

COMPUTATIONAL MODELLING AND SIMULATION OF
PRENYLTRANSFERASES AND CORRESPONDING PRENYLATION

By

YUE YANG

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2011

© 2011 Yue Yang

To Amelia

ACKNOWLEDGMENTS

I would like to express my gratitude to everyone that has been in my life in the past five years. To pursue a doctoral degree is never an easy task. All of your help and support have made this road much easier to walk through. It is five years of hard working and all of your assistance that make this dissertation here today. Please accept my sincerely grateful.

A special thank you is due to my family - especially my dearest daughter, Amelia - for all the happiness, hope, faith and courage they gave me. Another special thank you should be given to my advisor Kennie, thank you for giving me this precious opportunity to work with you and offering me help throughout my entire graduate career, it has been an enjoyable and memorable five years that I will always remember. Thank you Adrian, Erik, Nicole and Joanna for joining my supervisory committee and giving me countless help advices. And thank you also to Yuning, Yilin and all my friends in Merz group, QTP, Department of Chemistry and University of Florida.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS	4
LIST OF TABLES	8
LIST OF FIGURES	9
LIST OF ABBREVIATIONS.....	12
ABSTRACT	15
CHAPTER	
1 INTRODUCTION	17
1.1 Prenylation and Prenyltransferase	17
1.2 Protein Prenyltransferase	18
1.3 Aromatic Prenyltransferase.....	22
1.3.1 Bacterial Aromatic Prenyltransferase NphB	22
1.3.2 Fungal Indole Prenyltransferase FtmPT1	25
1.4 Computational Chemistry/Biochemistry	27
2 THEORY AND METHODS	29
2.1 Quantum Mechanics and the Born Oppenheimer Approximation.....	29
2.2 Molecular Mechanics	31
2.2.1 Bond Stretching	33
2.2.2 Angle Bending.....	34
2.2.3 Torsion and Improper Torsion	35
2.2.4 Van der Waals Interaction	36
2.2.5 Electrostatic Interaction	38
2.2.6 Cross Terms	39
2.2.7 Force Field Parameterization	41
2.3 Energy Minimization and Molecular Simulation.....	41
2.3.1 Energy Minimization.....	42
2.3.1.1 Steepest descent method	43
2.3.1.2 Conjugated gradient method	43
2.3.1.3 Newton-Raphson method.....	43
2.3.2 Molecular/Computational Simulation	44
2.3.2.1 Phase space.....	45
2.3.2.2 Monte Carlo method	45
2.3.2.3 A brief introduction to molecular dynamics	46
2.4 Molecular Dynamics	46
2.4.1 Finite Difference Method: Verlet and Leap-frog Algorithm	47
2.4.2 Time Step and SHAKE.....	48

2.4.3	Molecular Dynamics at Constant Temperature.....	50
2.4.4	Pressure Regulation in Molecular Dynamics	52
2.4.5	Solvent Effects and Water Models	53
2.4.5.1	Implicit solvent models	54
2.4.5.2	Explicit water models	55
2.4.5.3	Periodic boundary conditions and particle-mesh Ewald.....	57
2.5	Free Energy Calculations in Molecular Dynamics	59
2.5.1	Steered Molecular Dynamics	60
2.5.2	Umbrella Sampling.....	61
2.6	Simulating Chemical Reactions and QM/MM	63
2.6.1	QM/MM.....	63
2.6.2	Semiempirical Theory in QM/MM	65
2.6.3	Semiempirical DFT – SCC-DFTB	66
2.6.4	Transition State Theory	68
2.6.5	Kinetic Isotope Effect	71
3	APPLY MOLECULAR DYNAMICS SIMULATION TO PREDICT THE BINDING OF MAGNESIUM CATION IN FANESYLTRANSFERASE ACTIVE SITE	74
3.1	Background.....	74
3.2	Methods and Theories	78
3.2.1	Building Initial Model	78
3.2.2	Modelling and Simulating Mg ²⁺ in FTase	79
3.2.3	Modelling Dβ352A Mutation.....	82
3.2.4	Modeling Removal of Mg ²⁺ from Active Site.....	82
3.2.5	PMF Validation	83
3.2.6	Modelling and Simulation of GGTase I	83
3.3	Results and Discussion	84
3.3.1	Modeling and Simulation of FTase/CVIM/FPP ³⁻ /Mg ²⁺ Complexes	84
3.3.2	Dβ352 Mutant and Mg ²⁺ /H ₂ O Swap Simulation.....	91
3.3.3	PMF Studies	92
3.3.4	Modelling and Simulation of GGTase I	96
3.4	Summary	98
4	SYSTEMATICALLY MECHANISTIC STUDY OF FTASE COMPLEXED WITH DIFFERENT PEPTIDE SUBSTRATES	101
4.1	Background.....	101
4.2	Methods and Simulation Details	103
4.2.1	FTase/CVIM/FPPH ²⁻ and Yβ300F Mutant.....	103
4.2.2	FTase/CVLM/FPPH ²⁻	105
4.2.3	FTase/CVLS/FPPH ²⁻	106
4.2.4	QM/MM Validation of FTase/CVIM/FPP ³⁻ /Mg ²⁺ Model	107
4.3	Results and Discussions.....	108
4.3.1	A Glimpse of Entire Product Formation Process of FTase/CVIM/FPPH ²⁻	108
4.3.1.1	Transition state and free energy barrier	108

4.3.1.2 Comparison of QM/MM and MM PMF of conformational transition step.....	111
4.3.2 Effect of Mutations.....	112
4.3.2.1 Impact of Y β 300F mutation	113
4.3.2.2 Interesting context-dependent effect.....	116
4.3.2.3 Interpretation of α -SKIE results of FTase/CVLS complex.....	117
4.3.3 Mg ²⁺ in Chemical Step	122
4.4 Summary.....	125
5 COMPUTATIONAL APPROACH TO UNDERSTAND AROMATIC PRENYLTRANSFERASE CATALYTIC MECHANISMS	127
5.1 Background.....	127
5.2 Methods and Simulation Details	130
5.2.1 QM/MM Study of NphB Prenylation	130
5.2.2 Modelling and Simulation of Proton Transfer Step in NphB Catalysis	132
5.2.3 Classical and QM/MM study of FtmPT1	134
5.3 Results and Discussions.....	136
5.3.1 Orientation of Substrate Orientation in QM/MM Equilibration of NphB	136
5.3.2 Free Energy Surface of NphB Catalyzed Prenylation	138
5.3.2.1 S1 state.....	138
5.3.2.2 S3 state and S4 state.....	142
5.3.3 Proton Transfer Step of NphB Catalysis.....	145
5.3.4 Substrate Protonation State Preference of FtmPT1	148
5.3.5 A Look at FtmPT1 Catalysis.....	150
5.4 Summary.....	152
LIST OF REFERENCES	155
BIOGRAPHICAL SKETCH.....	168

LIST OF TABLES

<u>Table</u>		<u>page</u>
2-1	Parameters of several water models.	57
3-1	Zinc parameters used in MD simulations	81
4-1	Free energy parameters obtained from QM/MM and MM PMF studies.....	114

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
1-1 Chemical structure of isoprene and prenyl group.	17
1-2 Protein farnesyltransferase (FTase) catalyzed farnesylation.	19
1-3 Protein geranylgeranyltransferase (GGTase) catalyzed geranylgeranylation.	19
1-4 NphB (middle up) catalyzed geranylation.	22
1-5 Indole prenyltransferase FtmPT1 (middle up) catalyzed attachment of dimethylallylic group to brevianamide F (left bottom).	25
2-1 Schematic representations of five components in general MM force field:	32
2-2 Torsion (or dihedral) angles in protein structure.	36
2-3 Schematic representation of some multipole examples.	39
2-4 Schematic representation of angle closing coupling with bonds stretching.	40
2-5 Simple interaction site water models.	56
2-6 Part of 2-D periodic boundary conditions (PBC).	58
2-7 Schematic representation of Ewald summation.	59
2-8 Schematic representation of building PMF with umbrella samplings.	62
2-9 The effect of k_w on umbrella sampling (US).	62
2-10 Schemetic representation of QM and MM partition.	65
2-11 Reaction coordinate diagram for the biomolecular nucleophilic substitution (S_N2) reaction between bromomethane and the hydroxide anion.	69
2-12 Reaction mechanism of S_N1 reaction between S-3-chloro-3-methylhexane and iodide ion that yields both R- and S- 3-iodo-3-methylhexane.	73
2-13 Reaction mechanism of S_N2 reaction between bromomethane and hydroxide that yields ethanol and bromide.	73
3-1 Snapshot of FPP binding pocket. Two reacting centers are labeled as C1 (C_1 of FPP) and SG (S_Y of Cys1p).	75
3-2 Active site snapshot of GGTase I. C_1 of GGPP and S_Y of Cys1p are labeled as C1 and SG.	77

3-3	Snapshots of FTase active site for different systems.....	79
3-4	Distribution of d_{RC} in the original 6 ns MD simulation of WT2 model.	90
3-5	Snapshot of active site from the extend 10 ns MD simulation of WT2 model.	90
3-6	Free energy profiles of the modeled systems.....	94
3-7	Comparison of FPP ³⁻ orientation found in WT2 MD simulation and from previous FTase/CVIM/FPPH ²⁻ (without Mg ²⁺) MD simulation.	95
3-8	Distances variation in the simulation of GGTase/CVIL/GGPP ³⁻ (top) and GGTase/CVIL/GGPPH ²⁻ (bottom).....	97
4-1	QM region selection: side chains of Asp297 β , Cys299 β , His362 β , Cys1p, Zn ²⁺ and diphosphate and the first isoprene unit of FPPH.....	104
4-2	Free energy profile of FPPH model obtained from a set of 63 windows of umbrella sampling. Active site snapshots are given at key points.....	110
4-3	Snapshot of the active site of FPPH at the transition state. O ₁ and C ₁ of FPPH, and S _{γ} of Cys1p are labeled in red.	111
4-4	Free energy profiles of all five models included in this study.	113
4-5	Snapshot of the active site of Y300F mutant at the transition state. O ₁ and C ₁ of FPPH, and S _{γ} of Cys1p are labeled in red.....	115
4-6	Snapshot of the active site of CVLM mutant at the transition state. O ₁ and C ₁ of FPPH, and S _{γ} of Cys1p are labeled in red.....	117
4-7	Free energy profile associated with the conformational transition step of CVLS model.	118
4-8	Active site snapshot of CVLS model at the transition state. O ₁ and C ₁ of FPPH, and S _{γ} of Cys1p are labeled in red.	119
4-9	H ₁ -H ₂ -C ₁ -C ₂ dihedral variation in the farnesylation of FTase/FPPH/CVIM (red) and FTase/FPPH/CVLS (green) complexes.....	120
4-10	Ramachandran plot of a ₂ residue of the Ca ₁ a ₂ X motif throughout the farnesylation of different systems.....	121
4-11	The active site snapshot of WT2 at the transition state. C ₁ and O ₁ of FPP, and S _{γ} of Cys1p are labeled in red.....	124
5-1	QM partition of NphB: GPP and 1,6-DHN. Important atom specifications are given....	131

5-2	Active site configuration of the initial structure selected for proton transfer step simulation.....	133
5-3	Active site configuration of PDB ID 3O2K. DMAPP and Brevianamide F are treated with QM in QM/MM simulation. Important atoms are labeled in red.	135
5-4	1,6 DHN orientation in S1 (left) and S3 (right) State.....	138
5-5	Free energy profile associated with prenylation step for both S1 and S3 model.	139
5-6	Active site configuration at the intermediate state of S1 model. O _{1P} of PPi, C _{1P} of carbocation and C _{5A} of 1,6-DHN are labeled in red.....	140
5-7	H _{1(P)}} -H _{2(P)}} -C _{1(P)}} -C _{2(P)}} dihedral variation in the chemical reaction of NphB catalyzed 5-prenylation (red) and FTase/FPPH/CVIM catalyzed farnesylation (blue).	141
5-8	Active site conformation in S3 model at the intermediate state. O _{1P} of PPi, C _{1P} of carbocation, and C _{2A} of 1,6-DHN are labeled in red.....	144
5-9	Free energy profile associated with S4 model prenylation.	145
5-10	Active site snapshot of S4 prenylation. O _{1P} of PPi, C _{1P} of carbocation and C _{4A} of 1,6-DHN are red labeled.	146
5-11	Free energy profile associated with the S1 proton transfer step.	147
5-12	Active site configuration after proton abstraction from 5-prenyl-1,6-DHN to H3O (WAT in Figure 5-2).....	148
5-13	Carbocation formed in FtmPT1 catalysis. Important atoms are labeled in red.....	151

LIST OF ABBREVIATIONS

2(/n)-D	2(/n)-dimensional
AA	Amino Acid
APTase	Aromatic prenyltransferase
C-terminus	Carboxyl-terminus or COOH-terminus
DFT	Density functional theory
DHN	Dihydroxynaphthalene
DMAPP	Dimethylallyl diphosphate
DMAPP	Dimethylallyl diphosphate
FES	Free energy surface
FPP	Farnesyl diphosphate
FTase	Protein prenyltransferase
G proteins	Guanine nucleotide-binding proteins
GAFF	Generic AMBER force field
GB	Generalized Born
GGA	Generalized gradient approximation
GGPP	Geranylgeranyl diphosphate
GGTase	Protein geranylgeranyltransferase
GPP	Geranyl diphosphate
GPU	Graphics Processing Unit
GTPase	Guanine triphosphate hydrolase
HF	Hartree-Fock
HPPi	Monoprotonated diphosphate leaving group ($\text{P}_2\text{O}_7\text{H}^{3-}$)
IRC	Intrinsic reaction coordinate
KIE	Kinetic isotope effect

KS	Kohn-Sham
LDA	Local density approximation
LES	Local enhanced sampling
LJ	Lennard-Jones
MC	Monte Carlo
MD	Molecular Dynamics
MEP	Minimum-energy path
MM	Molecular mechanics
PB	Poisson-Boltzmann
PBC	Periodic boundary conditions
PDB	Protein data bank
PES	Potential energy surface
PKIE	Primary kinetic isotope effect
PME	Particle-mesh Ewald
PMF	Potential of mean force
PPi	Diphosphate leaving group ($P_2O_7^{4-}$)
PPTase	Protein prenyltransferase
PT barrel	Prenyltransferase barrel
QM	Quantum mechanics
QM/MM	Quantum mechanics/molecular mechanics
RC	Reaction coordinate
RESP	Restrained electrostatic potential
RLS/RDS	Rate limiting(/determining) step
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation

SCC-DFTB	Self-consistent charge density functional tight-binding
SKIE	Secondary kinetic isotope effect
SMD	Steered-MD
TS	Transition state
TSS	Transition state structure
TST	Transition state theory
USP	Umbrella sampling
VDW	Van der Waals
VTST	Variational transition state theory
WHAM	Weighted histogram analysis method
WT	Wild type
ZPE	Zero point energy
ZPVE	Zero point vibrational energy

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

COMPUTATIONAL MODELLING AND SIMULATION OF
PRENYLTRANSFERASES AND CORRESPONDING PRENYLATION

By

Yue Yang

August 2011

Chair: Kenneth M. Merz, Jr.

Major: Chemistry

Nowadays, with the power of supercomputers, a lot of labor extensive work especially in scientific research can be accomplished efficiently. Such a good example is the application of computational simulations into chemistry and biology research, e.g., protein folding, high throughput virtual screening and so on. In fact, due to the limitation of experimental instruments or techniques, many aspects of enzyme and their associate enzymatic reactions have not been understood completely or even clearly, therefore, the advantage of computational simulation makes itself an excellent candidate to solve these problems. In our group, computational studies have been conducted on a variety of enzymes and biological systems, such as human carbonic anhydrase, β -lactamase, prenyltransferase and even interactive enzyme-RNA systems. Among these functional important enzymes, prenyltransferase has attracted more and more attention in recent years. As an important candidate for anti-cancer drug design, protein prenyltransferases play a role in the posttranslational modification of Ras superfamily of enzymes and many others that have appeared in about 30% of human cancers¹. In this dissertation, we focus on our effort of conducting computational study to obtain knowledge from various prenyltransferases and corresponding enzymatic reactions catalyzed by them.

Chapter one gives an overall introduction of terms such like prenyltransferase, prenylation, as well as the prenyltransferase enzymes that will be discussed in this dissertation. The background and (pharmaceutical) significance of this class of enzymes will also be explained. Chapter two is the method section, where computational theories, techniques mentioned in this dissertation, along with difficulties of our approach and the strategy employed to tackle the problem will be introduced.

The next three chapters describe the work we have already accomplished in this big project, including our prediction of magnesium ion binding in the FTase and related validation, the catalytic mechanism of FTase catalyzed farnesylation along with impact of key mutation and the prenylation mechanism and the proton transfer pathway of NphB.

With computational modeling and sampling, we have solved these chemical/biological problems that experiments cannot explain and revealed a lot of interesting facts of how enzymes adopt different strategies based on different environments or situations.

CHAPTER 1 INTRODUCTION

1.1 Prenylation and Prenyltransferase

Prenylation, also called isoprenylation, refers to a type of attachment of a hydrophobic function group called isoprenoid or terpenoid. Isoprenoid group is the derivative of a five-carbon isoprene unit. Such a group is usually called a prenyl group (Fig. 1-1) and commonly appears in this kind of lipid modification reactions. The prenylation is of great importance to researchers because this reaction enables certain enzymes, such like Ras superfamily of enzymes, to anchor into the inner side of plasma cellular membrane², as well as modifies the biological activity of certain aromatic complexes³⁻⁵.

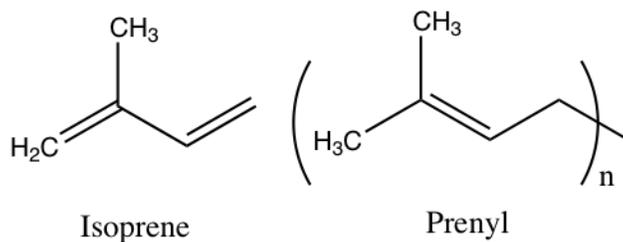


Figure 1-1. Chemical structure of isoprene and prenyl group.

Ubiquitous in natural products, isoprenoids are very important chemicals that can be found in almost every class of living things, including plants, fungi, animals and organisms⁶. Among them, plant terpenoids are extensively used because of their aromatic qualities. A number of plant terpenoids are well known, such as citral, menthol, camphor, cannabinoids and Salvinorin A. Their contributions to nature, such as the color of yellow flowers and the flavor of cinnamon and ginger, have been well recognized and studied. In addition, this kind of plant terpenoids plays an important role in traditional herbal remedies. Recently, investigation showed that some isoprenoids have certain pharmaceutical desired characteristics, such like antibacterial, antineoplastic, anticancer effects, making this kind of chemicals more important.

More importantly, several members of this isoprenoid family play a crucial role in the signal transduction pathway. Examples of these terpenoids include the 15-carbon farnesyl and the 20-carbon geranylgeranyl^{2,7}. The prenylation that attaches one of these two groups to a cysteine residue at or near the C terminus of the target intracellular enzyme is usually referred to as protein prenylation. Many proteins including various members of Ras superfamily require such a posttranslational attachment for certain functions.

