calculation.

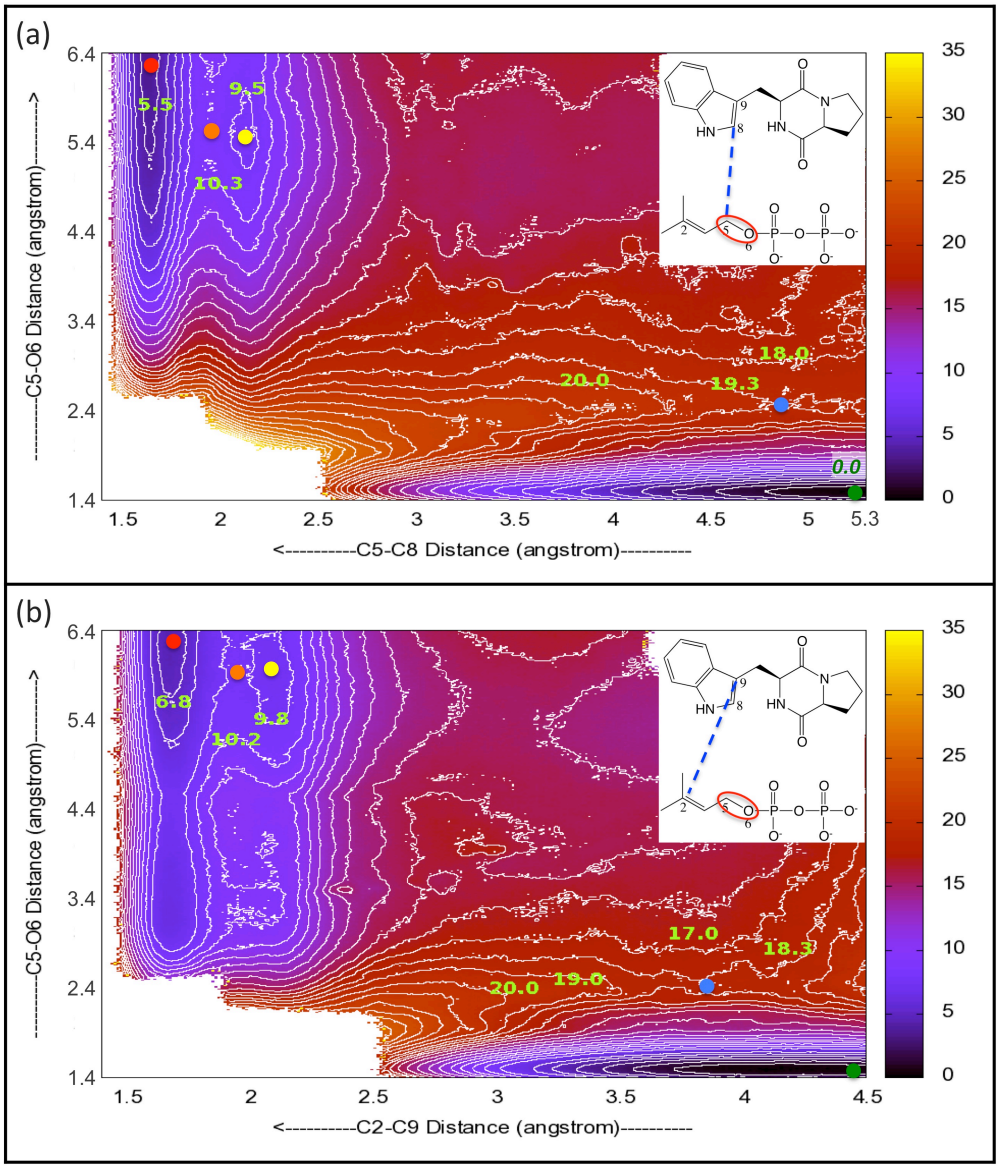


Figure 3.6 Free energy profiles (FEP) in kcal/mol for (a) C5-C8 and (b) C2-C9 prenyl transfer steps. The free energies of starting states were set to zero. Map colors correspond to the energy scale shown to the right of the images. Lime-colored values represent free energies of iso-contour lines. Dots represent the stationary points for the reactions. Dots from right to left: green – starting state; blue – first transition state; yellow – intermediate state; orange – second transition state; red – prenylation product.

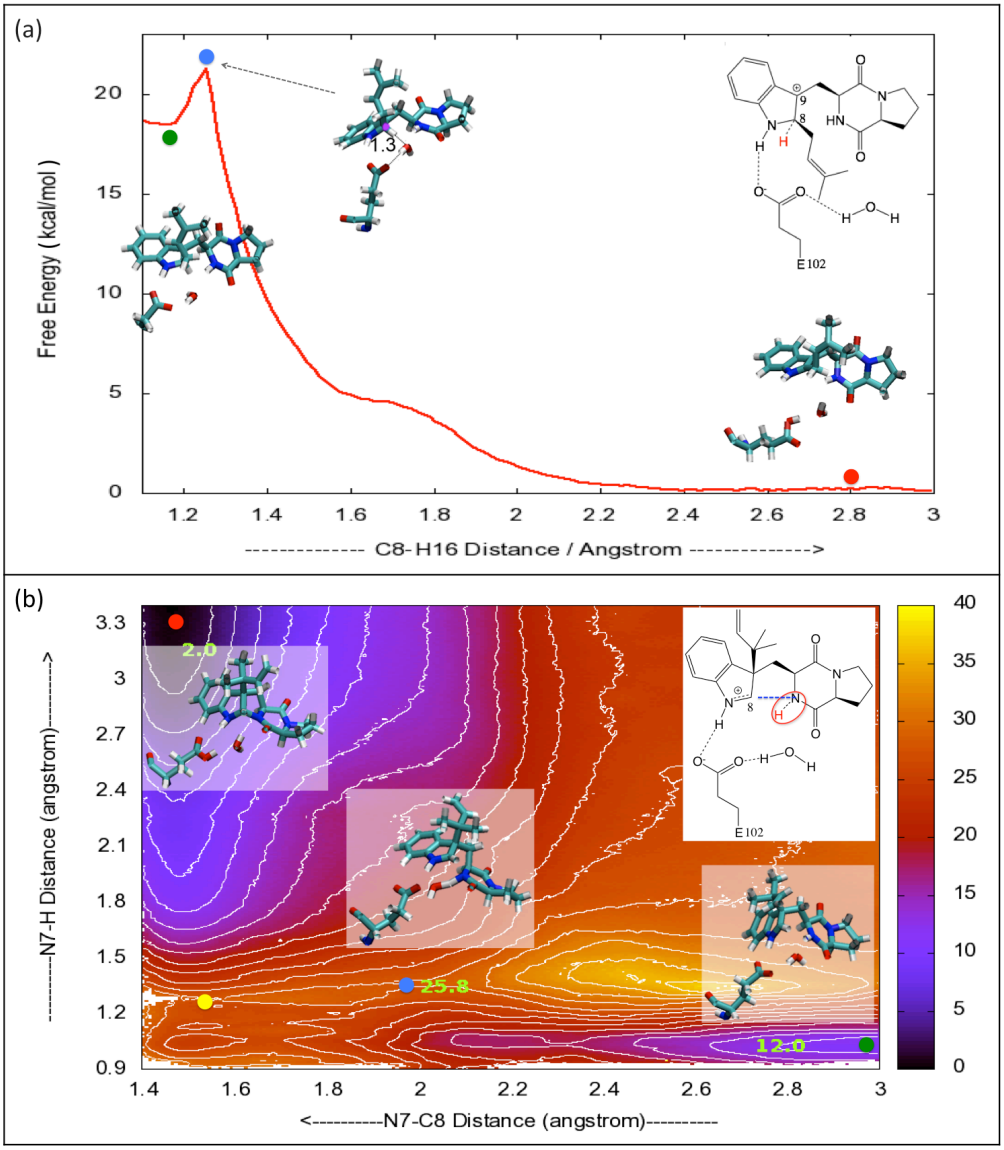


Figure 3.7 Free energy profiles (FEP) in kcal/mol for (a) the proton transfer for the C5-C8 reaction pathway, and (b) the deprotonation and cyclization steps for the C2-C9 reaction pathway. Map colors correspond to the energy scale shown to the right of the images. Lime-colored values represent free energies of the iso-contour lines. Dots represent the key points for the reactions: green – starting state; blue – transition state; red – final product. Note that for (b) yellow dot also represents another potential transition state for the reaction.

3.3 RESULTS AND DISCUSSION

3.3.1 C5-C8 Reaction Pathway

3.3.1.1 The Prenyl Transfer Process

Starting from a MD and QM/MM MD equilibrated starting complex, Steered MD and umbrella sampling simulations were conducted to obtain the 2D free energy profiles (FEPs). Figure 3.6(a) shows the energy-contoured map for the corresponding FEP, with the C5-C8 distance as the X-axis and the C5-O6 distance as the Y-axis. From the figure I can identify two saddle points and three wells, *i.e*. the starting state, two transition states, one intermediate state, along with the product carbocation state. Starting at a C5-C8 distance of 5.25Å and a C5-O6 distance of 1.42Å, I proceed to the first transition state at a C5-C8 distance of 4.88Å and a C5-O6 distance of 2.45Å where the C5-O6 bond is initially broken. The next step for this reaction is the formation of the relatively stable dimethylallyl carbocation (DMA^+) at a C5-C8 distance of 2.15Å and C5-O6 distance of 5.45Å. The carbocation is then attacked by the C8 carbon on the indole ring of Brevianamide F to reach the second transition state of the reaction at a C5-C8 distance of 1.95Å and C5-O6 distance of 5.50Å. The final prenylation product, which is the protonated form (@C8) of Tryprostatin B, is reached at a C5-C8 distance of 1.58Å and a C5-O6 distance of 6.35Å.

Other than the five states described above, no other local minima or saddle points could be identified for this reaction. Therefore, I conclude that this reaction follows a S_N1 reaction mechanism rather than a S_N2 reaction mechanism. The relative free energies are shown in Table 3.1. From this table the rate-determining first transition state for C5-O6 bond cleavage has a

relative free energy of 19.2 kcal/mol. This result matches the estimated experimental free energy barrier of 17.5 kcal/mol, which is approximated from its turnover rate 5.57s^{-1}.[12]

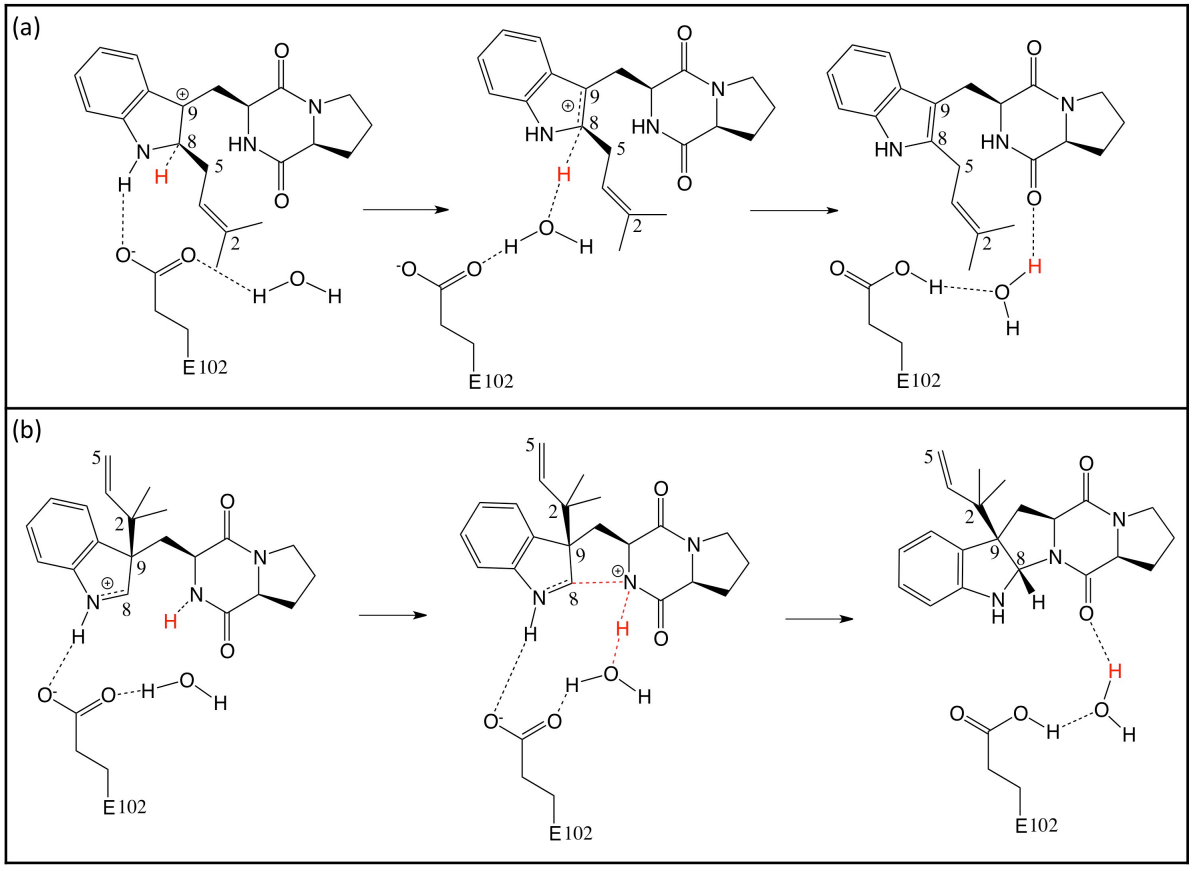


Figure 3.8 Reaction mechanisms for the proton transfer steps in (a) the C5-C8 and (b) the C2-C9 reaction pathways.

Table 3.1 Relative free energies (in kcal/mol) for each state in the C5-C8 and the C2-C9 reaction pathways. Note that the energies of the starting states are set to zero.

Free Energies (kcal/mol)	Starting Point	Prenyl Transfer Step				Proton Transfer / Cyclization Step		
		Transition State 1	Intermediate State	Transition State 2	Prenylation Product	Resting Starting State	Transition State 3	Final Product
C5-C8	0.0	19.3	9.2	10.0	5.3	-2.5	0.5	-20.8
C2-C9	0.0	18.5	9.5	10.0	6.5	0.8	16.8	-9.2

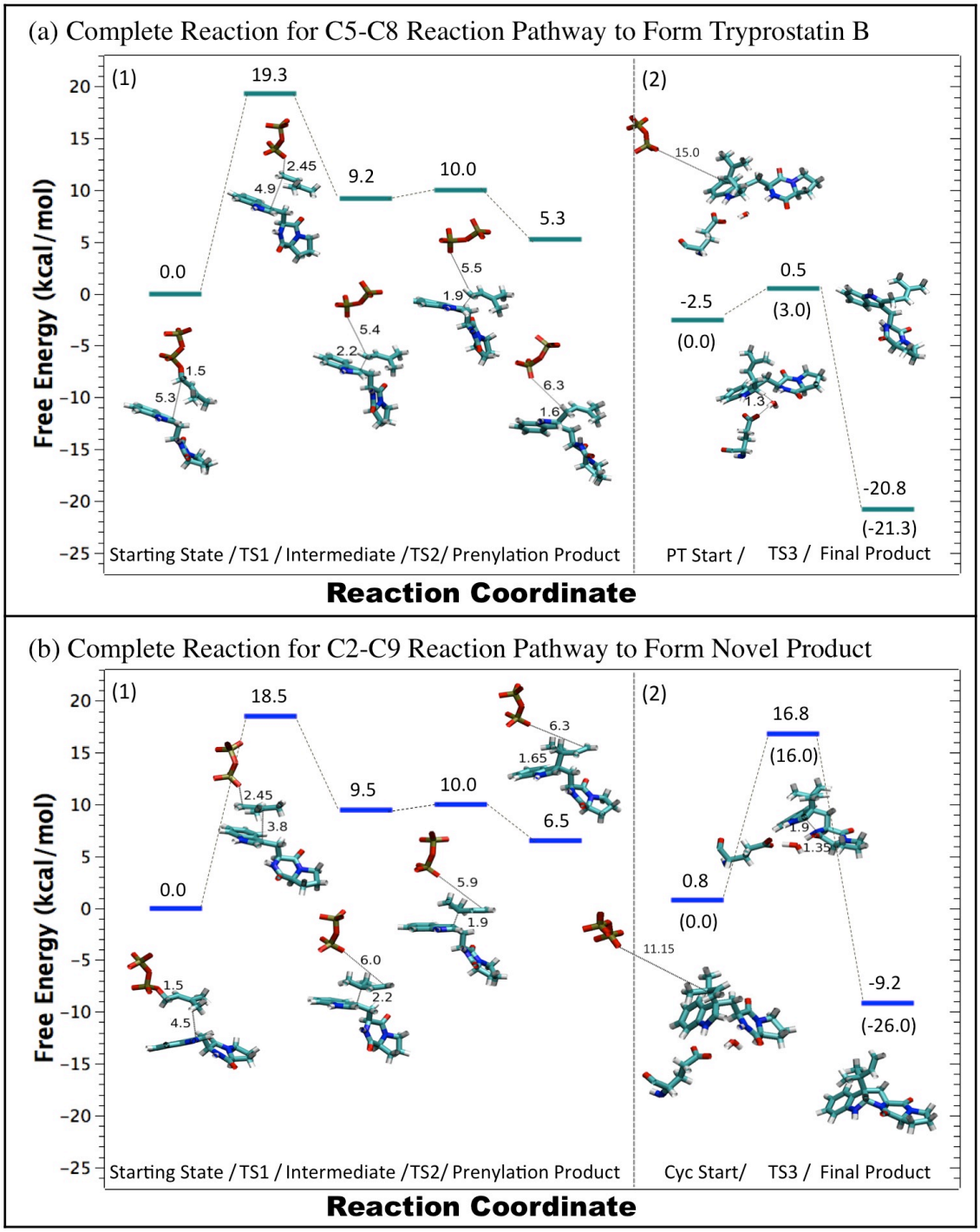


Figure 3.9. Complete reaction profiles with relative free energies and the associated structures for the C5-C8 and C2-C9 pathways – (1) prenyl transfer reaction; (2) proton transfer reaction. Distances for the relevant reaction coordinates are given on the structures. Free energies

compared to the starting state are given above the green and blue lines. Note the energies in parentheses for the proton transfer step that are compared to their own starting state minimum (for the diphosphate and prenylated carbocation). The starting state for reaction 2 (proton transfer) is the relaxed state of the product from reaction 1 (prenylation reaction).

3.3.1.2 Proton Transfer Process

To obtain the final product I need to deprotonate the C8 carbon from the protonated prenylation product. Prior to the deprotonation process I estimate that the relaxation of the protonated prenylation product (with a C5-O6 distance of 6.35 Å) to the starting state of the deprotonation process (with C5-O6 distance of 15.00 Å) has a free energy decrease of 7.8 kcal/mol (see Figure 3.4). This energy decrease was also counted in the total reaction, as shown in Figure 3.9a and Table 3.1. For the deprotonation process, with both Glu102 and a nearby water molecule serving as proton acceptor candidates I generated a 1D free energy profile for the deprotonation step along the H(@C8)-C8 reaction pathway, as shown in Figure 3.7(a). From the figure I find that the deprotonation process reaches a transition state at a H(@C8)-C8 distance of 1.25Å with an energy barrier of 3.0 kcal/mol compared to the starting protonated state. The reaction reaches the product state at a H(@C8)-C8 distance of 2.4Å with a free energy of 18.3 kcal/mol lower than the starting state.

From the mechanisms shown in Figures 3.8(a) and 7(a), the starting state for the protonated intermediate has two proton accepting groups: Glu102 and a water molecule. Initially, Glu102 forms a hydrogen bond with the nitrogen atom on the indole ring of the protonated prenylation

product, and simultaneously forms another hydrogen bond with the water molecule. At the transition state I find that the proton transfers initially to the water molecule. Right after the transition state, the extra proton transfers to the water molecule and then without activation to Glu102. Hence, Glu102 is the final transfer point for the proton as proposed reference[12], but an intervening water molecule facilitates the transfer.

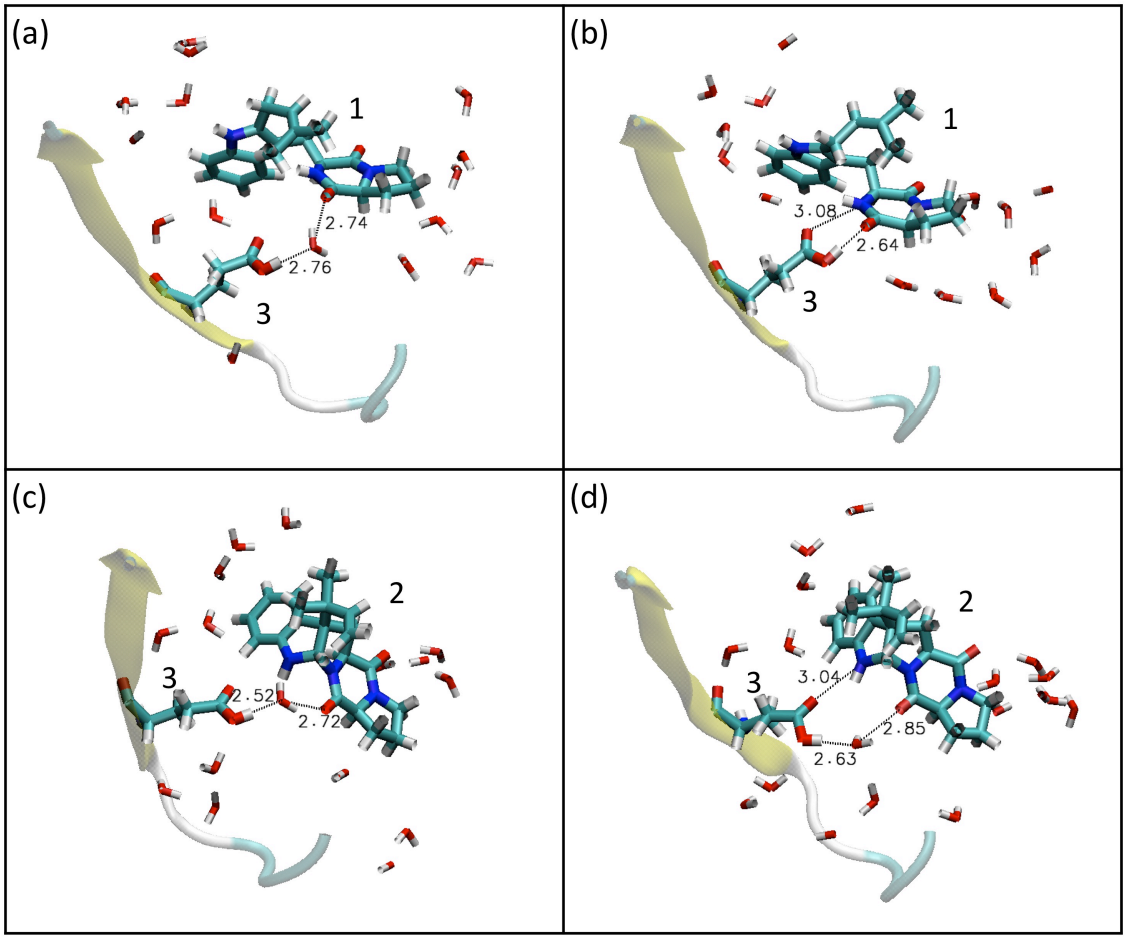


Figure 3.10 Final product hydrogen binding information for C5-C8 (a&b) and C2-C9 (c&d) reaction pathways. 1 – Tryprostatin B; 2 – the novel product from C2-C9 prenylation; 3 – Glu102. The distances (in Å) for each hydrogen bond are labeled in the images.

3.3.1.3 Final Product Equilibration

To observe dynamics of the final product, Tryprostatin B, in the binding pocket of FtmPT1, ran a 22ns classical MD simulation. In the MD simulation I could find that the product remained bound in the active site with two different hydrogen bonding situations: initially the original two hydrogen bonds formed at the end of the proton transfer reaction remained: the water molecule, which was involved in the proton transfer, was forming bridge between the oxygen atom on Tryprostatin B's six-membered ring and via protonated Glu102 through two hydrogen bonds (see Figure 3.10(a)). In the later stages of the MD simulation, the water molecule escapes the active site and two other hydrogen bonds are formed: one from the protonated oxygen on Glu102 and the oxygen atom on Tryprostatin B's six-membered ring, the other from the oxygen atom on Glu102 and the nitrogen atom on Tryprostatin B's six-membered ring (see Figure 3.10(b)). At least at the short timescale of these simulations Tryprostatin B remains in the active site of FtmPT1, and in order to observe product loss lengthier simulations will be required.

3.3.2 C2-C9 Reaction Pathway

In order to understand the regioselectivity of native FtmPT1 I decided to also examine the reaction pathway (C2-C9) that generates the product observed in the G115T mutant of FtmPT1. The initial step in both the C5-C8 and C2-C9 pathways are identical (cleavage of the DMAPP C-O bond), so the origin of the regioselectivity arises from another step in the reaction. From a study of the C2-C9 pathway I will garner insights into how the enzyme controls its product distribution once the carbocation intermediate is formed.

3.3.2.1 Prenyl Transfer Process

For the prenyl transfer part or computed 2D free energy profile is given in Figure 3.6(b). From this 2D FEP map I find that the C2-C9 pathway is similar to the C5-C8 pathway. For the formation of the carbocation the cleavage of the DMAPP C-O bond is again rate-limiting, while the resultant minima on the surface between C2-C9 and C5-C8 are very similar. Hence, the C2-C9 prenyl transfer reaction is also following a S_N1 reaction mechanism, which supports the mechanistic hypothesis in Figure 3.3(b). This means that the enzyme is not favoring one product or another from a thermodynamic perspective along the pathway generating the appropriately positioned carbocation intermediate.

Looking into the details of the C2-C9 reaction pathway based on the 2D FEP shown on Figure 3.6(b), I find that the reaction starts at a C2-C9 distance of 4.49Å and a C5-O6 distance of 1.42Å, with the first transition state at a C2-C9 distance of 3.81Å and a C5-O6 distance of 2.45Å which cleaves the C5-O6 bond on DMAPP. A relatively stable dimethylallyl carbocation is formed reaching the intermediate state at a C2-C9 distance of 2.15Å and a C5-O6 distance of 6.05Å. The carbocation then forms a C-C bond with C9 on the indole ring of Brevianamide F via a second transition state at a C2-C9 distance of 1.95Å and a C5-O6 distance of 5.95Å. The final prenylation product, which is the protonated state of the novel product, was finally formed at a C2-C9 distance of 1.65Å and a C5-O6 distance of 6.35Å.

The energy barriers for each state are given in Table 3.1. Compared to the C5-C8 prenylation reaction, I find that the first transition state has a similar free energy barrier of 18.5 kcal/mol with

respect to the starting state. As noted above this is because the first transition state only involves the cleavage of the C5-O6 bond on DMAPP and has nothing to do with the C-C bond formation.

For C-C bond formation the two free energy barriers for C2-C9 and C5-C8 are essentially identical at ~10.0 kcal/mol with respect to the starting state. The free energy for the C5-C8 prenylated product (5.3 kcal/mol) was slightly lower than that of the C2-C9 (6.5 kcal/mol). This is because the C2-C9 prenylated product is less stable due to steric effects associated with the formation of a highly substituted C-C bond (see Figure 3.3(b)).

3.3.2.2 Cyclization and Proton Transfer Process

Prior to the deprotonation process for the C2-C9 reaction pathway, similarly, I find that the relaxation from the protonated prenylation product (with a C5-O6 distance of 6.35 Å) to the deprotonation starting state (with a C5-O6 distance of 11.15 Å) experiences a free energy decrease of 5.7 kcal/mol. (see Figure 3.5) This energy decrease was included for the total reaction, as shown in Figure 3.9(b) and Table 3.1.

For the deprotonation process, the C2-C9 reaction pathway was more complicated than the C5-C8 one, as the protonated prenylation product had to undergo another cyclization process in addition to the deprotonation step to obtain the final product. The cyclization takes place via C8-N7 bond formation yielding a new five-membered ring between indole and the six-membered ring of Brevianamide F generating Tryprostatin B (see Figure 3.3). In order to fully characterize this pathway, I used a 2D FEP approach. A 2D graph is shown in Figure 3.7(b) with the C8-N7

distance as X-axis and the H(@N7)-N7 distance as the Y-axis. From this 2D map I was able to fully characterize the details of this deprotonation/cyclization process.

From the starting state at the bottom right with a C8-N7 distance of 3.0Å and a H(@N7)-N7 distance of 1.0Å, the transition state is reached at a C8-N7 distance of 2.05Å and a H(@N7)-N7 distance of 1.27Å, where both the N7-C8 is starting to be formed while the H(@N7)-N7 bond is starting to break. The H@N7 then left N7 and transferred the extra proton to Glu102 by bridging of the water molecule closed to them, while the final product was finally formed by bond formation of N7-C8 at distance of 1.55 Å. At this time the H atom was 3.4 Å away from N7. One more possibility for the transition state was observed at a C8-N7 distance of 1.65Å and a H(@N7)-N7 distance of 1.25Å. In both cases the H(@N7)-N7 bond breaks and the proton ends up bound to Glu102 via an activationless transfer through the active site water molecule. The transition state free energy barrier for the first pathway is 16.0 kcal/mol, while for the second one the free energy barrier is 16.5 kcal/mol. The two transition states are along a ridge between the starting state and the final product and differ by the amount by which the N7-C8 bond is formed with the former being longer than the later. The first pathway is favored and will be used through out the remaining discussions.

With the energy barriers shown in Table 3.1 I find that the rate-determining step for both the C5-C8 and C2-C9 reaction pathway are the O-C bond cleavage step. Since the prenylation reaction for both pathways are following a S_N1 reaction mechanism, and start from the same starting state, the O-C bond cleavage process should be identical as observed. The relative free energies for C5-C8 and C2-C9 prenylation pathways are also nearly identical. Hence, the final proton

transfer/cyclization step determines the final product distribution. The cyclization/deprotonation step (16.8 kcal/mol) for the C2-C9 pathway has a much higher barrier than that for the C5-C8 deprotonation step (0.5 kcal/mol). Thus, once the intermediate prenylation product is formed, the C5-C8 pathway transitions to product nearly activationlessly, while the C2-9 pathway has a substantial barrier to overcome, which is higher than the intermediate state itself. Equilibration between the C2-C9 and C5-C8 can then funnel away the C2-C9 pathway yielding the exclusive C5-C8 product for the reaction.

3.3.2.3 Final Product Equilibration

To further explore the binding interactions for the C2-C9 product, I performed a 22ns classical MD simulation. Hydrogen bonds between the active site water and Glh102 (protonated Glu) stabilized the final product within the active site where it remained over the relatively short timescale of the final classical MD simulation (see Figure 3.10(c)).

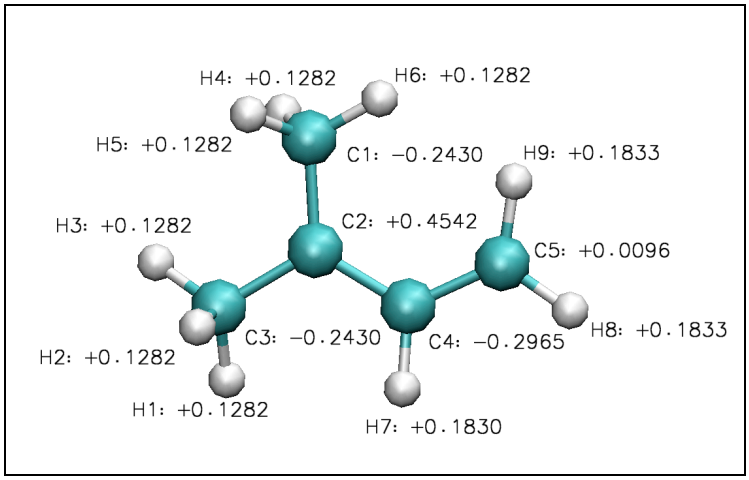


Figure 3.11 RESP charge of dimethyl allyl carbocation.