1.2 Protein Prenyltransferase

Commonly, enzymes that catalyze the prenylation reaction are called prenyltransferase. Consequently, the enzyme that catalyzes the protein prenylation is usually termed as protein prenyltransferase^{2,8-14}. There are three members in this family: protein farnesyltransferase (FTase)^{7,15-18}, protein geranylgeranyltransferase type I (GGTase I)¹⁹⁻²³ and type II (GGTase II)^{8,23-25}. Protein prenyltransferases can be further categorized into two subclasses: the CaaX prenyltransferase subclass that includes FTase (Figure 1-2) and GGTase I (Figure 1-3), and Rab GGTase subclass, which contains GGTase II. Suggested by the name, CaaX prenyltransferases are in charge of the catalysis of attaching a farnesyl group or a geranylgeranyl group to a structurally fixed cysteine residue at the fourth position from the C terminus of the target enzyme. In this four-letter term, C stands for cysteine, letter 'a' refers to aliphatic AA, while X is usually taken by alanine, serine, methionine or phenylalanine in FTase targets and occupied by either leucine or phenylalanine for GGTase I catalysis^{26,27}. On the other hand, GGTase II is distinct from both FTase and GGTase I due to the fact that it mainly attaches two geranylgeranyl groups to two C-terminus cysteine residues in a Cys-Cys, Cys-X-Cys or Cys-Cys-X-X motif in Rab family of enzymes^{13,28}.

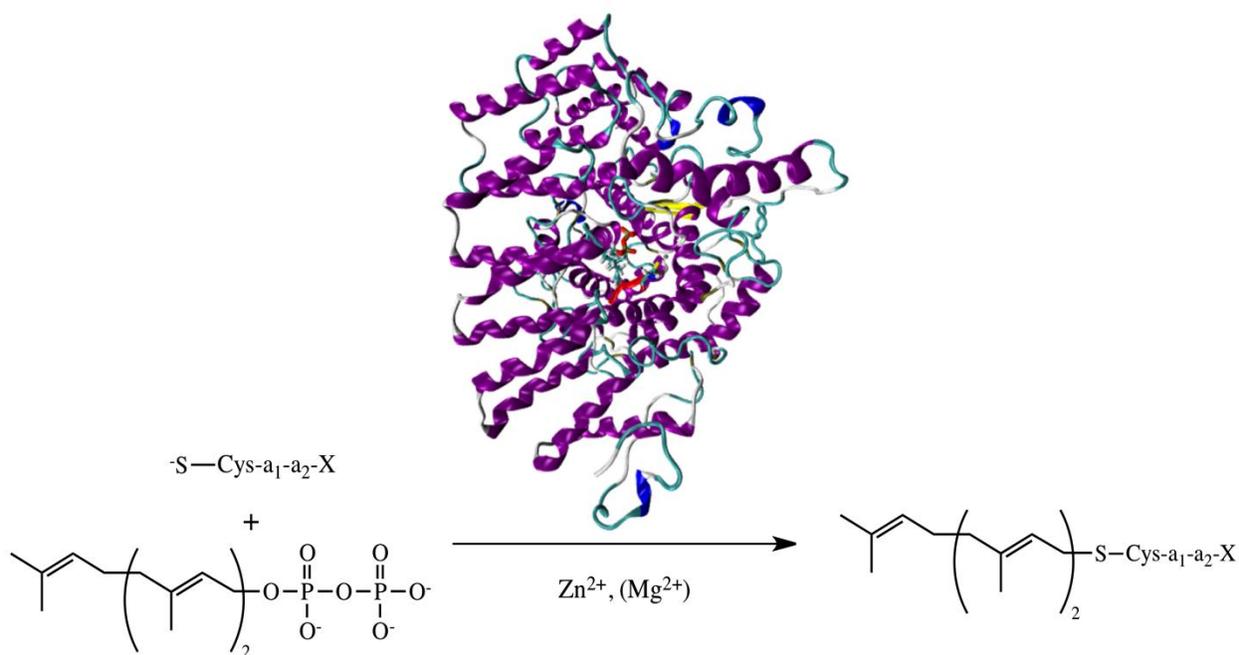


Figure 1-2. Protein farnesyltransferase (FTase) catalyzed farnesylation.

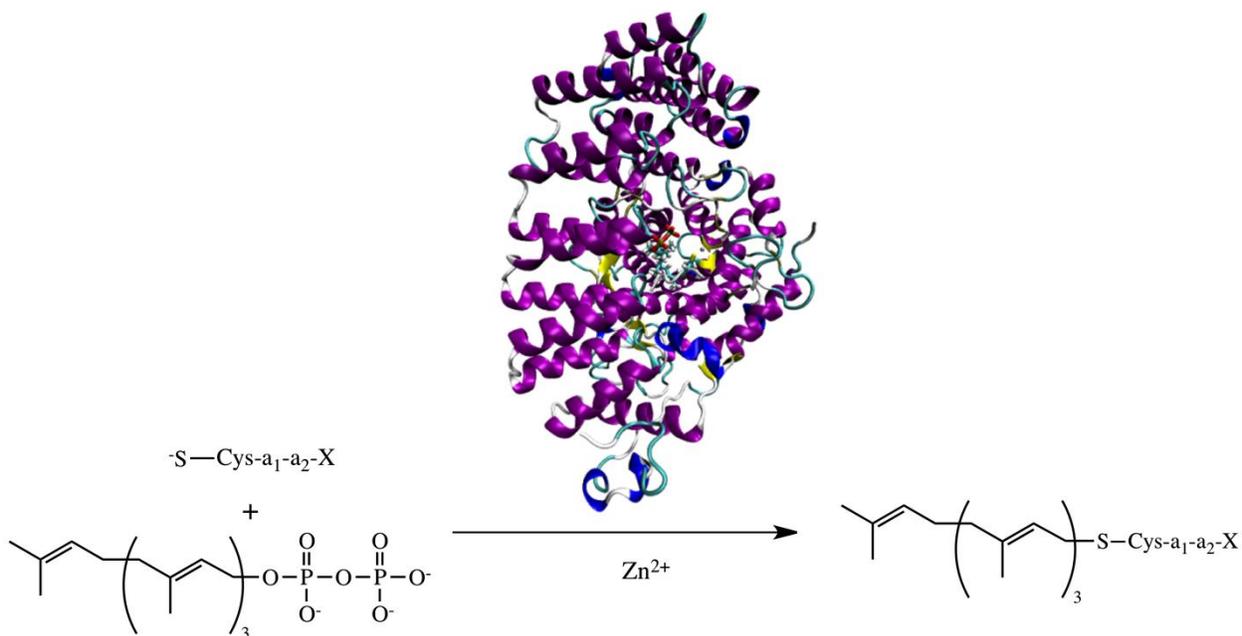


Figure 1-3. Protein geranylgeranyltransferase (GGTase) catalyzed geranylgeranylation.

Many experimental studies, in particular those sequence studies, has uncovered some secrets of the CaaX motif: first, Reiss and coworkers made further classification of Ca₁a₂X and revealed that a charged residue at a₁ position would slightly reduce the affinity while at a₂ position or X position would be associate with more drastic reduction²⁹; later Troutman and coworkers have discovered that different Ca₁a₂X motifs would give different stimulate to the product release, which would further give FTase different target selectivity³⁰; more recently, Houglund and coworkers investigated the dependency of FTase catalysis rate constant on variant Ca₁a₂X motifs and claimed the recognition of target substrate by FTase is not invariant but context-dependent³¹. In addition, although FTase and GGTase I primarily target on the Ras superfamily of enzymes possessing this CaaX motif, various short peptides with this same motif can also be recognized and catalyzed by these two prenyltransferases^{2,9}.

Because of their important pharmaceutical potential, many studies have been conducted to disclose the nature of these enzymes. As a result, a great deal of useful structural information of these PPTases has been exposed. FTase is a heterodimer composed of a 48-kDa α subunit and a 46-kDa β subunit. GGTase is also a heterodimer composed of two subunits. Interestingly, it shares an identical α subunit with FTase while possesses a β subunit of 43-kDa that is slightly smaller than FTase β subunit. GGTase II is distinctly different to FTase and GGTase I, although it is a heterodimer too. This enzyme possesses a much bigger α subunit which is 60-kDa and the smallest β subunit which is only 38-kDa.

All of three transferases are metalloenzymes, however, with different metal dependency. FTase requires a zinc ion for substrate binding and chemical reactivity, and a second magnesium cation for optimal reactivity^{18,32}. GGTase I also requires a zinc ion for substrate binding and catalysis activity, but does not need a magnesium ion, which is possibly due to the substitute role

played by Lys311 β ^{18,20,21}. GGTase II, on the other hand, appears to require a single Mg²⁺ for reactivity but not showing any dependency on Zn²⁺, making the metal dependency of this family of enzymes more intricate and complicated.

Each of three prenyltransferases binds a corresponding isoprenoid substrate and an enzyme (or peptide) substrate prior to catalysis. The substrates in FTase are farnesyl diphosphate (FPP) and an enzyme among Ras GTPases, β lamins or several enzymes involved in the visual signal transduction pathway, or short peptides with the CaaX motif. GGTase I adopts geranylgeranyl diphosphate (GGPP) as isoprenoid substrate and an enzyme substrate of either a Ras related GTPase or one of the γ subunit of heterotrimeric G proteins. GGTase II has the same isoprenoid substrate as GGTase I but targets primarily on Rab family of enzymes that possess those specific double Cys motifs.

Although the binding of diphosphate substrate does not require the binding of zinc ion in both FTase and GGTase I, such a metal binding is the prerequisite condition for either enzyme in order for the binding of enzyme or peptide substrate^{19,33}. Therefore, it explains why a zinc ion is indispensable to the activity of CaaX prenyltransferase, despite the fact that FTase also requires a magnesium ion to fully restore its activity. However, from experimental results, it is still not clear whether Zn²⁺ has the same function in both enzymes. It is also unclear that whether the zinc ion play a solo structural role or it also directly involves in the catalysis. An assumption is made that the zinc ion activates the sulfur or sulfhydryl of the substrate cysteine residue in CaaX motif, making it more nucleophilic, as observed in Ada, an enzyme that is known for DNA repair via a catalysis similar to the prenylations discussed here.³⁴.

In this dissertation, effort will be made focusing on FTase and GGTase I. GGTase II, on the other hand, will not be widely discussed.

1.3 Aromatic Prenyltransferase

An incredible increase in the amount of studies focused on prenylation has been seen in the past ten to twenty years. Though a fairly large portion attributed to PPTases that catalyzed the posttranslational S-prenylation required by certain enzymes in order for the later penetration into the plasma membrane, other prenyltransferase, such as those capable of the lipidic modification of various aromatic substrates, have also been identified and investigated^{5,35}. This type of prenyltransferases is known as aromatic prenyltransferases (APTase)³⁶⁻⁴⁰. Examples of such APTase include NphB (also called Orf2 before rename), a 33 kDa soluble monomeric bacterial prenyltransferase with 307 residues expressed in *Escherichia coli*³, and FtmPT1, a 464-residues fungal indole prenyltransferase⁴.

1.3.1 Bacterial Aromatic Prenyltransferase NphB

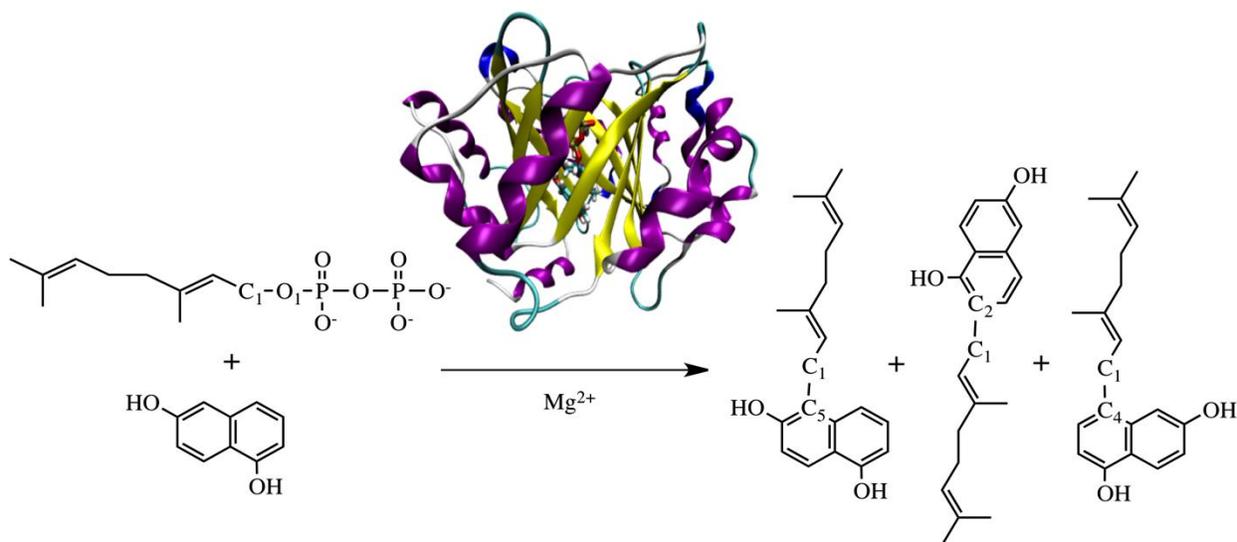


Figure 1-4. NphB (middle up) catalyzed geranylation. C₅-prenylation corresponds to the major product. C₂-prenylation represents the minor product, which is in a product ratio of 1:10 to the major product. C₄-prenylation refers to the extra-minor product.

NphB (Figure 1-4) is obtained from *Streptomyces* sp. Strain CL190³. It is originally called Orf2 that is named after being revealed as one of three new open reading frames (ORFs) in the

upstream region of MVA pathway genes containing gene clusters. It shows fairly high degree homologies to three other bacterial prenyltransferases: HypSc (GenBank accession # AF939130), CloQ (GenBank accession # AF329398) and NovQ (GenBank accession # AF170880)⁵. However, with regard to 3-dimensional (3-D) structure, substrate selectivity, and metal catalyst dependency, it exhibits distinct difference to the other three APTases mentioned above, not to mention any other PPTases^{7,18,41-45}.

CloQ, NovQ and HypSc display strong selectivity of five-carbon DMAPP substrate but show little activity for isoprenoid diphosphate with longer carbon chain, such like ten-carbon geranyl diphosphate (GPP) or fifteen-carbon FPP. In addition, all three are independent to magnesium catalysis^{3,5}. To the contrary, NphB shows significant Mg^{2+} dependency. It exhibits no reactivity to dimethylallyl diphosphate (DMAPP), but strong activity to GPP and relative low but observable reactivity to FPP. Additionally, NphB also displays flexible aromatic substrate selectivity. Its known substrates include flaviolin, 1,6-dihydroxynaphthalene (1,6-DHN), 2,7-DHN, apigenin and many others.

Structurally, NphB consists of a single domain. The 3-D structure highlights a novel barrel fold, which is termed as prenyltransferase barrel (PT barrel), consisting of ten anti-parallel β -sheets surrounded by a circle of α -helices that are exposed to solvent. Although this barrel looks similar to the TIM barrels, secondary structure of NphB shows no sharing with TIM barrels⁴⁶⁻⁴⁹. The ternary complexes of NphB with Mg^{2+} , GSPP, a GPP analog and 1,6-DHN or other aromatic substrate reveals a solvent accessible core and a spacious binding pocket inside provided by this PT barrel. Such a voluminous binding pocket allows the binding of a large variety of aromatic substrates, thus partially explaining the flexible substrate selectivity exhibited by NphB.

Although this spacious binding pocket is capable of binding various prenyl diphosphates including DMAPP, GPP and FPP, NphB displays highest reactivity to GPP³. Inside the binding pocket, GPP is encircled by several positively charged residues, including Lys119, Arg228 and Lys284. These positive charge residues, along with Tyr216 and Asn173, tether the diphosphate group through hydrogen bond interactions, helping to anchor the isoprene substrate. Magnesium provides additional stabilization to GPP by interacting with a non-bridge α -diphosphate oxygen atom through an octahedral coordination. The coordinating ligands also include Asp62 and four water molecules. However, a (N/D)DXDD signature motif observed in many other Mg²⁺-dependent isoprenoid diphosphate containing enzymes is absent in NphB.

Unlike CloQ, NovQ and HypSc, NphB shows significant Mg²⁺ dependency. Whether this metal cation directly involved in the catalysis is yet known. What has not been explained well also is the regioselectivity displayed in the system of NphB complexed with Mg²⁺, GPP and 1,6-DHN. Two products were originally isolated with the isoprenoids group attaching to different carbon atoms each neighbouring to one of the hydroxide connecting carbon atoms. Under the condition that each substrate reaches its saturation, the production ration is 10:1. Later, more geranylation location of 1,6-DHN has been discovered, although the production ratio was even lower.

Evidences showed that certain polyketide-based aromatic complexes, such as the anti-oxidant natural product naphterpin, have quite different biological activities to their isoprenoid-modified derivatives⁵⁰⁻⁵³. In addition, these aromatic derivatives that possess either a 5-carbon dimethylallyl, a 10-carbon geranyl or 15-carbon farnesyl displayed desired pharmaceutical properties such as anti-microbial, anti-oxidant, anti-inflammatory, anti-viral and anti-cancer

effect. Therefore, it is of great importance to study NphB chemistry computationally to complete our knowledge of this enzyme.

1.3.2 Fungal Indole Prenyltransferase FtmPT1

FtmPT1 (Figure 1-5) is another APTase that has just been purified from *Aspergillus fumigatus* in recent years⁴. Unlike NphB or CloQ that was obtained from bacteria, FtmPT1 and several other indole prenyltransferases were characterized from fungi, thus they were usually referring to as fungal indole prenyltransferase. In general, these fungal indole prenyltransferases share very few sequence similarities with other APTases and show no divalent metal ion dependency^{54,55}. This type of prenyltransferases catalyzes the isoprenoid attachment of tryptophan and its derivatives. The product of such catalysis, such like tryprostatin B and tryprostatin A, display high cytotoxicity for several cancer cell lines, making themselves promising anti-tumor agents⁵⁶⁻⁵⁸.

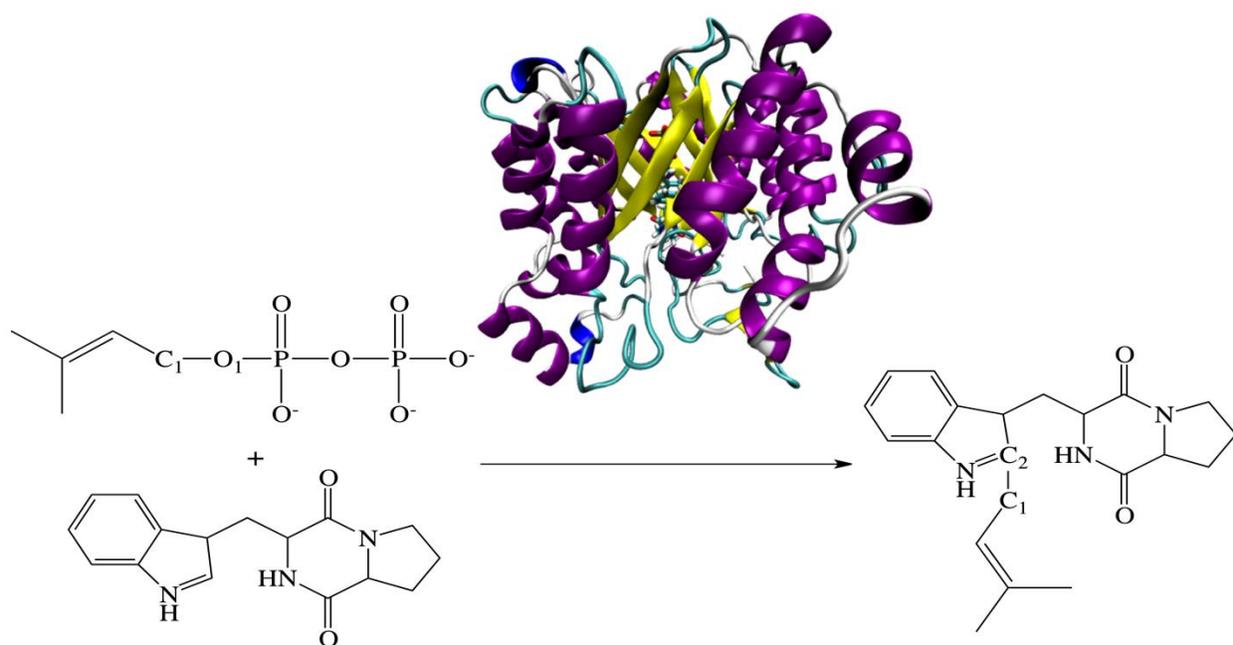


Figure 1-5. Indole prenyltransferase FtmPT1 (middle up) catalyzed attachment of dimethylallylic group to brevirianamide F (left bottom). The product is tryprostatin B (right).