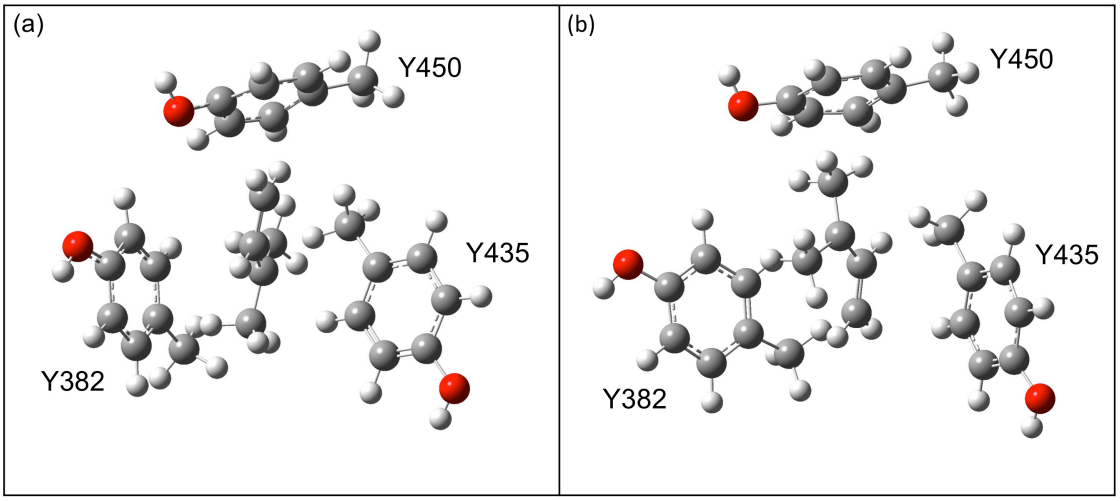


Figure 3.12 Truncated representations of the intermediate state π-chambers observed along the (a) C5-C8 prenylation and (b) C2-C9 prenylation pathways. The center fragment represents the carbocation, and the phenol fragments represent truncated tyrosine residues.

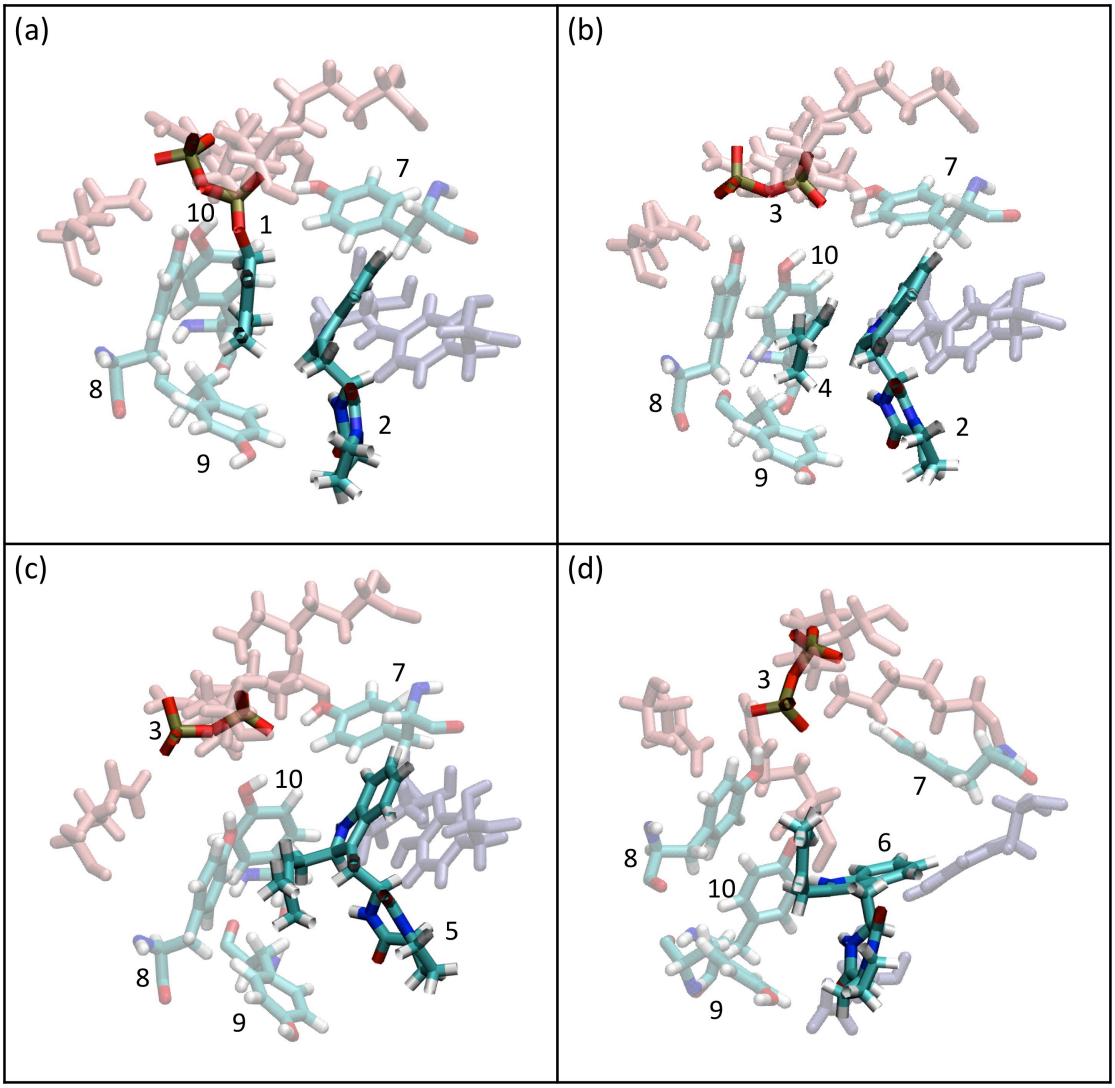


Figure 3.13 Key snapshots from FtmPT1 wild type catalysis representing the C5-C8 reaction pathway. (a) Starting state; (b) Intermediate state; (c) Prenylation product (intermediate) State; (d) Final product state. Numbers represent the substrates and key residues in the active site of FtmPT1. 1-DMAPP; 2-Brevianamide F; 3-OPP; 4-DMA+ carbocation; 5-protonated product; 6-final product Tryprostatin B; 7- Tyr203; 8- Tyr382; 9-Tyr435; 10-Tyr450. Pink residues (Arg113, Lys201, Lys294, Gln380) are hydrogen bound to OPP; ice blue residues (Glu102, Tyr205) are hydrogen bound to Brevianmaide F.

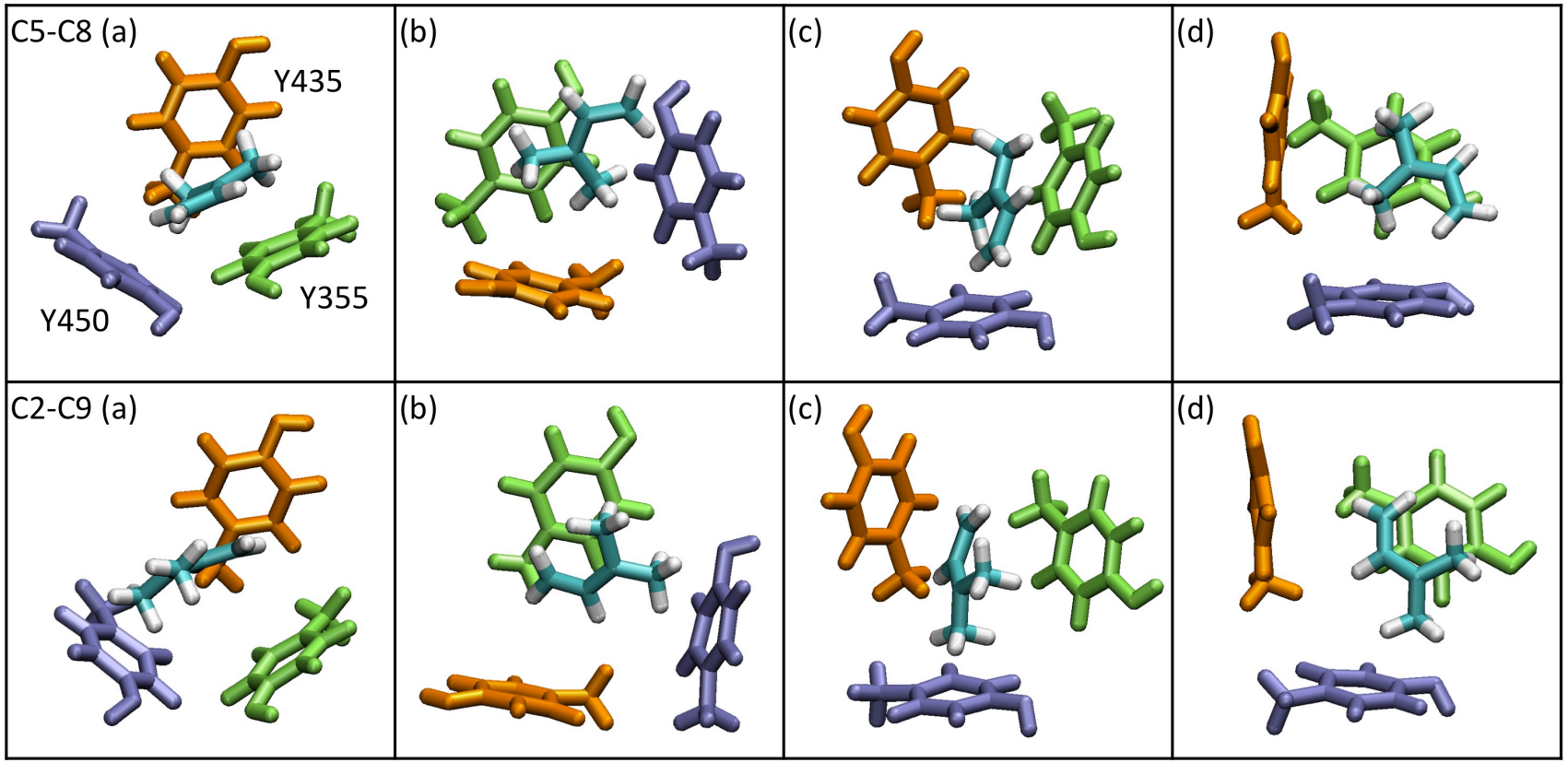


Figure 3.14 Truncated snapshots for intermediate state p-chambers observed along the C5-C8 and C2-C9 prenylation pathways. (a) to (d) presents different views of the tyrosine phenol rings and the intermediate carbocation.

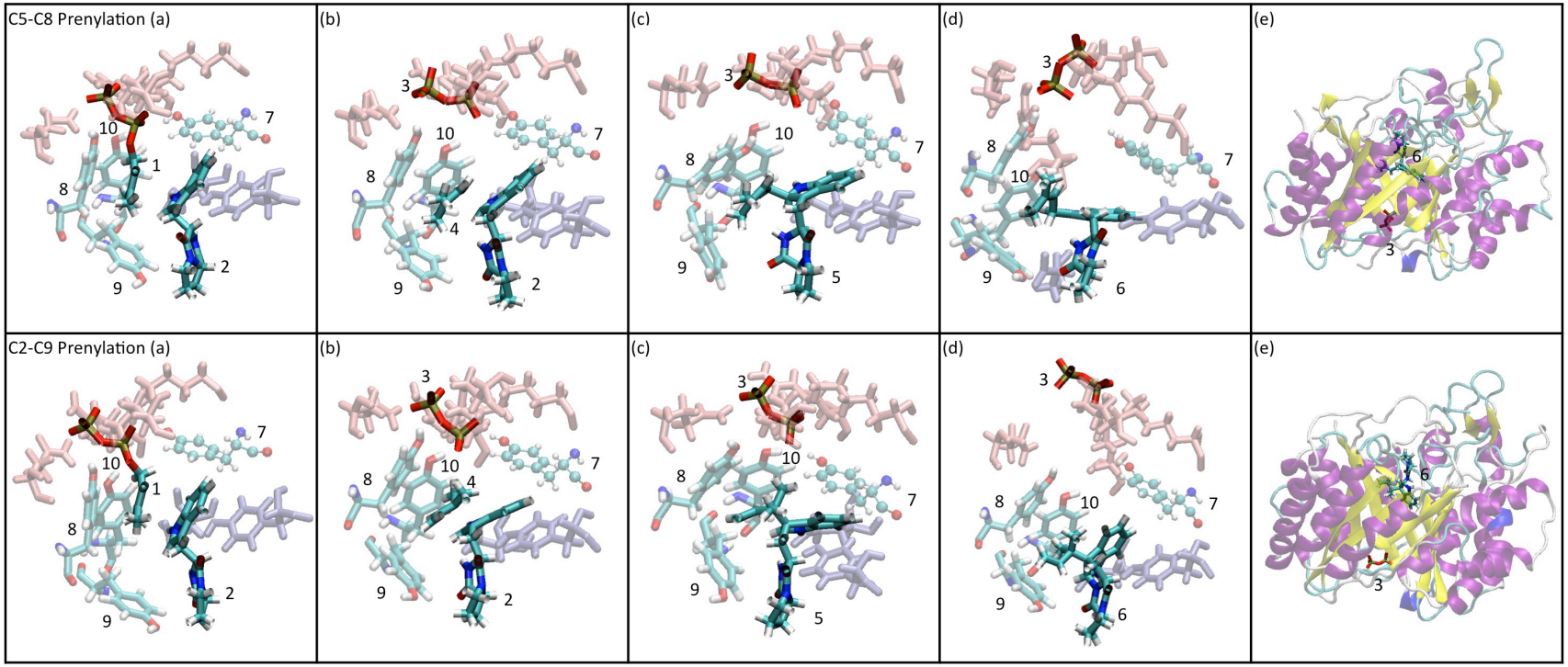


Figure 3.15 Snapshots from FtmPT1 wild type catalytic pathway along the C5-C8 and C2-C9 reaction pathways: (a) starting/resting state, (b) prenylation intermediate state, (c) prenylation product (intermediate) state, (d) final product, (e) final product state inside FtmPT1. Numbers represent ligands and active site residues. 1- DMAPP; 2- Brevianamide F; 3- OPP; 4- DMA+ carbocation; 5- protonated prenylation product; 6- final product; 7- Tyr205; 8- Tyr382; 9- Tyr435; 10- Tyr450. Pink residues (Arg113, Lys201, Lys294, Gln380) are hydrogen bound to the diphosphate anion; ice blue residues (Glu102, Tyr205) are hydrogen bound to Brevianamide F.

3.3.3 π-Chamber Binding Analysis

To further look into the interactions between the formed carbocation and the aromatic groups from the active site tyrosine residues during the prenyl transfer reaction, I carried out QM calculations (Figure 3.12) and the results are summarized in Table 3.2. From Table 3.2 I observe that Tyr435, which has a T-stacking interaction with the carbocation (see Figure 3.14(b) & 3.15(b)), contributed similar interaction energies to the C5-C8 (-9.81 kcal/mol) and C2-C9 (-9.67 kcal/mol) reaction pathways, respectively. Tyr382 had a -13.17 kcal/mol interaction energy with THE C5-C8 intermediate carbocation and -9.82 kcal/mol to the C2-C9 intermediate carbocation. Tyr450 contributed -12.57 kcal/mol to the C5-C8 intermediate carbocation and -10.58 kcal/mol to the C2-C9 intermediate carbocation. Both Tyr382 and Tyr450 had a >2.0 kcal/mol stronger interaction with the C5-C8 intermediate carbocation. In the case of Tyr450 the π-chamber placed the least substituted end of the carbocation into the face of these aromatic ring, while the C2-C9 intermediate placed the methyl groups from the more substituted end into the face of this aromatic ring. This placed more charge into the face of the rings for C5-C8 than C2-C9 ultimately leading to a greater stabilization in this instance. For Tyr382 the situation is less obvious, but it does appear that the intermediate carbocation has a closer interaction with the aromatic ring of Tyr382, which might afford the differential stabilization that is observed.

In order to further study the factors stabilizing the phosphate dianion and the carbocation I carried out an energy decomposition analysis. Figure 3.13 shows the four states involved in the C5-C8 pathway with their interacting residues and water molecules. From the figure I observe that the diphosphate group is stabilized by hydrogen-bonding interactions with the side chains of Arg113, Lys201, Tyr205, Gln380, Lys294, Tyr450 and nine crystal water molecules (not shown

in Figure) in all states examined. It also reveals the reason for stabilization of the DMA^+ carbocation: the surrounding sidechains of the tyrosine residues (Try201, TrY382, Tyr435 and Tyr450) play an important role in stabilizing the carbocation from beginning to end via π-cation interactions. Specifically, at the intermediate state, Tyr382 was parallel to the carbocation while Tyr435 and Tyr450 were perpendicular to the carbocation. These three residues are perfectly situated to stabilize of the carbocation. Table 3.3 also shows the decomposition free energies for these four tyrosine residues paired with the DMA^+ carbocation using MMPBSA. The results show that the tyrosine residues have strong van de Waals interactions with the carbocation. Hence, all four tyrosine residues help to stabilize DMA^+ from bond cleavage to new bond formation.

Table 3.2 π-cation interaction energies (kcal/mol) for the carbocation intermediate states along the C5-C8 and the C2-C9 reaction pathways.

Pathways	Y382	Y435	Y450
C5-C8	-13.17	-9.81	-12.57
C2-C9	-9.82	-9.67	-10.58

Table 3.3 Decomposition of the free energies (kcal/mol) between the DMA^+ and several Tyrosine side chains.

Free Energy / kcal/mol	van der Waals	Electrostatic	Polar Solvation	Non-Polar Solvation	TOTAL Effects
Y203	-0.611	0.157	-1.609	-0.736	-2.799
Y382	-0.686	-0.896	-0.238	-0.672	-2.491
Y435	-0.331	-0.304	0.059	-0.232	-0.808
TYR 450	-0.982	0.618	-2.447	-1.167	-3.978

3.4 CONCLUSION

I have successfully carried out QM/MM simulations to further elucidate the catalytic mechanism for the prenyltransferase reaction catalyzed by wild type FtmPT1. I explored the reaction of Brevianamide F with DMAPP via both the C5-C8 and C2-C9 reaction channels, and found, as expected, that C5-C8 and C2-C9 had very similar activation parameters for the prenylation step. Hence, the final proton transfer/cyclization step was the product determining step for these reactions, with the deprotonation of the C5-C8 intermediate being nearly activationless (0.5 kcal/mol), while the deprotonation process for the C2-C9 pathway had a substantial barrier (higher than the C5-C8 and C2-C9 intermediate states) of 16.8 kcal/mol, resulting in the exclusive formation of the wild-type product Tryprostatin B. A key feature of the final deprotonation step was the involvement of a bridging water molecule between the transferring proton and Glu102, which was found in both cases to be the proton receptor.

An energy decomposition analysis analyzed the stabilizing effects of active site tyrosine residues on the DMA^+ carbocation at QM level. The surrounding tyrosine residues formed a π-chamber resulting in cation-π interactions between the formed carbocation intermediate and the π-chamber. The C5-C8 carbocation was stabilized by 5.48 kcal/mol more than the C2-C9 carbocation intermediate. The origin of this comes from the placement of the methyl groups into the aromatic faces for the C2-C9 intermediate, while the hydrogen atoms from the primary end of the allyl cation are placed into the aromatic faces for C5-C8. Finally, the diphosphate anion leaving group was stabilized throughout by Arg113, Lys201, Tyr205, Gln380, Lys294, Tyr450 and nine crystal water molecules in the active site of FtmPT1.

REFERENCES

REFERENCES

(1) Gordon, L. B., Kleinman, M. E., Miller, D. T., Neuberg, D. S., Giobbie-Hurder, A., Gerhard-Herman, M., Smoot, L. B., Gordon, C. M., Cleveland, R., Snyder, B. D., Fligor, B., Bishop, W. R., Statkevich, P., Regen, A., Sonis, A., Riley, S., Ploski, C., Correia, A., Quinn, N., Ullrich, N. J., Nazarian, A., Liang, M. G., Huh, S. Y., Schwartzman, A., and Kieran, M. W. (2012) Clinical trial of a farnesyltransferase inhibitor in children with Hutchinson-Gilford progeria syndrome, *P Natl Acad Sci USA 109*, 16666-16671.

(2) Casey, P. J., and Seabra, M. C. (1996) Protein prenyltransferases, *J Biol Chem 271*, 5289-5292.

(3) Spindler, S. R., Li, R., Dhahbi, J. M., Yamakawa, A., Mote, P., Bodmer, R., Ocorr, K., Williams, R. T., Wang, Y. S., and Ablao, K. P. (2012) Statin Treatment Increases Lifespan and Improves Cardiac Health in Drosophila by Decreasing Specific Protein Prenylation, *Plos One 7*.

(4) Tantillo, D. J. (2010) The carbocation continuum in terpene biosynthesis--where are the secondary cations?, *Chem Soc Rev 39*, 2847-2854.

(5) Amatore, C., Thiebault, A., and Verpeaux, J. N. (1989) Unexpected Single Electron-Transfer Catalyzed Cyclization of Prenyl Sulfone Dimer - Evidence for Radical-Anion Coupling in the Outer-Sphere Oxidation of Prenyl Sulfone Carbanion, *J Chem Soc Chem Comm*, 1543-1545.

(6) Bonitz, T., Alva, V., Saleh, O., Lupas, A. N., and Heide, L. (2011) Evolutionary Relationships of Microbial Aromatic Prenyltransferases, *Plos One 6*.

(7) Brajeul, S., Delpech, B., and Marazano, C. (2007) Sulfonium salts as prenyl, geranyl, and isolavandulyl transfer agents towards benzoylphloroglucinol derivatives, *Tetrahedron Lett 48*, 5597-5600.

(8) Chen, A. J., Kroon, P. A., and Poulter, C. D. (1994) Isoprenyl Diphosphate Synthases - Protein-Sequence Comparisons, a Phylogenetic Tree, and Predictions of Secondary Structure, *Protein Sci 3*, 600-607.

(9) Davisson, V. J., Neal, T. R., and Poulter, C. D. (1993) Farnesyl-Diphosphate Synthase - Catalysis of an Intramolecular Prenyl Transfer with Bisubstrate Analogs, *J Am Chem Soc 115*, 1235-1245.

(10) Heide, L. (2009) Prenyl transfer to aromatic substrates: genetics and enzymology, *Curr Opin Chem Biol 13*, 171-179.

(11) Hong, B. C., Hong, J. H., and Tsai, Y. C. (1998) Regio- and enantioselective prenyl anion transfer: Application to the total synthesis of (-)-rosiridol, *Angew Chem Int Edit 37*, 468-470.

(12) Jost, M., Zocher, G., Tarcz, S., Matuschek, M., Xie, X. L., Li, S. M., and Stehle, T. (2010) Structure-Function Analysis of an Enzymatic Prenyl Transfer Reaction Identifies a Reaction Chamber with Modifiable Specificity, *J Am Chem Soc 132*, 17849-17858.

(13) Kuzuyama, T., Noel, J. P., and Richard, S. B. (2005) Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products, *Nature 435*, 983-987.

(14) Li, S. M. (2009) Evolution of aromatic prenyltransferases in the biosynthesis of indole derivatives, *Phytochemistry 70*, 1746-1757.

(15) Liang, P. H., Ko, T. P., and Wang, A. H. J. (2002) Structure, mechanism and function of prenyltransferases, *Eur J Biochem 269*, 3339-3354.

(16) Nurenberg, G., and Volmer, D. A. (2012) The analytical determination of isoprenoid intermediates from the mevalonate pathway, *Anal Bioanal Chem 402*, 671-685.

(17) Poulter, C. D. (1996) Mechanistic studies of the prenyl transfer reaction with fluorinated substrate analogs, *Biomedical Frontiers of Fluorine Chemistry 639*, 158-168.

(18) Saleh, O., Haagen, Y., Seeger, K., and Heide, L. (2009) Prenyl transfer to aromatic substrates in the biosynthesis of aminocoumarins, meroterpenoids and phenazines: The ABBA prenyltransferase family, *Phytochemistry 70*, 1728-1738.

(19) Sinko, W., de Oliveira, C., Williams, S., Van Wynsberghe, A., Durrant, J. D., Cao, R., Oldfield, E., and McCammon, J. A. (2011) Applying Molecular Dynamics Simulations to Identify Rarely Sampled Ligand-bound Conformational States of Undecaprenyl Pyrophosphate Synthase, an Antibacterial Target, *Chem Biol Drug Des 77*, 412-420.

(20) Tello, M., Kuzuyama, T., Heide, L., Noel, J. P., and Richard, S. B. (2008) The ABBA family of aromatic prenyltransferases: broadening natural product diversity, *Cell Mol Life Sci 65*, 1459-1463.

(21) Yu, X., Liu, Y., Xie, X. L., Zheng, X. D., and Li, S. M. (2012) Biochemical Characterization of Indole Prenyltransferases Filling the Last Gap pf Prenylation Positions by a 5-DimethylAllylTryptophan Synthase from Aspergillus Clavatus, *J Biol Chem 287*, 1371-1380.

(22) Yu, X., Xie, X. L., and Li, S. M. (2011) Substrate promiscuity of secondary metabolite enzymes: prenylation of hydroxynaphthalenes by fungal indole prenyltransferases, *Appl Microbiol Biot 92*, 737-748.

(23) Wollinsky, B., Ludwig, L., Xie, X. L., and Li, S. M. (2012) Breaking the regioselectivity of indole prenyltransferases: identification of regular C3-prenylated hexahydropyrrolo[2,3-b]indoles as side products of the regular C2-prenyltransferase FtmPT1, *Organic & biomolecular chemistry 10*, 9262-9270.

(24) Chen, J., Morita, H., Wakimoto, T., Mori, T., Noguchi, H., and Abe, I. (2012) Prenylation of a Nonaromatic Carbon of Indolylbutenone by a Fungal Indole Prenyltransferase, *Org Lett 14*, 3080-3083.

(25) Luk, L. Y. P., Qian, Q., and Tanner, M. E. (2011) A Cope Rearrangement in the Reaction Catalyzed by Dimethylallyltryptophan Synthase?, *J Am Chem Soc 133*, 12342-12345.

(26) Mahmoodi, N., and Tanner, M. E. (2013) Potential Rearrangements in the Reaction Catalyzed by the Indole Prenyltransferase FtmPT1, *Chembiochem 14*, 2029-2037.

(27) Case, D. A., Darden, T. A., T.E. Cheatham, III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B. P., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Hsieh, M. J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2012) *AMBER11,University of California, San Francisco*.

(28) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *The Journal of chemical physics 79*, 926.

(29) Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1996) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995), *J Am Chem Soc 118*, 2309-2309.

(30) Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2005) Development and testing of a general amber force field (vol 25, pg 1157, 2004), *J Comput Chem 26*, 114-114.

(31) Elstner, M., Porezag, D., Jungnickel, G., Elsner, J., Haugk, M., Frauenheim, T., Suhai, S., and Seifert, G. (1998) Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties, *Phys Rev B 58*, 7260-7268.

(32) Cui, G. L., Li, X., and Merz, K. M. (2007) Alternative binding modes in Orf2 catalyzed prenylation: Identification and validation, *Abstr Pap Am Chem S 233*.

(33) Yang, Y., Chakravorty, D. K., and Merz, K. M. (2010) Finding a Needle in the Haystack: Computational Modeling of Mg2+ Binding in the Active Site of Protein Farnesyltransferase, *Biochemistry-Us 49*, 9658-9666.

(34) Yang, Y., Wang, B., Ucisik, M. N., Cui, G. L., Fierke, C. A., and Merz, K. M. (2012) Insights into the Mechanistic Dichotomy of the Protein Farnesyltransferase Peptide Substrates CVIM and CVLS, *J Am Chem Soc 134*, 820-823.

(35) Yang, Y., Miao, Y. P., Wang, B., Cui, G. L., and Merz, K. M. (2012) Catalytic Mechanism of Aromatic Prenylation by NphB, *Biochemistry-Us 51*, 2606-2618.

(36) Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model, *J Phys Chem-Us 97*, 10269-10280.

(37) M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. (2009) Gaussian 09, Revision A.02, *Gaussian, Inc., Wallingford CT*.

(38) Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems, *J Chem Phys 98*, 10089-10092.

(39) Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A Smooth Particle Mesh Ewald Method, *J Chem Phys 103*, 8577-8593.

(40) Sagui, C., Pedersen, L. G., and Darden, T. A. (2004) Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations, *J Chem Phys 120*, 73-87.

(41) Toukmaji, A., Sagui, C., Board, J., and Darden, T. (2000) Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions, *J Chem Phys 113*, 10913-10927.

(42) Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes, *J. Comput. Phys. 23*, 327–341.

(43) Miyamoto, S., and Kollman, P. A. (1992) SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models, *J. Comput. Chem. 13*, 952-962.

(44) Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Debolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules, *Comput Phys Commun 91*, 1-41.

(45) Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *J Comput Chem 26*, 1668-1688.

(46) http://www.ks.uiuc.edu/Research/vmd/.

(47) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *J Mol Graph Model 14*, 33-38.

(48) http://gnuplot.info.

(49) Jensen, M. O., Park, S., Tajkhorshid, E., and Schulten, K. (2002) Energetics of glycerol conduction through aquaglyceroporin GlpF, *P Natl Acad Sci USA 99*, 6731-6736.

(50) Crespo, A., Marti, M. A., Estrin, D. A., and Roitberg, A. E. (2005) Multiple-steering QM-MM calculation of the free energy profile in chorismate mutase, *J Am Chem Soc 127*, 6940-6941.

(51) Jarzynski, C. (1997) Nonequilibrium equality for free energy differences, *Phys Rev Lett 78*, 2690-2693.

(52) Hummer, G., and Szabo, A. (2001) Free energy reconstruction from nonequilibrium single-molecule pulling experiments, *P Natl Acad Sci USA 98*, 3658-3661.

(53) Hummer, G., and Szabo, A. (2003) Kinetics from nonequilibrium single-molecule pulling experiments, *Biophys J 85*, 5-15.

(54) Grossfield, A. WHAM: an implementation of the weighted histogram analysis method, http://membrane.urmc.rochester.edu/content/wham *version 2.0.8*.

(55) Dougherty, D. A. (1996) Cation-pi interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp, *Science 271*, 163-168.

(56) Dougherty, D. A. (2007) Cation-pi interactions involving aromatic amino acids, *J Nutr 137*, 1504S-1508S.

(57) Jenson, C., and Jorgensen, W. L. (1997) Computational investigations of carbenium ion reactions relevant to sterol biosynthesis, *J Am Chem Soc 119*, 10846-10854.

(58) Hong, Y. J., and Tantillo, D. J. (2013) C-H center dot center dot center dot pi interactions as modulators of carbocation structure - implications for terpene biosynthesis, *Chem Sci 4*, 2512-2518.

(59) Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1995) Pairwise Solute Descreening of Solute Charges from a Dielectric Medium, *Chem Phys Lett 246*, 122-129.

(60) Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1996) Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium, *J Phys Chem-Us 100*, 19824-19839.

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

(62) Weiser, J., Shenkin, P. S., and Still, W. C. (1999) Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO), *J Comput Chem 20*, 217-230.

CHAPTER 4. MECHANISM OF THE FORMATION OF A NON-STANDARD PRODUCT IN THE G115T MUTANT OF FTMPT1: A CASE OF DYNAMICS CALLING THE SHOTS?

4.1 INTRODUCTION

The prenyltransferases (PTs) are a broad class of enzymes that catalyze prenyl transfer reactions generating a broad range of compounds of nutritional and medicinal value.[1-3] The fungal indole PT, is a sub-class of the PTs where the α helices and β strands have the 'αββα' topology.[4-9] FtmPT1, *i.e. Aspergillus fumigatus prenyltransferase*, an example of the fungal indole PTs, catalyzes the prenyl transfer reaction between dimethyl allyl pyrophosphate (DMAPP) and tryptophan and its derivatives, via a S_N1 mechanism[9-15] (see Figure 4.1)**.** The wild type reaction of DMAPP and Brevianamide F (cyclo-L-Trp-L-Pro) yields Tryprostatin B, a cell growth inhibitor, while the Gly115Thr mutant leads to the formation of a novel product.[11] Both reactions are initiated by C-O bond cleavage to form the dimethyl allyl carbocation, which undergoes nucleophilic attack by the indole ring of Brevianamide F for the native and mutated systems (see Figure 4.1).[16] I recently rationalized the observation of Tryprostatin B formation in native FtmPT1 by showing that the rate-determining step is the C-O bond cleavage step and that the final proton acceptor for the deprotonation step was Glu102, both of which were consistent with available experimental results.[11, 12]

The different products formed by native and mutant FtmPT1 follow distinct reaction channels. The formation of the "regular" prenylation product, Tryprostatin B, involves carbon-carbon bond formation between the C-5 carbon of DMAPP (C5@DMAPP) and the C8 carbon of the indole ring (C8@indole) of Brevianamide F. The formation of the "reverse" or novel prenylation product involves a carbon-carbon bond formation between the C-2 carbon of DMAPP (C2@DMAPP) and the C9 carbon of the indole ring (C9@indole) (see Figure 4.1). The

mechanistic difference between these two pathways arises because the reverse reaction involves a cyclization/deprotonation step to complete the reaction, while the regular pathway is terminated by a simple deprotonation reaction after the prenyl transfer reaction (see Figure 4.1). The mutation of Glycine115 to Threonine facilitates the formation of the reverse product, while for the wild type the reverse product was not observed. Besides tryptophan, its derivatives can be prenylated as well and a recent publication further demonstrates that non-aromatic substrates can react with DAMPP in FtmPT1.[17]

The regioselectivity of prenylation reactions between DMAPP and tryptophan (and its derivatives) has been investigated.[14, 18, 19] Individual PTs catalyze the formation of specific prenyl transfer products; for example, FtmPT1 favors C5-C8 bond formation, CdpC3PT[20]/AnaPT[21] favor C2-C9 bond formation, while other PTs give yet other regioselectivities.[10] Hence, the C2-C9 product, while not unprecedented, is not observed in wild-type FtmPT1.[11] The mutation of Gly115 to Ala or Ile in FtmPT1 deactivates the enzyme, while Thr prefers the C2-C9 reaction channel over the "regular" or C5-C8 one while simultaneously increasing the reaction barrier for the overall reaction.[11] It has been hypothesized that the change in regioselectivity is due to a reduction in the reaction coordinate distance, which favors the C9@indole-C2@DMAPP reaction channel over the C8@indole-C5@DMAPP reaction channel,[10, 14] following the Near Attack Conformation (NAC) argument.[22, 23] The regioselectivities of prenyltransfereases have been studied experimentally[18, 19, 24] as well as computationally[25] during the past two decades.

In my previous paper, I studied the regioselectivity of the reactions catalyzed by wild type FtmPT1.[12] Using potential of mean force simulations I succeeded in describing the reaction

mechanisms for both the C5-C8 and C2-C9 reaction pathways. I found that both reaction channels gave essentially identical barrier heights for the prenyl transfer process (*e.g.*, cleavage of the DMAPP C-O bond), but for C2-C9 the final cyclization/deprotonation step was rate-determining, while carbocation formation was rate-limiting for the regular C5-C8 pathway. Moreover, the barrier for the former was higher than for the latter, thereby, preferring C5-C8 bond formation and ultimately Tryptostation B as product. In this paper, however, exploring the catalysis of the Gly115Thr mutant proved more complex requiring a more detailed examination of the thermodynamics of the process as well as the reaction dynamics from the rate-determining DMAPP C-O bond cleavage transition state. According to the hypothesis of Li and co-workers,[14] due to the increase of the residue size afforded by the G115T mutation, the C9@indole to C2@DMAPP distance is reduced in the resting state giving preference to the formation of the reverse product, which is akin to "starting point compression" as described in the NAC argument.[22, 23, 26] In my work, I also find that in the G115T mutant the C2-C9 (~3.7 Å) distance is shorter than the C5-C8 (~5.0 Å) distance in the pre-reactive complex for the prenyl transfer reaction at the resting states, but this did not affect the C-O bond cleavage barrier, suggesting other factors are playing the defining role.

The role of reaction dynamics in catalysis has been explored in a number of cases and has been proposed to govern the observed reaction product distribution.[27-31] In these analyses, direct chemical dynamics simulations were applied to the transition states using electronic structure calculations,[32-34] and the reaction preference(s) of the various available reaction channels were elucidated through investigation of the reaction dynamics.[29, 35-39] Recently, reaction dynamics analyses for the determination of relatively complex reaction mechanisms has been reported.[35, 38,

[40] In these papers, external conditions, such as temperature and pressure, were potential inducers of the formation of unusual product outcomes.[29, 31, 41-43] Kinetic analyses were applied as needed to quantify the reaction outcomes.[36, 37] In these reaction dynamics simulations, vibrational states of the reaction transition state are used to generate the temperature distributions for the reactions.[32-35, 38] With the initial conditions applied, temperature-released dynamics simulations were done to explore how reaction dynamics affects the selection of the various reaction channels. Although this method affords challenges relative to thermodynamics based analyses, it does enlighten reaction channel preferences that can further rationalize experimental preferences of certain reactions.[42] Recent publications have established that dynamics has the potential to drive biomolecular reactions.[37, 38, 40, 42, 43] In an interesting recent paper[44] the authors have shown that the energy released during catalytic turnover can ultimately cause molecular motion in enzymatic systems. In the present case (the G115T mutant) two bifurcating reaction channels are present, which are thermodynamically competitive, but they yield only one major product. To explain this behavior I have used reaction dynamics analyses using an initial temperature distribution applied to my transition state structures observed in free energy calculation. I then complete a series of free MD simulations, to determine how dynamics influences the observed product distribution.

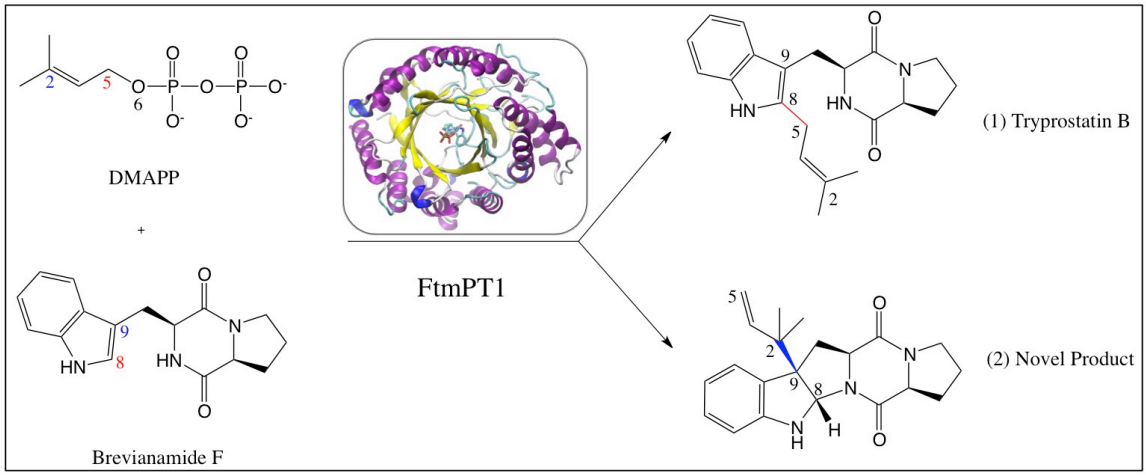


Figure 4.1 Overview of the FtmPT1 reactions. (1) Reaction product Tryprostatin B by bond formation of C5@DMAPP and C8@indole ring; (2) Reaction product five-member ring compound by bond formation of C2@DMAPP and C9@indole ring.

4.2 METHODS

4.2.1 General Aspects

The FtmPT1 X-ray complex crystal structure, which has dimethylallyl S-thiolodiphosphate ($C_5H_{14}O_6P_2S$; a substrate mimetic) and Brevianamide F ($C_{16}H_{17}N_3O_2$) bound in its active site, was used as the starting geometry (PDB code 3O2K).[11] It was modified by replacement of the substrate mimetic $C_5H_{14}O_6P_2S$ with dimethyl allyl pyrophosphate ($C_5H_{14}O_7P_2$), and Glycine115 was mutated to Threonine. The resulting complex structure was then processed by the LEaP module of the Assisted Model Building with Energy Refinement (AMBER, version 12) suite of programs[45] with all missing residues and hydrogen atoms added.

The complex system was solvated by TIP3P[46] water molecules in truncated octahedral simulation cell. The ff99sb force field[47] was employed for the protein system, and the generalized amber force field (GAFF)[48] was used to model the substrates in order to construct the parameter topology files. The restrained electrostatic potential (RESP) methodology[49] in the Antechamber module of the AMBERTools suite of programs was used to evaluate the charges of the substrates. Prior to running RESP the geometries of the substrates were optimized and then their electrostatic potentials were both obtained at the M06L/6-31+G(d,p)//HF/6-31+G(d,p) level of theory[50-52] using the Gaussian 09 electronic structure program[53]. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) method,[54-57] and an 8.0 Å cutoff was used for non-bonded interactions. All bonds with hydrogen atoms were constrained using SHAKE[58, 59] except for the studies involving proton transfer. All molecular dynamics (MD) simulations were performed using SANDER and PMEMD from the AMBER12 suite of programs.

Prior to carrying out MD simulations the simulation cell was first minimized to remove close contacts using restraints that were gradually decreased from 10.0 kcal/(mol·Å^2) to 0.0 kcal/(mol·Å^2) within 4000 minimization cycles. A 500ps restrained MD simulation was subsequently run using a 1fs time step, while the temperature was gradually increased from 0 K to 300 K. Then the restraints were slowly removed over 100ps from 10.0 kcal/(mol·Å^2) to 0.0 kcal/(mol·Å^2) prior to a 450 ps constant NPT MD simulation, which was followed by a 22 ns constant NVT MD simulation. Both of the latter two MD simulations were used a 2 fs time step, with snapshots saved every 2 ps. The restraints used in the energy minimization and NVT simulation were introduced using the NMR-derived energy restraints available in the SANDER

program[60, 61]. After the 22ns classical MD simulation, a QM/MM minimization followed by a 1ns QM/MM MD equilibration were carried out using SANDER. In all QM/MM MD simulations, unless otherwise noted, the self-consistent charge density functional tight-binding (SCC-DFTB) method,[62] as implemented in AMBER12, was the QM level of theory employed. All reaction related substrates (Brevianamide F, DMAPP and their intermediate products), Glu102 and two surrounding water molecules were included in the QM region.

The Visual Molecular Dynamics (VMD) program[63, 64] was used to analyze the simulation results. GNUPlot[65] was used to plot the free energy surfaces.

4.2.2 Prenyl Transfer Reaction Calculations

4.2.2.1 C2-C9 Bond Formation

QM/MM MD studies to obtain the prenylation reaction free energy profiles were performed along two dimensional (2D) reaction coordinates, with both DMAPP and Brevianamide F in the QM region, while the remainder of the system was treated classically**.** In the C2-C9 reaction type, the two reaction coordinates were the cleavage of the C5-O6 bond and the formation of the C2-C9 bond. Steered Molecular Dynamics (SMD)[66-70] was first applied to propagate the trajectory along the reaction coordinates of C2-C9 with a distance change from 3.7 Å to 1.4 Å, using a force constant of 5000 kcal/(molÅ^2) and a pulling speed of 0.02 Å/ps. This was followed by an umbrella sampling calculation where the SMD snapshots were used as the starting structure where the C2-C9 distance was fixed and the C5-O6 distance was scanned. An extended umbrella sampling was done with C2-C9 distance region of [3.7 Å, 4.2 Å] to ensure the coverage of the transition state around the C2-C9 distance of 3.7 Å. The complete umbrella sampling

calculation was performed along C2-C9 distances in the interval of [1.4 Å, 4.2 Å] and the C5-O6 distance along the interval [1.4 Å, 7.0 Å] with a spacing of 0.1 Å, resulting in a total of 1653 windows. The umbrella sampling was done using NVT with an initial 50 ps of equilibration followed by 50 ps of data collection. The force constants for C2-C9 and C5-O6 were 500-600 kcal/(molÅ^2) and 300-500 kcal/(molÅ^2) respectively, where higher force constants were used for the higher energy regions to insure thorough sampling. The two-dimensional weighted histogram analysis method (WHAM-2D)[71] was implemented to analyze the probability density and obtain the free energy profiles (FEP, 2D) for the unbiased system along the two reaction coordinates.

In addition to the 2D umbrella sampling run, a 1D umbrella sampling simulation was also performed along the C5-O6 reaction pathway and is shown in the supporting information. The starting configurations for the umbrella sampling calculations were obtained from a steered MD calculation along the O5-C6 reaction coordinate. The calculation covered spanned C5-O6 diatnaces along [1.4 Å, 4.6 Å] with a spacing of 0.1 Å between each adjacent window, which resulted in 33 total windows. A 100 ps constant NVT equilibration run was followed by a 150 ps constant NVT data collection run for each of the windows. The 1-D weighted histogram analysis method (WHAM) was used to obtain the free energy profiles (FEP, 1D) for the unbiased system along the C5-C8 reaction coordinate.

4.2.2.2 C5-C8 Bond Formation

A comparison study of carbon-carbon bond formation at C5-C8 was performed to determine its free energy profile in mutant FtmPT1. C5-C8 bond formation represents the favored reaction pathway in the wild type system. A full 2D PMF calculation was done to assemble a 2D free

energy map, while a 1D PMF calculation of C5-O6 bond breakage was not done because it would be identical to that for the C2-C9 reaction pathway in Section 2.2.1. The 2D umbrella sampling calculation was also performed using 50 ps of constant NVT equilibration and 50 ps of constant NVT data collection. The calculation windows were set at a C5-C8 distance interval of [1.4 Å, 5.5 Å] and a C5-O6 distance interval of [1.4 Å, 9.0 Å] to cover all possible reaction pathways and their associated transition states. This resulted in a total of 3234 windows. For the steered MD simulation the force constant was 5000 kcal/(molÅ^2) with a pulling speed of 0.02 Å/ps; for the PMF calculation the force constants were 500-600 kcal/(molÅ^2) and 300-500 kcal/(molÅ^2) for the C5-C8 bond and C5-O6 bond, respectively. The larger force constants were used in the high energy region to ensure uniform sampling. DMAPP and Brevianamide F were the sole species included in the QM region.

We carried out a subsequent NVT MD umbrella sampling calculation for both the C5-C8 and C2-C9 reaction pathways to connect the prenylation product with the starting point for the de-protonation process. my target was to determine the relative free energy changes from the unrelaxed prenylation product to the relaxed state which serves as the starting point for deportation. Starting from the prenylation product, an umbrella sampling calculation was carried out along the C5-O6 reaction pathway, covering an interval of [6.0 Å, 15.0 Å] with a spacing between points of 0.25 Å. A force constant of 100 kcal/(molÅ^2) was used in these simulations. The distance region was selected to fully coverage the potential distance change between the two states. Since this calculation was for a non-reactive pathway, the sampling was carried out using a classical force field model using a constant NVT MD simulation with a time step of 1fs. 1 ns of equilibration was followed by 1 ns of data collecting to assemble the profile.

4.2.3 Formation of the Final Products

4.2.3.1 Cyclization of Penta-Ring and Deprotonation to the Final Product for C2-C9 Reaction Pathway

After the C2-C9 prenylation reaction a cyclization reaction between C8-N7 combined with a deprotonation step involving H(@N7)-N7 bond breakage yields the final product. To affect this reaction the protonated prenylation product along with a proton acceptor in the QM region was needed. The only two possible proton acceptors were Glu102 and a nearby water molecule. The diphosphate anion was more than 12 Å away from the prenylation product (after relaxation structural relaxation), which eliminated it as a possible proton acceptor. Hence, the QM region included the protonated C2-C9 prenylation product, the side chain of Glu102 as well as the nearby water molecule.

We did a 20 ns constant NTP pure MD equilibration prior to a 1ns constant NVT QM/MM MD equilibration to fully relax the system described above. An initial steered MD run along the C8-N7 cyclization reaction pathway was done to generate starting configurations for a subsequent series of 2D PMF calculations using umbrella sampling. Starting at a fixed C8-N7 distance, the H(@N7)-N7 distance was change to assemble the 2D free energy profile. This resulted in 456 windows covering C-N distances of [3.2 Å, 1.4 Å] and N-H distances of [0.9 Å, 3.2 Å]. The umbrella sampling simulations were done with a 50 ps constant NVT equilibration step followed by A 50 ps constant NVT data collection step. The Steered MD force constant was 5000 kcal/(molÅ^2) with a pulling speed of 0.02 Å/ps. The PMF force constants for the C-N bond was 300-500 kcal/(molÅ^2) and for the N-H bond it was 500-1000 kcal/(molÅ^2), with higher force

constants used for windows close to the saddle points. WHAM-2D was used to obtain the final 2D free energy map.

A 100 ns NVT MD simulation was done in order to further analyze the structure and of dynamics of the final product and its interactions with the enzyme active site.

4.2.3.2 Proton Transfer to the Final Product for C5-C8 Reaction Pathway

For the C5-C8 reaction pathway, in order to complete the catalytic mechanism, the hydrogen atom on the C8 atom of the indole ring of Brevianamide F is transferred to a proton acceptor yielding Tryprostatin B. Similar to my study of the C2-C9 pathway, the protonated C5-C8 prenylation product, Glu102 and the water molecule were included in the QM region.

We also did a 20 ns constant NTP classical MD equilibration followed by a 1ns constant NVT QM/MM MD equilibration to relax the complex system prior to subsequent QM/MM Steered MD and Umbrella Sampling studies. The Steered MD and Umbrella Sampling calculations for the H(@C8)-C8 bond cleavage were done over a distance interval of [1.1 Å, 3.5 Å]. The Steered MD run used a force constant of 5000 kcal/(molÅ^2) and a pulling speed of 0.02 Å/ps. The Umbrella Sampling PMF calculation used a constant NVT equilibration of 100 ps and then a constant NVT data collection step of 150 ps, using force constants of 300 kcal/(molÅ^2) for pre- and post-transition state windows, 600 kcal/(molÅ^2) for windows close to transition state, and 100 kcal/(molÅ^2) for windows far away from the transition state region. WHAM-1D was used to obtain the 1D-FEP curve for this deprotonation step.

A 100 ns NVT classical MD simulation was completed for the final product to further analyze the products for the catalytic mechanism.

4.2.4 Free QM/MM MD Simulation Near the C5-O6 Bond Breakage Transition State

We did separate unrestrained QM/MM MD simulations at select critical points along the reaction pathways during the prenylation reaction for both the C2-C9 and C5-C8 reaction pathways to examine the role reaction dynamics plays in the product distribution. I call these MD simulations "free MD" in the following discussion, because all restraints were released along the C5-C8, C2-C9, and C5-O6 reaction coordinates. my goal was to observe the product preferences of the C5-C8, C2-C9, and C5-O6 bond formation, at the transition state region between the starting materials and the formation of the prenylation product. Hence, similar to the application of vibrational transition state theory (VTST) to reactions using electronic structure theories,[32, 34, 37, 39] I selected the transition state I region involving the C5-O6 bond cleavage and the C5-C8/C2-C9 bond formation. For the C2-C9 reaction pathway, windows for C5-O6 at [3.5 Å, 3.8 Å] and C2-C9 at [2.4 Å, 2.9 Å] were retained; for the C5-C8 pathway, windows for C5-O6 at [4.8 Å, 5.1 Å] and C5-C8 at [2.4 Å, 2.9 Å] were retained. For these regions, an additional 100 ps NVT QM/MM equilibration was done for each window based on the original 100 ps equilibration from the previous PMF calculations. A follow on 100 ps NVE simulation was done to obtain initial snapshots for the free MD simulations, where each free MD window was selected every 1 ps. In this NVE simulation, an initial temperature of 300 K was assigned to obtain a Boltzmann distribution of energy at this temperature. Thus, for each of the windows, 100 snapshots were picked for 5ps NVE MD simulations, resulting in 2400 windows in total for each of the reaction pathways. Another set of free MD simulations was done for comparison to the prenylation

process in the native FtmPT1 system. The same window coverage for the C2-C9 and C5-C8 reaction pathways were used as described above for G115T. Note that the free MD simulations are QM/MM MD simulations, with the reactive substrates in the QM region and the remainder in the classical MM region. The time step used was 1fs for all calculations.

4.3 RESULTS AND DISCUSSION

A typical prenyltransferase catalyzed reaction includes a prenyl transfer step followed by a proton transfer step.[11, 12] Below I analyze these mechanistic steps using both thermodynamic methods as well as an approach that examines how reaction path dynamics might affect the reaction outcome. Further studies attempt to better understand the affect of QM Hamiltonian choice has on the outcome of the proton transfer step. Integrating all of these approaches allows us to develop a hypothesis of how the G115T of FtmPT1 affords the novel product at the expense of the native product.

4.3.1 C2-C9 Reaction Pathway Thermodynamics Analysis

4.3.1.1 Prenyl Transfer

For the prenyl transfer reaction, using a range of computational tools described in the method section, I have generated sampling points on a 2-D surface describing the breaking/formation of the C2-C9 and C5-O6 bonds. Using WHAM-2D **I** give the resultant 2-dimensional potential of mean force (PMF) profiles for this reaction in Figure 4.3(a).

In Figure 4.3(a) the C2-C9 distance is the X-axis and the C5-O6 distance is the Y-axis. On this energy-contoured map I find three saddle points and four wells, *i.e*. the starting state, three transitions states, two intermediate states, and the prenylation product state. With the starting state arbitrarily set at 0.0 kcal/mol, I can estimate the relative free energies of these key states. Starting at a (C2-C9, C5-O6) distance of (3.7 Å, 1.5 Å), I directly proceed to the first transition state, *i.e*. C5-O6 bond cleavage at (3.7 Å, 2.45 Å) with a free energy of 24.6 kcal/mol. The newly formed carbocation can reach a stationary state at position (3.7 Å, 5.3 Å), with a slightly lower free energy of 21.5 kcal/mol. This "intermediate" carbocation rotates itself slightly and overcomes a small barrier of 1.0 kcal/mol (relative energy of 22.5 kcal/mol) at (2.75 Å, 5.35 Å) to proceed to the most favorable intermediate state at (2.1 Å, 5.5 Å), which has a relative free energy of 17.5 kcal/mol.

The second transition state, *i.e*. the C2-C9 bond formation, has a free energy barrier of 0.9 kcal/mol (relative energy of 18.4 kcal/mol) at (1.95 Å, 5.6 Å), and the prenylation product is found at (1.65 Å, 5.8 Å) with a relative free energy of 16.4 kcal/mol. While I could not find an extra saddle point at the bottom left part of the 2D PMF map, which would correspond to a S_N2 reaction pathway, to determine the relative energy of the S_N1 and S_N2 pathways, I am comfortable predicting that G115T follows a S_N1-like reaction mechanism. This means that O-C bond cleavage is separate from C-C bond formation with both having a carbocation as the common intermediate. The difference from a typical S_N1 pathway is that the G115T mutant pathway contains an additional shallow carbocation stationary state at (3.7 Å, 5.3 Å), which is slightly less stable (by 4 kcal/mol) than the final carbocation state at (2.1 Å, 5.5 Å) (See Figure 4.6(a)). Nevertheless, when looking into the wild type C2-C9 reaction pathway, the carbocation

"stationary state" at (3.7 Å, 5.3 Å) is only 0.5 kcal/mol lower in free energy than the rotational state at (3.0 Å, 5.35 Å), which is not as evident as that for the mutant system. (See Figure 4.9(a)) Hence these stationary and rotational states were not discussed in my wild type paper.[12]

4.3.1.2 Intermediate Products Relaxation

To complete the carbocation formation for the C2-C9 reaction pathway requires a relaxation step along the C5-O6 bond distance (which is constrained in the 2D PMF study) prior to examining the proton transfer/cyclization step. Here again, akin to what was done in the wild type study[12], I carry out an MD PMF calculation to estimate the free energy released when the C5-06 bond is relaxed. Figure 4.4 shows the computed profile. The free energy released when I relax the C5-O6 bond was 10.7 kcal/mol and involved a bond distance change of 5.8 Å to 14.0 Å. This relaxation step sets the relative free energy of the starting state for the proton transfer/cyclization reaction to 5.6 kcal/mol for C2-C9, as shown in Table 4.1.

4.3.1.3 Proton Transfer and Cyclization to Obtain Final Product

C2-C9 proton transfer is coupled with a cyclization step, which I also examined for the wild-type enzyme.[12] Because two bond are forming/breaking I again will make use of a 2-dimensional reaction coordinate involving the breaking N-H and the forming C-N bond. Figure 4.5(a) shows the resultant 2D PMF map with important states indicated by dots. The reaction starts at a C8-N7/N7-H distance of (3.2 Å, 1.0 Å), which is 13.2 kcal/mol above the product state (see Figure 4.5(a)), the transition state is found at (2.05 Å, 1.25 Å) with a free energy barrier of 13.5 kcal/mol which is 26.7 kcal/mol in above product state. The final cyclized deprotonated product is formed at (1.5 Å, 3.4 Å) and is 13.2 kcal/mol lower in free energy (it is set as the zero of

energy in Figure 4.5(a)) than the starting state for the deprotonation/cyclization reaction. I find a secondary transition state at (1.5 Å, 1.2 Å) where the N7-C8 bond is formed first followed by proton transfer (see the yellow dot in Figure 4.5(a)). The free energy barrier is slightly higher (1.0 kcal/mol) than the former transition state, which I use in all subsequent discussion. Combining the depronation reaction with the prenylation and relaxation processes I obtain the complete reaction free energy profile as shown in Table 4.1. Detailed structures associated with their free energies for each important state is shown in Figure 4.6(a).

Looking at the structures of Figure 4.6(a) I again find that Glu102, as found in my study of the wild type reaction[12], hydrogen bond with an active site water molecule initially before becoming the final proton donor. Experimentally, via mutagenesis experiments, Glu102 has also been found to be important for the catalytic reaction, because its elimination resulted in the formation of no products.[11] Once the proton is transferred to Glu102 a new hydrogen bond is formed between protonated Glu102 and the water molecule involved in the proton shuttle (see Figures 3a and 4a).

We have also carried out a NVT classical MD simulation for 100 ns at the end-point of the reaction, which shows the product remains bound inside the binding pocket at this timescale. Figure 4.7(a) shows a snapshot of the equilibrated binding site. I find that the binding site is relatively narrow and the product is tightly packed and hydrogen bound to the surrounding active site residues.

4.3.2 C5-C8 Reaction Pathway Analysis

Experimentally, not only are C2-C9 based products observed (the major product) but also the "native" C5-C8 based reaction product is observed (the minor product). Hence, in order to complete the story I have done the same analysis for the C5-C8 reaction pathway to gain further insights into the origin of the product distribution observed in the G115T mutant.

4.3.2.1 Prenyl Transfer

The prenyl transfer step for C5-C8 bond formation for the G115T mutant is akin to the wild type reaction[12] with the only difference being a longer C5-O6 distance in the prenylation product state. Figure 4.3(b) shows the detailed 2-dimensional PMF map for the reaction free energy profile. The reaction starts at a C5-C8/C5-O6) distance of (5.0 Å, 1.5 Å), passes through the first transition state at (5.0 Å, 2.45 Å) involving C5-O6 bond breakage with a free energy barrier height of 24.9 kcal/mol (the starting state is at the zero of energy). Similar to the C2-C9 reaction pathway, the resultant carbocation passes through a shallow well as its initial stationary state at (5.2 Å, 5.0 Å) with a relative free energy of 21.5 kcal/mol, and then rotates to give a transition state at (3.75 Å, 6.5 Å) via a barrier of 1.3 kcal/mol (relative free energy of 22.8 kcal/mol). (See Figure 4.6(b)) This is different from the wild type C5-C8 reaction pathway, which only has ~0.1 kcal/mol rotational free energy barrier from the stationary state at (4.25 Å, 5.45 Å) to the rotational state at (3.0 Å, 5.0 Å) (see Figure 4.9(b)). The carbocation then traverses to its most stable intermediate state at (2.15 Å, 8.0 Å) with a relative free energy of 18.0 kcal/mol. The carbocation then passes through the second transition state at (1.95 Å, 8.05 Å) with a free energy barrier of 1.8 kcal/mol (relative energy of 19.8 kcal/mol) forming the prenylation product at (1.6 Å, 7.8 Å) with a relative free energy of 15.0 kcal/mol. Again the PMF simulations indicate I am

dealing with a classic S_N1-like reaction process. Again I was unable to locate the S_N2 transition state in the regions I examined with my PMF study, but expect it to have a much higher free energy that the corresponding S_N1 process.

Table 4.1 gives the detailed free energies for the C5-C8 prenylation reaction. I observe that it is comparable with the C2-C9 reaction pathway, with only slight modifications. This result indicates that from the computational perspective the prenylation reaction does not determine the final product distribution of the C2-C9 or C5-C8 reaction pathways.

To garner further structural insights into the G115T mutant reaction I compare it to the previously studied wild type reaction.[12] Experimentally, when compared to the wild type system, the G115T mutant only yielded 0.4% Tryprostatin B.[11] Hence by the Boltzmann distribution equation, I can estimate that the G115T mutant has a ~3.3 kcal/mol higher free energy barrier relative to the wild type system at the rate determining step, which is the O-C bond cleavage step. Meanwhile, in my calculations on the formation of Tryprostatin B, I obtained a 19.4 kcal/mol free energy barrier for the wild type system[12] (Table 4.2), while herein, I report 24.9 kcal/mol barrier for the mutant protein (Table 4.1) which represents a 5.5 kcal/mol increase in the free energy barrier. I find that both the G115T mutant and the wild type systems have similar O-C and C-C distances at the first transition state (TS I) for each pathway – O-C = 2.45 Å, C2-C9 = 3.7/3.8 Å, C5-C8 = 4.9/5.0 Å. This suggests that the G115T mutant active site is more crowded because of the presence of the Thr residue relative to the compact Gly residue resulting in a shifting of the transition state coordinates. This crowding can be readily seen in Figure 4.2. Furthermore, in the G115T mutant, I observe that the C5-C8 pathway has a longer starting C-C

distance (5.0 Å) and finishing C-O distance (7.8 Å) when compared to that of the C2-C9 pathway (3.7 Å for C-C distance and 5.8 Å for C-O distance). So, even though these two pathways have similar activation free energies for the C-O bond cleavage there are noticeable structural differences along the reaction pathway.

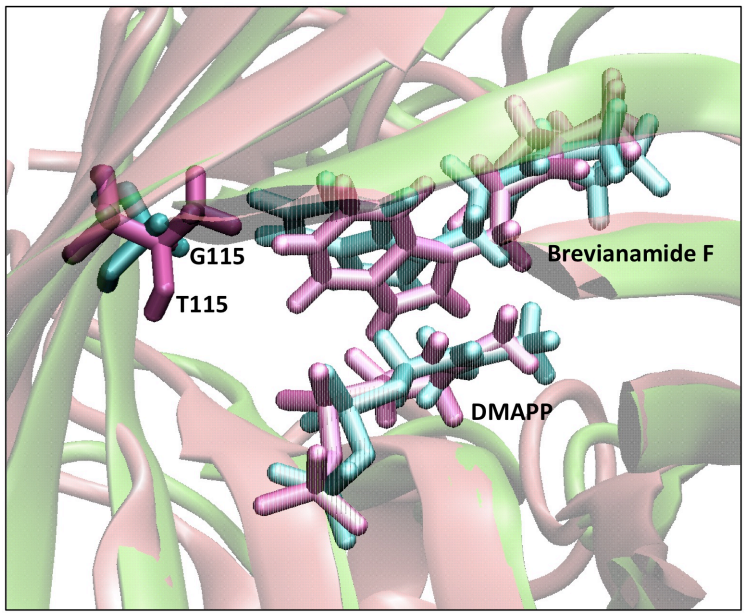


Figure 4.2 Snapshots of wild type FtmPT1 and G115T mutated FtmPT1 systems at their reaction starting states. Note all pink tone structures and protein represent the mutated system, and all green tone ones represent the wild type system. Key structures are shown as labeled.

4.3.2.2 Intermediate Products Relaxation

An additional intermediate state relaxation step was done as well for the C5-C8 reaction pathway after obtaining the prenylation product to obtain the starting point for the proton transfer reaction. Here again, I carry out a classical MD PMF simulation to estimate the free energy change as a result of the relaxation process. Figure 4.4 shows the simulation results. my simulations estimate that the C5-O6 distance relaxes from 7.8 Å to 15.3 Å and that the free

energy decreases by 6.2 kcal/mol. This relaxation step sets the relative free energy of the starting state for the proton transfer step to 8.8 kcal/mol (relative to the starting state) for C5-C8 pathway, as indicated in Table 4.1.

4.3.2.3 Proton Transfer to Obtain Final Product

To complete the C5-C8 catalytic mechanism requires a proton transfer to the surrounding environment. The C5-C8 one-dimensional PMF calculation along the C8-H distance was carried out to estimate its energetic cost. The resultant profile is shown in Figure 4.5(b). Here I find that this reaction has a lower activation free energy barrier than the for C2-C9 deprotonation reaction relative to their relaxed starting states. At a C8-H distance of 1.23 Å, the deprotonation reaction reaches its transition state with a free energy barrier of 7.1 kcal/mol, which when referenced to the starting state is 20.3 kcal/mol as indicated in Figure 4.5(b). An active site water molecule is again connecting Glu102 and the transferring proton from the prenylation product at the transition state. The reaction is completed at a C-H distance of 3.15 Å, with 13.2 kcal/mol lower free energy relative to the deprotonation starting state. The extra proton is also finally transferred from the prenylation product, passing through the water molecule, and protonating Glu102. The final product of this reaction, however, does not form a new hydrogen bond with the water molecule, as shown in Figure 4.5(b) and 4.6(b). Figure 4.4(b) shows that after 100ns of classical MD simulation the product remains inside the binding pocket. Combining the proton transfer reaction with the previous prenylation and relaxation processes, I obtain a complete reaction free energy profile with regards to the starting state. This is summarized in Table 4.1, and the structures with their associated free energies are shown in Figure 4.6(b).