The recently characterized 3-D structure of FtmPT1 features a PT barrel consisting 10 anti-parallel β -strands surrounded by 10 α -helices, similar to that observed in NphB. The secondary structure of FtmPT1 conforms a classical $(\alpha\alpha\beta\beta)_5$ representation with two α -helices coupled two β -sheets in each $(\alpha\alpha\beta\beta)$ turn. Two substrate, dimethylallyl diphosphate (DMAPP) and brevianamide F, situate in the center of the PT barrel. Similar to NphB, the active site in FtmPT1 is also spacious and solvent accessible.

Despite similarities in the core structures, the active sites of FtmPT1 and NphB are distinct from each other. A signature “tyrosine shield” in the center of PT barrel has been observed in FtmPT1. The hydroxyl groups of Tyr203, Tyr296, Tyr382 and Tyr450 make up a hydrophobic circle therein binds brevianamide F and DMAPP. Although NphB also possesses an aromatic rich environment inside its binding pocket, this “tyrosine shield” seems more interesting. With the dimethylallyl group completely embedded, it seems like specifically designed to stabilize and engineer the potential carbocation through cation-aromatic π interaction.

Comparing AA sequence of FtmPT1 and homologue enzyme FgaPT2, which is another fungal indole prenyltransferase possessing a PT barrel and represents an excellent superposition (RMSD of C_α based on 402 residues measures 1.3) with FtmPT1, differences at several key residues are discovered and worth noting. For instance, Gly115 and His279 (in FtmPT1) are replaced by Thr102 and Arg244 in FgaPT2, respectively. These two substitutions spare more space in the binding pocket of FtmPT1, allowing the enzyme to bind larger indole substrate, such like brevianamide F. On the other hand, Glu102 of FtmPT1 has been found conserved in nearly all fungal indole prenyltransferases. This negative charged residue forms hydrogen bond with indole nitrogen, stabilizing the substrate. It also possibly functions to abstract the abundant proton after the prenylation step, in order to reestablish the aromatic quality of the prenylated

indole substrate. In addition, it is interesting to notice that a small modification of AA sequence can result in the attachment of dimethylallyl group from a different DMAPP carbon to a different carbon atom in indole substrate.

Fungal indole prenyltransferases display interesting diversity and play an important role in biosynthesis of a variety of chemically important compounds^{54,55}. In this dissertation, progress in the study of FtmPT1 is reported.

1.4 Computational Chemistry/Biochemistry

The past 30 years have seen significant improvement in computer science and technology. The introduction of workstations and supercomputers has opened the door to ‘huge’ calculations, generating a new branch of chemistry that utilizes the assistance from computer science to solve chemical problems. Such branch of chemistry is usually known as computational chemistry.

Computational chemistry has been well known for its capability of calculating properties such as absolute or relative energies, electronic charge distributions and vibrational frequencies, based on the application of quantum mechanics (QM). Back into the 1940s, the idea of solving complicated and labor-intensive wave functions for atomic systems has already come out. In the 1950s the first semi-empirical atomic orbital calculation and Hartree-Fock (HF) calculation of diatomic molecule have been carried out. Later, in the 1960s, the empirical methods based on Hückel method were introduced. Coming into the 1970s, methods of molecular mechanics (MM) were developed. As of today, more advanced QM calculations and more complex hybrid QM/MM simulations have been widely implemented resulting in many exciting results.

The implementation of computational chemistry used to be strictly limited by several factors such as the size of the system, due to the rapidly increasing in computational expense associated with the implementation of higher level theory or the calculation of larger system. In present, with supercomputers and parallel computing, it is common to perform highly accurate

coupled cluster calculations on large organic molecules, or carry out molecular dynamics (MD) simulations on large enzymes, DNA's, RNA's and even bigger systems. The implementation of graphics processing unit (GPU) allows 100 ns simulation per day for classical MD simulations of a system composed of tens of thousands atoms, making *in silico* investigation of microsecond level or even bigger time scale biological activity possible. Moreover, with super fast machines specifically optimized for MD simulation, such as Anton introduced by D. E. Shaw Research Lab, a 10 μ s/day of classical simulation of a 20,000+ atoms system has become feasible⁵⁹.

More and more powerful computer techniques have pushed computational chemistry into diverse research areas, such as bioinformatics. The ability to build and analyze huge databases with high efficiency has made computational approach a very important part of modern drug design in pharmaceutical industries. By providing an alternative and more environmental way to study the properties of drug molecules, the correlations between structures and properties, as well as the possibility of finding or even designing a more suitable substrate, computational chemistry has saved countless labors and other lab expenses, making huge impact in industry.

Today, with less expensive computational expense and higher accuracy, computational chemistry has already become an indispensable branch of chemistry and biochemistry. It is able to investigate many areas that are not accessible by traditional experimental approaches. In addition, it can also make valuable predictions that can in turn guide experiments designing.

In this dissertation, computational effort has been made on mechanistic study of several important prenyltransferases. Great results and interesting predictions are reported.

CHAPTER 2 THEORY AND METHODS

Before computers became popular in scientific research, molecular modelling was used to associate with the utilization of pens and paper to make simple calculations of small molecules mimicking their behaviors. In present, with the rapid development of computer science and technologies, molecular modelling becomes rarely independent to computational chemistry. Utilizing the power of supercomputers and parallel computing, molecular modelling has been able to apply onto macromolecules and even bigger systems like enzymes. Thus it becomes more and more important in the researches of chemistry, biology, drug design and many other areas. Today, molecular modellers have implemented all available computational chemistry methods, including QM, MM, MD, minimizations, simulations, conformational analysis, protein-ligand interactions, chem/bioinformatics and other computer-based techniques, to elucidate and predict molecular behaviors.

In the rest of this chapter, many important concepts and methods in molecular modelling will be discussed.

2.1 Quantum Mechanics and the Born Oppenheimer Approximation

The fundament and heart of computational chemistry based molecular modelling is QM, which was introduced to describe the behavior of microscopic particles. Today, although molecular modelling does not necessarily mean QM, every successful model must be able to find its basis in quantum mechanics.

It is well known that in classical world, macroscopic systems obey Newtonian mechanics. Thus, the change of their energies is continuous. However, rules in the microscopic world are quite different: particles that form matter have a characteristic of quantization. As a result, their energies are discrete. In fact, microscopic systems possess both wavelike and particle-like

properties. This characteristic makes no pre-existing theory, including wave mechanics that possesses the quantization characteristic, capable to explain their behavior. Therefore, a new mechanics capable to interpret this dichotomy is required. It was under such a circumstance that QM was introduced.

The base of QM is the assumption that wave functions can be applied to microscopic systems and can describe all their properties. In detail, the properties are characterized by applying corresponding operators onto the wave functions. Each observable physical property is assumed to have one such operator.

The most important operator in QM theory is the Hamiltonian operator, H , which describes and returns the system energy. The mathematical equation of such a relationship is

$$H\Psi = E\Psi \tag{2-1}$$

Equation 2-1 is commonly known as Schrödinger equation. Hamiltonian operator generally consists of five parts: the 1st and 2nd part describes the kinetic of electrons and nuclei, the 3rd part accounts for the attraction between electrons and nuclei, the 4th part attributes to interelectronic repulsion while the 5th part corresponds to the internuclear repulsion.

Simply speaking, by solving the Schrödinger equation, the energy of the system will be obtained. However, in practice, this is difficult to accomplish since Schrödinger equation cannot be exactly solved for any molecular systems due to the interdependency between electronic motion and nuclear motion. Therefore, another fundamental postulate is required to simplify the case, which is the famous Born-Oppenheimer approximation.

Considering the fact that either protons or neutrons are more than 1000 times heavier than electronics, it is rational to assume that the motion of electrons is instantaneous compared to

nuclear motion. Thus, to calculate the electronic energies is possible since the nuclear motion “has been fixed” and the electronic Schrödinger equation has a reform of:

$$(H_{EL} + V_N)\Psi_{EL} = E_{EL}\Psi_{EL} \quad (2-2)$$

Here, the subscript ‘EL’ and ‘N’ mean electronic and nuclear, respectively.

As Born-Oppenheimer approximation is unconditionally applied in any case of molecular modelling, a system’s energy of its ground electronic state can be considered as a function of the nuclear coordinate. In such a case, any change of nuclei positions varies the energy of the system. The changes of energy against the all of the possible nuclear or atomic positions construct a hyper surface denoted as potential energy surface (PES). The stationary points on the PES are of particular interest to computational chemists because at those points the first derivative of energy with respect to the coordinate is zero, the physical meaning of which demonstrates the forces at certain points are zero. Both local minima and saddle points are stationary points, however, in different types.

2.2 Molecular Mechanics

QM deals with electrons in molecular system, usually giving the highest accuracy among all molecular modelling methods. However, it is not occasional for molecular modellers to find that the problems they are going to solve are too large to be considered quantum mechanically. Molecular mechanics (also called force field methods) are introduced in such a situation to handle those “oversized” systems.

With the implementation of Born-Oppenheimer approximation, it is possible to compute the energy of a system as a function of nuclear coordinates with electronic motion ignored, which is the basis of MM. Comparing with QM, obviously MM uses a simpler model to mimic interactions with generally five components included, amongst bonding stretching, angle bending

and bond torsion that all represent the energy penalties associated with deviation from their reference position are considered bonded interactions, whereas the other two terms, van der Waals (VDW) and electrostatics are categorized into non-bonded interactions. Thus, the potential energy of a MM modeled system can be written in the following form:

$$\begin{aligned}
 U = & \sum_{\text{bonds}} \frac{k_i}{2} (l_i - l_{i0})^2 + \sum_{\text{angles}} \frac{k_i}{2} (\theta_i - \theta_{i0})^2 + \sum_{\text{torsions}} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) \\
 & + \sum_{i=1}^N \sum_{j=i+1}^N 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}
 \end{aligned}
 \tag{2-3}$$

In equation 2-3, the first three terms on the right are bond stretching, angle bending and bond torsions separately, while the latter two terms are VDW and electrostatics, respectively. The schematic representations of these five terms are given in Figure 2.1.

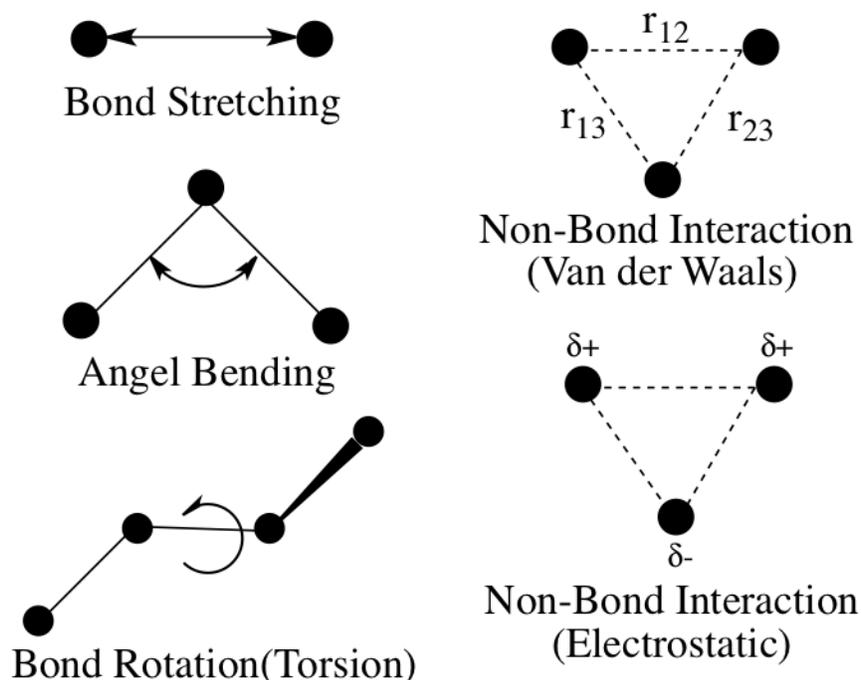


Figure 2-1. Schematic representations of five components in general MM force field: bond stretching, angle bending, bond torsion, non-bonded VDW and electrostatic.

Today, more advanced MM force field such as MM4⁶⁰⁻⁶² have been introduced, however, the five contributions above nearly exist in every known MM force field. It is very important to bear in mind that all MM force fields are empirical, which means there is absolutely no correct form for any force fields. Also, it is worth noting that the transferability, which means whether the same set of parameters can be extended to successfully model a set of different molecules or enzymes, is the most important feature for every force field. Otherwise, although a certain force field can well model the system that it was parameterized against, another desired function of molecular modelling, the ability to make predictions (to other systems), is lost.

2.2.1 Bond Stretching

Taking a careful examination at equation 2-3 again, the bond stretching term is in the form of Hooke's law, as shown in equation 2-4, however, it is in fact the second order truncation of Taylor expansion about the potential energy function a bond at length l_i based on the assumption that the potential at equilibrium bond length l_{i0} is zero.

$$U_{bond} = \frac{k_i}{2} (l_i - l_{i0})^2 \quad (2-4)$$

This truncation makes the energy function working excellent to bond stretching near the reference points but brings physically unrealistic energy when the bond is largely stretched. A practical solution is to include the cubic term:

$$U_{bond} = \frac{1}{2} [k_i + k_i^{(3)}(l_i - l_{i0})] (l_i - l_{i0})^2 \quad (2-5)$$

Here the superscript '(3)' refers to the cubic force constant. The negative cubic force constant can then counteract the unrealistic high potential energy associated with large stretching.

However, equation 2-5 also leads to a chaotic result that the lowest potential energy represents bonds dissociation. Again, the solution is to include the quartic term in the Taylor expansion:

$$U_{bond} = \frac{1}{2} [k_i + k_i^{(3)}(l_i - l_{i0}) + k_i^{(4)}(l_i - l_{i0})^2](l_i - l_{i0})^2 \quad (2-6)$$

Equation 2-6 is used in the general organic force field MM3⁶³⁻⁶⁵. However, including higher order terms in Taylor expansion increases the accuracy as well as the computational cost distinctly, thus equation 2-4 is still very popular in today's MM force fields.

In fact, there is an alternative way to provide higher accuracy of describing potential energy contribution associated with bond stretching, the Morse function,

$$U_{bond} = D_i [1 - e^{-\alpha_i(l_i - l_{i0})}]^2 \quad (2-7)$$

Here D_i is the dissociation energy of bond and α_i is a fitting constant. However, this approach has been proved much computationally less efficient than those truncations of Taylor expansion.

2.2.2 Angle Bending

The second term in equation 2-3 is to calculate the potential energy deviation associated with angle bending.

$$u_{angle} = \frac{k_i}{2} (\theta_i - \theta_{i0})^2 \quad (2-8)$$

Again, in spite of the form of Hooke's law, equation 2-8 indeed comes from another second order truncation of Taylor expansion. It is worthy noting that the force constant of angle bending is significantly smaller than that of bond stretching, due to the energy required to disturb an angle away from its equilibrium pose is much less than to compress or withdraw a bond. Like any other truncation of Taylor expansion, including higher order terms associates with greater accuracy. One such example is MM3, which includes up to sextic term in the expansion for certain angles. Certainly, the computational expense increases when more terms in the expansion are included.

2.2.3 Torsion and Improper Torsion

The third term in equation 2.3 reflects the potential energy change associated with the deviation of dihedral angles or torsion angles, as shown in equation 2-9.

$$u(\omega)_{torsion} = \sum_{n=0}^N \frac{V_n}{2} [1 + \cos(n\omega - \gamma)] \quad (2-9)$$

ω in equation 2-9 is the torsion angle; n is the multiplicity that means how many energy minimum can be located through a 360 ° scan; γ is called the phase factor, defining the location of energy minimum; and V_n is the barrier height, which refers to the relative barrier of rotation.

The torsion term might be considered as the most important term among the bonded interaction components, because a large amount of energy is required to alter the reference conformation for both bond stretching and angle bending terms while torsion term and non-bonded interaction terms accounts for most variation in potential energy and system structure.

Sometimes, another type of torsion is also required to computationally reproduce experimental structures. This type of torsion is always defined as improper torsion (as shown in equation 2-10, which means a torsion angle is not bonded in the order of 1-2-3-4. One good example is cyclobutanone. Experimental results exhibited that the oxygen atom is in the same plane as the cyclobutane ring, whereas computing without such an improper torsion results in an out of plane oxygen atom. The reason for this discrepancy is simple, in real world the C-C=O angle is 133 ° in order to have oxygen atom in the plane, however, in calculation this angle still adopts a value close to its reference value of 120 °.

$$u(\omega)_{improper} = k(1 - \cos 2\omega) \quad (2-10)$$

In proteins, dihedrals play very important role in determining their structures. There are three torsion angles in protein structure (Figure 2.2): the dihedral of C-N-C $_{\alpha}$ -C is called ϕ , the dihedral of N-C $_{\alpha}$ -C-N is called ψ , while the dihedral of C $_{\alpha}$ -C-N-C $_{\alpha}$ is denoted as ω . Among them,

ω is usually referred as the rigid peptide dihedral angle since its value is always close to 180° , while the other two torsions have certain range of values thus they can control the protein's conformation.

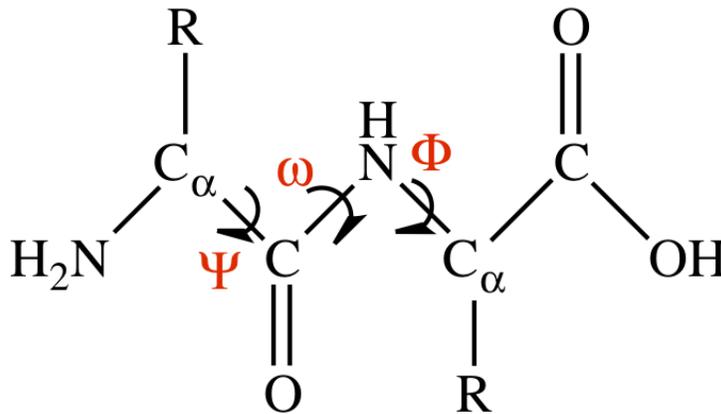


Figure 2-2. Torsion (or dihedral) angles in protein structure.