Comparing the free energy data in Table 4.1, I observe that the two reactions have relative similar reaction profiles. Starting from the prenyl transfer reaction, the C2-C9 and C5-C8 pathways have essentially identical free energy values for transition state I, which is not unreasonable due to the identical O-C bond being broken within DMAPP. Moreover, transition state I is also the rate-determining step for both reactions. C2-C9 has 1.4 kcal/mol lower free energy barrier for transition state II than C5-C8, while for the proton transfer steps C2-C9 has a higher transition state (19.1 kcal/mol, TS III) than C5-C8 (15.6 kcal/mol). Finally, C2-C9 has a 3.2 kcal/mol lower final product (-7.6 kcal/mol) compared to the C5-C8 pathway (-4.4 kcal/mol). Therefore, according to my PMF calculation, the C2-C9 pathway generates a more stable final product compared to that of the C5-C8 pathway, though the complete reaction procedures are highly competitive.

Given these profiles I am unable to explain why the C2-C9 cyclized product is favored over the "native" C5-C8 product. The rate-determining step is identical in both cases, unlike what I found in the native system, where there was a change in rate-determining step for the C2-C9 product involving the proton transfer/cyclization step, while for C5-C8 the cleavage of the O-C bond in DMAPP was rate-limiting. Hence, in order to explain the experimentally observed product distribution I have to consider other alternatives. Since both pathways favor O-C bond cleavage as rate-limiting perhaps dynamical affects as transition state I collapses affects the formation of the final carbocation intermediate and governs the final product distribution? I explore this possibility below. Another possibility is the methods employed are not of high enough "resolution" to accurately evaluate the G115T mutant reaction profile and this is explored as well in what follows.

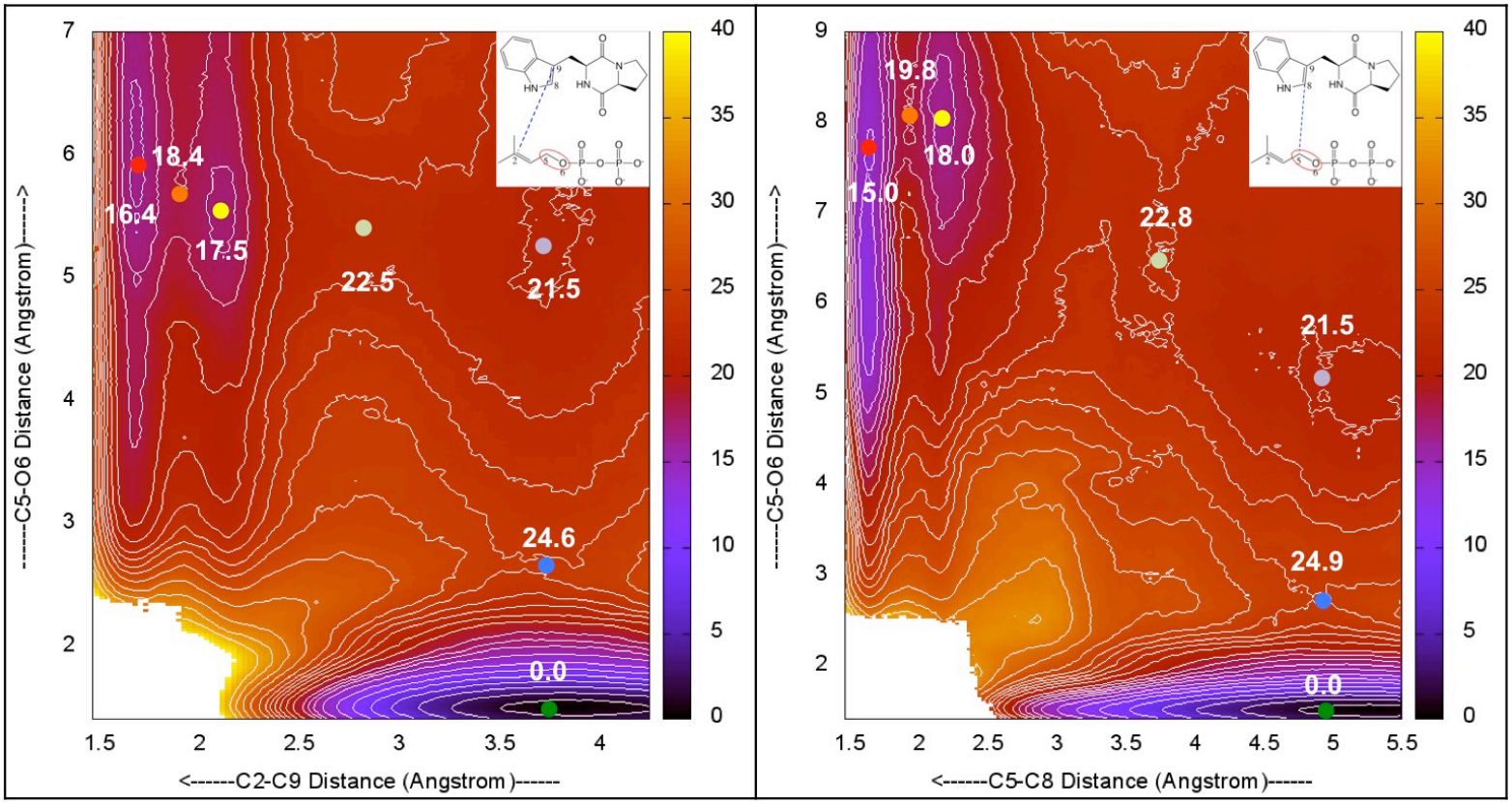


Figure 4.3 Contoured free energy profiles (FEP) in kcal/mol for (a) C2-C9 and (b) C5-C8 prenyl transfer steps in G115T mutant FtmPT1. The free energies of the starting points are both set to zero. Map colors match the energy scales shown to the right of the images. White colored values represent the free energies of the iso-contour lines of the states. Dots represent important states from left to right: green – starting state; blue – transition state I; iceblue – stationary state; lime – rotational transition state; yellow- intermediate state; orange – transition state II; red – prenylation product.

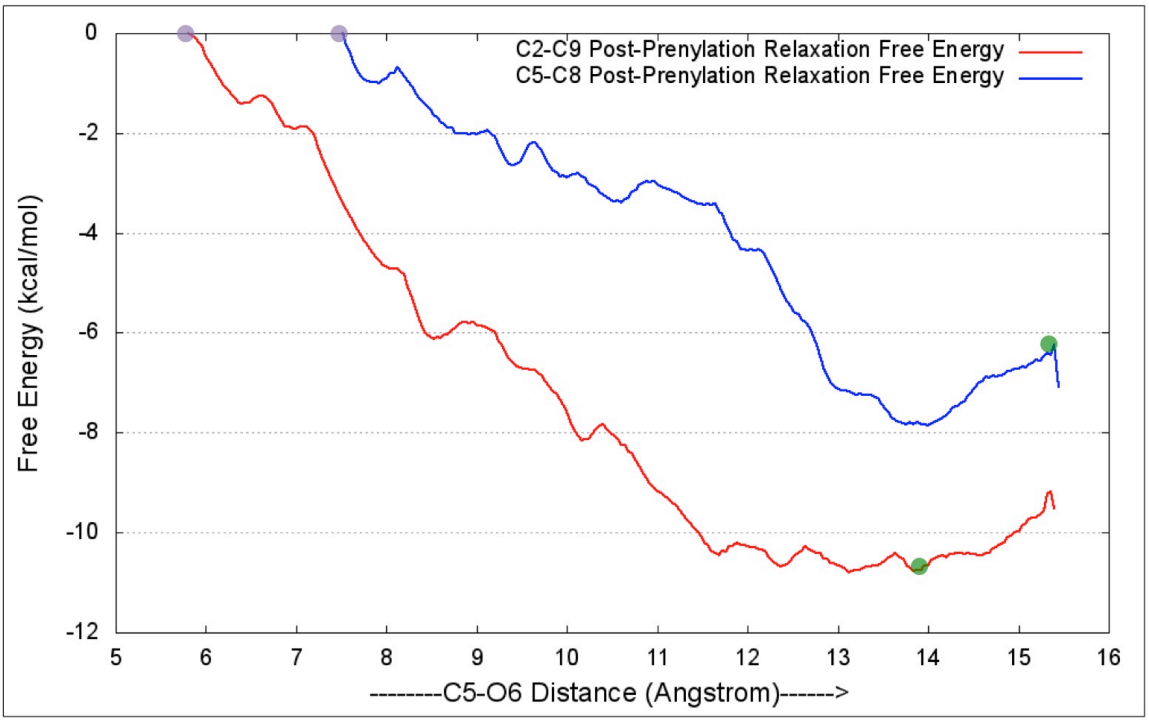


Figure 4.4 Potential of mean force calculation for relaxed free energy by MD simulation after the prenylation reaction. Red curve – C2-C9 reaction pathwy; blue curve – C5-C8 reaction pathway; purple dots – finishing points of the previous prenylation reaction; green dots – starting points of the following deprotonation reaction.

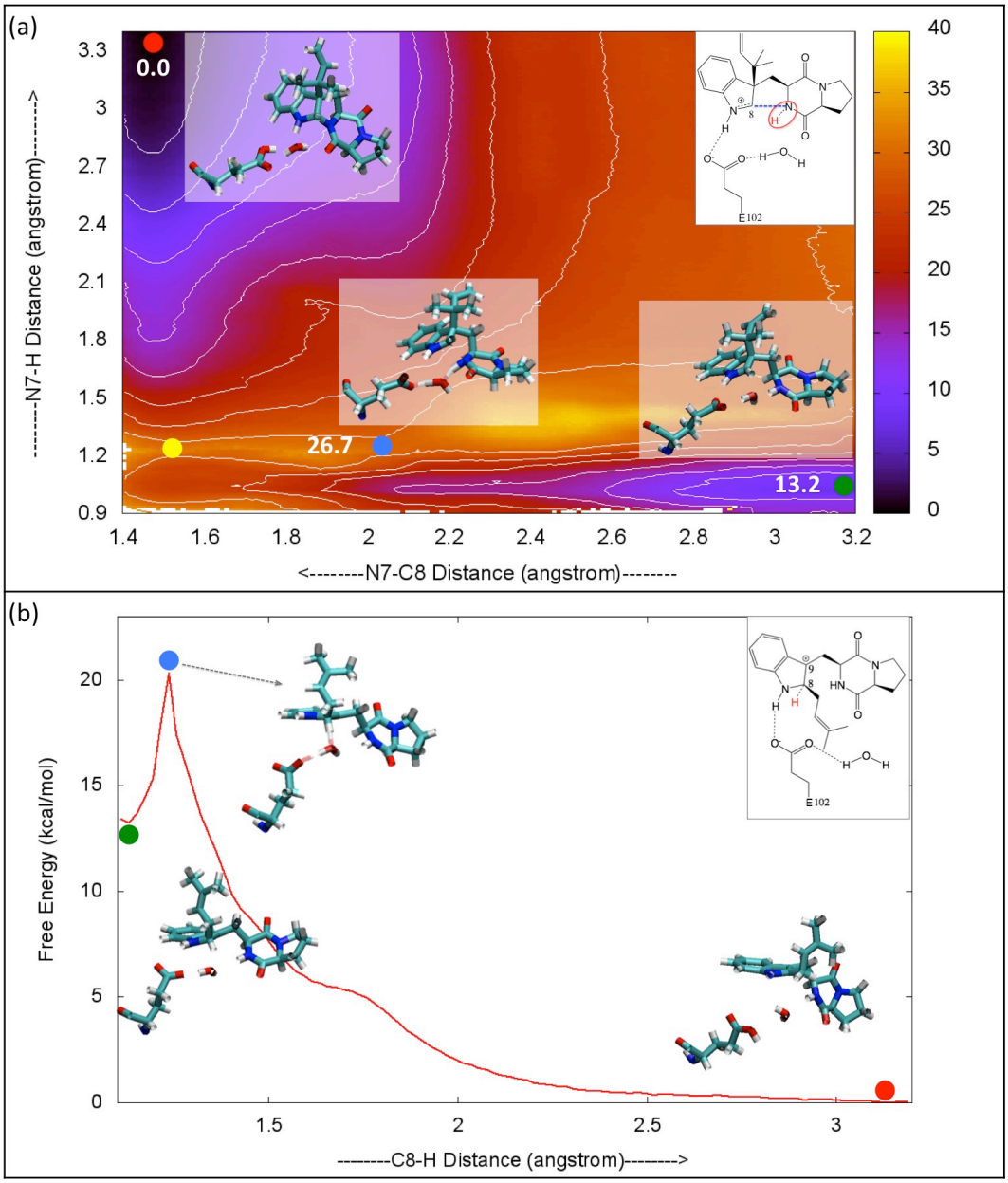


Figure 4.5 Free energy profiles in kcal/mol for (a) C2-C9 (2-dimensional) deprotonation and cyclization reaction; (b) C5-C8 (1-dimensional) deprotonation reaction to obtain final products. In both reactions, dots represent reaction states from left to right: red – starting state; blue – transition state; green – product state. Note in (a), yellow dot stands for another potential transition state; white lines represent selected iso-contoured free energies; white numbers represent the three states in this reaction.

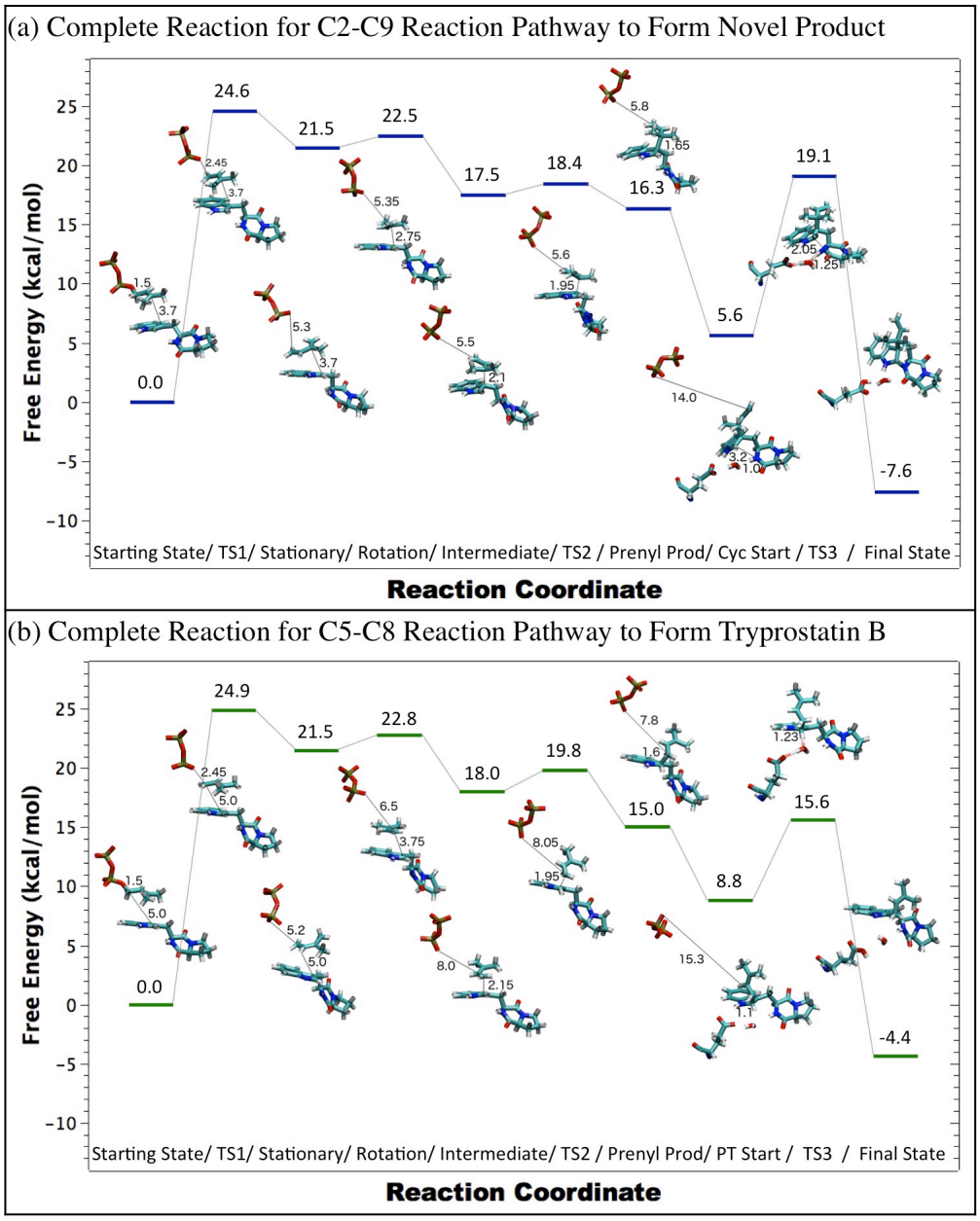
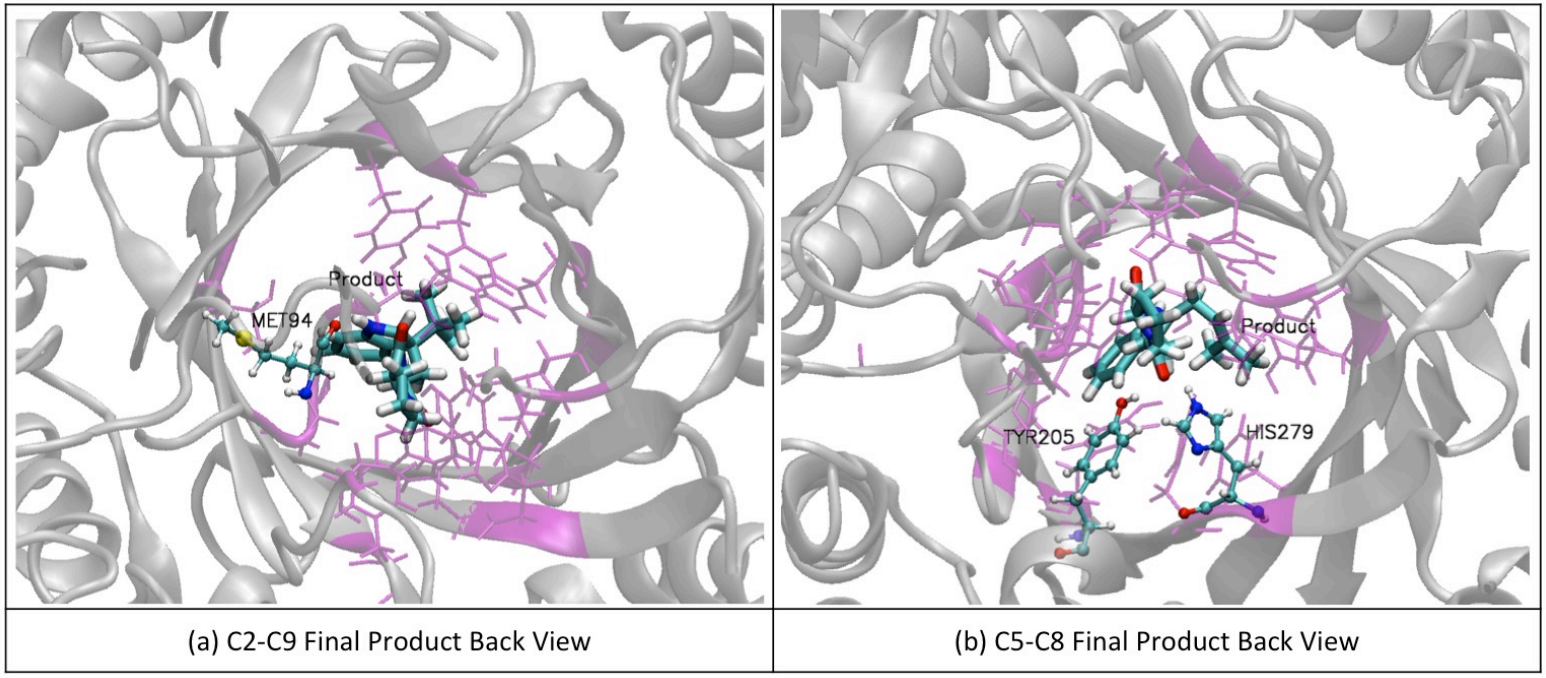


Figure 4.6 Complete reaction profiles including both prenyl transfer and proton transfer reactions associated with structures for the (a) C2-C9 and (b) C5-C8 reaction pathways. Relative free energies the reactions are given at the blue and green stages of each state. Distances of the reaction coordinates are labeled with the structures. Note the starting states of the proton transfer reactions are the relaxed states of the products for the prenyl transfer reactions.



(a) C2-C9 Final Product Back View | (b) C5-C8 Final Product Back View

Figure 4.7 Snapshots of final products after 20ns NVT MD simulation. Figures are listed both from back view (bottom view). Products are in multi-colored stick as labeled, and residues forming hydrogen bonds with the products are in multi-colored ball and stick mode as labeled. Purple colored thin stick structures are residues and water molecules within 3.0 Å distance of the final products, with their backbone ribbon conformations painted in purple color as well. The silver ribbon structures are the mutated enzymes.

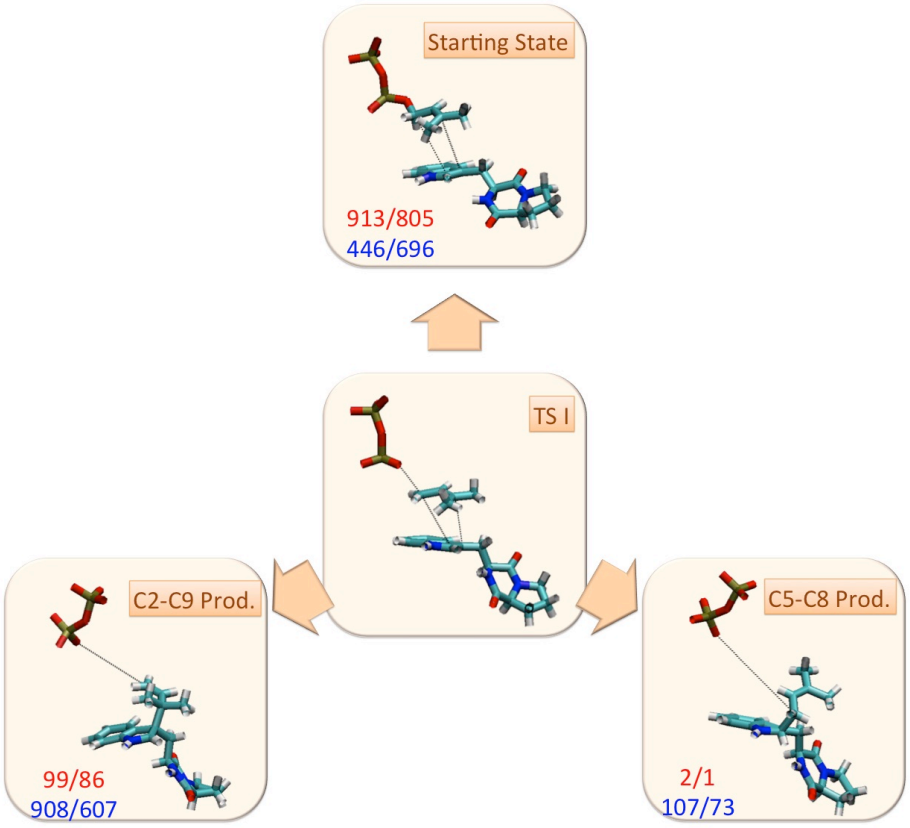


Figure 4.8 Free MD results for prenylation reaction. Here both prenylation product (top & bottom) and intermediate state (left & right) are counted. The numbers at bottom left of each structure stand for the count of the formed states from calculations of C2C9/C5C8 transition state. Red color stands for mutated system results, and blue color stands for the wild type system results, in C2C9/C5C8 forms for each.

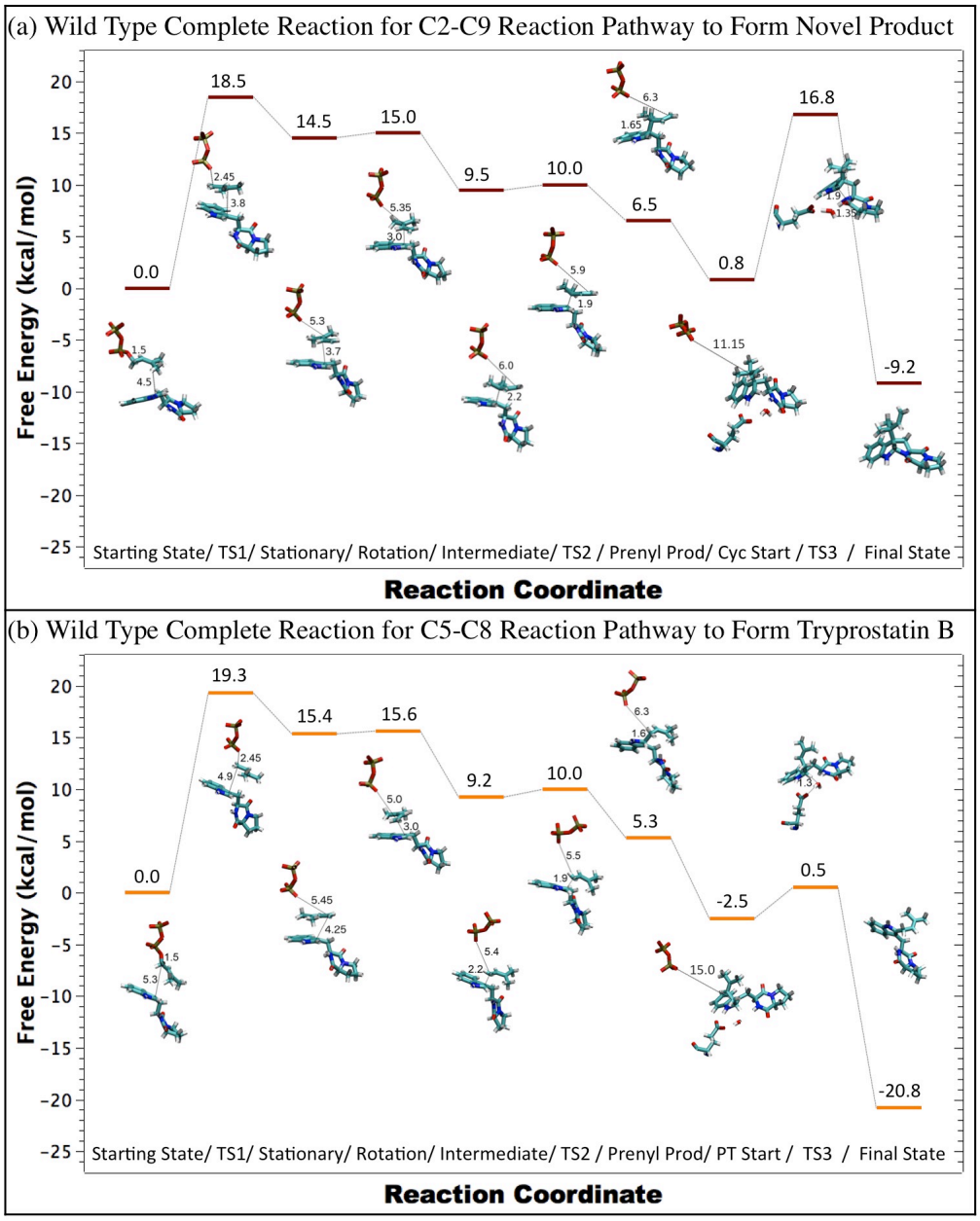


Figure 4.9 Complete reaction profiles for the wild type FtmPT1 system including both prenyl transfer and proton transfer reactions associated with structures for the (a) C2-C9 and (b) C5-C8 reaction pathways. Distances of the reaction coordinates are given with the structures. Note the starting state of the proton transfer reactions are the relaxed state of the product for the respective prenyl transfer reaction.

Table 4.1 Relative free energies (in kcal/mol) for each state in the C2-C9 and C5-C8 reaction pathways for the G115T mutated system. Note that the free energies of the starting states are set to zero.

Free Energies (kcal/mol)	Starting Point	Prenyl Transfer Step						Proton Transfer / Cyclization Step		
		Transition State I	Rotation Stationary State	Rotation Transition State	Intermediate State	Transition State II	Prenylation Product	Resting Starting State	Transition State III	Final Product
C2-C9	0.0	24.6	21.5	22.5	17.5	18.4	16.3	5.6	19.1	-7.6
C5-C8	0.0	24.9	21.5	22.8	18.0	19.8	15.0	8.8	15.6	-4.4

Table 4.2 Relative free energies (in kcal/mol) for each state in native C5-C8 and C2-C9 reaction pathways. Note that the free energies of the starting states are set to zero.

Free Energies (kcal/mol)	Starting Point	Prenyl Transfer Step						Proton Transfer / Cyclization Step		
		Transition State I	Rotation Stationary State	Rotation Transition State	Intermediate State	Transition State II	Prenylation Product	Resting Starting State	Transition State III	Final Product
C2-C9	0.0	18.5	14.5	15.0	9.5	10.0	6.5	0.8	16.8	-9.2
C5-C8	0.0	19.3	15.4	15.6	9.2	10.0	5.3	-2.5	0.5	-20.8

Table 4.3 Free MD calculation on behalf of reaction dynamics during the prenyl transfer step. Neighbored windows are set with interval of 0.1 Å. Note here the distances count for bond formations are eligible bonding distances smaller than 2.0 Å.

Reaction Pathways	Ranges of windows (Å)	MD time per window	Count of windows w/ C2-C9 Bond Formation	Count of windows w/ C5-C8 Bond Formation	Count of windows w/ C5-O6 Bond Formation
MT[1] C2-C9	C5-O6=[2.4, 2.9] C2-C9=[3.5, 3.8]	5 ps	99	2	913
MT[1] C5-C8	C5-O6=[2.4, 2.9] C5-C8=[4.8, 5.1]	5 ps	86	1	805
WT[2] C2-C9	C5-O6=[2.4, 2.9] C2-C9=[3.6, 3.9]	5 ps	908	107	446
WT[2] C5-C8	C5-O6=[2.4, 2.9] C5-C8=[4.8, 5.1]	5 ps	607	73	696

[1] MT – G115T mutated system; [2] WT – wild type system.

4.3.3 Dynamics Discussion

As Li and coauthors mentioned in their review published in 2012 for fungal indole prenyltransferases,[14] the shortened distance of the reactive substrates DMAPP and indole ring of Brevianamide F in the G115T mutated FtmPT1 could possibly lead to the observed preference of the C2-C9 reaction pathway presumably by dramatically reducing the starting C2-C9 distance for this pathway. However, this is not what I observe herein leading us to hypothesize that dynamics associated with the collapse of TS I favors the generation of the C2-C9 forming carbocation intermediate state over the C5-C8, which I test in detail below.

In order to assess the role of reaction dynamics in affecting the product choice of the C2-C9 and C5-C8 reaction pathways, I tested along these reaction coordinates, starting from the C5-O6 bond breaking transition states, and then followed the evolution of the systems as they collapsed either to the starting state or the specific carbocation intermediate states (preferring either the C2-C9 or C5-C8 product). Using an interval step size of 0.1 Å, the selected regions for the C5-O6 distances (Y-axis) were [2.4 Å, 2.9 Å], for the C2-C9 distance (X-axis) were [3.5 Å, 3.8 Å] and for the C5-C8 distance (X-axis) were [4.8 Å, 5.1 Å]. For a better comparison I also include the wild type system in the reaction dynamics investigation as well (this was not done previously[12]). Since TS I, *i.e.* O-C bond breakage, has a similar O-C and C-C distances for both mutant and wild type systems as discussed above, I used the same distance intervals for the two pathways in the wild type system. Using NVE QM/MM simulations in AMBER, I obtained two sets of data for wild type and G115T mutant system, representing both the C2-C9 and C5-C8 reaction pathways. Selecting random starting states sampling (0.1Å interval) along [2.4 Å, 2.9 Å] for C5-O6, [3.5 Å, 3.8 Å] for C2-C9 (with C5-C8 freed) and [4.8 Å, 5.1 Å] for C5-C8 (with C2-C9

freed) I generated 2400 starting states. I then released the constraints and carried out 5 ps of NVE QM/MM MD simulations for each of the four systems. The results are summarized in Table 4.3 and Figure 4.8.

I bin the trajectories as favoring C5-O6, C2-C9, and C5-C8 bond formation (the remainder could not be categorized as being clearly in any one of these states) in Table 4.3. From the table I find that for the G115T mutant system, more than half of the windows, *i.e.* 913 (C2-C9) / 805 (C5-C8) out of 2400 windows result in C5-O6 bond re-formation, *i.e.* recovery of the starting state, making it the most favored outcome among the three processes monitored. For the carbocation favoring the formation of the C5-C8 product 2 (C2-C9) / 1 (C5-C8) out of 2400 windows favored the carbocation leading to Tryprostatin B, making it the most unfavorable state; 99 (C2-C9) / 86 (C5-C8) out of 2400 windows form the carbocation favoring the C2-C9 reaction channel. Moreover, since I am releasing all degrees of freedom in these simulation all possible outcomes are possible. Therefore, in my reaction dynamics calculations, my runs can be viewed as parallel calculations since they are far from the C-C bond formation regions and, hence, can sample any of the three reaction channels. In light of this, I can sum up the two sets of runs and obtain a less ambiguous result: 1718 out of 4800 windows result in C5-O6 bond re-formation, 185 out of 4800 windows result in C2-C9 bond formation, and 3 out of 4800 widows result in C5-C8 bond formation. While the favored outcome is regeneration of the starting material, this tells us nothing about the ultimate product formation preferences. Based on the formation of the carbocation intermediates I find that the C2-C9 reaction pathway channel is favored over that of C5-C8 by a ratio of 61:1. This is because the C2-C9 distance is shorter than C5-C8 yielding more opportunities for the formation of the cyclized product. In other words the carbocation has to

travel a longer distance to arrange itself in such as way to favor the formation of the carbocation leading to C5-C8 bond formation. The ratio 61:1 for C2-C9 versus C5-C8 bond formation is close to the experimental product ratio of 40:1 (17% Brevianamide F consumed with 0.4% Tryprostatin B formed)[11], indicating that reaction pathway dynamics plays a significant role in determining the final product distribution**.** Even though the free energy barrier for the following cyclization/deprotonation step is higher for C2-C9 than that for C5-C8, the reaction dynamics strongly prefers the formation of the carbocation leading to C2-C9 in prenylation. Once at the C2-C9 carbocation a barrier of 13.5kcal/mol leads to the final product, while retracing its steps to the initially formed carbocation allowing for the possible formation of the C5-C8 carbocation costs 16.9kcal/mol. *Hence I argue that reaction dynamics is the key differentiator in the final product formation in this system*.

For comparison, I have also carried out the identical analysis on the wild type system. C5-O6 bond reformation was observed 446 (C2-C9) / 696 (C5-C8) out of 2400 windows; C5-C8 bond formation preference was observed 107 (C2-C9) / 73 (C5-C8) out of 2400 windows; and C2-C9 bond formation preference was observed 908 (C2-C9) / 607 (C5-C8) out of 2400 windows. Again, similar to the mutant system, I can combine the two reaction dynamics runs: 1142 out of 4800 windows result in C5-O6 bond re-formation, 1715 out of 4800 windows result in C2-C9 bond formation, and 180 out of 4800 windows result in C5-C8 bond formation. Interestingly, the G115T mutant had a smaller number of productive bond formation preferences when compared to the wild type system, presumably due to the introduction of the Thr115 point mutation. Moreover, in light of the reaction dynamics analysis, both the mutant and wild type show a preference of C2-C9 bond formation relative to C5-C8, due to the shorter C-C distance for

formation of the carbocation favoring C2-C9. For the native protein the ratio of C2-C9 versus C5-C8 bond formation is 9:1, which is smaller that what I observed for G115T (61:1), but still suggests that reaction dynamics has a preference for the product arising from C2-C9 bond formation. However, in the case of wild type protein thermodynamics drives the formation of the preferred C5-C8 derived product. For the C5-C8 reaction pathway the rate-determining step is C-O bond cleavage (18.5kcal/mol) and once it proceeds along this pathway each step is lower in energy with the final product being -20.8kcal/mol below the starting materials For C2-C9 the rate-determining step is C-O cleavage again (19.3kcal/mol), but it costs less energy to retrace the reaction pathway back to the first formed cation (15.3kcal/mol) that to surmount the proton transfer/cyclization transition state (16.0 kcal/mol). See Figure 4.9 and Table 4.2 for details. *Hence, I hypothesize that thermodynamics drives product formation in the native protein, while dynamics plays a bigger role in the G115T mutant.*

4.4 CONCLUSION

We have successfully carried out QM/MM calculations to examine two competitive reaction pathways for the prenylation reaction catalyzed by the FtmPT1 G115T mutant. Via thermodynamic analyses, I have found that the two pathways, *i.e.* C2-C9 and C5-C8, have very similar reaction free energy values for their respective prenylation reaction pathways. In both instances O-C bond breakage as the rate-determining step for the reaction pathway. Hence, my thermodynamic analysis predicts that the mutant should form an equal distribution of the two products arising from C5-C8 and C2-C9 prenylation, but experimentally the major product arises from the C2-C9 pathway. To further rationalize this observation I turned to free MD or reaction

dynamics studies to see if the shape of the potential energy surface governed aspects of this reaction process.

In the prenyl transfer reaction dynamics analyses, I have done free MD calculations for the systems from transition state I to the rotation stationary state. The results favor the C2-C9 prenylation product for the mutant system and indicate it as the major product choice from a reaction dynamics perspective. This observation points to how difficult it will be to *a priori* design novel catalytic agents because both thermodynamics and reaction dynamics need to be fully considered to obtain robust predictions.[28, 37, 42, 72] However, the detailed analysis described herein is very time consuming and, at least with current generation technology, makes it an inefficient way in which to design novel enzymes in a timely manner.[31, 35, 36, 73]

REFERENCES

REFENRENCES

1. Gordon, L. B.; Kleinman, M. E.; Miller, D. T.; Neuberg, D. S.; Giobbie-Hurder, A.; Gerhard-Herman, M.; Smoot, L. B.; Gordon, C. M.; Cleveland, R.; Snyder, B. D.; Fligor, B.; Bishop, W. R.; Statkevich, P.; Regen, A.; Sonis, A.; Riley, S.; Ploski, C.; Correia, A.; Quinn, N.; Ullrich, N. J.; Nazarian, A.; Liang, M. G.; Huh, S. Y.; Schwartzman, A.; Kieran, M. W., Clinical trial of a farnesyltransferase inhibitor in children with Hutchinson-Gilford progeria syndrome. *P Natl Acad Sci USA* **2012**, 109, 16666-16671.

2. Spindler, S. R.; Li, R.; Dhahbi, J. M.; Yamakawa, A.; Mote, P.; Bodmer, R.; Ocorr, K.; Williams, R. T.; Wang, Y. S.; Ablao, K. P., Statin Treatment Increases Lifespan and Improves Cardiac Health in Drosophila by Decreasing Specific Protein Prenylation. *Plos One* **2012**, 7.

3. Casey, P. J.; Seabra, M. C., Protein prenyltransferases. *J Biol Chem* **1996**, 271, 5289-5292.

4. Tello, M.; Kuzuyama, T.; Heide, L.; Noel, J. P.; Richard, S. B., The ABBA family of aromatic prenyltransferases: broadening natural product diversity. *Cell Mol Life Sci* **2008**, 65, 1459-1463.

5. Liang, P. H.; Ko, T. P.; Wang, A. H., Structure, mechanism and function of prenyltransferases. *Eur J Biochem* **2002**, 269, 3339-54.

6. Bonitz, T.; Alva, V.; Saleh, O.; Lupas, A. N.; Heide, L., Evolutionary Relationships of Microbial Aromatic Prenyltransferases. *Plos One* **2011**, 6.

7. Chen, A. J.; Kroon, P. A.; Poulter, C. D., Isoprenyl Diphosphate Synthases - Protein-Sequence Comparisons, a Phylogenetic Tree, and Predictions of Secondary Structure. *Protein Sci* **1994**, 3, 600-607.

8. Heide, L., Prenyl transfer to aromatic substrates: genetics and enzymology. *Curr Opin Chem Biol* **2009**, 13, 171-179.

9. Li, S. M., Evolution of aromatic prenyltransferases in the biosynthesis of indole derivatives. *Phytochemistry* **2009**, 70, 1746-1757.

10. Yu, X.; Xie, X. L.; Li, S. M., Substrate promiscuity of secondary metabolite enzymes: prenylation of hydroxynaphthalenes by fungal indole prenyltransferases. *Appl Microbiol Biot* **2011**, 92, 737-748.

11. Jost, M.; Zocher, G.; Tarcz, S.; Matuschek, M.; Xie, X. L.; Li, S. M.; Stehle, T., Structure-Function Analysis of an Enzymatic Prenyl Transfer Reaction Identifies a Reaction Chamber with Modifiable Specificity. *J Am Chem Soc* **2010**, 132, 17849-17858.

12. Pan, L. L.; Yang, Y.; Merz, K. M., Jr., Origin of product selectivity in a prenyl transfer reaction from the same intermediate: exploration of multiple FtmPT1-catalyzed prenyl transfer pathways. *Biochemistry-Us* **2014**, 53, 6126-38.

13. Poulter, C. D.; Rilling, H. C., Prenyl Transfer-Reaction - Enzymatic and Mechanistic Studies of 1'-4 Coupling Reaction in Terpene Biosynthetic-Pathway. *Accounts Chem Res* **1978**, 11, 307-313.

14. Wollinsky, B.; Ludwig, L.; Xie, X. L.; Li, S. M., Breaking the regioselectivity of indole prenyltransferases: identification of regular C3-prenylated hexahydropyrrolo[2,3-b]indoles as side products of the regular C2-prenyltransferase FtmPT1. *Organic & biomolecular chemistry* **2012**, 10, 9262-9270.

15. Yu, X.; Liu, Y.; Xie, X. L.; Zheng, X. D.; Li, S. M., Biochemical Characterization of Indole Prenyltransferases Filling the Last Gap of Prenylation Positions by a 5-DimethylAllyTryptophan Synthase from Aspergillus Clavatus. *J Biol Chem* **2012**, 287, 1371-1380.

16. Poulter, C. D.; Rilling, H. C., Prenyltransferase: the mechanism of the reaction. *Biochemistry-Us* **1976**, 15, 1079-83.

17. Chen, J.; Morita, H.; Wakimoto, T.; Mori, T.; Noguchi, H.; Abe, I., Prenylation of a Nonaromatic Carbon of Indolylbutenone by a Fungal Indole Prenyltransferase. *Org Lett* **2012**, 14, 3080-3083.

18. Liu, C. W.; Minami, A.; Noike, M.; Toshima, H.; Oikawa, H.; Dairi, T., Regiospecificities and Prenylation Mode Specificities of the Fungal Indole Diterpene Prenyltransferases AtmD and PaxD. *Appl Environ Microb* **2013**, 79, 7298-7304.

19. Thandavamurthy, K.; Sharma, D.; Porwal, S. K.; Ray, D.; Viswanathan, R., Regioselective Cope Rearrangement and Prenyl Transfers on Indole Scaffold Mimicking Fungal and Bacterial Dimethylallyltryptophan Synthases. *J Org Chem* **2014**, 79, 10049-10067.

20. Yin, W. B.; Yu, X.; Xie, X. L.; Li, S. M., Preparation of pyrrolo[2,3-b]indoles carrying a beta-configured reverse C3-dimethylallyl moiety by using a recombinant prenyltransferase CdpC3PT. *Organic & biomolecular chemistry* **2010**, 8, 2430-2438.

21. Yin, W. B.; Grundmann, A.; Cheng, J.; Li, S. M., Acetylaszonalenin Biosynthesis in Neosartorya fischeri IDENTIFICATION OF THE BIOSYNTHETIC GENE CLUSTER BY GENOMIC MINING AND FUNCTIONAL PROOF OF THE GENES BY BIOCHEMICAL INVESTIGATION. *J Biol Chem* **2009**, 284, 100-109.

22. Hur, S.; Bruice, T. C., The near attack conformation approach to the study of the chorismate to prephenate reaction. *P Natl Acad Sci USA* **2003**, 100, 12015-12020.

23. Strajbl, M.; Shurki, A.; Kato, M.; Warshel, A., Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. *J Am Chem Soc* **2003**, 125, 10228-10237.

24. Hong, B. C.; Hong, J. H.; Tsai, Y. C., Regio- and enantioselective prenyl anion transfer: Application to the total synthesis of (-)-rosiridol. *Angew Chem Int Edit* **1998**, 37, 468-470.

25. Cui, G.; Li, X.; Merz, K. M., Jr., Understanding the substrate selectivity and the product regioselectivity of Orf2-catalyzed aromatic prenylations. *Biochemistry-Us* **2007**, 46, 1303-11.

26. Ranaghan, K. E.; Mulholland, A. J., Conformational effects in enzyme catalysis: QM/MM free energy calculation of the 'NAC' contribution in chorismate mutase. *Chemical communications* **2004**, 1238-1239.

27. Carpenter, B. K., Trajectories through an Intermediate at a Fourfold Branch Point - Implications for the Stereochemistry of Biradical Reactions. *J Am Chem Soc* **1985**, 107, 5730-5732.

28. Carpenter, B. K., Dynamic Matching - the Cause of Inversion of Configuration in the [1,3] Sigmatropic Migration. *J Am Chem Soc* **1995**, 117, 6336-6344.

29. Carpenter, B. K., Dynamic behavior of organic reactive intermediates. *Angew Chem Int Edit* **1998**, 37, 3341-3350.

30. Carpenter, B. K., Intramolecular Dynamics for the Organic Chemist. *Accounts Chem Res* **1992**, 25, 520-528.

31. Rehbein, J.; Carpenter, B. K., Do I fully understand what controls chemical selectivity? *Phys Chem Chem Phys* **2011**, 13, 20906-20922.

32. Bolton, K.; Schlegel, H. B.; Hase, W. L.; Song, K. Y., An ab initio quasi-classical direct dynamics investigation of the F+C2H4 -> C2H3F+H product energy distributions. *Phys Chem Chem Phys* **1999**, 1, 999-1011.

33. Wu, H.; Rahman, M.; Wang, J.; Louderaj, U.; Hase, W. L.; Zhuang, Y., Higher-accuracy schemes for approximating the Hessian from electronic structure calculations in chemical dynamics simulations. *J Chem Phys* **2010**, 133.

34. Paranjothy, M.; Sun, R.; Zhuang, Y.; Hase, W. L., Direct chemical dynamics simulations: coupling of classical and quasiclassical trajectories with electronic structure theory. *Wires Comput Mol Sci* **2013**, 3, 296-316.

35. Lourderaj, U.; Park, K.; Hase, W. L., Classical trajectory simulations of post-transition state dynamics. *Int Rev Phys Chem* **2008**, 27, 361-403.

36. Zheng, J. J.; Papajak, E.; Truhlar, D. G., Phase Space Prediction of Product Branching Ratios: Canonical Competitive Nonstatistical Model. *J Am Chem Soc* **2009**, 131, 15754-15760.

37. Siebert, M. R.; Manikandan, P.; Sun, R.; Tantillo, D. J.; Hase, W. L., Gas-Phase Chemical Dynamics Simulations on the Bifurcating Pathway of the Pimaradienyl Cation Rearrangement: Role of Enzymatic Steering in Abietic Acid Biosynthesis. *J Chem Theory Comput* **2012**, 8, 1212-1222.

38. Hong, Y. J.; Tantillo, D. J., Biosynthetic consequences of multiple sequential post-transition-state bifurcations. *Nat Chem* **2014**, 6, 104-111.

39. Bolton, K.; Hase, W. L.; Doubleday, C., A QM/MM direct dynamics trajectory investigation of trimethylene decomposition in an argon bath. *J Phys Chem B* **1999**, 103, 3691-3698.

40. Biswas, B.; Collins, S. C.; Singleton, D. A., Dynamics and a Unified Understanding of Competitive [2,3]- and [1,2]-Sigmatropic Rearrangements Based on a Study of Ammonium Ylides. *J Am Chem Soc* **2014**, 136, 3740-3743.

41. Horn, B. A.; Herek, J. L.; Zewail, A. H., Retro-Diels-Alder femtosecond reaction dynamics. *J Am Chem Soc* **1996**, 118, 8755-8756.

42. Black, K.; Liu, P.; Xu, L.; Doubleday, C.; Houk, K. N., Dynamics, transition states, and timing of bond formation in Diels-Alder reactions. *P Natl Acad Sci USA* **2012**, 109, 12860-12865.

43. Jimenez-Oses, G.; Liu, P.; Matute, R. A.; Houk, K. N., Competition Between Concerted and Stepwise Dynamics in the Triplet Di-pi-Methane Rearrangement. *Angew Chem Int Edit* **2014**, 53, 8664-8667.

44. Riedel, C.; Gabizon, R.; Wilson, C. A. M.; Hamadani, K.; Tsekouras, K.; Marqusee, S.; Presse, S.; Bustamante, C., The heat released during catalytic turnover enhances the diffusion of an enzyme. *Nature* **2015**, 517, 227-U288.

45. Case, D. A.; Darden, T. A.; T.E. Cheatham, III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A., *AMBER11,University of California, San Francisco*. **2012**.

46. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983**, 79, 926.

47. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *J Am Chem Soc* **1996**, 118, 2309-2309.

48. Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field (vol 25, pg 1157, 2004). *J Comput Chem* **2005**, 26, 114-114.

49. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J Phys Chem-Us* **1993**, 97, 10269-10280.

50. Zhao, Y.; Truhlar, D. G., A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J Chem Phys* **2006**, 125.

51. Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A., Self-Consistent Molecular-Orbital Methods .20. Basis Set for Correlated Wave-Functions. *J Chem Phys* **1980**, 72, 650-654.

52. Mclean, A. D.; Chandler, G. S., Contracted Gaussian-Basis Sets for Molecular Calculations .1. 2nd Row Atoms, Z=11-18. *J Chem Phys* **1980**, 72, 5639-5648.

53. M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian 09, Revision A.02. *Gaussian, Inc., Wallingford CT* **2009**.

54. Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J Chem Phys* **1993**, 98, 10089-10092.

55. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J Chem Phys* **1995**, 103, 8577-8593.

56. Toukmaji, A.; Sagui, C.; Board, J.; Darden, T., Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions. *J Chem Phys* **2000**, 113, 10913-10927.

57. Sagui, C.; Pedersen, L. G.; Darden, T. A., Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. *J Chem Phys* **2004**, 120, 73-87.

58. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, 23, 327–341.

59. Miyamoto, S.; Kollman, P. A., SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, 13, 952-962.

60. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput Phys Commun* **1995**, 91, 1-41.

61. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J Comput Chem* **2005**, 26, 1668-1688.

62. Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G., Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys Rev B* **1998**, 58, 7260-7268.

63. http://www.ks.uiuc.edu/Research/vmd/

64. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J Mol Graph Model* **1996**, 14, 33-38.

65. http://gnuplot.info

66. Jarzynski, C., Nonequilibrium equality for free energy differences. *Phys Rev Lett* **1997**, 78, 2690-2693.

67. Hummer, G.; Szabo, A., Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *P Natl Acad Sci USA* **2001**, 98, 3658-3661.

68. Jensen, M. O.; Park, S.; Tajkhorshid, E.; Schulten, K., Energetics of glycerol conduction through aquaglyceroporin GlpF. *P Natl Acad Sci USA* **2002**, 99, 6731-6736.

69. Hummer, G.; Szabo, A., Kinetics from nonequilibrium single-molecule pulling experiments. *Biophys J* **2003**, 85, 5-15.

70. Crespo, A.; Marti, M. A.; Estrin, D. A.; Roitberg, A. E., Multiple-steering QM-MM calculation of the free energy profile in chorismate mutase. *J Am Chem Soc* **2005**, 127, 6940-6941.

71. Grossfield, A., WHAM: an implementation of the weighted histogram analysis method. http://membrane.urmc.rochester.edu/content/wham version 2.0.8.

72. Baker, D.; Agard, D. A., Kinetics Versus Thermodynamics in Protein-Folding. *Biochemistry-Us* **1994**, 33, 7505-7509.

73. Pearson, A. D.; Mills, J. H.; Song, Y. F.; Nasertorabi, F.; Han, G. W.; Baker, D.; Stevens, R. C.; Schultz, P. G., Trapping a transition state in a computationally designed protein bottle. *Science* **2015**, 347, 863-867.

CHAPTER 5. A FREE ENERGY BASED CONFORMATIONAL SEARCH ALGORITHM USING THE "MOVABLE TYPE" SAMPLING METHOD

5.1 INTRODUTION

The conformational search problem has been of interest over the past several decades for the prediction of the conformations of "small" molecules.[1-3] The molecular conformers generated using these methods are useful for virtual screening,[4-6] 3D-QSAR,[7] pharmacophore modeling,[8-10] protein-ligand binding,[2, 11] *etc*.. In light of the importance of conformational searching, a number of algorithms have been developed over the past twenty years.[1, 3, 12-32] These algorithms efficiently generate realistic molecular conformers according to predefined sets of criteria.

In general, a conformational searching algorithm uses either a systematic or stochastic searching strategy.[33] The systematic search approach exhaustively explores regular variations of the torsion angles.[15, 16, 19, 25] Each systematically generated conformation has its single point energy calculated typically with a force field. One drawback of systematic searching is the number of variables dramatically increases with the number of atoms,[33, 34] thereby, increasing the cost of conformation generation. In a stochastic search algorithm,[12, 13] conformations are randomly generated and then minimized using an energy function. Structure generation is terminated once the desired number of structures is generated, the program reaches a predefined stopping point or all conformers are sampled. Currently, Genetic Algorithms,[21, 25, 35-39] Monte Carlo Simulated Annealing[18, 20, 39, 40] and distance geometry[15, 31, 41] methods are three widely used stochastic searching methods. The stochastic search method offers significant computational efficiency when compared to traditional systematic search methods. However, the global minimum energy conformer cannot be guaranteed to be identified using stochastic searching methods. Moreover,

each conformer generated has to be energy minimized using classical or quantum mechanical (QM) methods which adds additional computational expense to the procedure.[42, 43]

Hence, locating the key minima (preferably the free energy minima) efficiently while covering a variety of energy states in the output conformations is a challenging task for any conformational searching method. Inspired by a novel energy sampling method, the "Movable Type" (MT) sampling method[44, 45] devised in my group, I have developed a conformation generation algorithm, the MT conformational search algorithm (MT-CS). The terminology of "Movable Type" is traced back to an ancient Chinese printing method later pioneered by Gutenberg in the western world. The MT algorithm collects and combines different atom pairwise distances according to a pre-built database of atom pairwise energies as a function of distance (r), in terms of bonds, angles, dihedrals, and non-bonded interactions. In the MT-CS method, for simplicity, and due to the limited distance variation, the bond and angle distances for all atom pair types are fixed to their corresponding energy minimum values. Based on their relative strengths, different torsion and non-bonded interactions are given different weights in the selection of distance values during the course of a MT-CS calculation. The MT-CS algorithm translates all collected atom pairwise distances into Cartesian coordinates to yield the final conformations. A standard MT sampling calculation[44] is then performed against each generated conformer to collect an energy ensemble of the conformers. Overall, using the MT-CS approach relevant conformers can be generated rapidly (0.002 sec per conformer) and, importantly, on the gas-phase free energy surface.

Besides exploring the gas-phase free energy surface for a given molecule, I also want to obtain the solvation free energy as well to understand conformational preferences in aqueous solution.[46-50] The further addition of the solvation free energy using the MT method[51] offers another level of sophistication in my conformer generation modeling. Hence, the MT based conformational search approach offers notable differences over traditional conformer generation approaches.

Our study focuses on the molecular conformation generation of "drug-like" small molecules on the corresponding free energy (gas and aqueous phase) surface. The MT-CS approach is validated against a variety of "drug-like" compound structures via the examination of the match between the generated conformations with crystal structure conformations, either in the free or protein-bound state. Two validation benchmarks were studied herein. (1) Evalidation of the energy/free energy surface calculations was accomplished using nine thrombin inhibitors[52, 53] selected from the work of Klebe's laboratory.[54] For the purpose of a more thorough analysis, MT-CS generated conformations are compared to *ab initio* optimized structures.[55] (2) Validation of the robustness of the conformer generation of MT-CS was carried out by examining another 299 organic compounds from the Cambridge Structural Database (CSD).[56, 57] The major difference between the PDB and CSD compounds is that the CSD compounds are unbound ligands, while the PDB compounds were all protein-bound. In all cases the gas-phase free energy surface was generated and then the individual solvation free energies of each conformer were added to arrive at the predicted global minimum.

5.2 METHODS

5.2.1 The Movable Type Sampling Algorithm

The Movable Type (MT) sampling method utilizes Monte Carlo integration to estimate the canonical partition function. A molecular energy is sampled by the sum of all the atom pairwise energies within the canonical ensemble. The sampling space is created by allowing random matching of the atom pairwise energies as a function of distance (r) within certain cutoff or "blurring" ranges, thereby, sampling the region of interest in order to assemble and estimate the canonical ensemble. A database is first constructed containing interaction energies between all classes of atom pairs found in the chemical space under investigation as a function of distance. The molecular energy is then calculated simply by assembling atom pairwise energies selected from the database. This database can be generated from any pairwise potential, but, herein, I use the knowledge or statistical potential KECSA developed in my group.[45]

Formally, the MT sampling method generates extremely unwieldy matrices describing the molecular interactions, but in order to keep the sampling size at a manageable level and to enhance computational speed, I introduced a scramble/tiling strategy for the atom pairwise energy assembly. First the vector of each atom pairwise interaction as a function of r is selected from the energy database and the energy cutoff or "blurring" distance is then applied to this vector. This can result in extensive sampling of the vector (with a long cutoff of ~6Å) or a local level of sampling (with a cutoff of 1Å). The resultant matrix is then scrambled according to the distance distributions. An example of a scrambled atom pairwise energy vector is illustrated in equation 1.

$$\not{z}_\alpha = \begin{bmatrix} z_\alpha(r_1) \\ z_\alpha(r_2) \\ z_\alpha(r_3) \\ \vdots \\ z_\alpha(r_n) \end{bmatrix} = \begin{bmatrix} e^{-\beta E_\alpha(r_1)} \\ e^{-\beta E_\alpha(r_2)} \\ e^{-\beta E_\alpha(r_3)} \\ \vdots \\ e^{-\beta E_\alpha(r_n)} \end{bmatrix} \Rightarrow scramble\left(\not{z}_\alpha\right) = \begin{bmatrix} z_\alpha(r_{n-2}) \\ z_\alpha(r_3) \\ z_\alpha(r_1) \\ \vdots \\ z_\alpha(r_2) \end{bmatrix} \tag{1}$$

The vector goes through multiple different random scramble steps and is then tiled into a fixed size matrix, the Z-matrix (shown in equation 2). In the MT calculation, identically sized Z-matrices are established for all atom pairwise energies in the molecular system under investigation. The molecular energy assembly is performed by point-wise multiplication through all the Z-matrices in the molecular system (shown in equation 3).

$$\not{Z}_\alpha = \underbrace{\left.\begin{bmatrix} scramble(\not{z}_\alpha)_1 & scramble(\not{z}_\alpha)_{i+1} & \cdots & scramble(\not{z}_\alpha)_{n-i+1} \\ scramble(\not{z}_\alpha)_2 & scramble(\not{z}_\alpha)_{i+2} & & scramble(\not{z}_\alpha)_{n-i+2} \\ \vdots & \vdots & \ddots & \vdots \\ scramble(\not{z}_\alpha)_i & scramble(\not{z}_\alpha)_{2i} & \cdots & scramble(\not{z}_\alpha)_n \end{bmatrix}\right\} m \text{ rows}}_{n \text{ columns}} \tag{2}$$

$$\not{Z}_M = \not{Z}_1 \bullet \not{Z}_2 \bullet \cdots \bullet \not{Z}_n \tag{3}$$

The random disordered permutations to each Z-matrix is meant to maximize the variety of energy combinations at different distances, and the fixed-size matrix multiplication is to maintain a computationally acceptable sampling size. The row number *m* of the Z-matrices is defined as the least multiplier of all the atom pairwise vector sizes in the molecular system under study, in

order to fit all atom pairwise energy vectors with different distance distribution ranges, while the column number *n* is a user defined number defining the sampling size. Hence the final Z-matrix after assembly includes *m*×*n* molecular conformational energies.

In the Z-matrix, atom pairwise energies at different distances are assembled to simultaneously represent different energy states in the canonical ensemble. The free energy is then calculated by incorporating the average of the Boltzmann factors into:

$$A = -RT \ln\left[Z_M\right] \approx -RT \ln\left[V_M \left\langle e^{-\beta E_M(\tau)} \right\rangle\right] \tag{4}$$

where $E_M(\tau)$ is the molecular energy as a function of the geometric variable τ and V_M indicates the sampling volume.

The free energy is computed directly from the NVT ensemble avoiding issues related to the additivity of the free energy.[54] In particular, I assemble the interaction energies using equation 3 and then place this into equation 4 to directly compute the free energy, thereby, affording the ability to avoid issues related to the decomposition of the free energy into entropic and enthalpic components. This is a real advantage of the MT method relative to other conformational sampling approaches.

The MT energy sampling method incorporates my newly developed implicit water model for the solvation free energy calculation.[51] In this model the water molecules are modeled as isotropic rigid balls with van der Waals radii of 1.6 Å and placed into isometric solute-surrounding solvent

layers. The solute-solvent interaction sampling starts from the solute's water accessible surface until 8 Å away from the solute's van der Waals surface using an increment of 0.005 Å per layer. The number of water molecules was limited by comparing their maximum cross-sectional areas with the solvent accessible surface area at each solvent layer for each atom in the solute. The number of water molecules (N_w) accessible to each atom at distance R away from the atomic center of mass is rounded down using the maximum cross-sectional area (S_w) of water and the atomic solvent accessible surface area (S_a) in the solvent layer at distance R.

$$N_w(r) = \text{floor}\left(\frac{S_a(r)}{S_w}\right) \tag{5}$$

The maximum cross-sectional areas (S_w) of a water molecule is calculated as:

$$\begin{aligned} S_w &= \int_{\frac{\pi}{2}-\theta}^{\frac{\pi}{2}} 2\pi\left(R_a + R_w\right) R_w \sin\left(\frac{\pi-\theta}{2}\right) d\left(\frac{\pi}{2}-\theta\right) \\ &= 2\pi\left(R_a + R_w\right) R_w \cos\left(\frac{\pi-\theta}{2}\right) \end{aligned} \tag{6}$$

where R_w and R_a are the van der Waals radii for water and the atom in the solute molecule respectively.