2.2.4 Van der Waals Interaction

The interatomic force is usually divided into two parts: the long-range attractive force and the short-range repulsive force. The former is commonly known as dispersive forces, as well as London force in order to honor London's achievement on explaining the dispersive force with quantum mechanics⁶⁶.

The origin of such a dispersive force is the instantaneous electrical dipole moments developed by atoms. In general, without a permanent charge, the dipole-dipole interaction resulted from the attraction between an instantaneous dipole and another dipole in the neighbor atoms induced simultaneously by the first dipole, is the strongest interaction of such a dispersive force. Drude's model provides a simple description of such a dispersion interaction with the Schrödinger equation of this model is in the same form as that of a simple harmonic oscillator. Thus the dispersive interaction energy can be approximately given by equation 2-11,

$$u(r)_{dispersion} = -\frac{\alpha^4 \hbar \omega}{2(4\pi\epsilon_0)^2 r^6} \quad (2-11)$$

Obviously, the dispersion energy is dependent on $1/r^6$.

The dispersion force describes the attractive interaction when two atoms (or particles) are quite away from each other. When they approaching each other into a certain range, for example, within 3 Å, even a very small decrease in distance will result in large increase in potential energy. The quantum mechanical basis of this phenomenon is any two identical fermions may not occupy the same set of quantum numbers, as described in Pauli exclusion principle (though some people believe it is actually attribute to nuclear repulsion). Thus the repulsive force in short range is usually known as exchange force.

It is possible to calculate the dispersive force and exchange force quantum mechanically, however, certainly with a high computational cost since electron correlation must be included in the treatment. In MM, the VDW potential is modeled using a much simpler empirical expression, the Lennard-Jones 12-6 function.

$$u(r)_{VDW} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (2-11^*)$$

In equation 2-11*, σ is the collision parameter that physically corresponds to the distance associated with a zero potential energy, and ϵ is referred to the well depth that represents the minimum potential energy. Clearly the sixth power term inside the square bracket is referring to the dispersive force that has been shown to have an exponentially dependence of $1/r^6$, and the other term is corresponding to the exchange repulsive force that varies as $1/r^{12}$. However, it is worth noting that this twelfth exponential does not have a significant physical meaning, but instead has a favor in calculation since it is simply the square of $1/r^6$.

There are other VDW energy functions introduced to improve on reproducing the experimental data, such as Halgren buffered 14-7 potential⁶⁷⁻⁷². However, Lennard-Jones (LJ) 6-12 potential remains the most popular choice of calculating potential energy associated with VDW interactions.

2.2.5 Electrostatic Interaction

The electrostatic interaction is caused by the unequal distribution of charge in a molecular system. By sacrificing the accurate electron-electron interactions, MM usually adopts a so-called point charges approach to represent the charge distribution of a molecule. Restricted to atomic centers, the point charges are commonly known as partial atomic charges or net atomic charges. Then it is straightforward to apply Coulomb's law to calculate the electrostatic energy, as shown in equation 2-12, where N_A and N_B are the number of points charges in two molecules.

$$U_{electrostatic} = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2-12)$$

Central multipole expansion provides another way to calculate intermolecular electrostatic interaction energy. This method considers a molecule as an entity and separate electrostatic interaction into different electric moments such as charge, dipole, quadrupole, octopole and so on (Figure 2-3). With multipole can be described as a proper distribution of charges, dipole, quadrupole and octopole can be represented by two, four and eight charges correspondingly, making calculations simpler. One good example of using such an expansion is the benzene model⁷³, which demonstrates significant advantage of including quadrupole. However, on the other hand, this kind of multipole expansion can be troublesome when it is applied to large systems and cannot be employed to describe intramolecular interactions.

The implementation of fixed point charges provides a rather simple and straightforward electrostatic function based on Coulomb's law. However, it is important to notice that the change

of distribution of an atom or a molecule can induce dipole to neighboring atoms or molecules. Therefore, in order for better description of electrostatic interactions, it is necessary to include not only interactions between fixed charge distributions but also interactions caused by polarization. To describe polarization is not a straightforward task, but somewhat tricky and complex, because one must bear in mind that when a molecule's induced dipole affects the charge distribution of a neighbored molecule its own charge distribution is in turn be modified by the induced dipole of the neighboring molecule. Many approaches have been proposed to model polarization effect⁷⁴⁻⁷⁷. However, due to the significant increase of computational cost, the polarization effect is not included in all force field but instead considered only in the case that it is indispensable.

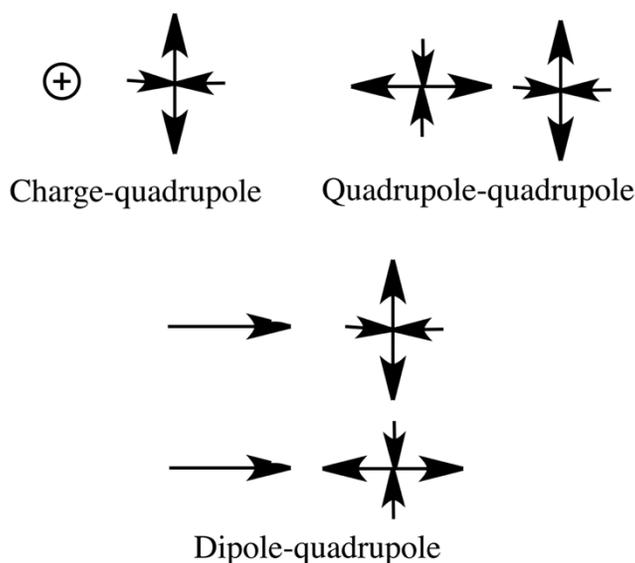


Figure 2-3. Schematic representation of some multipole examples.

2.2.6 Cross Terms

It has been mentioned that in electrostatic interaction, the charge distribution of two molecules are ‘coupled’ and affect each other. Similar to that, in bonded interactions, the

deviations of bond stretching, angle bending or bond torsion are not independent but in fact ‘coupled’ as well. The nomenclature of ‘cross terms’ represents these coupling terms. Mathematically, the cross terms can be revealed by expanding the potential energy of a molecular system in a multi-dimensional Taylor expansion. For example, if we consider the bonds stretching when an angle is closing, as shown in Figure 2-4.

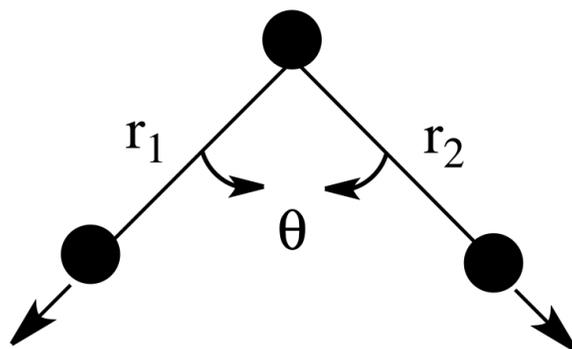


Figure 2-4. Schematic representation of angle closing coupling with bonds stretching.

Such a scheme can be formulated in the following form.

$$u(r_1, r_2, \theta)_{\text{cross term}} = \frac{k_{l_1, l_2, \theta}}{2} [(l_1 - l_{1,0}) + (l_2 - l_{2,0})](\theta - \theta_0) \quad (2-13)$$

Equation 2-13 has been applied in MM force field MM2, MM3 and MM4. Besides this example, there are many others cross terms, such as stretch-stretch, stretch-bend, stretch-torsion, bend-torsion and so on.

Hwang and coworkers have proposed an interesting categorizing scheme for MM force fields⁷⁸⁻⁸⁰: based on whether a force field includes cross terms, force fields are divided into class I that does not have cross terms integrated or class II that should have both anharmonic terms and cross terms. Later, Allinger and coworkers further proposed that more advanced class III

force field should also have electronegativity, hyperconjugation and other features taken into account^{81,82}.

2.2.7 Force Field Parameterization

A force field is not simply a function but a potential energy function coupled with a set of parameters. Developing force field parameters, or force field parameterization, is always an intricate task requiring a great deal of effort.

Once the function form of a force field is chosen, the next step is to choose a series of data. The parameters will then be developed to reproduce, or fit, these data. Ideally, experimental data are preferred, however, they are not always available. Thus, QM calculations using higher level *ab initio* theory is required to provide missing data. When both force field function and selected set of data are decided, the next step is to carry out the parameterization. Two major approaches have been reported, one is the traditional ‘trial and error’ which gradually refine the parameters round by round, the other is the ‘least squares approach’ that uses least-squares fitting to obtain the parameters that best fit to the selected data⁸³⁻⁸⁵.

An important implementation of MM force field is to model large biological systems such as proteins, DNA’s and RNA’s. It is not rare to see these systems have more than 10,000 atoms, especially when explicit solvents are included. Therefore, to parameterize against the entire system is almost prohibitive. In fact, most popular force fields today, such as AMBER⁸⁶⁻⁸⁹, CHARMM⁹⁰⁻⁹², GROMOS^{93,94}, OPLS^{95,96} and AMOEBA⁹⁷⁻⁹⁹ are all parameterized against small organic molecules. In this manner, the transferability of such force fields is very important.

2.3 Energy Minimization and Molecular Simulation

In molecular modelling, one of the major goals is to calculate the potential energy and reproduce the PES. With potential energy and/or PES available, many molecular properties and

thermodynamic properties can thus be computed. There are two totally different approaches to achieve this goal: energy minimization and molecular/computational simulation.

2.3.1 Energy Minimization

Ideally, with a PES available, it is simple and straightforward to locate energy minimum of a molecular system. However, in practice, this is only working with very simple molecules. For a large system with N atoms, there are $3N$ Cartesian coordinates and $(3N - 6)$ internal coordinates; thus, this system's potential energy represents a function of many coordinates and the PES is complicated. In such a case, locating the minimum points along the entire PES, which is referred to as global minimum, can be troublesome. Energy minimizations are designed to search the global minimum and other energy minimums along the PES.

Considering the potential energy as a function F of a set of variables r_1, r_2, \dots, r_i , the energy minimums have the following mathematical characteristics,

$$\frac{\partial F(r)}{\partial r_i} = 0; \frac{\partial^2 F(r)}{\partial^2 r_i} > 0 \quad (2-14)$$

Equation 2-14 means the first derivative of function F with regard to any variables equals to zero, while the second derivatives are positive. It is worthy noting that saddle points, representing those stationary points on the PES but not local extremum points, also have their first derivatives equal to zero.

Although minimization algorithms with no derivatives of potential energy considered, most of today's popular minimization engines prefer to use such derivatives since they can provide many useful information such as the shape of PES. In fact, experienced molecular modeler might find the application of energy derivatives can boost the efficacy of locating the important points. Examples of most common minimization algorithms includes, but not limited to, the steepest descent method, conjugated gradient method and Newton-Raphson method.

2.3.1.1 Steepest descent method

This algorithm calculates the first derivative of energy function, determining the direction of force. Then the coordinates of the system are gradually modified to make the system move in a direction that is parallel to the net force to approach the minimum point. In this method, the structure of the system obtained at one step is used as the starting structure for the next step. Also, it is important to note that both the gradients and the direction of two connecting steps are orthogonal. Practically, steepest descent method is good at rapidly finding out a structure that is very close to the minimum but rather inefficient in the final approaching, thus, this method is always carried out in the beginning stage of minimization if multiple minimization methods are carried out in order to accurately locate the minimum.

2.3.1.2 Conjugated gradient method

This algorithm also approaches the minimum by consequently moving the system just like what steepest descent method does. However, different to steepest descent method, conjugated gradient method does not require the orthogonal directions for two consecutive steps. In fact, within this approach, the gradients of two successive steps are orthogonal but the directions are rather ‘conjugated’.

Generally speaking, this method is better at locating the exact minimum and in some cases can achieve this goal with fewer steps than steepest descent method when a rather good structure is given. However, it is interesting to notice that the move of first step of both steepest descent method and conjugated gradient method should be identical.

2.3.1.3 Newton-Raphson method

Different to steepest descent or conjugated gradient method that only use the gradient (the first derivative), Newton-Raphson method uses both the first and the second derivatives to search the minimum. The inclusion of the second derivatives provides more information, e.g. the

curvature of the energy function, however, on the other hand, increases computational cost by possibly a significant margin because within this algorithm the Hessian matrix must be calculated and inverted.

Newton-Raphson method is always applied to small systems due to the high computational expense. In addition, this approach might not be a good choice given a structure not even close to its minimum. On the other hand, this method is able to move ‘uphill’ as both steepest decent and conjugated gradient method can only deal with ‘downhill’ move. One important application of such feature is to search for saddle points.

2.3.2 Molecular/Computational Simulation

Although energy minimization can discover the minimum points along a PES and thus provide information for predicting properties of a system, it is not reliable that the information provided is sufficient to make predictions especially to those thermodynamic properties. This is especially true given a large biological system with thousands of atoms. In such cases, the huge number of degrees of freedom and multiple minima separated by energy barriers make it extremely difficult to gather information covering the entire energy surface. In addition, most experimentally measurable properties especially those thermodynamic properties are indeed time average (equation 2-15), which cannot be theoretically reproduced or predicted using only energy minimizations. Thus, it is clear that a technique different to energy minimization is required to accompany those attempts to understand the nature of the PES of a system.

Computational simulation, or call it molecular simulation, is a good candidate for better understanding of the intrinsic characteristics of energy surface of molecular systems. Integrated Newton’s second law, it is possible for computational simulation to study the motion of a system evolving in time and thus to calculate those properties averaged over time using certain numerical equations. However, the biggest problem encountered is that it is rarely possible to

find the initial configuration. To solve this problem, statistical mechanics was introduced by Boltzmann and Gibbs in order to convert studying a time evolving system into investigating a large number of system replicas simultaneously. Whereafter in associate with the ergodic hypothesis, ensemble average properties (equation 2-16) can substitute time average properties.

$$P_{timeaverage} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t P(\mathbf{p}^N(t), \mathbf{r}^N(t)) dt \quad (2-15)$$

$$\langle P \rangle = \int \int d\mathbf{p}^N d\mathbf{r}^N P(\mathbf{p}^N, \mathbf{r}^N) \rho(\mathbf{p}^N, \mathbf{r}^N) \quad (2-16)$$

In both equations, \mathbf{p} and \mathbf{r} corresponds to momentum and coordinate separately, while in equation 2-16 the angle bracket means ensemble average and $\rho(\mathbf{p}^N, \mathbf{r}^N)$ is the probability density.

Monte Carlo (MC) method and molecular dynamics (MD) method are the two most popular simulation techniques applied in today's computational simulations.

2.3.2.1 Phase space

For a classical system containing N atoms, each of its conformation has $3N$ coordinates and $3N$ momenta for. Hence, the state of this system can be characterized by a set of $6N$ values. Such a state that is defined by these $6N$ values is called phase space. All computational simulations sample the phase space of the given molecular system. Sampling enough phase space is the key to calculate important system properties especially thermodynamic properties.

2.3.2.2 Monte Carlo method

In MC simulation, each configuration is generated by making a random modification, such like moving an atom or rotating a bond to its predecessor, thus this kind of simulation stores no time information. For each configuration, the potential energy is calculated. By applying a certain criteria, usually Metropolis criterion¹⁰⁰ in today, whether the new configuration is accepted or rejected is decided.

After sufficient simulation of configurations, a certain property can be calculated using equation 2-17,

$$\langle P \rangle = \frac{1}{M} \sum_{i=1}^M P(\mathbf{r}^N) \quad (2-17)$$

where M refers to the number of configuration simulated.

It is noteworthy that traditional MC simulations are carried out in canonical ensemble. In such ensemble, the number of particles (N), the volume (V) and the temperature (T) remain constant.

2.3.2.3 A brief introduction to molecular dynamics

Unlike MC, MD derives new configurations based on Newton's second law (equation 2-18). In detail, the force of each particle is first calculated by differentiating the potential energy function, then applying the equation 2-18 will give in the acceleration and velocity information, thus a real trajectory over time is generated.

$$F = m \cdot a = m \cdot \frac{d^2 x}{dt^2} \quad (2-18)$$

$$\langle P \rangle = \frac{1}{M} \sum_{i=1}^M P(\mathbf{p}^N, \mathbf{r}^N) \quad (2-19)$$

In MD simulations, thermodynamic properties can be obtained via averaging over time, as shown in equation 2-19. Comparing equation 2-19 and equation 2-17, clearly the only difference is that the function of a certain property in MD simulations has momentum dependence that is absent in the function of MC simulations.

2.4 Molecular Dynamics

MD generates a series of continuous configurations based on Newton's equation of motion, propagating the system in phase spaces and generating a trajectory varies with time. For

a simple system such as an one dimensional classical harmonic oscillator or a hard-sphere model moving straightly with a constant velocity¹⁰¹, the position and momentum information can be obtained analytically as functions of time. However, for more realistic molecular models, the force on each atom (or particle) is not changing independently and the motion of the system can hardly be solved analytically. Hence, in such situation, the equation of motion must be solved using certain integration algorithms.

2.4.1 Finite Difference Method: Verlet and Leap-frog Algorithm

Finite difference method is commonly used to integrate the equation of motion when dealing with large and complicated molecular systems. The basic idea of this technique is that the integration can be separated into different small parts with a very small time interval of Δt and the force on each particle is assumed to be constant during interval. Thus, the force, coordinate, velocity and acceleration can be solved for each step and the simulation can be propagated.

One popular integrator based on finite difference method is the Verlet algorithm¹⁰². The central idea of this algorithm is to use the position and acceleration information at time t and the position information from previous step ($t - \Delta t$) to determine the position of new step at ($t + \Delta t$), as demonstrated in equation 2-20.

$$\begin{aligned}
 r(t + \Delta t) &= r(t) + v(t)\Delta t + \frac{1}{2} a(t)\Delta t^2 + \dots \\
 r(t - \Delta t) &= r(t) - v(t)\Delta t + \frac{1}{2} a(t)\Delta t^2 - \dots \\
 r(t + \Delta t) &= 2r(t) - r(t - \Delta t) + a(t)\Delta t^2 \\
 v(t) &= \frac{[r(t + \Delta t) - r(t - \Delta t)]}{2\Delta t}
 \end{aligned}
 \tag{2-20}$$

However, there are several clear disadvantages of this algorithm: the possible precision loss associated with the use of $2\mathbf{r}(t)$ and $\mathbf{r}(t - \Delta t)$ along with much smaller term $\mathbf{a}(t) \Delta t^2$, the lack of an explicit velocity term and the trouble of getting positional information of $\mathbf{r}(t - \Delta t)$.

Leap-frog algorithm is actually a variant form of Verlet method¹⁰³. The mathematical expression is shown as:

$$\begin{aligned} \mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \mathbf{v}(t + \frac{1}{2} \Delta t) \Delta t \\ \mathbf{v}(t + \frac{1}{2} \Delta t) &= \mathbf{v}(t - \frac{1}{2} \Delta t) + \mathbf{a}(t) \Delta t \end{aligned} \quad (2-21)$$

Here $\mathbf{a}(t)$ can be solved using Newton's second law. Thus, $\mathbf{v}(t)$ can be calculated from

$$\mathbf{v}(t) = \frac{1}{2} \left[\mathbf{v}(t + \frac{1}{2} \Delta t) + \mathbf{v}(t - \frac{1}{2} \Delta t) \right] \quad (2-22)$$

The major advantage of leap-frog algorithm over Verlet algorithm is that velocity term is included explicitly, however, the coordinates and velocities are unsynchronized, making this method unable to count the kinetic energy contribution at the same step that the coordinates are defined.