The Boltzmann factor matrix for the kth solute atom-water ($\mathcal{Z}_k^{A-S}$) interaction is defined as a Boltzmann weighted solute atom-water energy multiplied by the number of accessible water molecules at the different distances. Multiplication of the Z-matrices for all solute atom-water

interactions composes the final solute molecule-water Z-matrix ($\mathcal{Z}_{total}^{L-S}$), which when multiplied by the Z-matrix for the intra-solute molecular interactions ($\mathcal{Z}_{total}^{L}$) derives the final Z-matrix for the solute-solvent complex system ($\mathcal{Z}_{total}^{LS}$). Multiplication of the final Z-matrix with its corresponding normalized Q-matrix generates the Boltzmann-weighted energy ensemble ($\mathcal{C}_{total}^{LS}$). With the energy ensembles for the solute molecule ($\mathcal{C}_{total}^{L}$) and solute-solvent complex ($\mathcal{C}_{total}^{LS}$), the solvation free energy is calculated using equation 10.

$$\mathcal{Z}_{k}^{A-S}=\begin{bmatrix} e^{-\beta E_k^{A-S}(r_1)N_w(r_1)} & e^{-\beta E_k^{A-S}(r_{i+1})N_w(r_{i+1})} & \cdots & e^{-\beta E_k^{A-S}(r_{n-i+1})N_w(r_{n-i+1})} \\ e^{-\beta E_k^{A-S}(r_2)N_w(r_2)} & e^{-\beta E_k^{A-S}(r_{i+2})N_w(r_{i+2})} & \cdots & e^{-\beta E_k^{A-S}(r_{n-i+2})N_w(r_{n-i+2})} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\beta E_k^{A-S}(r_i)N_w(r_i)} & e^{-\beta E_k^{A-S}(r_j)N_w(r_j)} & \cdots & e^{-\beta E_k^{A-S}(r_n)N_w(r_n)} \end{bmatrix} \tag{7}$$

$$\mathcal{Z}_{total}^{L-S}=disorder\left(\mathcal{Z}_{1}^{A-S}\right)\bullet disorder\left(\mathcal{Z}_{2}^{A-S}\right)\bullet\cdots\bullet disorder\left(\mathcal{Z}_{n}^{A-S}\right) \tag{8}$$

$$\mathcal{C}_{total}^{LS}=\bar{\mathcal{Q}}_{total}^{LS}\bullet\mathcal{Z}_{total}^{LS}=\bar{\mathcal{Q}}_{total}^{L-S}\bullet\bar{\mathcal{Q}}_{total}^{L}\bullet\mathcal{Z}_{total}^{L-S}\bullet\mathcal{Z}_{total}^{L} \tag{9}$$

$$\begin{aligned}\Delta G_{solv}^{L}\approx -RT\ln\left[\frac{Z_{LS}}{Z_{L}}\right]=-RT\ln\left[\frac{\left\langle e^{-\beta E_{LS}(r)}\right\rangle}{\left\langle e^{-\beta E_{L}(r)}\right\rangle}\right]= \\ -RT\ln\left[\frac{\mathrm{sum}\left(\mathcal{C}_{total}^{LS}\right)}{\mathrm{sum}\left(\mathcal{C}_{total}^{L}\right)}\right]=-RT\ln\left[\frac{\mathrm{sum}\left(\bar{\mathcal{Q}}_{total}^{LS}\bullet\mathcal{Z}_{total}^{LS}\right)}{\mathrm{sum}\left(\bar{\mathcal{Q}}_{total}^{L}\bullet\mathcal{Z}_{total}^{L}\right)}\right]\end{aligned} \tag{10}$$

5.2.2 Generating Conformers

5.2.2.1 Database preparation for bonds, angles, torsion angles, and non-covalent interactions

There is one database used by the MT sampling algorithm: the pairwise distance dependent Boltzmann energies. All bonds and angles were fixed at the distances associated with their maximum energies (*i.e., s*trongest bond or angle preference), while for the torsion angles distances were selected based on the Boltzmann energy curves as a function of distance. Non-covalent interactions were evaluated after conformation generation via the torsion angles and ring structures were treated as fixed units.

5.2.2.2 Torsion angle driven 3D structure generation

The MT conformational search program requires an initial *mol2* input file with complete atom connection information. Torsion angle parameters for every rotatable bond were used to generate the different conformations, according to their local maximum energies. The structure generation started with the first atom placed at a fixed position and the molecule was built from there by adding an atom (or a fixed ring unit) to the system at a time. The molecular construction details are given in Figure 5.1. In this way an ensemble of molecule conformers were generated that represented the conformational space favored by the torsion angles. The torsion energies were then recorded respectively for each conformation. Simultaneously, all bond and angle energies were also recorded for future reference.

5.2.2.3 Filtering Conformers and Free Energy Determination

The initial ensemble of conformations was further screened to eliminate those with internal clashes. Each member of the final conformer ensemble had its associated free energy computed using the full set of Boltzmann factors (Equation 4). The conformation that had the lowest gas-phase free energy was selected as the reference (zero free energy) to obtain the relative gas-phase free energies for the remaining conformations in the ensemble.

Each member of the resultant ensemble was then solvated and had its solvation free energy determined using KMTISM.[51] Adding these free energies to the gas-phase free energies, yields the relative aqueous phase free energy. Note that both the gas and aqueous phase free energies are relative values, while the solvation free energy is an absolute value.

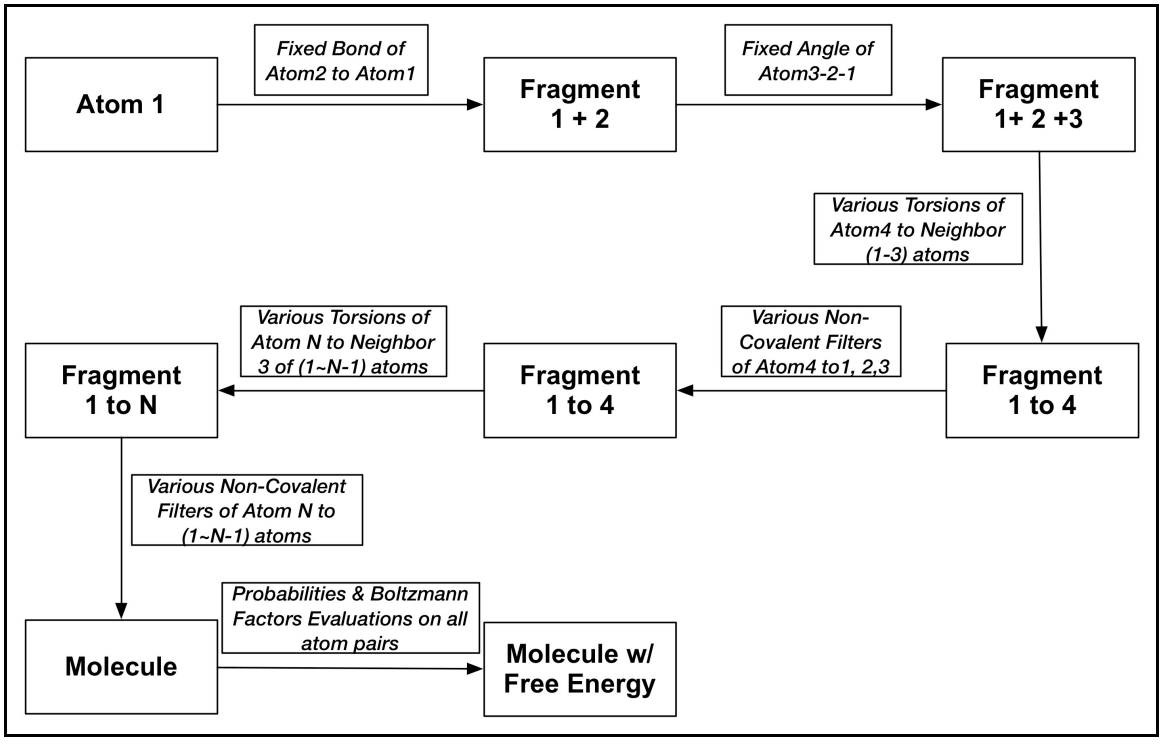


Figure 5.1 Scheme for conformer growing flow using MT conformational search algorithm.

5.2.3 Small Molecules from the CSD & PDB

We selected 299 molecules from the CSD, which represented a culled subset (299 out of 492) of molecules examined in a validation study of OMEGA.[19] The CSD ID codes for the selected molecules are given in the SI. The 299 molecules were selected to have R-factors smaller than 10.0, with at least 4 rotatable bonds and a heavy atom count of larger than 10. Furthermore, I also examined 104 molecules out of 197 PDB ligands explored for validation previously. For each molecule thousands to tens of thousands of conformers were generated depending on the exact nature of the individual molecules. The resultant conformations were then analyzed relative to the original CSD and PDB structure.

5.2.4 Thrombin Ligands

Besides the 299 CSD and 104 PDB structures, 9 thrombin ligands from Klebe's work were examined in more detail (PDB ID's 2ZFP, 2ZGB, 2ZC9, 2ZI2, 2ZIQ, 2ZHQ, 2ZGX, 2ZNK, and 2ZDA). Moreover, I picked the 2ZI2 thrombin ligand (3c) out of the 9 total ligands to compare to an OMEGA conformational search and to further calculate the conformer energies at the M06-2X/6-31G* level of theory (using Gaussian 09[55]). The resultant global minimum and local energy minima were compared to the conformations generated by MT conformational search both structurally and thermodynamically. The absolute free energies for the DFT optimized structures was obtained from a vibrational analysis carried out using the Gaussian 09 package.

We also carried out MM-PB/GBSA calculations on the nine thrombin ligands to obtain their solvation free energies. The nine ligand structures stripped from their respective PDB files were optimized at the M06-2X/6-31G* in Gaussian09 prior to a 5ns MD simulation using AMBER12.

Trajectory data points were collected every 1ps, resulting in 5000 frames for later analysis. These 5000 frames were put into the MM-PBSA.py program[58] in AMBERTools for post-MD analysis. The solvation free energies using both the linearized Poisson Boltzmann (MM-PB/SA)[59, 60] and Generalized Born (MM-GB/SA)[61] models were calculated and stored for later comparison to MT results.

The MT conformational sampling code was written in Python (version 2.7) and the Visual Molecular Dynamics (VMD)[62, 63] program was used for structure analysis. GNUPlot[64] was used for all plots.

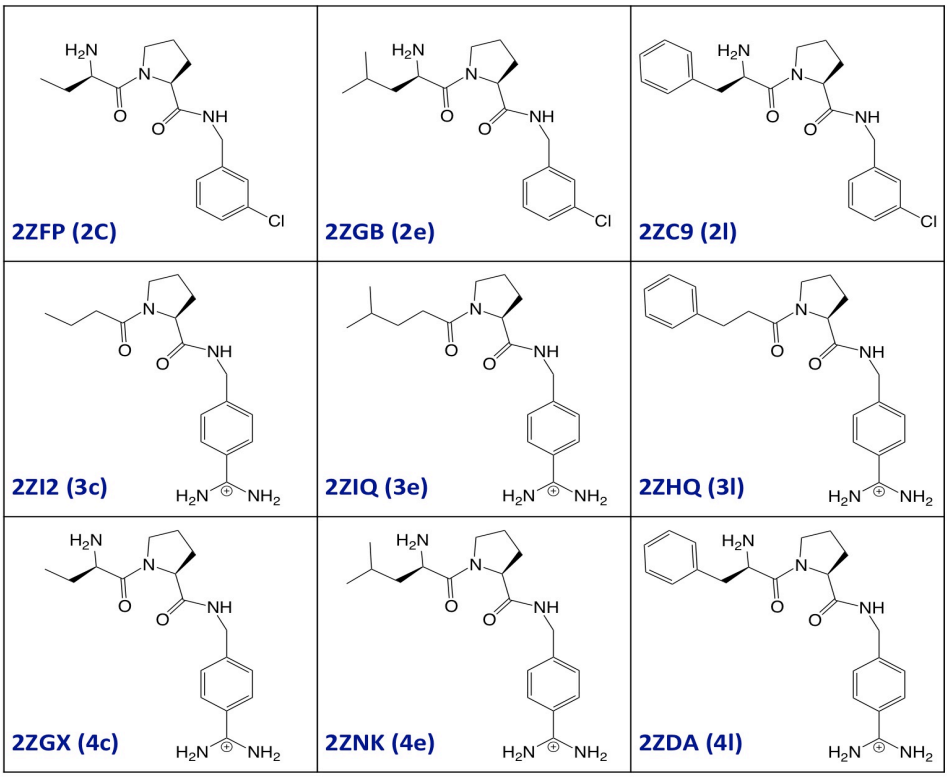


Figure 5.2 Thrombin ligands from the work of Klebe and co-workers.[54] For 2c, 2e, 2l and 4c, 4e, 4l the $-NH_2$ group at low pHs would become $-NH_3^+$.

5.3 RESULTS AND DISCUSSION

We generated over 1,458,000 total conformers for the 299 CSD and 9 thrombin structures according to the criteria delineated in the methods section. The conformational ensembles were compared to their original crystal structures and in the case of one thrombin inhibitor to M06-2X/6-31G* structures.

5.3.1 Structure Validation of the Thrombin Ligands

The thrombin ligands represented a typical challenge for conformational searching programs. In generating these conformational ensembles one must overcome several challenges including chirality, ring formation, and intramolecular hydrogen bonding. Below I discuss, in detail, structural RMSDs, free energies and the global minimum structure generated using MT sampling for 3c (PDB ID 2ZI2) as a representative example.

5.3.1.1 The Conformational Ensemble

We generated over 30,000 conformers for the 9 thrombin inhibitors that were structurally and thermodynamically characterized by Klebe and co-workers (see Figure 5.2; I have retained their original numbering).[54]

The small molecules in the thrombin ligand set have between 8-11 rotatable bonds that need to be considered in the MT conformational search. Every single torsion angle (rotatable bond) has 2-4 local maximum energy values from which I can set the torsion angle. Thus, this could lead to $2*2^8$ - $2*4^{11}$ total conformations, which is reduced to 1800 ~ 31,000 because of non-physical

close-contacts, which were eliminated. my choice of the non-bonded distance criteria (~2.8Å) eliminates unrealistic MT conformations.

Chirality considerations also reduce the conformational space examined and this is also true for the thrombin ligands. These ligands all have a carbon atom with the S configuration, which connects the amide to the five member ring (see Figure 5.2). This S-chirality eliminates all of the conformations with R stereochemistry. Furthermore, 6 ligands (2c, 2e, 2l and 4c, 4e, 4l) have a carbon atom with R stereochemistry where the amino (ammonium) group binds (see Figure 5.2), which eliminates all conformations with S stereochemistry. Overall, through various cutoffs and stereochemical considerations the resultant sizes of the thrombin inhibitors ensembles were of manageable size (see Table 5.1).

3c (2ZI2) is a typical ligand among the 9 thrombin ligands. It has one five-member ring with one S-chiral carbon atom, one potential intramolecular hydrogen bond, and one benzamidine functional group. The 10 rotatable bonds have 3~5 energy maxima which result in between $(3*2)^{10} \sim (5*2)^{10}$ unique conformations (each choice for a rotatable bond generates two positions for the targeted atom), but this is reduced to 7201 ensemble members by my filters (see Table 5.1). All conformations were generated within 2 minutes, including their gas phase free energies, and RMSDs. The subsequent solvation calculations take ~3sec per conformation on a laptop computer.

5.3.1.2 Structure Validation

Table 5.1 shows the RMSD results for the conformations generated for the 9 ligands compared to their corresponding protein-bound PDB structures. Here I provide not only the minimum RMSDs in the conformational ensembles, but also the RMSDs for the lowest free energy structures, both in the gas and aqueous phase. From the table I can see that the MT conformational search approach yields geometries with average RMSDs between 1.521 ~ 2.094 Å, and minimum RMSDs between 0.660 ~ 0.990 Å, respectively. Figure 5.3 summarizes the minimum RMSD conformations (pink), the lowest gas phase free energy conformations (blue), as well as the lowest aqueous phase free energy conformations (orange), with all structures aligned to the corresponding PDB crystal structure (multi-colored). The RMSDs for the aqueous phase and gas phase lowest free energy conformers are also listed in the table, with their values ranging from 1.128 to 2.794 Å for the former and from 1.656 to 2.644 Å for the latter. The increased RMSDs for the gas and aqueous phase structures arise because these are free structures rather that protein-bound. The detailed RMSD curves are shown in Figure 5.4.

Figure 5.4 shows the RMSD curves for 3c associated with its three key conformers. From this figure I observe that the RMSD range for 3c is 0.866 ~ 2.679 Å. The minimum RMSD conformation has a 0.866 Å RMSD value and the minimum gas phase structure is 1.985 Å, with the main geometric differences being found in the five-member ring. The aqueous phase structure has a 2.300 Å RMSD and differs mostly in the region connecting the five and six-membered rings. The lower right corner of this figure gives an overlay of the four different structures.

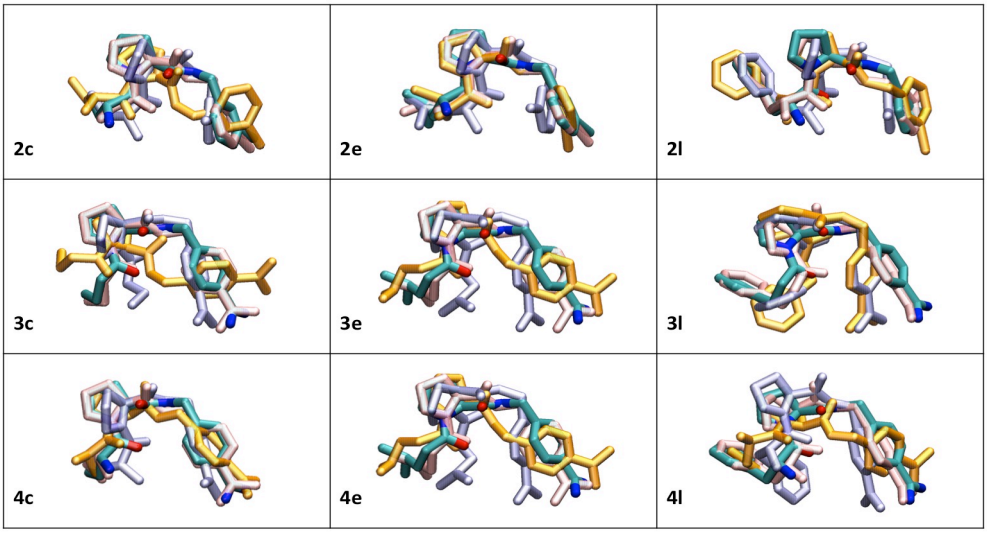


Figure 5.3 Geometry comparisons for the 9 thrombin ligands. Multi-color – PDB structures; pink – minimum RMSD conformation; blue – lowest gas phase free energy conformation; orange – lowest aqueous phase free energy conformation. All conformations are aligned to their corresponding PDB structures.

Table 5.1 MT Conformational Ensembles and the Associated RMSDs (in Å) for the Thrombin ligands of Figure 5.2.

	Conformer Amount	Mean RMSD	Min. RMSD	1st Aq. Free E. RMSD	1st Gas Free E. RMSD
2c	2010	1.521	0.660	1.613	2.165
2e	2132	1.604	0.746	1.128	2.248
2l	6643	1.803	0.738	1.575	2.156
3c	7201	1.723	0.866	2.300	1.985
3e	18728	1.821	0.990	1.578	2.628
3l	31747	2.094	0.945	2.794	2.644
4c	1826	1.663	0.904	1.231	1.656
4e	1950	1.660	0.940	1.804	1.767
4l	6079	1.909	0.906	1.985	2.452

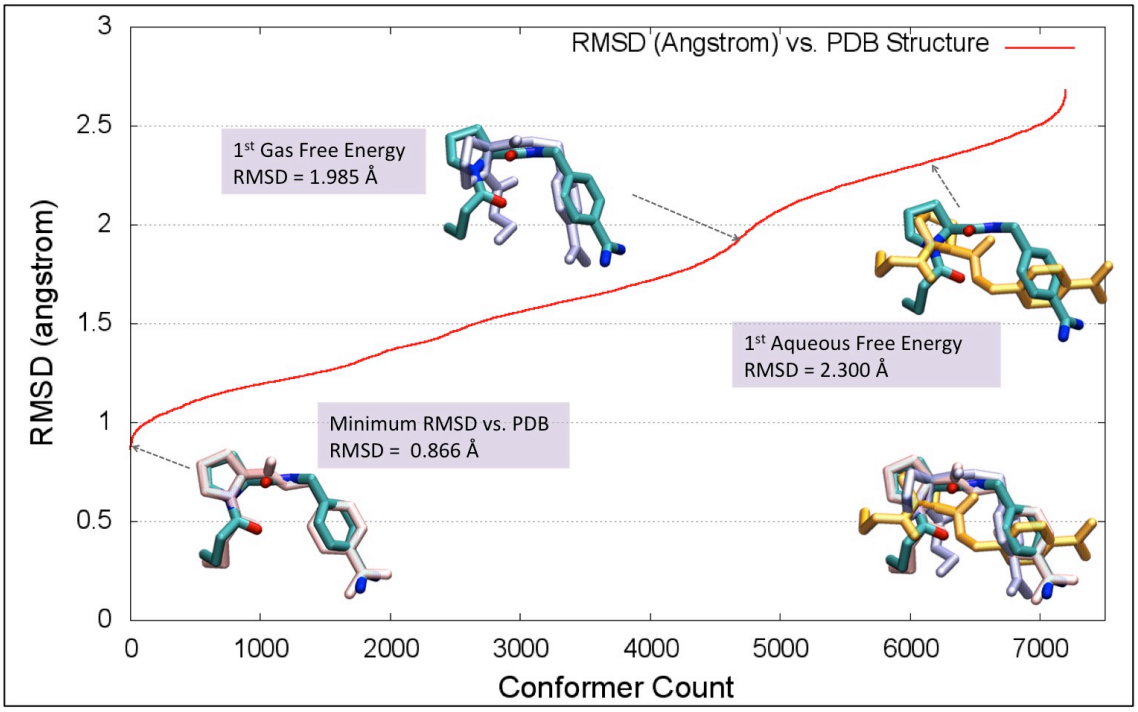


Figure 5.4 RMSD ranges for thrombin ligand 3c. The purple shadowed text identifies the three key conformations and their RMSD values. Here the multi-colored molecule is the PDB structure and the single colored molecules are MT generated conformations: purple – minimum RMSD conformation; blue – lowest gas phase free energy conformation; orange – lowest aqueous phase free energy conformation. All conformations are aligned to the PDB structure.

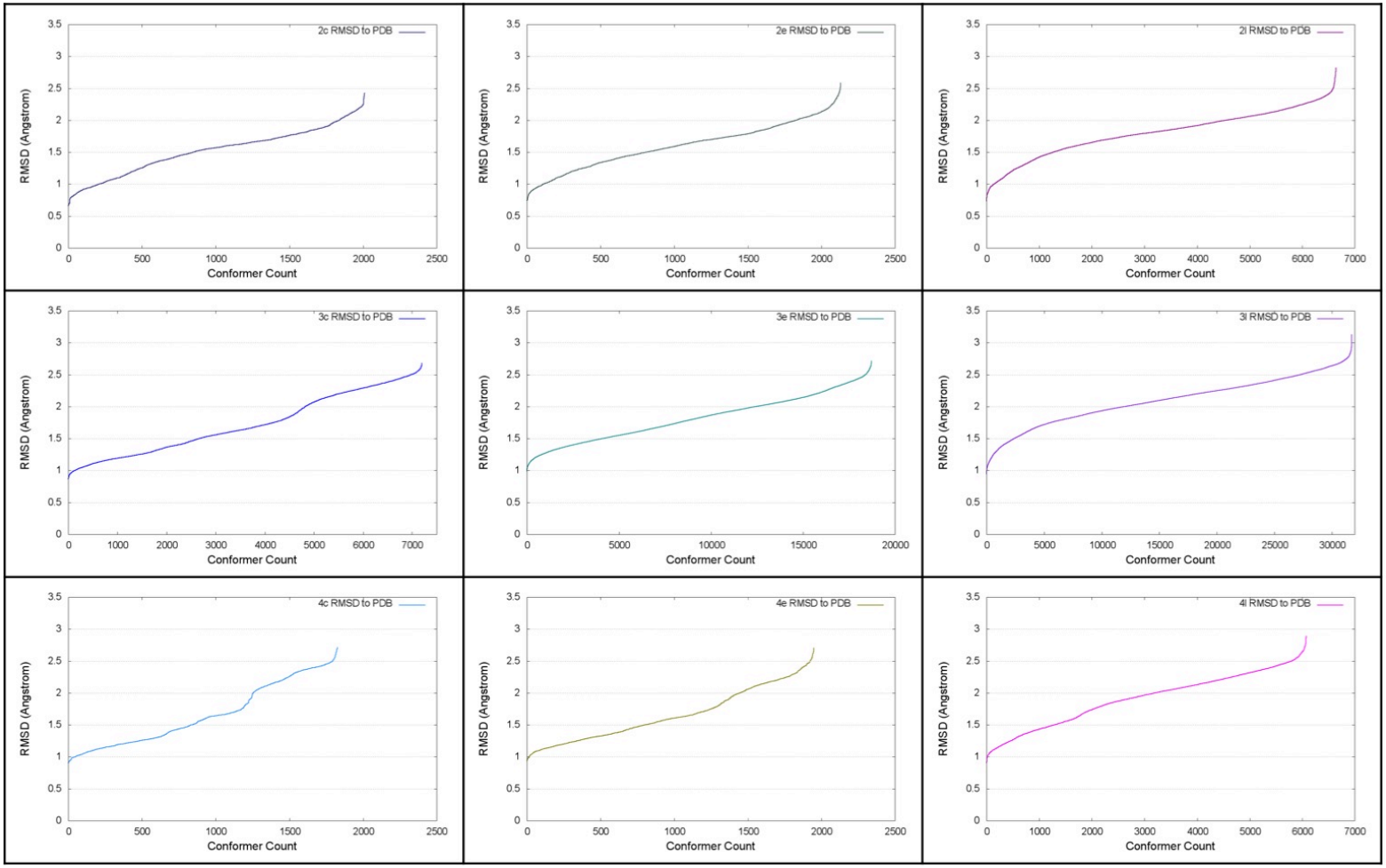


Figure 5.5 RMSDs (in Angstrom) for the 9 thrombin ligands (2c, 2e, 2l, 3c, 3e, 3l, and 4c, 4e, 4l), with alignment to their corresponding PDB structures.

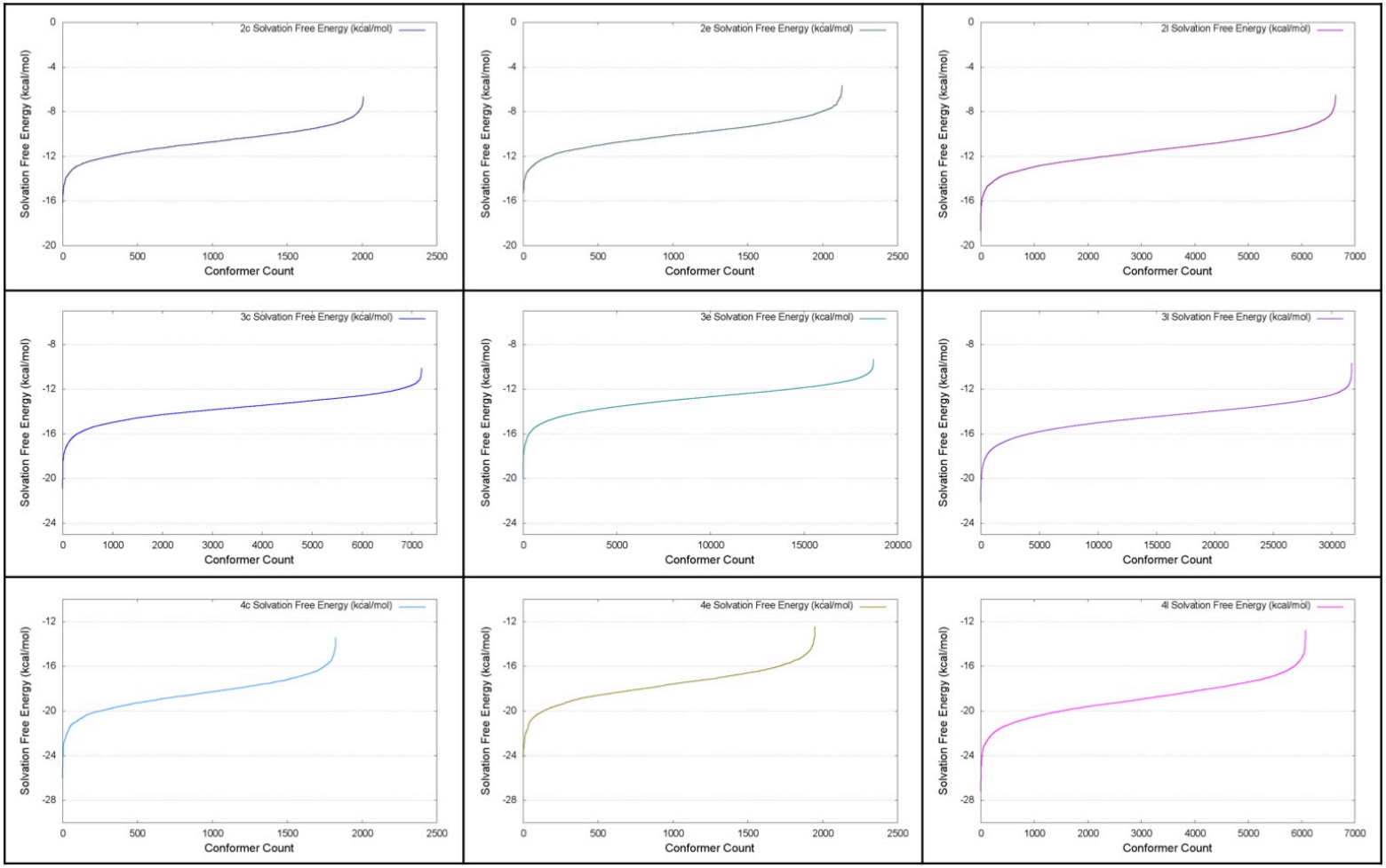


Figure 5.6 Solvation free energies (in kcal/mol) for the 9 thrombin ligands (2c, 2e, 2l, 3c, 3e, 3l, and 4c, 4e, 4l).

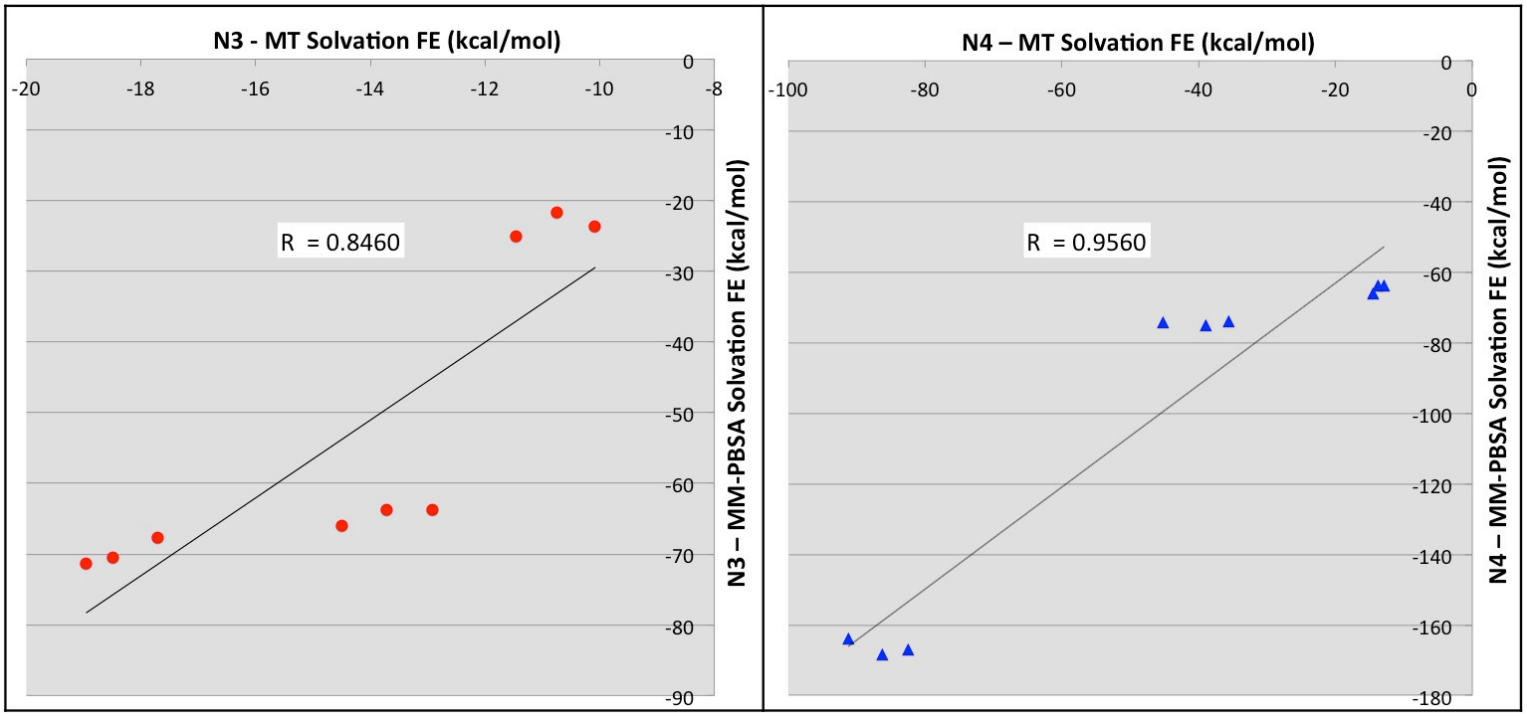


Figure 5.7 Data points and fitting results for MT vs. MM-PBSA solvation free energy calculation of Klebe thrombin ligands, with–NH_2 (left) and –NH_3^+ functional groups (right) for Series 2 and 4.

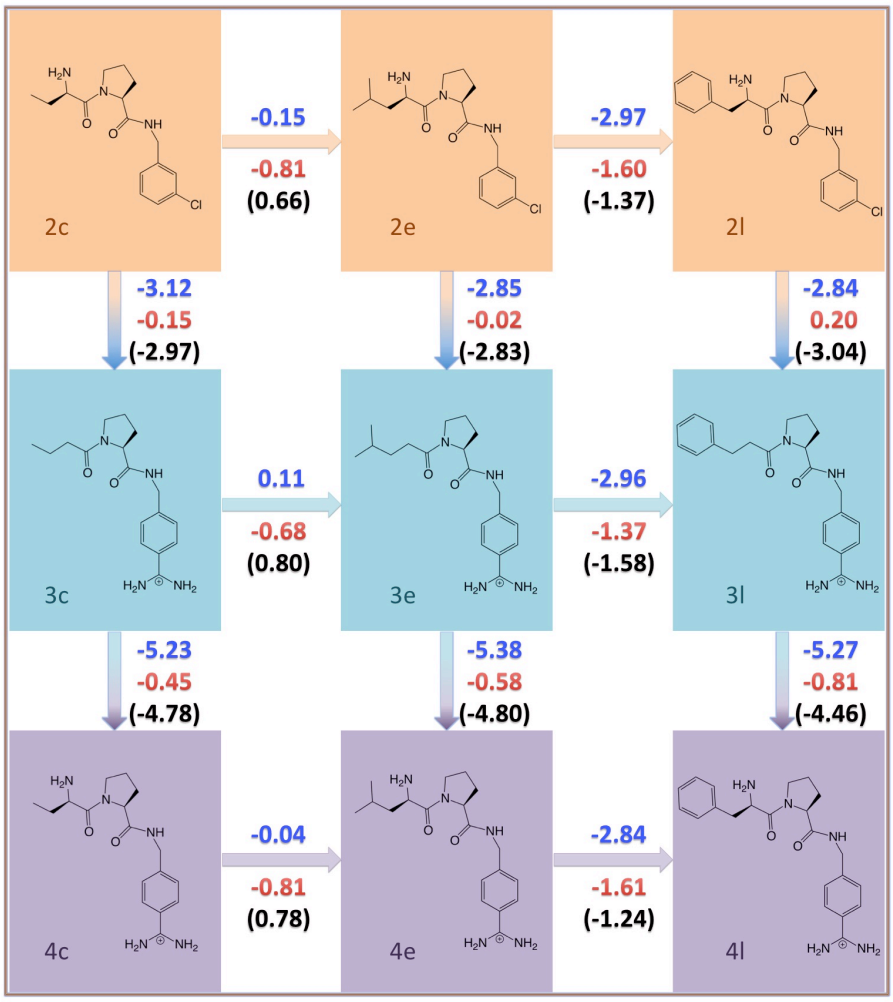


Figure 5.8 Free energy perturbations (FEPs) between the 9 thrombin ligands calculated by MT. Colored values are the free energy changes (in kcal/mol) in aqueous phase (blue) and gas phase (red) conditions. The values in black parentheses are the changes in solvation free energies.

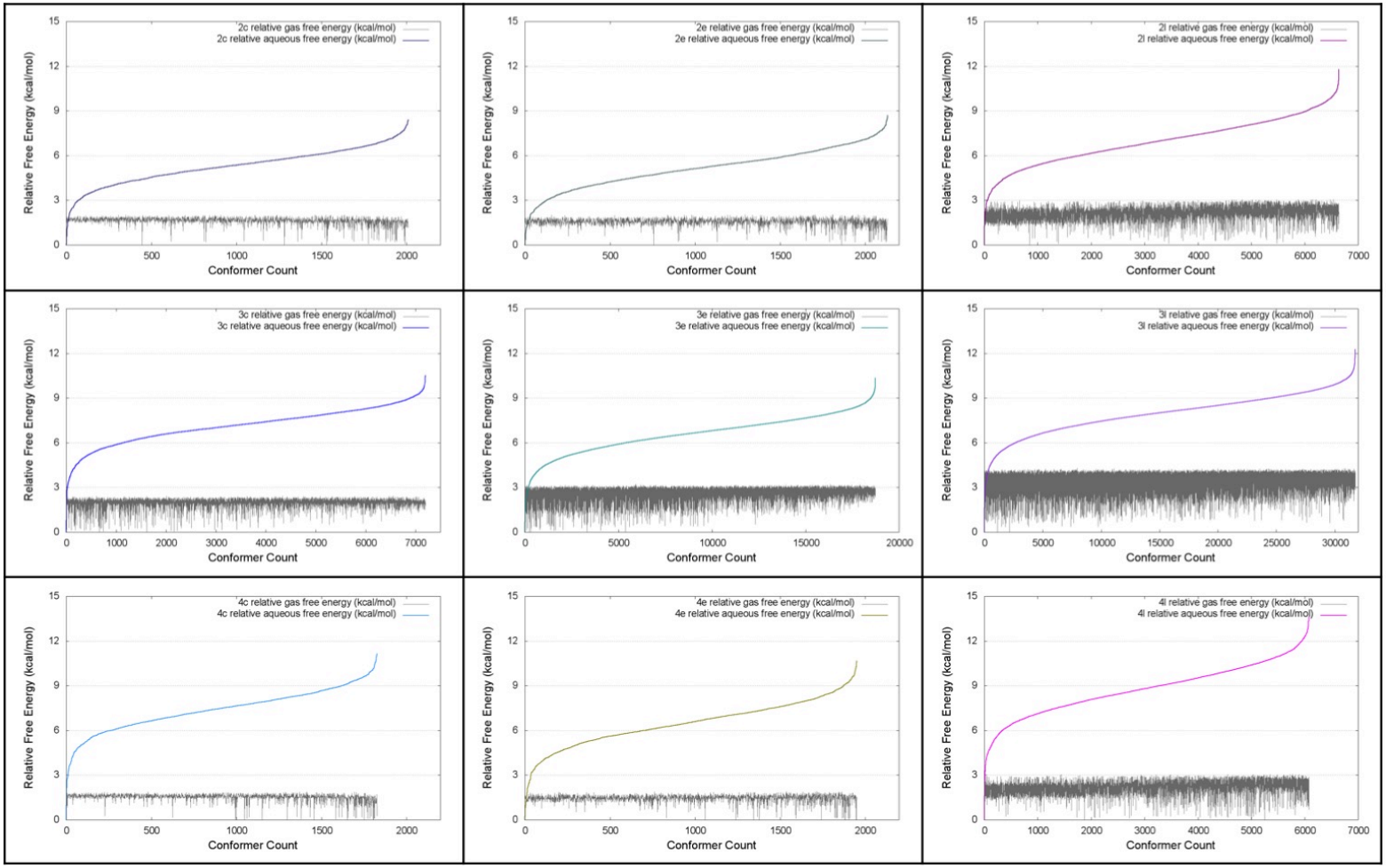


Figure 5.9 Relative aqueous free energies (in kcal/mol) (colored curves) and relative gas phase free energies (in kcal/mol) (gray curves) for thrombin ligands. The conformer count is sequenced in order of aqueous free energies.

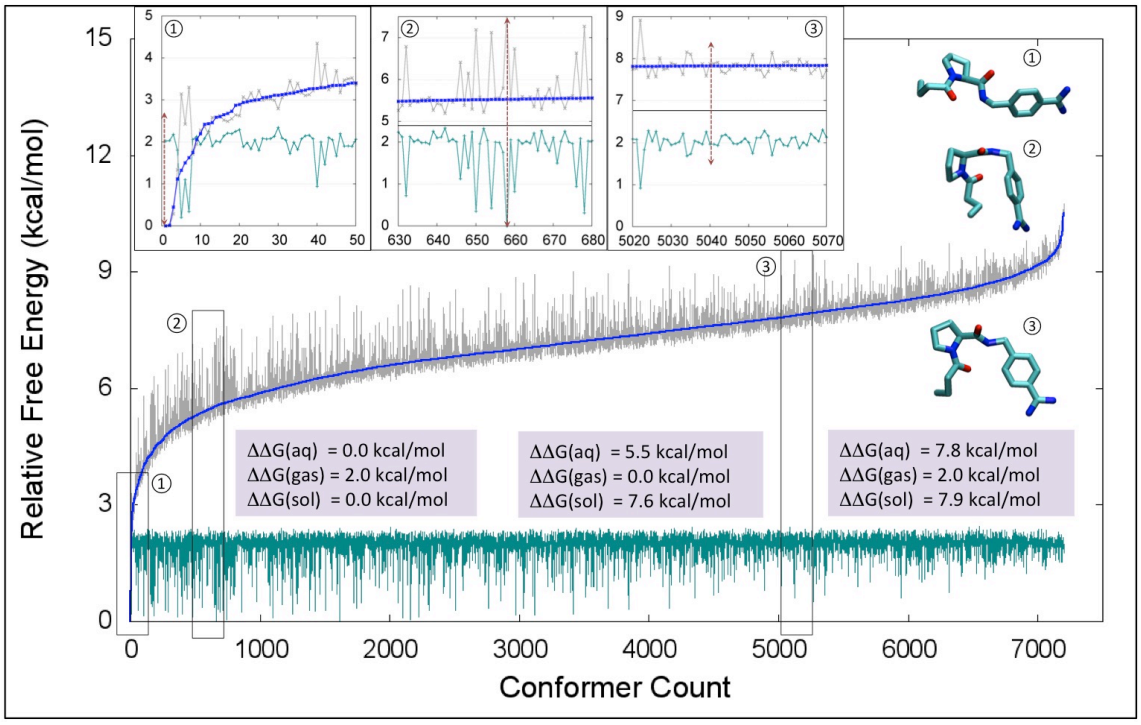


Figure 5.10 Relative free energies vs. conformation count for thrombin ligand 3c. Blue curve – aqueous phase free energy; grey curve – solvation free energy; cyan curve – gas phase free energy. The structures – (1) lowest aqueous phase free energy conformation; (2) lowest gas phase free energy conformation; (3) minimum RMSD conformation compared with the PDB structure. The small windows are enlarged versions of the free energy for every 50 conformers, where conformers (1)(2)(3) are included. The pink shadowed texts from left to right, represent the relative aqueous phase free energy ΔΔG(aq), relative gas phase free energy ΔΔG(gas), and relative solvation free energy ΔΔG(sol) for the three conformers (1) (2) (3) respectively. Note ΔΔG(aq), ΔΔG(gas), and ΔΔG(sol) do not have direct relationship with each other. They are only relative to their lowest aqueous, gas, and solvation free energy conformations, respectively.

Table 5.2 Solvation Free Energies for the Thrombin Ligands, Including MT Solvation Ranges, the MT Averages and the MM-PBSA averages.

Free E. (kcal/mol)	MT Average		MM-PBSA Average	
	N3[a]	N4[b]	N3[a]	N4[b]
2c	-10.74	-39.02	-21.72	-74.98
2e	-10.08	-35.58	-23.68	-73.94
2l	-11.45	-45.24	-25.01	-74.20
3c	-13.71	N/A	-63.73	N/A
3e	-12.91	N/A	-63.79	N/A
3l	-14.49	N/A	-65.95	N/A
4c	-18.49	-86.25	-70.53	-168.33
4e	-17.71	-82.51	-67.73	-167.02
4l	-18.95	-91.21	-71.27	-163.88

[a] "N3" indicates the $-NH_2$ functional group for 2c, 2e, 2l, and 4c, 4e, 4l. "N3" is not applicable for 3c, 3e and 3l since this functional group is absent.

[b] "N4" indicates the$-NH_3^+$ functional group for 2c, 2e, 2l, and 4c, 4e, 4l. "N4" is not applicable for 3c, 3e and 3l since this functional group is absent.

5.1.3 Free Energies

We obtained the gas phase free energies initially for every single conformer when doing the conformation generation. Subsequently I obtained their relevant solvation free energies via post-conformational search calculations on the corresponding geometries, as shown in Figure 5.6. I report calculations both on the protonated ($R–NH_3^+$) and unprotonated ($R–NH_2$) amine group and these results are shown in Table 5.2. It is not surprising that the resultant solvation free energies are more negative for the structures (2c, 2e, 2l & 4c, 4e, 4l) with the $–NH_3^+$ functional group than those with the $–NH_2$ functional group. Moreover, 4c, 4e and 4l are better solvated than 2c, 2e and 2l because the former are singly charged (benzamidine groups) and the latter set are neutral (see Figure 5.2). The negative solvation free energies for 3c, 3e and 3l arise due to the positive charge on the benzamidine moiety.

We also carried out MM-PBSA calculations to further analyze my solvation free energy results. The MM-PBSA results are also shown in Table 5.2. The MM-PBSA values are more negative than my average values, which was the trend I observed in a detailed analysis of the MT, MM-PB/SA and MM-GB/SA solvation methods.[51] In this paper I further compared my MT solvation free energy results with the MM-PBSA results with resultant Pearson's correlation coefficients of 0.846 for the N3 sets and 0.956 for the N4 sets. The plot is shown in Figure 5.7. Hence, I demonstrate that although I couldn't directly compare my solvation free energies of the thrombin inhibitors with the MM-PBSA results because of the latter's tendency to give more negative solvation free energies, then trends in solvation free energy match reasonably well.

To further analyze the nine ligands, I also computed the absolute free energies from the conformational ensembles couple with equation 4. Using the MT approach this is possible to do and, for example, on going from 2c -> 2e the gas-phase free energy change arises from 2c + 2 methane C atoms -> 2e. Moreover, this allows us to get the free energy changes among the nine thrombin ligands, as shown in Figure 5.8, in both solution and in the gas phase. Here I find that the free energy differences are larger in solution than in gas phase due to changes in the solvation free energies. Again using 2c -> 2e as an example I gain free energy from C-C bond formation, but pay an entropic penalty when 2 free methane carbon atoms attach to 2c forming 2e. The solvation free energy is slightly unfavorable (0.66 kcal/mol) as well. Using the data of Figure 5.8 similar thermodynamic analyses can be performed.

We also plot the aqueous phase free energies as well as the gas phase free energies for the 9 ligands in Figure 5.9. Herein I only discuss 3c as a representative example. Figure 5.10 shows the aqueous phase free energy curve (blue – lowest to highest free energy) with their corresponding gas phase free energies (cyan) and solvation free energies (grey). Note in Figure 5.10, all free energies including the aqueous phase, gas phase, and solvation free energies, are all originally carried out with absolute values – independent gas phase G(gas) and solvation G(sol), adding up of aqueous phase G(aq) = G(gas) + G(sol) – but here I am only showing the relative values $\Delta\Delta G(aq)$, $\Delta\Delta G(gas)$ and $\Delta\Delta G(sol)$ to their individual lowest free energy conformers respectively. There are no direct relationships between each other in the purple text boxes. In this relative aqueous phase free energy curve there is a range of 0.00 ~ 10.51 kcal/mol, interestingly the relative solvation free energy curve tracks the aqueous phase free energy curve, with the relative solvation free energy in a range of 0.00~10.74 kcal/mol and the relative gas phase free

energy is in a range of 0.00~2.44 kcal/mol, indicating that the aqueous phase conformer sequence is driven by the solvation free energy. For example, the 1st ranked structure in the aqueous phase is the top-ranked conformation in terms of its solvation free energy. Thus, in the case of 3c (in fact also for the remaining 8 thrombin ligands (see Figure 5.9)) the gas phase free energy only slightly modifies the relative stability of the conformations contained in the ensemble. This resultant plot has a S shape, where the first 50 conformers (Figure 5.10-1) have a sharp free energy increase (0.00 to 3.41 kcal/mol in the aqueous phase), while the middle two sets of 50 conformers (Figure 5.10-2 & 5.10-3) have a modest free energy increase (5.48 to 5.57 kcal/mol & 7.82 to 7.84 kcal/mol, respectively). Potential structural similarities are excluded, where each of the middle two sets of 50 conformers have different aqueous and gas phase free energies, as shown in the figure. Overall, the top ranked aqueous phase conformation is favored both because of its favorable solvation free energy (0.0 kcal/mol) and gas phase free energy (2.0 kcal/mol). The top ranked gas phase conformation has 5.5 kcal/mol aqueous phase free energy and a 0.0 kcal/mol gas phase free energy with a solvation free energy of 7.6 kcal/mol. Note that these energies are all relative, so adding the gas-phase free energy with the salvation free energy does not yield the aqueous free energy of the conformation. The minimum RMSD conformation has a less favorable solvation free energy (7.9 kcal/mol) but has a relatively unfavorable gas phase free energy (7.8 kcal/mol), leading to its position in the aqueous phase free energy plot.

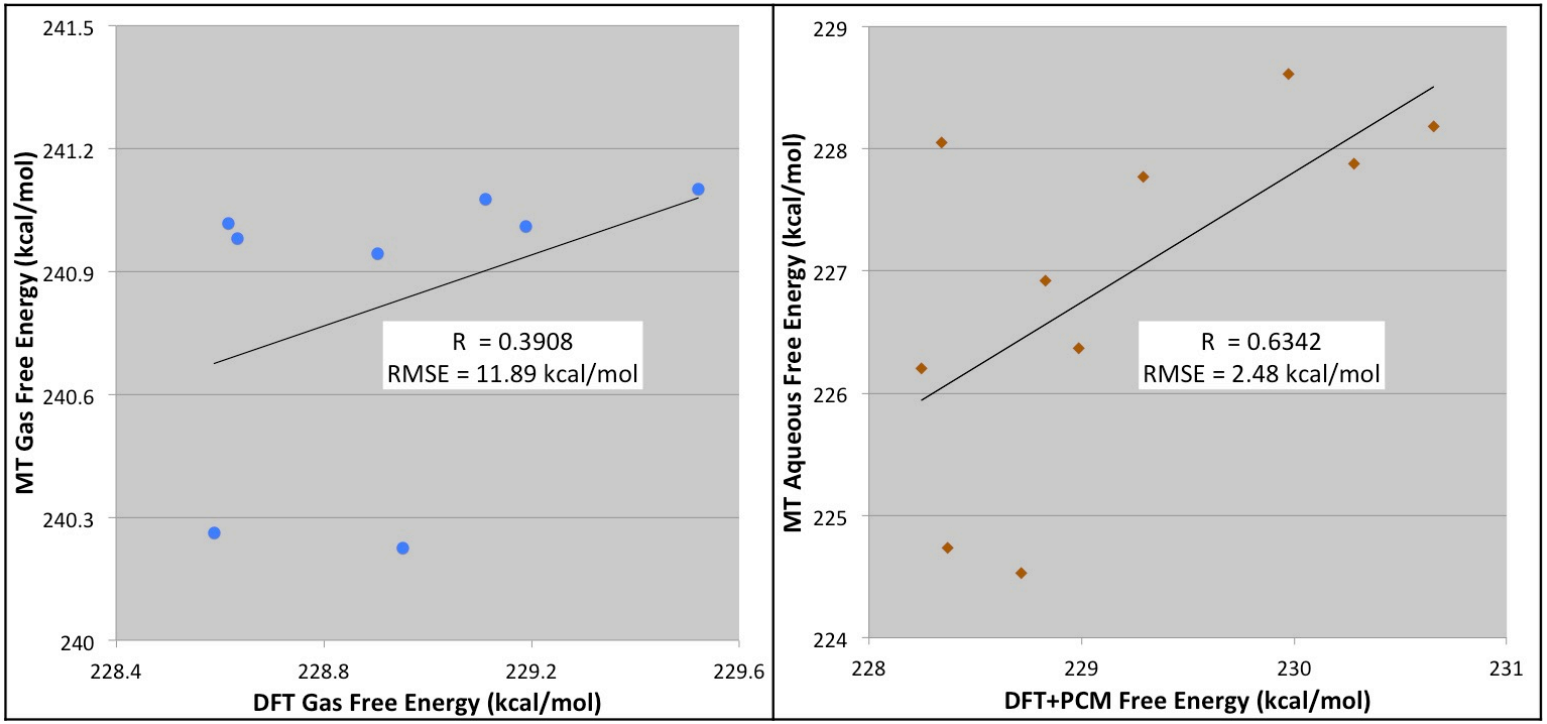


Figure 5.11 Comparison of MT-CS and DFT calculated free energy values for the 8 gas phase (left) and 10 PCM model (right) DFT conformations. The dots on each figure represent the positions of these conformations, and the line represents a linear fit. The Pearson's R values are given on the plots. The Root-Mean-Squared-Error (RMSE) is also included in the graph.

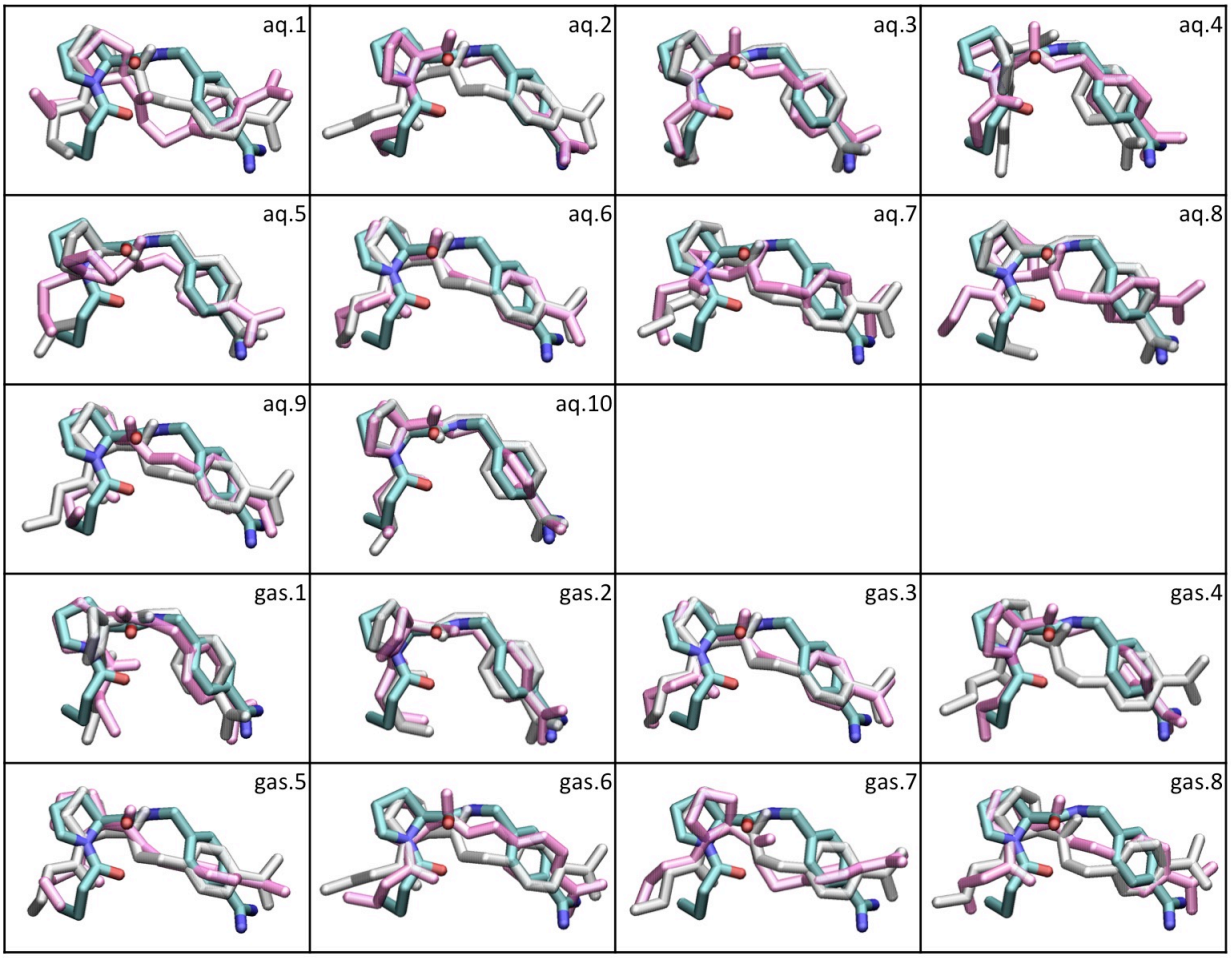


Figure 5.12 DFT low energy conformers of thrombin ligand 3c in PCM model and in gas phase sequenced by their MT free energies respectively. Multi-color – PDB crystal structure; silver – G09 optimized conformers; mauve – closest MT conformers to relevant DFT conformers.

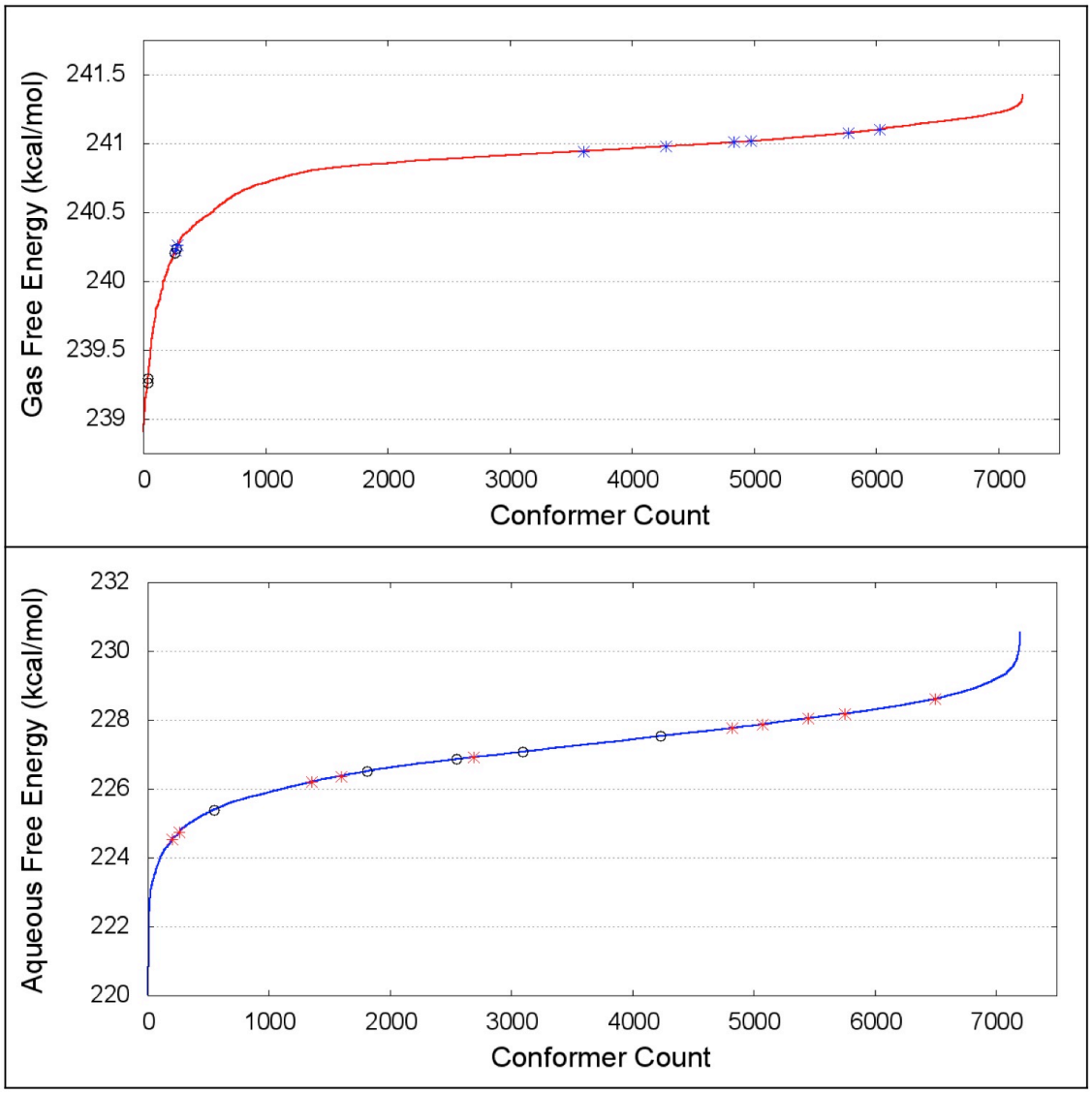


Figure 5.13 Free Energy for DFT optimized 3c conformers vs. MT calculated conformers, in both gas phase (upper) and aqueous phase (lower). The curves represent the free energy values for MT-CS generated conformers, and the points on the curve represent the 8 gas phase (upper) and 10 PCM model (lower) DFT conformations position on the MT free energy curves. Note here the free energies are absolute values. Black circles represent MT calculated free energies positions of the DFT energy-minimized lowest 5 gas/aqueous MT conformers respectively.

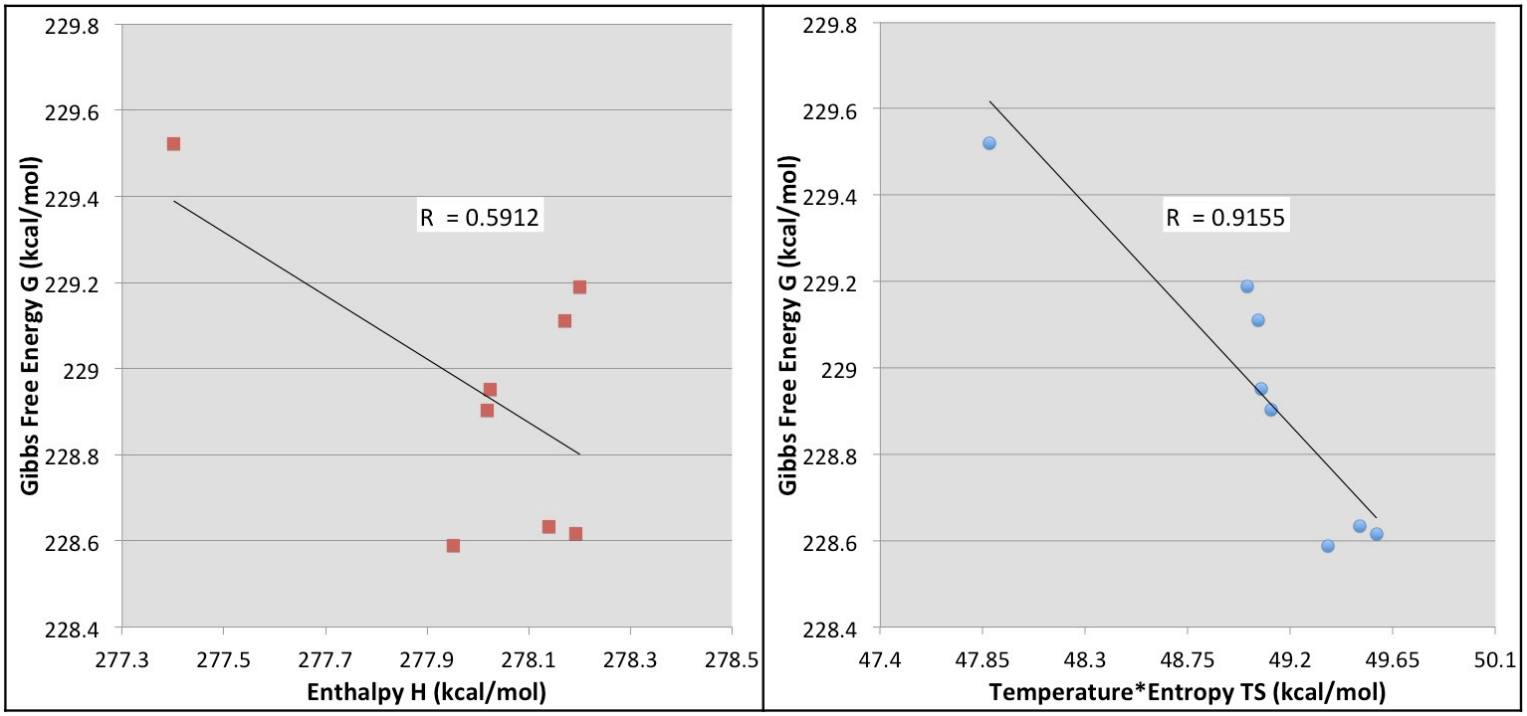


Figure 5.14 Linear correlation comparisons of H vs. G (left) and TS vs. G (right). The Pearson's R values are attached on the graphs.