2.4.2 Time Step and SHAKE

The time interval mentioned in 2.3.3.1 is usually referred to as time step. Choosing a proper time step is clearly important for setting up a MD simulation, giving the fact that an inappropriate time step choice can make the simulation either unable to sample sufficient phase spaces (when it is too small) or instable (when it is too large). Most biological systems are fairly flexible. In such systems, the highest frequency vibration corresponds to the bond stretching, especially for those heavy-atom-to-hydrogen bond. This kind of vibration is usually at 10 fs level. Hence, the appropriate time step should not exceed 1 fs. With such a small time step, it is

time consuming to carry out long MD simulation in order to understand a system's behavior at large time scale, e.g. ns or ms.

In order to accelerate MD simulations without making it instable, constraint dynamics are introduced. In such a scheme, bonds or angles are forced to adopt specific values thus the degrees of freedom associated with those terms are totally removed. The most popular constraint technique is SHAKE introduced by Berendsen and coworkers¹⁰⁴. Oppositely equal forces are applying on each of the two atoms of a constraint bond. Thus, a constraint σ_{ij} that freezes the bond length between atom i and j can be expressed as,

$$\sigma_{ij} = |\mathbf{r}_{ij}|^2 - d_{ij} = (\mathbf{r}_i - \mathbf{r}_j)^2 - d_{ij} = 0 \quad (2-23)$$

Here d_{ij} denotes the constraint distance. The force on each involving atom resulting from this constraint can be calculated base on,

$$F_x = \lambda \frac{\partial \sigma}{\partial r} \quad (2-24)$$

Here λ is the Lagrange multiplier. Applying equation 2-23 into equation 2-24, it is easy to get

$$\begin{aligned} F_i &= \lambda \frac{\partial (\mathbf{r}_i - \mathbf{r}_j)^2}{\partial \mathbf{r}_i} = 2\lambda(\mathbf{r}_i - \mathbf{r}_j) \\ F_j &= \lambda \frac{\partial (\mathbf{r}_i - \mathbf{r}_j)^2}{\partial \mathbf{r}_j} = -2\lambda(\mathbf{r}_i - \mathbf{r}_j) \end{aligned} \quad \text{hence } F_i = -F_j \quad (2-25)$$

In application, the constraints are applied iteratively, consequently, the Lagrange multipliers following incorporating equation 2-23 and 2-24 into propagating function such like equation 2-20 or 2-21. Practically, the SHAKE algorithm is usually applied on those bonds involving both heavy atom and hydrogen. Removing the degrees of freedom associated with such bonds will not affect the remaining degrees of freedom since they are only weakly coupled.

By doing so, the time step can be set to up to 2 fs that is twice as fast as simulation without invoking SHAKE.

2.4.3 Molecular Dynamics at Constant Temperature

In general, MD simulations are carried out in microcanonical ensemble where the number of particles N , the volume of system V and the total energy E remain constant. Ideally, thermodynamic properties obtained in NVE ensemble are able to transfer to other ensembles such as canonical (constant NVT) ensemble. However, the prerequisite condition of such transferability is that the size of system should be infinite. Therefore, it is important to apply MD in other ensembles such as constant NVT ensemble. For this reason, different thermostats for MD simulations are developed.

Kinetic energy in a MD simulation is closely related to its temperature, thus temperature can be calculated base on,

$$K = \sum_{i=1}^N \frac{|\mathbf{p}_i|^2}{2m_i} = \frac{k_B T}{2} (3N - n_c) \quad (2-26)$$

Here k_B is the Boltzmann constant, N is the total number of particles and n_c is the degrees of freedom being constraint. Obviously, temperature can be written as a function of momentum or velocity. Therefore, rescale velocity is the simplest way to regulate temperature¹⁰⁵.

Multiplying the velocities at time t by a factor of λ , the corresponding change in temperature can be expressed in the following form:

$$\Delta T = \frac{1}{2} \sum_{i=1}^N \frac{2}{3} \frac{m_i (\lambda v_i)^2}{Nk_B} - \frac{1}{2} \sum_{i=1}^N \frac{2}{3} \frac{m_i v_i^2}{Nk_B} = (\lambda^2 - 1)T(t) \quad (2-27)$$

Now, the scaling factor λ can be written as.

$$\lambda = \sqrt{\frac{T(t) + \Delta T}{T(t)}} = \sqrt{\frac{T_{aim}}{T_{curr}}} \quad (2-28)$$

Thus, the temperature of a system can be controlled.

There are several alternative approaches to regulate the temperature for MD simulations. One of them is the Berendsen thermostat that is acting like weakly coupling to an external heat bath¹⁰⁶. The mathematical expression of such algorithm is given in equation 2-29.

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_{bath} - T(t)) \quad \text{so} \quad \Delta T = \frac{\Delta t}{\tau}(T_{bath} - T(t)) \quad (2-29)$$

Here variable τ is actually functioning as a coupling constant. Bigger τ results in weaker coupling while smaller τ brings tighter coupling. Empirically, good performance is achieved by setting coupling constant to 40 times of the time step (Δt in equation 2-29).

The advantage of Berendsen thermostat is that the temperature of a system is allowed to fluctuate, however, there is no guarantee that the temperature is canonical averaged over the entire system that sometimes results in the temperature of solute is lower than that of solvent. In fact, this is the biggest disadvantage of any velocity scaling thermostats.

Stochastic collisions method is one of the thermostat algorithms developed to improve the temperature distribution thus to reach the canonical ensemble. Langevin dynamics¹⁰⁷ is classified into this category¹⁰⁸. By mimicking the Brownian movement, the force on each particle in Langevin dynamics is composed of three components, as shown in equation 2-30.

$$F = m_i \frac{d^2 r_i(t)}{dt^2} = F_i\{r_i(t)\} - \gamma_i m_i \frac{dr_i(t)}{dt} + A_i(t) \quad (2-30)$$

The first component F_i is resulting from interaction between particle i and other particles thereby depends on positions of particles, the second component is due to the frictional drag on particle i from solvent with a collision frequency of γ_i , while the third part A_i is a random force arisen from random fluctuation of particle i due to interactions with solvents. The random force A_i obeys Gaussian probability distribution with correlation function,

$$\langle A_i(t)A_j(t') \rangle = 2\gamma k_B T \delta_{ij} \delta(t-t') \quad (2-31)$$

Here δ_{ij} is the Kronecker delta function while $\delta(t-t')$ is the Dirac delta function. With equation 2-31, clearly the equation of motion of Langevin dynamics is temperature dependent, thus it can be applied to temperature regulation.

2.4.4 Pressure Regulation in Molecular Dynamics

In molecular modelling, it is frequently desired to simulate in isothermal-isobaric ensemble in which the number of particles N , the system temperature T and the system pressure P are reversed. This ensemble is actually very close to the real lab condition under which most experimental measurements are made. Pressure fluctuation is usually large cause by rapid change of virial, $r dU(r)/dr$, with r . The relation of system pressure and virial is demonstrated in equation 2-32.

$$P = \frac{1}{V} \left[Nk_B T - \frac{1}{3} \sum_{i=1}^N \sum_{j=i+1}^N r_{ij} \frac{dU(r_{ij})}{dr_{ij}} \right] \quad (2-32)$$

$U(r_{ij})$ is the interaction energy between atom i and j . Pressure regulation in MD simulation is usually accomplished by controlling system volume, because of clear dependence of pressure on volume from equation 2-32. While on the other hand, since N is fixed, regulating density can also be applied to barostat.

Volume fluctuation can be quantified by introducing isothermal compressibility κ defined by equation 2-33.

$$\kappa = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T \quad (2-33)$$

Under isobaric condition, κ can be expressed in another form.

$$\kappa = \frac{1}{k_B T} \frac{\langle V^2 \rangle - \langle V \rangle^2}{\langle V^2 \rangle} \quad (2-34)$$

Similar to Berendsen thermostat, by coupling the system to an external ‘pressure bath’ it is possible to use Berendsen barostat to regulate system pressure using equation 2-35.

$$\frac{dP(t)}{dt} = \frac{1}{\tau_p} (P_{bath} - P(t)) \quad (2-35)$$

Here τ_p is the pressure coupling constant. Supposing the volume is scaled by a factor of λ , which corresponds to scaling the coordinates by a factor of $\lambda^{1/3}$. Then the mathematical solution of λ can be expressed as:

$$\lambda = 1 - \kappa \frac{\Delta t}{\tau_p} (P - P_{bath}) \quad (2-36)$$

Constant NTP ensemble is widely used in MD simulations of biological systems. Important thermodynamic properties such as Gibbs free energy can only be determined under this ensemble (whereas under constant NTV canonical ensemble Helmholtz free energy is instead measured).

2.4.5 Solvent Effects and Water Models

Simulation in the gas phase is straightforward because of its simplicity. However, most chemical reactions and biological activities are taken place in solution. One good example is the Menshutkin reaction of ammonia and chloromethane, which will take place in solution (water) but instead show no reactivity in gas phase. Therefore, it is of great importance to including solvent effects into molecular simulations.

In MD simulations, the most important solvent model is the water model. In present, there are two very different types of approaches developed: the implicit solvent models that apply a

so-called reaction field to the simulation, while the explicit water model that impends a water box to the system.

2.4.5.1 Implicit solvent models

In implicit solvent models, solution water is treated as a dielectric continuum and the solvent effect is incorporated into the simulation as a reaction field, which is actually a continuous electric field. The free energy required to build the charge distribution can be calculated based on equation 2-37.

$$G = -\frac{1}{2} \int \rho(\mathbf{r})\phi(\mathbf{r})d\mathbf{r} \quad (2-37)$$

$\phi(\mathbf{r})$ is the electrostatic potential and $\rho(\mathbf{r})$ is the charge density. Depending on different models, the latter can be written in the form of either continuous functions of \mathbf{r} or discrete charges.

The Poisson-Boltzmann (PB) model is based on Poisson equation (equation 2-38), but practically the linearized PB equation (equation 2-39) is always used¹⁰⁹.

$$\nabla^2 \phi(\mathbf{r}) = -\frac{4\pi\rho(\mathbf{r})}{\epsilon} \quad (2-38)$$

$$\nabla \epsilon(\mathbf{r}) \cdot \nabla \phi(\mathbf{r}) - \epsilon(\mathbf{r})\lambda(\mathbf{r})\kappa^2 \phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \quad \text{and} \quad \kappa = \sqrt{8\pi q^2 I / \epsilon k_B T} \quad (2-39)$$

Here κ denotes Debye-Hückel parameter and I is the ionic strength of the electrolyte solution. It is worthy noting that the linearized PB equation cannot be solved analytically, thus, it has to be calculated iteratively, which is time consuming.

Suppose a charge q is diffused uniformly to a conducting sphere surface with a radius of d , then, at any point on the surface, the charge density is

$$\rho = \frac{q}{4\pi d^2} \quad (2-40)$$

Since the dielectric constant of a conductor is infinite and at anywhere inside the conductor the electrostatic potential is zero, the electrostatic potential at r on the surface can only be determined from outside by using equation 2-41.

$$\varphi(r) = -\frac{q}{\varepsilon|r|} = -\frac{q}{\varepsilon d} \quad (2-41)$$

Thus, equation 2-37 can be written as

$$G = -\frac{1}{2} \int \left(\frac{q}{4\pi d^2} \right) \left(-\frac{q}{\varepsilon d} \right) dx = \frac{q^2}{2\varepsilon d} \quad (2-42)$$

Since the solvation free energy $G_{solvation}$ is the difference of free energy in gas phase and in solution, it can thus be calculated.

$$G_{solvation} = -\frac{1}{2} \left(1 - \frac{1}{\varepsilon} \right) \frac{q^2}{d} \quad (2-43)$$

Equation 2-43 is the Generalized Born (GB) equation, which is the basis of GB model¹¹⁰⁻
¹¹⁴. Comparing to PB model, GB equation can be calculated analytically and thus computationally more efficient. By extending the GB equation to polyatomic system, equation 2-43 becomes

$$G_{solvation} = -\frac{1}{2} \left(1 - \frac{1}{\varepsilon} \right) \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{F_{GB}} \quad \text{where } F_{GB} = \sqrt{\left(r_{ij}^2 + d_i d_j e^{-r_{ij}^2/4d_i d_j} \right)} \quad (2-44)$$

Variable d_i is the effective Born radius of charge q_i .

2.4.5.2 Explicit water models

Under certain circumstances, for instance, no water or other solvent molecule directly involve in interactions with solute system, implicit solvent models are popular choices. However, in most cases, if computational expense is affordable, using explicit water models is more desired, since the essential nature of chemistry lies on molecules and atoms.

There are several types of explicit water models developed. The basic type of models adopts a rigid geometry with three to five interaction sites (Figure 2-5), thus it is also referred to as simple interaction site model. Because of rigid geometry, their potential energy functions have no bonded term. In addition, only VDW interactions between oxygen atoms are calculated.

TIP3P¹¹⁵ is one of the most popular water models in present. It falls into 3-site category with a 0.9572 Å O-H distance and a 104.52 °H-O-H angle. The 4-site model has an extra ‘M’ position with the negative charge transferred from oxygen. One such example is TIP4P¹¹⁵, which has exactly the same bond length and angle as TIP3P. ST2¹¹⁶ is the most commonly used 5-site water model, with negative charge of oxygen put on two lone pair sites. It is notable that two models within same number of sites do not have to have identical parameters, as demonstrated in Table 2-1.

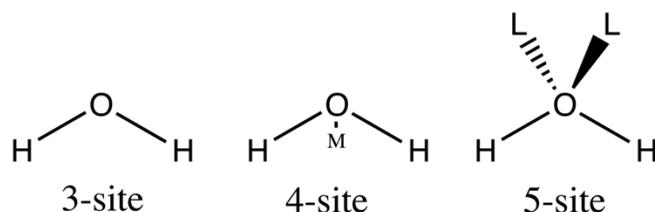


Figure 2-5. Simple interaction site water models.

Parameters developed for each water model should allow bulky water to reproduce their natural properties, such as to give a maximum density at 4 °C. Unfortunately, the most widely used TIP3P model failed to reproduce this property, but its successors TIP4P and TIP5P¹¹⁷ are capable of doing so. The implementation of rigid geometry in these water models is able to save computational cost, however, on the other hand unable predict some properties, such as vibrational spectrum due to the lack of internal flexibility. More advanced water models that

allow flexible geometry¹¹⁸ or include polarization effect^{77,119} have been developed to improve the simulation of water solvent effect in biological systems.

2.4.5.3 Periodic boundary conditions and particle-mesh Ewald

When explicit solvents are employed, it is important to set “boundaries” to keep the solvents away from diffusing into ‘vacuum’. Periodic boundary conditions (PBC) are one of such applications that allow measurable properties to be obtained from simulations with much smaller numbers of particles¹²⁰. The central idea of PBC is that a box is duplicated in all direction and a particle leaving the box from one side will ‘enter’ the box from the opposite side, like shown in Figure 2-6. In practice, a two-dimensional (2-D) PBC application will have 8 duplicated boxes while a 3-D application will have 26 such boxes. The duplicated box is also called image box and coordinates of a particle in the image box can be determined based on that of the corresponding in the original box. Today, there are several box choices with different shapes, amongst the rectangular and truncated octahedral box are the most widely used. The latter is especially good for simulating enzymes/proteins because its spherical shape is close to the real shape of enzymes.

Table 2-1. Parameters of several water models.

	SPC/E ¹²¹	TIP3P	BF ¹²²	TIP4P	ST2
# of sites	3.0000	3.0000	4.0000	4.0000	5.0000
d _{OH} (Å)	1.0000	0.9572	0.9500	0.9572	1.0000
HOH (°)	109.4700	104.5200	105.7000	104.5200	109.4700
q(O)	-0.8476	-0.8340	0.0000	0.0000	0.0000
q(H)	0.4238	0.4170	0.4900	0.5200	0.2375
q(M)/q(L)	0.0000	0.0000	-0.9800	-1.0400	-0.2375

In MM simulations, electrostatic decays as r^{-1} , thus it is important to extend the calculation to distant. However, to calculate this kind of long-range forces can be very time-consuming. Ewald summation¹²³ was developed to speedup such calculations by divide the original summation into two series, as demonstrated in equation 2-45 and Figure 2-7.

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1-f(r)}{r} \quad (2-45)$$

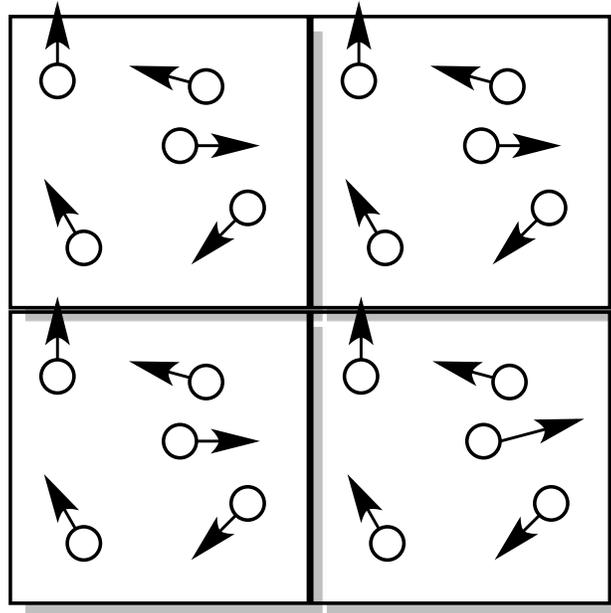


Figure 2-6. Part of 2-D periodic boundary conditions (PBC). Note in 2-D 8 image boxes should surround the original box while in 3-D this number should go up to 26.

By represent function $f(r)$ by using complimentary error function (equation 2-46).

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt \quad (2-46)$$

Both terms on the right side of equation 2-45 converge faster than the original form $1/r$.

Therefore, the summation is expedited. More recently, several variants of Ewald summation has

been developed, including particle-mesh Ewald (PME)¹²⁴ and particle-particle-particle-mesh Ewald^{125,126}. Amongst PME is more popular a choice at present.

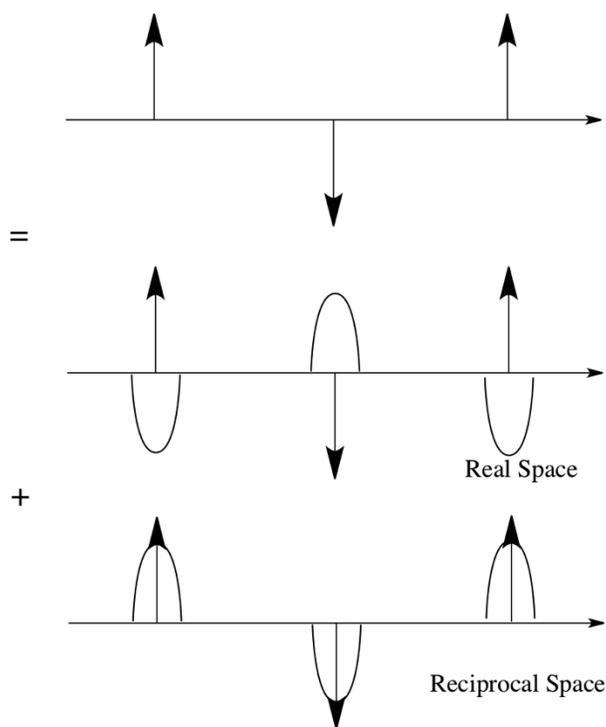


Figure 2-7. Schematic representation of Ewald summation.

2.5 Free Energy Calculations in Molecular Dynamics

As one of the most important property in thermodynamics, free energy calculation is always of computational chemist's great interest. However, it is not straightforward to calculate free energy in molecular simulations due to the insufficient sampling of certain phase spaces. In macromolecular of enzymatic systems, many energy minimums exist, separated by different energy barriers. MD (or MC) is usually unable to sample those barrier areas, therefore making it impossible to obtain an accurate free energy that requires sufficient sampling through the entire phase space.

In particular, people are interested in the potential of mean force (PMF), which represents the free energy surface along certain internal coordinates. Such an internal coordinate is also

known as reaction coordinate (RC). PMF provides valuable information with regard to the reaction. For example, the location of the free energy barrier on the surface usually corresponds to the transition state while the barrier height can be used to calculate the rate constant, an important kinetic property.

Various sampling techniques have been developed for such purposes. Umbrella sampling (US) and steered-MD (SMD) are two of them that are heavily implemented in studies discussed in this dissertation.

2.5.1 Steered Molecular Dynamics

The basic idea of SMD is that by putting an additional force onto the system, the simulation will be able to overcome the free energy barrier and sample the entire phase space. The force applied is usually in the form of a harmonic restraint (equation 2-47).

$$F = \frac{1}{2} k [r - r_0(t)]^2 \quad (2-47)$$

Here, k denotes the restraint force constant, while x is the RC that could be distance, angle, dihedral or more complexed linear combination of coordinates. By integrating the force over RC, the work, W , is obtained, however, it is from a non-equilibrium process. Jarzynsky relationship (equation 2-48) can convert this generalized work to an “equilibrium” free energy difference between the initial and final state.

$$e^{\left(\frac{-\Delta G}{k_B T}\right)} = \left\langle e^{\left(\frac{-W}{k_B T}\right)} \right\rangle_A \quad (2-48)$$

Many factors have been proved to play important roles in successful SMD simulations. It is particularly important to notice that if applied solo, several parallel SMD simulations must be carried out using different starting structures obtained from an equilibrium simulation^{127,128}. In addition, the ‘speed’ of moving the external force over RC is vital to SMD. For instance, fast

moving could result in higher barrier height than reality. SMD has been a perfect way to set up subsequent umbrella samplings throughout all the studies discussed in this dissertation.

2.5.2 Umbrella Sampling

Another popular choice to construct the PMF is USP. The implementation of USP also lies on an external potential term that in general in quadratic form,

$$W(\mathbf{r}) = \frac{1}{2} k_w (\mathbf{r} - \mathbf{r}_0) \quad (2-49)$$

However, unlike SMD, the external potential W here functions to bias the potential energy surface, making it ‘flat’ thus able to be sampled sufficiently. The biased potential energy can be expressed as equation 2-50.

$$U'(\mathbf{r}) = U(\mathbf{r}) + W(\mathbf{r}) \quad (2-50)$$

In practice, the first step is to divide the RC into a series of ‘windows’. Then, in the second step, an appropriate force constant k_w is chosen for each window to guarantee enough sampling of phase spaces through the RC covered by the window. The k_w for adjacent windows should be able to provide sufficient overlap that is required for collecting distribution information. The obtained distribution is certainly a non-Boltzmann distribution. Next, in the third step, the unbiased PMF as a Boltzmann average, can be extracted from the non-Boltzmann distribution by using equation 2-51¹²⁹ (Figure 2-8).

$$\langle A(\mathbf{r}) \rangle_{unbiased} = \frac{\langle A(\mathbf{r}) e^{W(\mathbf{r})/k_B T} \rangle_{biased}}{\langle e^{W(\mathbf{r})/k_B T} \rangle_{biased}} \quad (2-51)$$

The advantage of USP over SMD is that the obtained PMF obtained from a series USP usually does not have the ‘artificial’ effect on the barrier height increasing associated with an improper SMD moving speed. However, the success of USP strongly relies on the choice of

force constant k_w for each window: a too small k_w usually is not able to bias the energy surface enough for sufficient sampling over all phase spaces, while on the other hand, a too large k_w always results in a ‘narrow’ shape of distribution thus requires much longer time of simulation to converge (Figure 2-9).

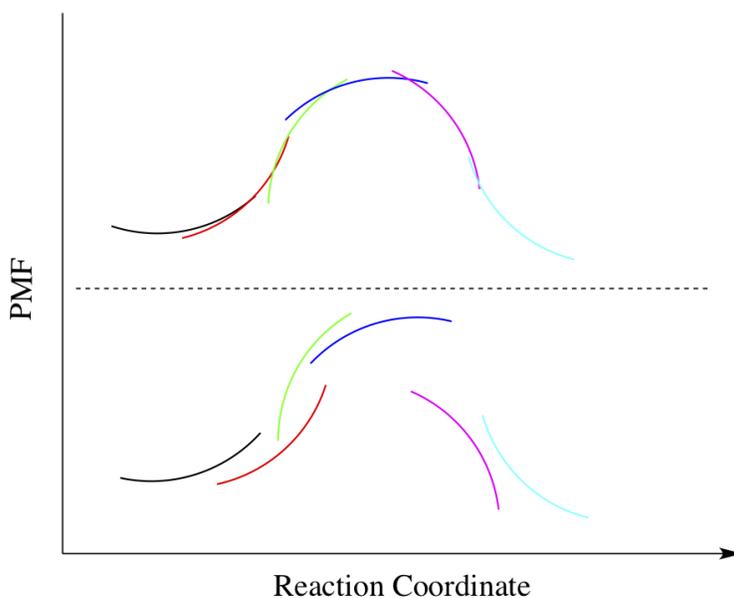


Figure 2-8. Schematic representation of building PMF with umbrella samplings.

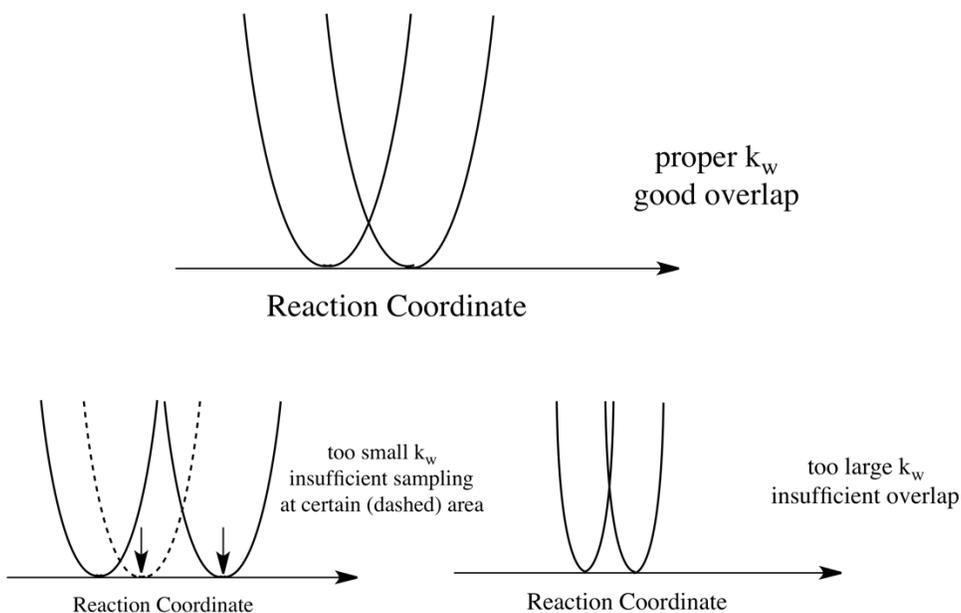


Figure 2-9. The effect of k_w on umbrella sampling (US).

The weighted histogram analysis method (WHAM)¹³⁰ is routinely used to extract the non-Boltzmann distribution information from USP followed by constructing the unbiased PMF. The WHAM equation is,

$$\langle \rho(\mathbf{r}) \rangle = \sum_{i=1}^{N_W} n_i \langle \rho(\mathbf{r}) \rangle_i \times \left[\sum_{j=1}^{N_W} n_j e^{\frac{-[W_j(\mathbf{r}) - F_j]}{k_B T}} \right]^{-1}$$

and

$$e^{-F_i/k_B T} = \int e^{-W_i(\mathbf{r})/k_B T} \langle \rho(\mathbf{r}) \rangle d\mathbf{r}$$
(2-52)

Both equations in equation 2-52 must be solved self-consistently, which is achieved by iteratively solving them in turn until both are satisfied.

2.6 Simulating Chemical Reactions and QM/MM

The implementation of USP or SMD in MM MD simulations enables computational chemists to study the free energy associated with certain transition processes, such as the *trans* and *gauche* conformational transition of butane^{131,132}. However, the limitation of MM force field impedes the free energy studies associated with chemical reactions. While on the other hand, the large size of molecular system always keeps pure QM calculations away from simulating such scenarios. As a result, hybrid QM/MM method arises as the promising candidate to handle the modelling of chemical reactions.

2.6.1 QM/MM

In general, QM/MM simulation divides the system into two parts: a small part that is of great important to research interest will be modeled using QM methods, while the large part of the system (most times including the solvent) will be modeled using MM force field. Thus, the Hamiltonian can be rewritten as:

$$H_{sys} = H_{QM} + H_{MM} + H_{QM/MM}$$
(2-53)

Here the first two parts on the right side apply to QM region and MM region separately while the third part, $H_{QM/MM}$ is in charge of the interactions between QM particles and MM nuclei. Since H_{MM} is governed by potential energy function of chosen force field only but not sensitive to QM particles at all, the energy of the system can be represented as:

$$E = \langle \Psi | H_{QM} + H_{QM/MM} | \Psi \rangle + E_{MM} \quad (2-54)$$

The H_{QM} is controlled by chosen QM theory, while the $H_{QM/MM}$ is the most complicated part, as demonstrated in equation 2-55.

$$H_{QM/MM} = - \sum_i^{QM} \sum_j^{MM} \left[q_j h_{elec}(r_e, r_j) + q_i q_j h_{core}(r_i, r_j) + \left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} \right) \right] \quad (2-55)$$

In equation 2-55, q is the charge while the subscript i, j, e refer to QM nuclei, MM nuclei and QM electrons, respectively.

When partition the system into QM and MM regions, it is often necessary to cut a covalent bond (although one can also cut polarized bond, it is not recommended). In such scenarios, the $H_{QM/MM}$ becomes even more complicated and requires special treatments. Approaches to tackle this problem include link atoms, frozen orbitals and others, amongst the link atoms is most closely related to this dissertation.

In implementation, a hydrogen atom will be placed on the boundary QM atom at default 1.09 Å along the QM-atom-MM-atom vector, as depicted in Figure 2-10.

For QM part, this hydrogen is ‘real’ hydrogen, but it interacts with MM region via only electrostatic interaction. The MM ‘link’ atom (C_α in Figure 2-10) will have zero charge in order to avoid double counting the electrostatics. 1-2 and 1-3 VDW interactions between QM and MM atoms are set to zero that means only non-bonded VDW interactions are considered, in addition, 1-4 interaction is sometimes rescaled to reproduce physical properties. Bond stretching, angle

bending and bond torsions with at least one atom from each side included are calculated with MM force field.

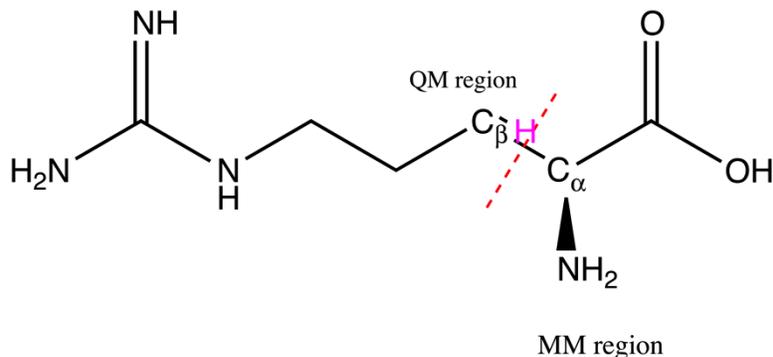


Figure 2-10. QM and MM regions are defined by cutting the covalent bond between C_{α} (MM) and C_{β} (QM). The hydrogen in pink is the capping link atom added to C_{β} along QM-atom-MM-atom vector.

In QM/MM simulation, when the PBC is applied, the number of interactions between QM and MM will be infinite. To handle such a case, a PME approach similar to that of MM has been developed. While for non-periodic system, such a worry is not necessary, but a large QM cutoff is always recommended.

2.6.2 Semiempirical Theory in QM/MM

The implementation of QM/MM in molecular simulations can provide greater accuracy at chemically interested area with moderate computational cost. However, even as of today, the application of *ab initio* level QM/MM simulation is still very expensive, thus prohibiting this approach to be extended to large molecular systems. For instance, the popular density functional theory (DFT) B3LYP coupled with MM potential has been rarely reported to apply on biological systems. Under such circumstances, semiempirical QM theories have become popular choices for QM/MM simulation of macromolecules and proteins.

Comparing to *ab initio* theories, apparently semiempirical methodologies have made several approximations. The first approximation is that the two-electron part is not explicitly included in the Hamiltonian. Instead, Hückel theory is employed to describe π -electrons while extended Hückel theory is applied on all other valence electrons where the core electrons are considered ignorable. The second approximation made is that the development of each semiempirical method includes some parameterizations to reproduce experimental data, thus these theories partially corrects the HF energy by including some electron correlation effects. As a result, the semiempirical methods usually have speed advantage over those *ab initio* theories and can provide comparable accuracy to those molecular systems that are similar to which the method is parameterized against. However, it is important to bear in mind that semiempirical methods could also lead to very wrong results on those systems that are not in the dataset used for parameterization.

There have been various semi-empirical methodologies developed. Amongst AM1 and PM3 are perhaps the most popular methods at present and they have been tested through a large variety of calculations. However, the transferability and the lack of *d* orbital are the major shortcomings of these two theories.

2.6.3 Semiempirical DFT – SCC-DFTB

Recently, a new semi-empirical theory called self-consistent charge density functional tight binding (SCC-DFTB) based on DFT has been developed and shown promising future.

Starting by picking up a reference density ρ_0 , the energy can be expressed as,

$$E = \sum_i^{\text{occupied}} \langle \Psi_i | H_{KS} | \Psi_i \rangle - \frac{1}{2} \iint \frac{\rho_0(\mathbf{r}_1)\rho_0(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 + E_{xc}[\rho_0(\mathbf{r})] - \int V_{xc}[\rho_0(\mathbf{r})]\rho_0(\mathbf{r})d\mathbf{r} + E_N \quad (2-56)$$

Here H_{KS} is the Kohn-Sham (KS) Hamiltonian,

$$H_{KS} = -\frac{1}{2}\nabla^2 - \sum_k^{nuclei} \frac{Z_k}{|\mathbf{r}-\mathbf{r}_k|} + \int \frac{\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + V_{XC}[\rho_0(\mathbf{r})] \quad (2-57)$$

ρ_0 is considered as the sum of neutral atomic densities, $\rho_0 = \sum_{atom} \rho_0^{atom}$.

In order to expedite the computation, the KS orbital Ψ_i can be expanded as $\Psi_i = \sum_u c_u^i \phi_u$.

By applying this basis to the first term of equation 2-56, it can be rewritten as,

$$\begin{aligned} \langle \mu | H_{KS} | \nu \rangle &= \langle \mu | T + v_{eff}[\rho_0^A(\mathbf{r}) + \rho_0^B(\mathbf{r})] | \nu \rangle, \quad \mu \in A, \nu \in B \\ \langle \mu | H_{KS} | \nu \rangle &= \varepsilon_\mu \end{aligned} \quad (2-58)$$

Here ε_μ is the KS eigenvalue of neutral, unperturbed atom, v_{eff} is the effective KS potential consist of Coulombic and exchange-correlation parts, while A and B define two neutral atoms.

The three-center terms are neglected, thus the off-diagonal elements are evaluated by the two-center terms. The Hamiltonian matrix element $H_{\mu\nu}^{KS}$ and overlap matrix element

$S_{\mu\nu} = \langle \langle \phi_\mu | \phi_\nu \rangle \rangle$ are calculated with respect to interatomic distance between A and B and

tabulated, thus making DFTB calculation faster.

However, with above assumption, the charge density of a system is expressed as the sum of the charge density of its component atoms, which is in contradict to the reality that polarization should occur. To address this issue, the contribution regarding to the second order density variation should be included,

$$E[\rho_0(\mathbf{r}) + \delta\rho(\mathbf{r})] = E[\rho_0(\mathbf{r})] + \frac{1}{2} \iint \left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta^2 E_{XC}}{\delta\rho\delta\rho'} \Big|_{\rho_0} \right) \delta\rho(\mathbf{r})\delta\rho(\mathbf{r}') d\mathbf{r}d\mathbf{r}' \quad (2-59)$$

And,

$$\delta\rho = \sum_A^{atoms} \delta\rho^A \approx \sum_A^{atoms} \Delta q^A = \sum_A^{atoms} (q^A - q_0^A) \quad (2-60)$$

In equation 2-60, q^A is the Mulliken charge while q_0^A is the number of valence electrons of neutral atom A. Equation 2-59 can also be written as.

$$E_{2nd} = \frac{1}{2} \sum_{A,B}^{atoms} \gamma_{AB} \Delta q_A \Delta q_B \quad (2-61)$$

γ_{AB} can be expressed by either Hubbard parameter U_A or chemical hardness η_A . With equation 2-59 or 2-61 included in the total energy term, the approach is now self-consistent in charge. With the repulsive potential, E_{rep} , is defined as,

$$E_{rep} = -\frac{1}{2} \iint \frac{\rho_0(\mathbf{r}_1)\rho_0(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 + E_{xc}[\rho_0(\mathbf{r})] - \int V_{xc}[\rho_0(\mathbf{r})]\rho_0(\mathbf{r})d\mathbf{r} + E_N \quad (2-62)$$

Thus, the total energy of SCC-DFTB can be rewritten as,

$$E_{SCC-DFTB} = \sum_i^{occupied} \langle \Psi_i | H_{KS} | \Psi_i \rangle + \frac{1}{2} \sum_{A,B}^{atoms} \gamma_{AB} \Delta q^A \Delta q^B + E_{rep} \quad (2-63)$$

It is worthy noting that strictly speaking SCC-DFTB is not a semi-empirical method because the parameterization includes no fitting to empirical data as other semi-empirical theories do. In addition, SCC-DFTB employs non-orthogonal basis set, which is the key for the improved transferability¹³³. Also, it is important to include an explicit dispersion term when apply SCC-DFTB, since it has been well known that most LDA and GGA classes of DFT functionals have a poor record on dispersion force.

With parameters for more and more atoms developed, SCC-DFTB has been widely applied to numerous molecular modelling and sampling studies and given good results.

2.6.4 Transition State Theory

Transition state theory (TST) was developed to help people qualitatively and quantitatively understand the process of chemical reactions. In order to understand TST, it is important to make clear several concepts first.