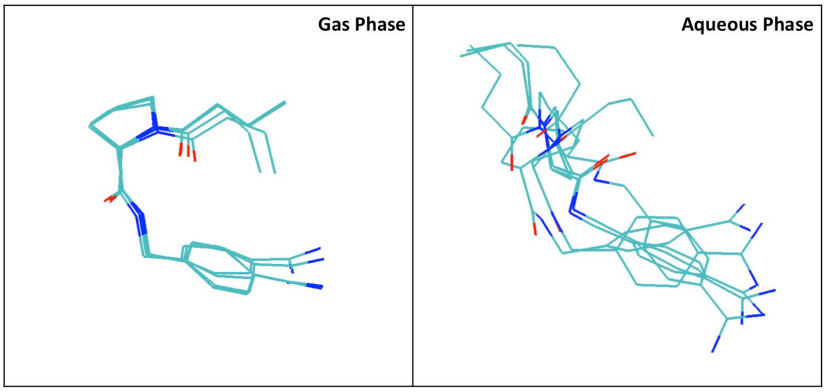


Figure 5.15 Alignment of top 5 gas and aqueous phase lowest MT free energy conformers that optimized in DFT gas phase and PCM models respectively.

Table 5.3 Geometric Comparisons of M06-2X/6-31G* (DFT) Generated Conformations and MT Generated Conformations (RMSD values in Å).

PCM Phase Rank	DFT vs. MT	Gas Phase Rank	DFT vs. MT
1	1.613	**1**	1.977
2	2.003	**2**	1.480
3	1.713	**3**	1.535
4	2.495	**4**	1.905
5	1.639	**5**	1.351
6	0.482	**6**	1.226
7	1.292	**7**	1.211
8	2.599	**8**	1.528
9	1.487		
10	0.754		

Table 5.4 RMSDs (in Angstrom), DFT free energies, and MT free energies for PCM Modeled DFT optimized structures.

PCM Model	RMSD (DFT to PDB)	RMSD [a] (MT to DFT)	DFT Free Energy [b] (kcal/mol)	MT Aq. Free Energy [c] (kcal/mol)
1	1.726	2.706	228.715	224.526
2	1.768	1.308	228.368	224.733
3	0.861	1.400	228.248	226.200
4	1.962	1.230	228.986	226.369
5	0.973	2.056	228.831	226.919
6	1.614	1.504	229.292	227.767
7	1.864	2.086	230.282	227.872
8	1.077	2.150	228.341	228.053
9	1.737	1.673	230.655	228.183
10	0.948	1.292	229.973	228.609

[a] "MT" here represents the conformers that has closest free energy MT conformer compared to the corresponding DFT conformers.

[b] Here the DFT free energy represent the free energies carried out by DFT frequency calculation.

[c] MT aq. free energies are absolute free energies for each conformer generated by MT program.

Table 5.5 RMSDs (in Angstrom), DFT free energies, and MT free energies for gas phase DFT optimized structures.

Gas Phase	RMSD (DFT to PDB)	RMSD [a] (MT to DFT)	DFT Free Energy [b] (kcal/mol)	MT Gas Free Energy [c] (kcal/mol)
1	1.765	1.332	228.952	240.225
2	1.180	1.756	228.589	240.263
3	1.815	2.338	228.903	240.945
4	1.952	1.218	228.634	240.980
5	1.707	1.837	229.189	241.010
6	1.883	1.241	228.616	241.017
7	1.834	2.286	229.111	241.078
8	2.146	1.481	229.521	241.102

[a] "MT" here represents the conformers that has closest free energy MT conformer compared to the corresponding G09 conformers.

[b] Here the DFT free energy represent the free energies carried out by DFT frequency calculation in gas phase.

[c] MT aq. free energies are absolute free energies for each conformer generated by the MT approach.

Table 5.6 Calculated Relative DFT Energies and Absolute MT Free Energies in kcal/mol for the Top 5 Lowest Free Energy Gas and Aqueous MT Conformations Further Optimized Using Gas and Aqueous Phase (PCM) DFT Calculations.[a]

Gas Phase			Aqueous Phase		
Ranking	Relative DFT Electronic Energy (kcal/mol))	Absolute MT Gas Phase Free Energy (kcal/mol)	Ranking	Relative DFT Electronic Energy (kcal/mol))	Absolute MT Aqueous Free Energy (kcal/mol)
1	0.000	239.262	1	0.000	225.393
2	1.031	239.294	2	1.069	226.863
3	3.421	240.240	3	3.084	227.533
4	3.428	240.240	4	3.254	227.077
5	3.446	240.203	5	3.374	226.511

a) Note here that the DFT energies are relative to the lowest energy conformation.

Table 5.7 M06-2X/6-31G* calculated enthalpy (H), entropy (S) and Gibbs free energy (G) in Gaussian 09. Note the temperature here is 298.15K.

Gas Phase	H (kcal/mol)	S (cal/mol-K)	TS (kcal/mol)	G (kcal/mol)
1	278.025	164.673	49.073	228.952
2	277.952	165.648	49.363	228.589
3	278.018	164.816	49.115	228.903
4	278.140	166.126	49.506	228.634
5	278.200	164.466	49.011	229.189
6	278.192	166.364	49.576	228.616
7	278.170	164.629	49.059	229.111
8	277.402	160.673	47.881	229.521

5.3.2 DFT Validation

We also compared the MT generated conformations with DFT optimized local minimum energy structures for 3c. First I used OMEGA[19, 65] to independently generate 341 conformations for 3c and then geometry optimized these conformations at the M06-2X/6-31G* level of theory using Gaussian 09. From this I obtained 8 low energy structures in the gas phase and 10 using the PCM model. I then used a vibrational analysis (as implemented in G09) to obtain the absolute free energies for the DFT generated conformations for further comparison to MT. I then used my MT method to estimate the absolute free energy values for all of the MT generated conformations using equation 4 for a head-to-head absolute free energy comparison. Figure 5.11 shows the comparison of the MT vs. DFT derived free energies, in both the gas and aqueous phases. The resultant DFT free energies range from 228.589~229.521 kcal/mol for the gas phase, and 228.248~230.655 kcal/mol for the PCM model, while the MT free energy for these conformations range from 240.226~241.102 kcal/mol for the gas phase, and 224.526~228.609 kcal/mol for solution. Details are shown in Table 5.4 & 5.5. The values of Pearson's correlation are shown in Figure 5.11. The aqueous phase gives the best match between the MT and G09 absolute free energies with a correlation of 0.6342, indicating that the MT and DFT absolute free energies at the same structure are in qualitative agreement. The gas phase gives a poorer match with a correlation of 0.3908 between the MT and DFT results.

The 18 DFT determined absolute free energy values were then compared to the MT absolute free energies and the comparison is shown in Figure 5.13. I find that the 8 gas phase and 10 PCM model structures have absolute free energy values that are on the same scale as my MT generated absolute free energy conformational curves. Table 5.3 shows the RMSD results for the 18 lowest

energy structures from the DFT calculations when compared to their closest partner in terms of free energy. In the table, I list RMSD values for the M06-2X/6-31G* conformation *vs*. the nearest (by free energy) MT conformation. I also visually compare the geometries in Figure 5.12. A geometric comparison of the DFT and MT conformations with the PDB structure is given in Table 5.4 & 5.5 as well. I find that the RMSDs between the DFT conformations and closest in free energy MT conformations range from 0.754~2.599 Å, while the lowest MT conformation vs. lowest DFT one had RMSDs of 1.509Å for the gas-phase and 1.656Å for the aqueous phase.

To further study the relationship of energy-minimized conformations to their corresponding MT free energies, I took the five lowest gas and aqueous phase free energy conformations from my MT studies and then optimized them at the M06-2X/6-31G(d) level of theory. I then took these 10 DFT energy-minimized MT (DFT-MT) conformations and calculated their new gas and aqueous phase MT derived free energies. The results are shown also in Figure 5.13 as black circles. I find that the 5 gas-phase conformations generated 3 novel conformations in the gas-phase and the 5 aqueous ones produced 5 in the aqueous phase. The aligned conformations are shown in Figure 5.15.

From Figure 5.13 I observe that in the gas-phase, 2/3 of my MT-DFT optimized conformations have lower free energies than the 8 OMEGA-DFT optimized conformations, while the remainder (1/3) has a free energy value close to the OMEGA-DFT lowest free energy conformation. In the aqueous phase, the MT-DFT lowest free energy conformation has a higher free energy value than the OMEGA-DFT one, while the remainder (4/5) were distributed among the remaining 10 OMEGA-DFT conformations. Detailed values are shown in Table 5.6 with their related absolute

MT free energies. In this table the trends of the absolute MT free energy values are consistent with that of the DFT electronic energy values, in both the gas and aqueous phase. Note in the gas phase No.'s 3 to 5 are nearly geometrically identical, so that they have close MT free energies and DFT electronic energies.

What is more, the gas-phase DFT energy-minimized conformations (8 total) did not yield low free energy conformations along the MT free energy curve (Figure 5.13 top panel), so in order to analyze this a bit further I examined the individual contributions of the entropy and enthalpy. In the DFT determination of the Gibbs free energy the Thermal Energy and Entropy at 298.15 K are each computed, allowing us to investigate the contributions of enthalpy and entropy. Hence, I did another series of linear correlations comparing H *vs*. G and TS *vs*. G, which is shown in Figure 5.14. Detailed H, S and G values are shown in Table 5.7. From this I observe that the TS *vs*. G graph (right) has a high Pearson's R value of 0.9155, while the H vs. G graph (left) has a lower Pearson's R value of 0.5912, implying that that the TS term has a higher influence than that of the H term in determining the energy landscape of the 3c conformations. That means the 3c molecule has an entropy driven free energy landscape.

It is this latter observation that, in part, explains why the OMEGA/DFT derived conformations appear well above the minimum of the MT derived free energy landscape – the DFT conformations are obtained via energy minimization calculations, not an optimization on the free energy surface. From the latter analysis I find that the preferred low free energy confomations for 3c arise due to entropy effects and MT conformation generation is done on the free energy

surface (both entropy and enthalpy (energy)), so it is not surprising that MT yields lower free energy conformations.

5.3.3 CSD & PDB Structure Validation

We also completed a CSD & PDB crystal structure validation with my MT conformational search program. Table 5.8 gives the general details including conformation counts and the observed RMSD ranges, as well as their median values. From this table I observe that the conformation counts, not surprisingly, vary significantly depending on the number of rotatable bonds found in the molecule being studied (from 2 to 228919 of CSD and from 2 to 32593 of PDB total conformations). However those with smaller conformation counts do not necessarily have a better match with the crystal structures since the diversity of conformations explored is simultaneously reduced.

Figure 5.16 shows the RMSD ranges compared to the crystal structures for all of the 299 CSD molecules, and Figure 5.17 shows this information for the PDB molecules. For the CSD RMSDs, as indicated by the blue line in the figure, I find that the minimum RMSDs differ quite a bit from one another, ranging from 0.014 ~ 1.720 Å, with a median value of 0.388 Å, which overall indicates at least one MT conformation has a high structure similarity with the CSD structure. The lowest free energy conformations in the gas phase reasonably match the CSD structures with 239 out of 299 structures having RMSDs < 2.0 Å. The median values are 1.270 Å. (Not listed here) Meanwhile, for the 104 PDB molecules, I obtained minimum RMSDs ranging from 0.009 ~ 1.663 Å, with a median value of 0.253 Å. The lowest free energy conformations in the gas phase qualitatively match the PDB structures for 82 out of the 104 structures with RMSDs < 2.0

Å. The median values are 1.342 Å. Using these metrics (RMSDs < 2.0 Å) I have ~78% match between the lowest free energy MT conformations for both CSD and PDB structures in comparison to their related crystal structures. Detailed values are shown in Table 5.8.

One difference between the CSD and the PDB structures is, although both are crystal structures, the CSD are for single molecules in a crystal, while the PDB is for protein-bound small molecules. One example, of some of the issues that can arise in structure validation, is illustrated by the molecule shown in Figure 5.18 (CSD ID BAJZOB10). The minimum free energy structure, in both the gas and aqueous phase, has a high preference to form two intramolecular hydrogen bonds, while in the CSD crystal structure this is not observed, but what is preferred are intermolecular hydrogen bonds forming in the crystalline lattice. Since I am not modeling a crystal, but a molecule in solution this observation is not surprising. Nevertheless, I still obtained a minimum RMSD of 0.806 Å and 1.002 Å for both the gas and aqueous phase low free energy conformation. The geometry comparisons are shown in Figure 5.18, where the complete free energy profile is also included.

Table 5.8 MT Conformer Counts and RMSD Values (in Å) for 299 CSD and 104 PDB Structures.

		Conformer Count	Average RMSD	Min. RMSD	1st Gas RMSD
CSD	**General Range**	2~228919	0.102~3.068	0.014~1.720	0.014~3.872
	Median Value	84	1.295	0.388	1.270
PDB	**General Range**	2~32593	0.166~3.512	0.009~1.663	0.069~4.107
	Median Value	142	1.376	0.253	1.314

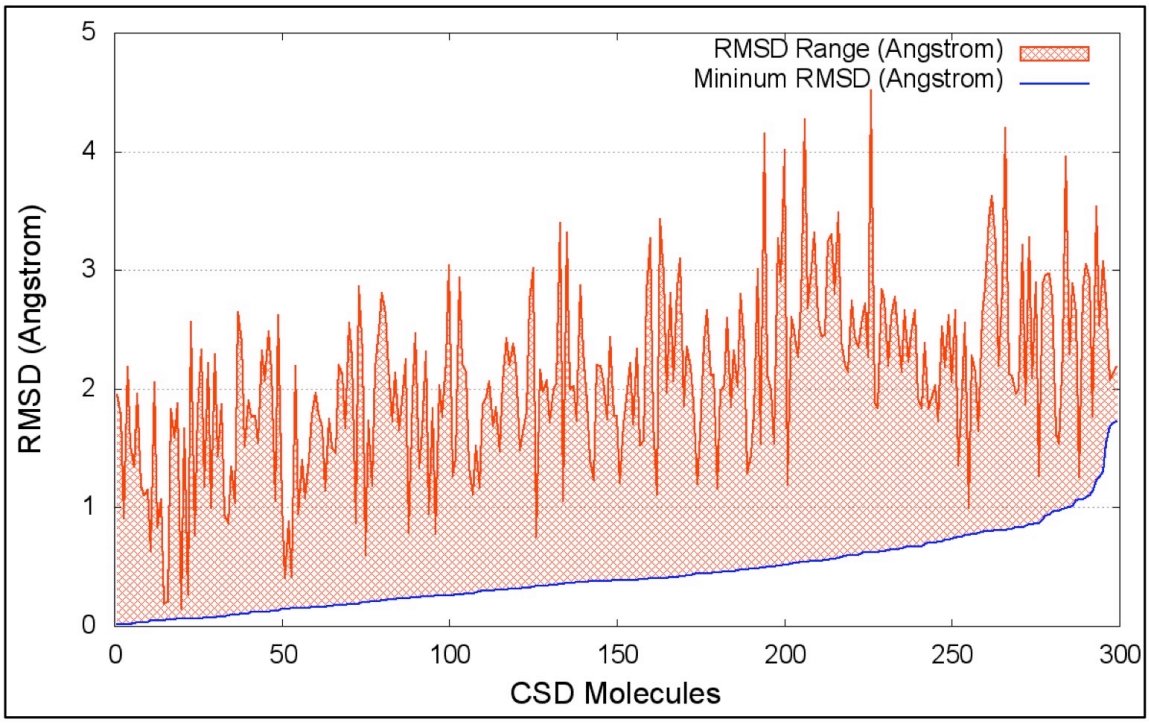


Figure 5.16 Minimum RMSD values and RMSD ranges for all CSD molecules. Blue curve – minimum RMSDs of the 299 CSD structures, red crosshatched regions – RMSD ranges of the 299 CSD structures.

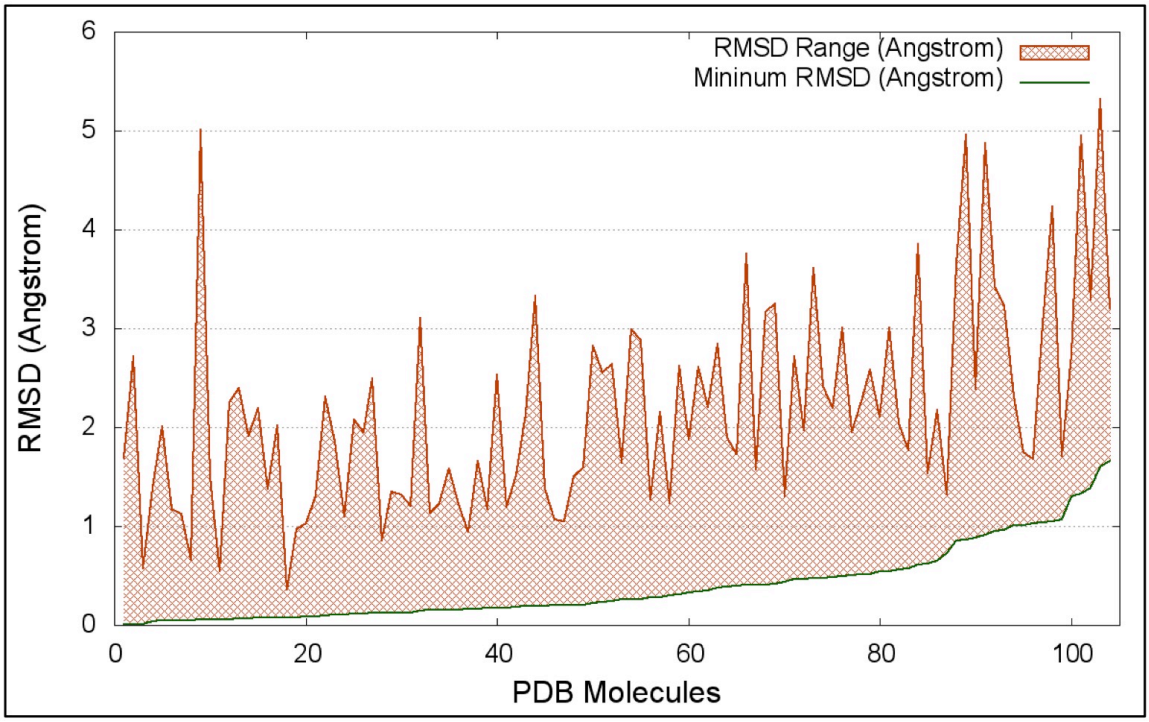


Figure 5.17 Minimum RMSD values and RMSD ranges for all PDB molecules. Green curve – minimum RMSDs of the 104 PDB structures, orange crosshatched regions – RMSD ranges of the 104 PDB structures.

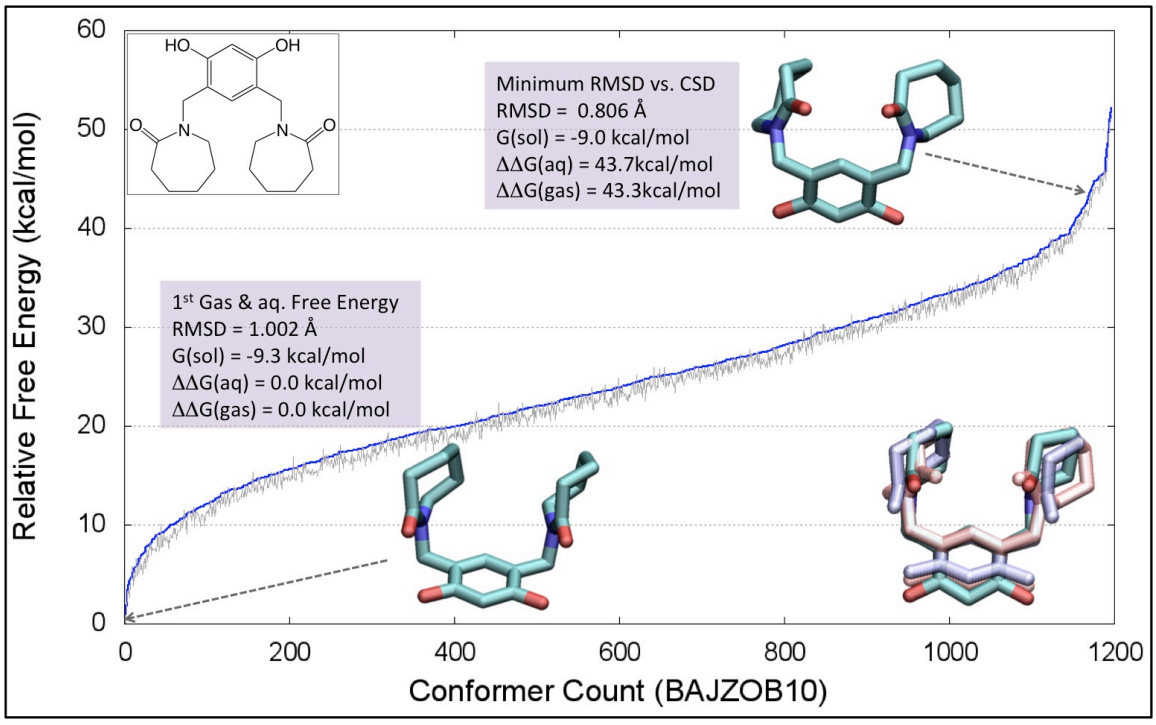


Figure 5.18 Relative free energies *vs*. conformer count for CSD molecule BAJZOB10. Blue curve – aqueous phase free energy; grey curve – gas phase free energy. The single multi-colored structures are the minimum RMSD conformation vs. the CSD and lowest gas/aqueous phase conformation with the requisite information given in the text boxes. The overlapped structures at the bottom right represent the minimum RMSD conformer (pink) and lowest gas/aqueous phase free energy conformer (ice blue) aligned to the CSD crystal structure (multi-colored). The sequence of the conformations is based on the order of the relative aqueous phase free energy.

5.4 CONCLUSION

In this paper I introduce a novel conformational search program based on the "Movable Type" sampling method, which directly gives conformational free energies in the gas and aqueous phase. I explain my conformation generation strategies, and validated the MT approach via comparisons to the crystal structures of 9 thrombin inhibitors found in the PDB, 299 CSD molecules and 104 PDB molecules found in the OMEGA validation effort.[19] The highlight of my approach is I use the MT method to focus my torsional space search to relevant regions, provide solvation free energies and gas and aqueous phase relative free energies for all conformations generated. I compared my results to experiment by examining RMSDs of my conformational ensembles and individual structures from the crystal structures. I also compared my solvation free energies to those computed using MM-PBSA. I also compared the best conformations obtained using DFT with the MT-CS generated conformations. Overall, the MT sampling approach rapidly and efficiently generates conformational ensembles of small molecules on the free energy surface opening up new avenues to understand the complex behavior of molecules in solution and when bound to receptors.

REFERENCES

REFERENCES

1. Sadowski, J.; Gasteiger, J.; Klebe, G., Comparison of Automatic 3-Dimensional Model Builders Using 639 X-Ray Structures. *J Chem Inf Comp Sci* **1994**, 34, 1000-1008.

2. Lorber, D. M.; Shoichet, B. K., Flexible ligand docking using conformational ensembles. *Protein Sci* **1998**, 7, 938-950.

3. Ebejer, J. P.; Morris, G. M.; Deane, C. M., Freely Available Conformer Generation Methods: How Good Are They? *J Chem Inf Model* **2012**, 52, 1146-1158.

4. Stahura, F. L.; Bajorath, M., New methodologies for ligand-based virtual screening. *Curr Pharm Design* **2005**, 11, 1189-1202.

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

6. Hahn, M., Three-dimensional shape-based searching of conformationally flexible compounds. *J Chem Inf Comp Sci* **1997**, 37, 80-86.

7. Verma, J.; Khedkar, V. M.; Coutinho, E. C., 3D-QSAR in Drug Design - A Review. *Curr Top Med Chem* **2010**, 10, 95-115.

8. Schwab, C. H., Conformations and 3D pharmacophore searching. *Drug discovery today. Technologies* **2010**, 7, e203-70.

9. Conformations and 3D pharmacophore searching. *Drug discovery today. Technologies* **2010**, 7, e203-70.

10. Kristam, R.; Gillet, V. J.; Lewis, R. A.; Thorner, D., Comparison of conformational analysis techniques to generate pharmacophore hypotheses using catalyst. *J Chem Inf Model* **2005**, 45, 461-476.

11. Makino, S.; Kuntz, I. D., Automated flexible ligand docking method and its application for database search. *J Comput Chem* **1997**, 18, 1812-1825.

12. Inc., A. S., Catalyst, Release 4.7. *San Diego: Accelrys Software Inc.* **2014**.

13. Macromodel, Version 10.4. *Schrödinger, Inc., New York, NY* **2014**.

14. Oda, A.; Yamaotsu, N.; Hirono, S.; Takano, Y.; Fukuyoshi, S.; Nakagaki, R.; Takahashi, O., Evaluations of the conformational search accuracy of CAMDAS using experimental three-dimensional structures of protein-ligand complexes. *Jounal of Physics: COnference Series* **2013**, 454.

15. Molecular Operating Environment (MOE), Version 2013.08. *Chemical Computing Group Inc. , 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7* **2013**.

16. O'Boyle, N. M.; Vandermeersch, T.; Flynn, C. J.; Maguire, A. R.; Hutchison, G. R., Confab - Systematic generation of diverse low-energy conformers. *Journal of cheminformatics* **2011**, 3.

17. Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C., ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers. *J Chem Inf Model* **2010**, 50, 534-546.

18. Miteva, M. A.; Guyon, F.; Tuffery, P., Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Res* **2010**, 38, W622-7.

19. Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T., Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J Chem Inf Model* **2010**, 50, 572-584.

20. Sperandio, O.; Souaille, M.; Delfaud, F.; Miteva, M. A.; Villoutreix, B. O., MED-3DMC: a new tool to generate 3D conformation ensembles of small molecules with a Monte Carlo sampling of the conformational space. *Eur J Med Chem* **2009**, 44, 1405-9.

21. Liu, X. F.; Bai, F.; Ouyang, S. S.; Wang, X. C.; Li, H. L.; Jiang, H. L., Cyndi: a multi-objective evolution algorithm based method for bioactive molecular conformational generation. *BMC bioinformatics* **2009**, 10.

22. Lagorce, D.; Pencheva, T.; Villoutreix, B. O.; Miteva, M. A., DG-AMMOS: a new tool to generate 3d conformation of small molecules using distance geometry and automated molecular mechanics optimization for in silico screening. *BMC chemical biology* **2009**, 9, 6.

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

24. Chen, I. J.; Foloppe, N., Conformational sampling of druglike molecules with MOE and catalyst: Implications for pharmacophore modeling and virtual screening. *J Chem Inf Model* **2008**, 48, 1773-1791.

25. Li, J.; Ehlers, T.; Sutter, J.; Varma-O'Brien, S.; Kirchmair, J., CAESAR: A new conformer generation algorithm based on recursive buildup and local rotational symmetry consideration. *J Chem Inf Model* **2007**, 47, 1923-1932.

26. Leite, T. B.; Gomes, D.; Miteva, M. A.; Chomilier, J.; Villoutreix, B. O.; Tuffery, P., Frog: a FRee Online druG 3D conformation generator. *Nucleic Acids Res* **2007**, 35, W568-W572.

27. Sadowski, J.; Bostrom, J., MIMUMBA revisited: Torsion angle rules for conformer generation derived from X-ray structures. *J Chem Inf Model* **2006**, 46, 2305-2309.

28. Hare, B. J.; Walters, W. P.; Caron, P. R.; Bemis, G. W., CORES: An automated method for generating three-dimensional models of protein/ligand complexes. *J Med Chem* **2004**, 47, 4731-4740.

29. Balducci, R.; Pearlman, R. S., CONFORT: A rational conformation analysis tool. *Abstr Pap Am Chem S* **1999**, 217, U644-U645.

30. Landrum, G., RDKit: Open-source cheminformatics. http://www.rdkit.org.

31. Rubicon, Version 4.7.1. *Daylight Chemical Information Systems, Inc. Aliso Viejo, CA*, http://www.daylight.com.

32. The 3D structure generator CORINA is available from Molecular Networks GmbH. *Erlangen, Germany*, http://www.molecular-networks.com.

33. Leach, A. R., Molecular Modeling - Principles and Applications, 2nd ed. *Prentice Hall* **2001**, Chaoter 9, p 457.

34. Leach, A. R., A survey of methods for searching the conformational space of small and medium-size molecules. *Reviews of Computational Chemistry, Wiley, New York* **1991**, 2, 1–55.

35. Vainio, M. J.; Johnson, M. S., Generating conformer ensembles using a multiobjective genetic algorithm. *J Chem Inf Model* **2007**, 47, 2462-2474.

36. Pavlov, T.; Todorov, M.; Stoyanova, G.; Schmieder, P.; Aladjov, H.; Serafimova, R.; Mekenyan, O., Conformational coverage by a genetic algorithm: Saturation of conformational space. *J Chem Inf Model* **2007**, 47, 851-863.

37. Mekenyan, O.; Dimitrov, D.; Nikolova, N.; Karabunarliev, S., Conformational coverage by a genetic algorithm. *J Chem Inf Comp Sci* **1999**, 39, 997-1016.

38. Chang, G.; Guida, W. C.; Still, W. C., An Internal Coordinate Monte-Carlo Method for Searching Conformational Space. *J Am Chem Soc* **1989**, 111, 4379-4386.

39. Wilson, S. R.; Cui, W.; Moskowitz, J. W.; Schmidt, K. E., Applications of Simulated Annealing to the Conformational-Analysis of Flexible Molecules. *J Comput Chem* **1991**, 12, 342-349.

40. Grebner, C.; Becker, J.; Stepanenko, S.; Engels, B., Efficiency of Tabu-Search-Based Conformational Search Algorithms. *J Comput Chem* **2011**, 32, 2245-2253.

41. Havel, T. F.; Kuntz, I. D.; Crippen, G. M., The Theory and Practice of Distance Geometry. *B Math Biol* **1983**, 45, 665-720.

42. Fu, Z.; Li, X.; Merz, K. M., Conformational Analysis of Free and Bound Retinoic Acid. *J Chem Theory Comput* **2012**, 8, 1436-1448.

43. Fu, Z.; Li, X.; Merz, K. M., Accurate Assessment of the Strain Energy in a Protein-Bound Drug Using QM/MM X-ray Refinement and Converged Quantum Chemistry. *J Comput Chem* **2011**, 32, 2587-2597.

44. Zheng, Z.; Ucisik, M. N.; Merz, K. M., The Movable Type Method Applied to Protein-Ligand Binding. *J Chem Theory Comput* **2013**, 9, 5526-5538.

45. Zheng, Z.; Merz, K. M., Jr., Development of the knowledge-based and empirical combined scoring algorithm (KECSA) to score protein-ligand interactions. *Journal of chemical information and modeling* **2013**, 53, 1073-83.

46. Tomasi, J.; Persico, M., Molecular-Interactions in Solution - an Overview of Methods Based on Continuous Distributions of the Solvent. *Chem Rev* **1994**, 94, 2027-2094.

47. Cramer, C. J.; Truhlar, D. G., Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem Rev* **1999**, 99, 2161-2200.

48. Orozco, M.; Luque, F. J., Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem Rev* **2000**, 100, 4187-4225.

49. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum mechanical continuum solvation models. *Chem Rev* **2005**, 105, 2999-3093.

50. Cramer, C. J.; Truhlar, D. G., A universal approach to solvation modeling. *Accounts Chem Res* **2008**, 41, 760-768.

51. Zheng, Z.; Wang, T.; Li, P. F.; Merz, K. M., KECSA-Movable Type Implicit Solvation Model (KMTISM). *J Chem Theory Comput* **2015**, 11, 667-682.

52. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000**, 28, 235-42.

53. Battle, G.; Kleywegt, G.; Velankar, S.; Oldfield, T.; Gore, S.; Mir, S.; Dana, J., Chemistry-related resources at the Protein Data Bank in Europe. *Abstr Pap Am Chem S* **2013**, 245.

54. Baum, B.; Muley, L.; Smolinski, M.; Heine, A.; Hangauer, D.; Klebe, G., Non-additivity of functional group contributions in protein-ligand binding: a comprehensive study by crystallography and isothermal titration calorimetry. *J Mol Biol* **2010**, 397, 1042-54.

55. M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V.

N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian 09, Revision A.02. *Gaussian, Inc., Wallingford CT* **2009**.

56. Henderson, S.; Bruno, I. J., Cambridge Structural Database: Moving with the times. *Abstr Pap Am Chem S* **2013**, 245.

57. Allen, F. H., The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr B* **2002**, 58, 380-388.

58. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J Chem Theory Comput* **2012**, 8, 3314-3321.

59. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G., Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem Phys Lett* **1995**, 246, 122-129.

60. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G., Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J Phys Chem-Us* **1996**, 100, 19824-19839.

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

62. http://www.ks.uiuc.edu/Research/vmd/

63. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J Mol Graph Model* **1996**, 14, 33-38.

64. http://gnuplot.info

65. Hawkins, P. C. D. S., A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T, OMEGA 2.5.1.4: OpenEye Scientific Software, Santa Fe, NM. .