Consider the reaction described in Figure 2-11, to illustrate the reaction path of a single set of molecules involved, the PES is a good choice. For such a S_N2 type reaction, the PES has a similar shape to the free energy surface (FES): the two minimums correspond to the reactant and product separately, while the saddle point implies the transition state structure (TSS). The paths that connect the TSS downward to the reactant or product are called the minimum-energy path (MEP) or intrinsic reaction coordinate (IRC). However, in macroscopic world, the equilibria and kinetics of reacting systems are consisted of copies of single set of reacting system, and the rules are governed by free energy instead of potential energy. The free energy path is not necessarily the MEP, while the point associated with the highest free energy on the FES, the transition state (TS), may not exactly be the TSS on the PES.

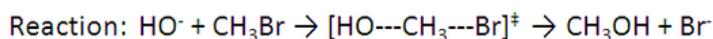
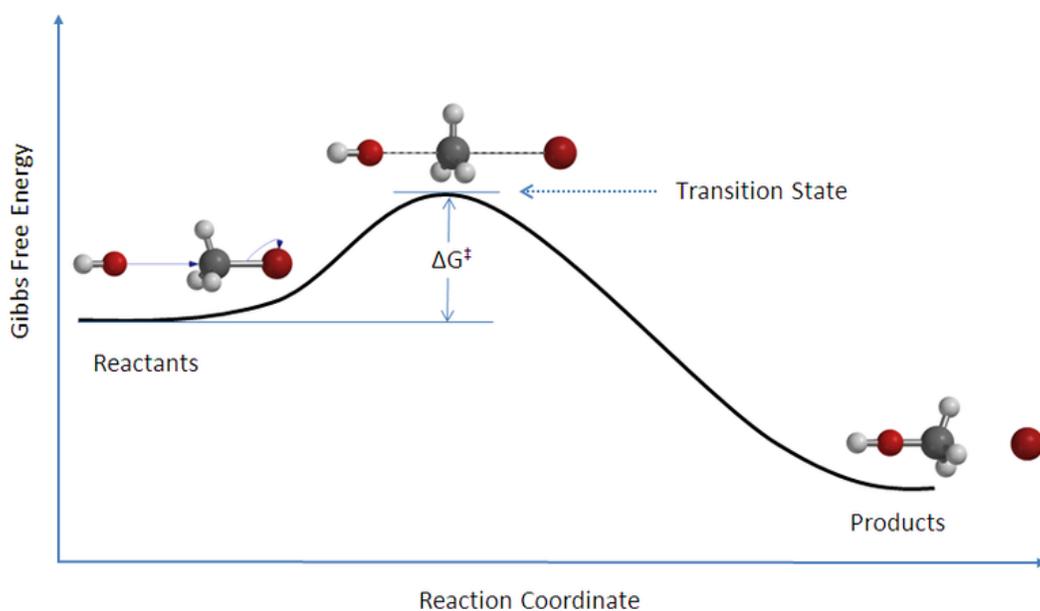


Figure 2-11. Reaction coordinate diagram for the bimolecular nucleophilic substitution (S_N2) reaction between bromomethane and the hydroxide anion. (Source: http://en.wikipedia.org/wiki/File:Rxn_coordinate_diagram_5.PNG)

TST invokes a so-called quasi-equilibrium assumption that considers a quasi-equilibrium is reached between the TS and the reactant even the reactants and products are not equilibrium with each other. Hence, the quasi-equilibrium constant K^{\ddagger} for the reaction in Figure 2-11 is

$$K^{++} = \frac{[\text{HO} - \text{CH}_3 - \text{Br}]^{++}}{[\text{HO}^-][\text{CH}_3\text{Br}]} \quad \therefore [\text{HO} - \text{CH}_3 - \text{Br}]^{++} = K^{++}[\text{HO}^-][\text{CH}_3\text{Br}] \quad (2-64)$$

Therefore the rate equation of product formation can be expressed as

$$\frac{d[\text{CH}_3\text{OH} + \text{Br}^-]}{dt} = k^{++}[\text{HO} - \text{CH}_3 - \text{Br}]^{++} = k^{++}K^{++}[\text{HO}^-][\text{CH}_3\text{Br}] \quad (2-65)$$

While the rate constant of product formation can be written as

$$k = k^{++}K^{++} \quad (2-66)$$

Statistical mechanics provides us a temperature dependent relationship between the equilibrium constant and free energy, which is

$$K^{++} = \frac{k_B T}{h\nu} e^{-\frac{(G^{++} - G^{react})}{k_B T}} = \frac{k_B T}{h\nu} e^{-\frac{\Delta G^{++}}{k_B T}} \quad (2-67)$$

Since k^{\ddagger} is proportional to the vibrational frequency associated with converting TS complex to product, rate constant of product formation, k , can be rewritten as

$$k = \kappa \frac{k_B T}{h} e^{-\frac{\Delta G^{++}}{k_B T}} \quad (2-68)$$

Here κ denotes the transmission coefficient referring to how much vibration contributes to the formation of product.

TST has several major drawbacks, such as it does not account for quantum tunneling and it fails at high temperatures. There are several variants of TST, for example, variational transition state theory (VTST) is a popular method that employs a variationally optimizing procedure to reach more accurate rate constant.

2.6.5 Kinetic Isotope Effect

The rate constant of product formation under certain circumstances can be altered. One such example is the quantum tunneling effect mentioned above, which allows the system to tunnel through the barrier even when the system does not have an energy higher than the saddle point. Hence, such an effect can increase the reaction rate constant. Another example is the kinetic isotope effect (KIE), which refers to the difference in the observed reaction rate constants of two differently isotopically labeled reactants.

There are two types of KIE. When the isotopical substitution is involved in a bond that will break or form in the rate-limiting step (RLS), it is referring to as primary KIE (PKIE); while in all other situation, the effect is denoted as secondary KIE (SKIE). Based on the location of substitution, SKIE can be further classified into α -SKIE, which means the isotope connects one of the reacting center, β -SKIE, which refers to isotope connecting to the neighbor atom to the reacting center and so on.

On a PES, the electronic energies of stationary points are independent of atomic mass, thus the difference of zero point energies (ZPE) associated with two isotopically labeled reactions is the origin of KIE. For PKIE, the ZPE is dominant by the zero point vibrational energy (ZPVE). This is easy to understand because in the case of either a bond is broken or formed, there is no such a vibrational mode in the TSS. If KIE is mathematically defined as the ratio of k_{light} and k_{heavy} , as shown in equation 2-69 (assume it is a first order reaction),

$$KIE = \frac{k_{light}}{k_{heavy}} \quad (2-69)$$

Thus, the PKIE can be expressed as,

$$PKIE = \frac{k_{light}}{k_{heavy}} \propto e^{\frac{1}{2} \frac{h(\omega_{heavy}^{++} - \omega_{light}^{++})}{k_B T}} \quad (2-70)$$

The proportionality symbol can be replaced by an equal sign if the bond is fully broken or formed in the TSS. In the case of SKIE, however, there is no such vibrational mode disappearing associated with bond broken for formation, thus the ZPVE difference of the TSS also plays an important role in determining the k_{light}/k_{heavy} , making the situation more complicated.

Numerically, the PKIE measurements in general are significantly larger than that of SKIE, while in SKIE, the measurements for S_N1 reactions are usually larger than S_N2 reactions. The former is easy to understand because of the contribution from ZPVE difference of TSS cancelling the ZPVE difference of reactants, while the latter is somewhat tricky.

Consider an S_N1 reaction of 3-chloro-3-methylhexane and iodide ion, as illustrated in Figure 2-12. The RLS is the formation of the carbocation intermediate. This process is associated with the hybridization of chloride connecting carbon change from sp^3 to sp^2 , which results in a variety of vibrational modes affected, especially the in-plane and out-of-plane bending. The typical SKIE measurement of an sp^3 to sp^2 hybridization change is 1.1 to 1.2, while that of an sp^2 to sp^3 change is typically 0.8 to 0.9.

For S_N2 reactions, such like bromomethane and hydroxide yielding ethanol and bromide as shown in Figure 2-11 and Figure 2-13, features a TSS with both the nucleophile (hydroxide) and exiphile (bromomethane) being present. Thus, the energies with regard to the old bond breaking and the new bond forming compensate each other, maintaining the ZPE difference throughout the reaction process. Depending on how well the nucleophile attacking and the exiphile retreating are synchronized, the magnitude of SKIE varies slightly, in general between 0.95 and 1.06. If the hydroxide entry and bromide removal take place exactly simultaneously, the measured SKIE could be close to a unity value. However, it is important to notice that a less than 1.07 of SKIE measurement should not be automatically linked to an S_N2 reaction. For instance,

in enzymatic reactions, a unity SKIE value could be implying that the RLS is other than the chemical step.

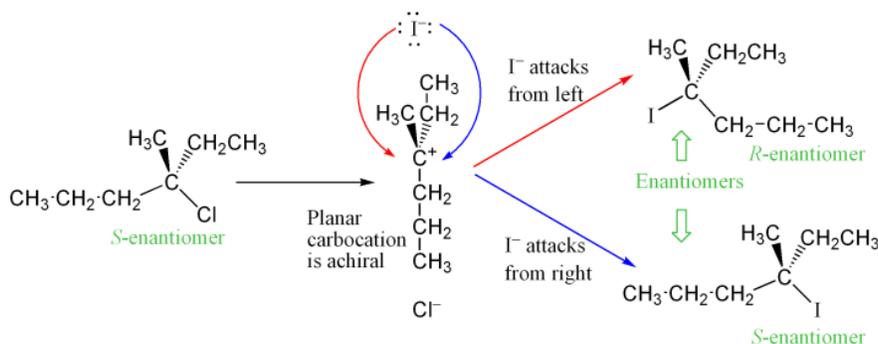


Figure 2-12. Reaction mechanism of S_N1 reaction between *S*-3-chloro-3-methylhexane and iodide ion that yields both *R*- and *S*- 3-iodo-3-methylhexane. (Source: <http://en.wikipedia.org/wiki/File:SN1Stereochemistry.png>)

KIE measurement provides useful insights of determining the reaction mechanism, especially for enzyme-catalyzed reactions. Several computational packages have been developed to enable computational chemists to reproduce experimental KIE values.

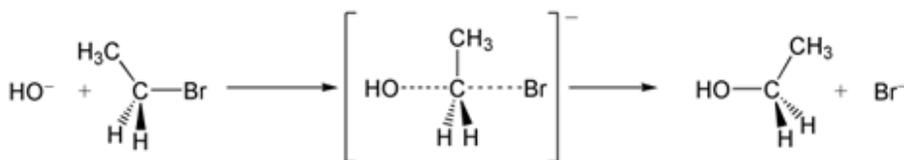


Figure 2-13. Reaction mechanism of S_N2 reaction between bromomethane and hydroxide that yields ethanol and bromide. (Source: <https://secure.wikimedia.org/wikipedia/en/wiki/File:BromoethaneSN2reaction-small.png>)

CHAPTER 3
APPLY MOLECULAR DYNAMICS SIMULATION TO PREDICT THE BINDING OF
MAGNESIUM CATION IN FARNESYLTRANSFERASE ACTIVE SITE

3.1 Background

FTase catalyzed the posttranslational attachment of a 15-carbon farnesyl group to a cysteine residue at a conserved CaaX motif at or near the carboxyl-terminus (C-terminus) of certain biological systems, enabling them to be recognized by the inner side of the plasma membrane. For many small G proteins including the Ras superfamily of enzymes, such a modification is essential to their attachment to the cell membrane prior to play their important roles in the signal transduction¹³⁴⁻¹³⁸. Mutations in Ras enzymes are responsible for nearly 30% of human cancers. Therefore, the potential of inhibiting mutant Ras activity makes FTase catalyzed farnesylation a popular research area.

FTase is a metalloenzyme that requires a zinc ion for activity and a second magnesium ion for optimal reactivity. The location of zinc binding site has been determined at the edge of β subunit near the subunits interface: the metal ion forms a tetrahedral coordination with three enzyme AA side chains, namely Asp297 β , Cys299 β and His362 β , while either a water molecule (before target peptide or enzyme binding) or a cysteine residue (Cys1p) from the substrate biological systems makes up the fourth coordination. Results from pKa experiments reveal that Cys1p interacts with zinc via a form of thiolate (Zn-S⁻) instead thiol (Zn-SH). In fact, a weak zinc-thiolate interaction is believed to a key to the catalysis cycle as such a bond can enhance the nucleophilicity of the bound thiolate¹³⁹. In contrast to its important role in chemical step, the zinc ion plays little role in the FPP binding, which is required for the later peptide binding. Thus, the substrate binding scheme for FTase can be described as follows: zinc ion first binds into FTase with coordinating to 3 AA side chains and a water molecule, then FPP binds into the highly positively charged pocket surrounded by Lys164 α , Arg291 β , Lys294 β and possibly a magnesium

ion if present, finally the peptide or enzyme substrate localizes itself with the Cys1p replacing the water molecule to fill up Zn^{2+} tetrahedral coordination. However, the resolved ternary crystal structure of FTase•FPP•CaaX complex shows an approximate 7 Å distance between the thiolate and another reacting center, C₁ atom of FPP. (The active site snapshot is given in Figure 3-1.) In order to bring two reacting atoms close enough into reacting range, a conformational rearrangement features two substrate approaching each other is necessary. Unfortunately, experiments could not give further support to reproduce this procedure.

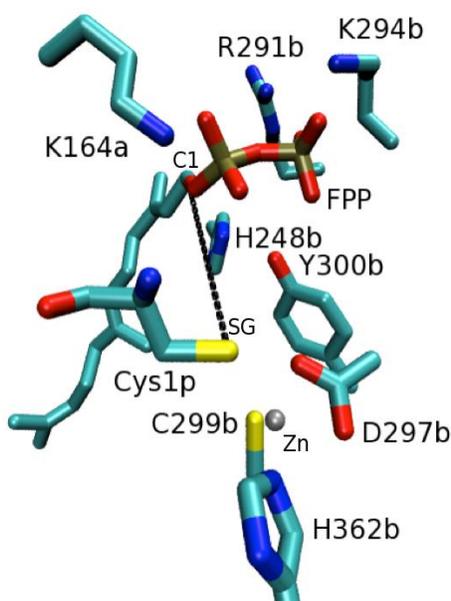


Figure 3-1. Snapshot of FPP binding pocket. Two reacting centers are labeled as C1 (C₁ of FPP) and SG (S_γ of Cys1p).

Magnesium cannot restore the full reactivity of FTase without zinc. However, its existence enhances the rate of product formation by up to 700-fold. In terms of Gibbs free energy, the bound of Mg^{2+} causes approximate 4 kcal/mol difference. Additionally, recent study reveals that (1) Mg^{2+} is interacting with Asp352 β and at least a non-bridging oxygen atom from β -phosphate group and (2) such a coordination can also contribute to the physical step that makes required conformation transition. However, unlike many other Mg^{2+} metalloenzymes, a rich aspartate

DDXXD motif is not present in FTase, making the localization of this ion somewhat difficult. In catalysis, a millimolar of Mg^{2+} is enough to make the huge impact on reaction rate enhancement, resulting in the lack of this ion in any of currently available FTase structures.

Computational effort has already been made to analyze the FTase structure and illustrate its conformational transition mechanism^{140,141}. With the Mg^{2+} ion excluded, the results showed the monoprotonated form of FPP ($FPPH^{2-}$) is capable of making such a transition, while the fully deprotonated FPP (FPP^{3-}) needs a very high energetic cost thus is unlikely able to undergo this procedure. Due to the limitation of classical MD simulation, this PMF study was carried out to move the RC ($d_{C1-S\gamma}$) from 8 Å to 4.5 Å. Such a range is sufficient to cover the transition from the resting state and also able to avoid drastic VDW clashes associated with closely placing two atoms. In the light of this success, a computational modelling of Mg^{2+} in FTase active site and a series of consequent validations have been conducted and will be presented here.

Also included in this chapter is the modelling and simulation of GGTase I. GGTase I and FTase share a great deal of similarities: a common α subunit and their β subunits are highly homologues, with approximately 25% of residues identical and 32% of similar AA's. The active sites of two zinc metalloenzymes are particularly similar. In FTase, FPP possesses a binding pocket surrounded by Lys164 α , His248 β , Arg291 β , Lys294 β and Tyr300 β . While in GGTase I (Figure 3-2), the binding pocket of GGPP is circled by Lys164 α , His219 β , Arg263 β , Lys266 β and Tyr272 β . In fact, these residues are highly conserved in CaaX prenyltransferases.

An interesting fact is that GGTase I shows no Mg^{2+} dependency for reactivity, whereas FTase requires this second metal ion for optimal activity. Experimental study of GGTase I revealed that Lys311 β plays a similar role to Mg^{2+} in FTase. Consider also this place in FTase is taken by Asp352 β , the proposed most important Mg^{2+} binding ligand, this fact becomes more

interesting since in such highly related enzymes this change may relate to important functional switch. In fact, in mutagenesis studies, the K β 311D mutant displayed Mg²⁺ dependency, indicating that Mg²⁺ binding site could be restored with an aspartic acid replacing the lysine. Nevertheless, in the crystal structure of GGTase I complex, this lysine is over 5 Å away from the nearest diphosphate oxygen, making its role in the catalysis process unclear. It is also unclear whether GGTase I binds a monoprotonated GGPP (GGPPH²⁻), as FTase does in the absence of Mg²⁺, or a fully deprotonated GGPP (GGPP³⁻) due to the existence of Lys311 β . Therefore, it is also of great interest to perform simulation with GGTase I and make comparison with FTase.

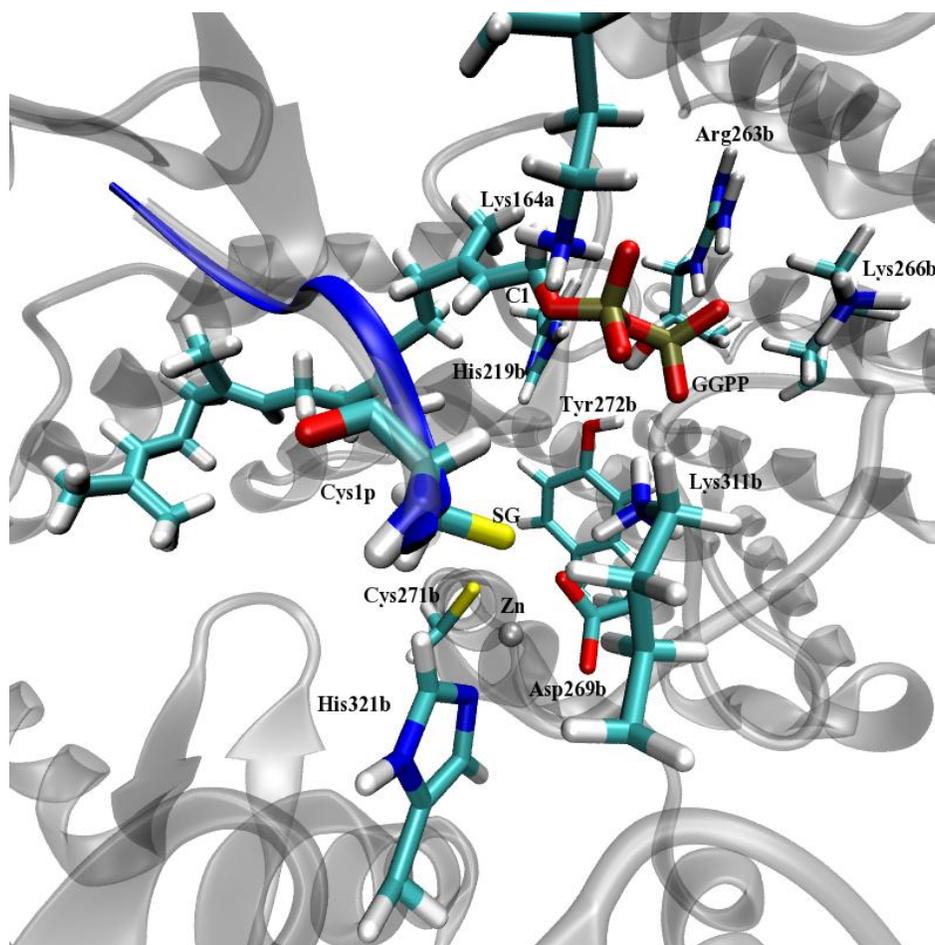


Figure 3-2. Active site snapshot of GGTase I. C₁ of GGPP and S_Y of Cys1p are labeled as C1 and SG.

3.2 Methods and Theories

3.2.1 Building Initial Model

Two crystallographic structures of FTase complex from protein data bank (PDB), PDB code 1QBQ and 1FT2, are superimposed to give an active FTase complex composed of FTase, FPP and acetyl-CVIM peptide¹⁴⁰. Crystal water molecules from PDB code 1QBQ are completely retained in the new complex. All charged AA's natural protonation states under physiological pH are adopted. Histidine residues are all modeled in the neutral form and Cys1p is modeled as thiolate, as suggested based on experimental observations¹⁴². Zinc ion is coordinated with Cys1p, Asp297 β , Cys299 β and His362 β . The enzyme and peptide are modeled with AMBER force field ff99SB⁸⁹ while the zinc site is modeled using the parameters developed by Cui et al. based on a bonded approach (Table 3-1). The zinc VDW parameters, $\epsilon=1.10$ and $r_0=0.00125$, developed by Merz et al., are adopted¹⁴³. FPP is modeled in its fully deprotonated state (FPP³⁻). FPPH²⁻ is excluded from this study since the protonation of diphosphate unit leads to a loss of magnesium binding affinity¹⁴⁴. Force field parameters applied to FPP are obtained from generic AMBER force field (GAFF)¹⁴⁵ for organic molecules, the torsion parameters for the farnesyl group developed by Merz and coworkers and the diphosphate parameters developed by Carlson lab are adopted¹⁴⁶. In order to obtain the atomic charges for FPP³⁻, a geometry optimization of FPP³⁻ is first performed using B3LYP/6-31G* level of theory, then the electrostatic potential derived atomic charges are computed using HF/6-31G* level of theory, finally the atomic charges are derived using a two stage restrained electrostatic potential (RESP) fitting procedure^{147,148}. Within such a procedure, both QM calculations are conducted with Gaussian 03¹⁴⁹ while the fitting process is performed using ANTECHAMBER integrated in the AMBER 10 suite of programs¹⁵⁰. AMBER 10 is also chosen to carry out all MD simulations.

3.2.2 Modelling and Simulating Mg^{2+} in FTase

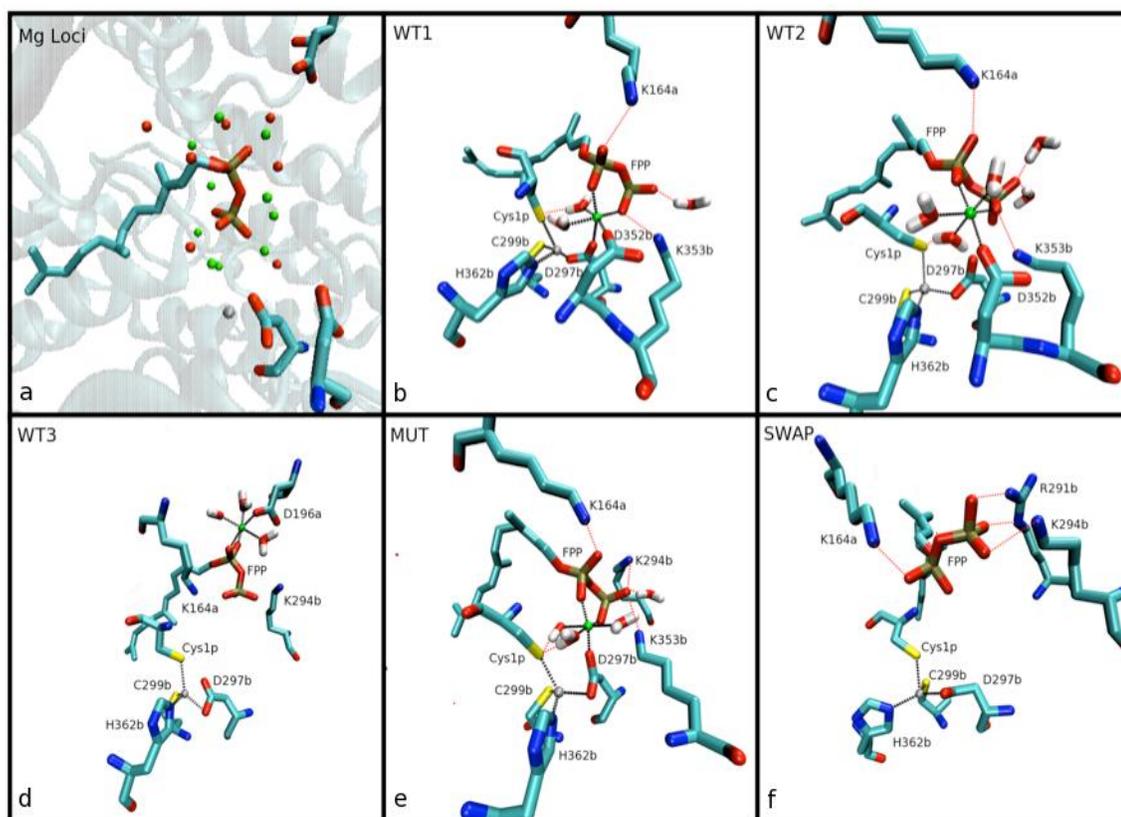


Figure 3-3. Snapshots of FTase active site from A) initial structure with 10 Mg^{2+} placement loci; B) WT1 model; C) WT2 model; D) WT3 model; E) MUT model and F) SWAP model. In each model, Zn^{2+} and Mg^{2+} are represented as grey sphere and green sphere separately. Mg^{2+} coordination is labeled in bold dashed line. Zn^{2+} coordination is shown in light dashed line. Important hydrogen bond interactions are labeled in red dashed lines. Residue from α and β subunit are labeled with a and b separately.

By carefully examining the FPP^{3-} binding pocket, three aspartate residues, namely Asp196 α , Asp297 β and Asp352 β separately, are determined as suitable candidates for magnesium binding. The latter two have been recognized as the most possible Mg^{2+} binding ligands based on available experimental evidence¹⁵¹. A total of 10 guess starting structures are created (Figure 3-3A). In each of them, Mg^{2+} is placed close to negatively charged diphosphate moiety and one of three candidate aspartic acid residues. Neighbouring crystal water molecules make up the octahedral coordination of Mg^{2+} , which has been identified in many Mg^{2+}

metalloenzymes. AMBER default VDW parameters for magnesium ion, $\epsilon=0.7926$, $r_0=0.8947$ ¹⁵², is applied. Unlike Zn^{2+} which parameters are derived through a bonded approach, Mg^{2+} is modeled as a free ion, since success has been achieved with such a strategy in previous MD simulations of NphB¹⁵³.

Each of new generated FTase/CVIM/FPP³⁻/ Mg^{2+} complexes is solvated in a truncated octahedron periodic water box. Each side of the box is at least 10 Å away from the closest solute atom. The TIP3P water model is employed and counter ions are added to neutralize the system. PME method¹²⁴ is applied to compute long-range electrostatic interactions with 8 Å and 1 Å set as cutoff of the nonbonded interactions in real-space and reciprocal space grid spacing separately. The translational movement of the system is removed. The motions of covalent bonds with at least one hydrogen atom involved are constrained with SHAKE algorithm.

Each system is first relaxed with 5000 steps of steepest descent minimization followed by 5000 steps conjugate gradient minimization in order to remove any close contacts. Within both stages of minimization, a weak positional restraint is applied to the enzyme, peptide and zinc, while free movements of magnesium and FPP³⁻ are allowed. Then the system is slowly heating up to 300 K over 100 ps time period using Langevin dynamics with a 1 ps⁻¹ collision frequency in canonical ensemble. A 1 fs time step is used in this stage to improve the stability of MD. All restraints are gradually removed over the same period of time. Finally, a 6 ns production MD simulation is carried out in an NPT ensemble with a 2 fs time step. For every picosecond, a snapshot is saved. The last 4 ns trajectory is used for analysis, validating the corresponding Mg^{2+} binding pose.

One of the Mg^{2+} binding scheme, which features the Asp297 β bridged Zn-Mg site (WT1 model in Figure 3-3B), proposed by Fierke and coworkers¹⁵¹ is artificially constructed by applied

strong NMR restraint in the minimization stages to force the Mg^{2+} -Asp297 β coordination in place of a Mg^{2+} -water coordination. The remaining stages of simulations are carried out using the same strategy as described previously.

Table 3-1. Zinc parameters used in MD simulations¹⁴⁰.

Atom/Bond/Angle Types	Atomic Mass/Force Constant (kcal/mol $\cdot\text{\AA}^2$ / kcal/mol $\cdot\text{radian}^2$)	Atomic Charges (e)/Equilibrium Value (\AA /degree)
ZN	65.3700	0.6928
S1	32.0600	-0.6797
S2	32.0600	-0.6666
ZN-O2	180.9400	1.9000
ZN-NB	94.8200	2.2400
ZN-S1	167.6000	2.2100
ZN-S2	154.5300	2.4800
CT-S1	237.0000	1.8100
CT-S2	237.0000	1.8100
ZN-O2-C	88.0300	101.5300
ZN-S1-CT	80.8800	111.2800
ZN-S2-CT	77.3400	124.2400
ZN-NB-CV	42.2900	112.6900
ZN-NB-CR	41.8300	133.6800
O2-ZN-S1	12.5700	99.9500
O2-ZN-S2	10.5400	90.7300
S1-ZN-NB	34.4800	131.5900
S2-ZN-NB	25.1500	97.1700
S1-ZN-S2	11.8400	111.2900

An extend MD simulation in addition to the 6 ns MD simulation has been performed later with the most successful model to evaluate the FPP³⁻ interactions and Mg^{2+} coordination. With such a purpose, canonical ensemble that allows more flexible conformational change is adopted,

while Langevin dynamics with a 1 ps^{-1} collision frequency is again selected. This simulation lasts 10 ns. The time step is 2 fs.

3.2.3 Modelling D β 352A Mutation

Based on experimental evidences, the mutation of Asp352 β to Ala352 β in the presence of Mg^{2+} would result in significant loss in reactivity. In the light of this, a D β 352A mutant (MUT model in Figure 3-3E) is prepared based on the final snapshot of a successful Mg^{2+} -Asp352 β binding motif. All counter ions along with water molecules that are more than 6 Å away from FPP $^{3-}$ or Mg^{2+} are removed. The side chain of Asp352 β is modified to represent an alanine residue. Such a change leaves Asp297 β the only Mg^{2+} binding candidate. The resulting system is then simulated following the same strategy as described in Chapter 3.3.2.

3.2.4 Modeling Removal of Mg^{2+} from Active Site

A good tactic to validate the Mg^{2+} binding motif is to test whether the removal of Mg^{2+} from FPP $^{3-}$ binding site will restore the behavior of FTase/CVIM/FPP $^{3-}$ complex. To achieve this goal, the most straightforward approach is to switch the position of Mg^{2+} and a solute water molecule far away from the binding pocket (SWAP model in Figure 3-3F). The advantage of this approach over the ‘authentic’ removal of Mg^{2+} from the system is the avoidance of rebuilding the system (topology) thus any associated artificial impacts will also be avoided. As a result, the optimal comparability between SWAP model and original Mg^{2+} binding model is reached. In practice, the coordinates and velocities of bound Mg^{2+} and a water oxygen atom are exchanged. The involved water is referred to as $\text{H}_2\text{O}_{\text{swap}}$. The coordinates of two hydrogen atoms of $\text{H}_2\text{O}_{\text{swap}}$ are modified accordingly and the velocities are maintained. The remaining portions of the system are totally preserved. This approach also saved much effort such like re-minimization and re-heating up. However, a longer equilibration (4 ns) is required to allow the system to fully relax itself. This is followed by a 4 ns production MD under NPT ensemble.

3.2.5 PMF Validation

In order to further validate the binding motifs, PMF for each reasonable binding motif is calculated as a function of C_1-S_γ , which is referred to as RC. To compare to previous results from Cui and Merz¹⁴¹, the same coverage of RC, from 8.0 Å to 4.5 Å, is chosen in this study. A MD scanning process is performed first to generate a set of initial structures for US. 15 such structures are prepared with an interval of 0.25 Å between adjacent two windows. A harmonic potential of $20 \text{ kcal/mol} \cdot \text{Å}^2$ is applied to each window throughout 2 ns simulation with a 2 fs time step. Data are collected every 5 steps for the latter 1 ns simulation, resulting in 100,000 data per window. More windows, longer samplings, and larger force constant ($100 \text{ kcal/mol} \cdot \text{Å}^2$) are employed upon necessity, in order to improve overlaps. Dr. Grossfield's WHAM code¹⁵⁴ is utilized to unbiased the distribution of RC and reconstruct the FES.

3.2.6 Modelling and Simulation of GGTase I

PDB ID 1N4S represents a heterohexamer of GGTase I complexed with GGPP and a LCVIL peptide²⁰. A monomer is selected from the original PDB file and prepared as the initial structure. Two models are generated, adopting GGPP³⁻ and GGPPH²⁻ separately. Each of the resulting models is solvated into a truncated octahedron water box. TIP3P water model is adopted. AMBER force field ff99SB is applied to model the protein while GAFF is employed to model the isoprenoid substrate. The charge of GGPP is derived following the two-stage RESP fitting procedure, as described in Chapter 3.2.1. Long-range electrostatic potential is handled by PME method and SHAKE is applied to constraint hydrogen involve bond motion.

The resulted structures are minimized carefully with a weak harmonic positional restraint applied. After that, each of the systems is heating to 300 K over 100 ps time span with 1 fs time step. Langevin dynamics is employed as thermostat. Positional restraint is gradually removed

during this process. Then each system is equilibrated in NPT ensemble for 900 ps with 1 fs time step, followed by a 4 ns production run.

3.3 Results and Discussion

3.3.1 Modeling and Simulation of FTase/CVIM/FPP³⁻/Mg²⁺ Complexes

A set of 10 possible Mg²⁺ binding motifs are generated by manually placing magnesium ion at different loci in the FPP³⁻ binding pocket. All binding schemes are tested through MD simulations, resulting in three possible Mg²⁺ binding models, namely WT1, WT2 and WT3, respectively (Figure 3-3B,C,D). Each of three candidate aspartate residues, Asp196 α , Asp297 β and Asp352 β has been detected involving in the Mg²⁺ coordination at least once in three models.

In WT1, an Asp297 β bridged bimetal site is identified. Mg²⁺ is found to interacting with a pair of oxygen atoms each from α - and β -diphosphate, a single carboxylic oxygen atom from Asp297 β , a single carboxylic oxygen atom from Asp352 β , and two oxygen atoms from two water molecules.

In WT2, Mg²⁺-Asp352 β is the featured coordination. The other two candidate aspartates are not participating in the binding scheme. A single carboxylate oxygen atom from Asp352 β , a phosphate oxygen atom from α -subunit and a phosphate oxygen atom from β -subunit of FPP³⁻, along with three oxygen atoms from three water molecules make up the octahedral coordination with Mg²⁺.

In WT3, Asp196 α becomes the featured aspartic acid in the Mg²⁺ coordination. In such a model, Asp196 α donates a single carboxylate oxygen atom to Mg²⁺. An α - diphosphate oxygen atom four water molecules fill up the remaining of the octahedral coordination.

Each of these three binding motifs appears rational upon visual inspection. Octahedral coordination is observed in all three models, in agreement the natural preference of such a cation in many biological systems. Each motif includes at least one Mg²⁺-Asp interaction with an

exception of WT1 model which has two such interactions. Despite the absence of an aspartate-rich DDXDX motif identified in many Mg^{2+} metalloenzymes, a solvent accessible FPP³⁻ binding pockets allows multiple water molecules to make up the coordination. In fact, not only crystal water molecules, but also those from solvent box, are identified in the Mg^{2+} coordination in three possible binding schemes.

Recently, locally enhanced sampling (LES) method has become popular in determining such binding locations¹⁵⁵⁻¹⁵⁷. However, we are not able to achieve success with such an approach. A rational explanation is that Mg^{2+} carries a fairly strong positive charge thus it is rather difficult to move it around.

All three binding schemes are validated through MD simulations. Stable octahedral coordination of Mg^{2+} is observed in each of these simulations. Hence, at least one other important characteristic is required for further validation. Such an example is the distance of $\text{C}_1\text{-S}_\gamma$. It is experimentally determined that FTase complexes possess an over 7 Å distance between two reacting centers and require a conformational transition prior to the chemical step. Under such circumstances, d_{RC} becomes an important property to ‘grade’ the binding motifs, because it represents the level of difficulty to perform such a transition. The d_{RC} is measured throughout MD simulations and an average value is calculated for each of three binding poses in order for comparison. In WT1 and WT2 models, such an average distance is below 6.5 Å. Comparing to the 7.2 Å of d_{RC} measured in FTase/CVIM/FPPH²⁻ system from previous studies^{140,141}, it is reasonable to draw a conclusion that both WT1 and WT2 models represent more favorable structures than the same FTase and peptide complex with Mg^{2+} in absence, since less distance in RC is required to overcome to reach the reacting range. On the other hand, however, in contrast to WT1 and WT2 models, the average distance of $\text{C}_1\text{-S}_\gamma$ is found to be over 10 Å for WT3

model. Such a distance represents an approximately 50% increasing in length from the FTase/CVIM/FPPH²⁻ complex, indicating WT3 is unlikely an active complex. In addition, Mg²⁺ coordination in WT3 features a Mg²⁺-Asp196 α interaction via an α -subunit non-bridging oxygen atom of FPP, although a similar Mg²⁺ scheme has been identified in aromatic prenyltransferase NphB, the absence of a β -subunit FPP oxygen atom in the octahedral coordination is contradict to the model proposed by Fierke and coworkers¹⁵¹. Therefore, further effort has been more focused on WT1 and WT2 models, though the possibility of WT3 binding scheme cannot be completely excluded.

In nearly every crystallographically resolved FTase complex structures, Asp297 β and Asp352 β have been found highly conserved. Such a conservation somewhat implies the possibly important roles of these two aspartates in the chemistry of FTase. Amongst, Asp297 β has been found at least responsible for the localization of the indispensable zinc ion, while Asp352 β has been found particularly important for the magnesium chemistry in this prenyltransferase. Mutagenesis experiments has revealed an approximate 27-fold decrease in catalytic activity associated with D β 352A mutation, hence it is fair to say Asp352 β is essential for the 700-fold increase in catalytic activity associated with Mg²⁺ binding¹⁵¹. Based on experimental observations, the assumption has been proposed that Mg²⁺ plays an active role not only in the chemical step but also in the physical step¹⁵¹. With the chemical step remaining unclear, results from this study support the presumption that including Mg²⁺ in the FPP³⁻ binding pocket makes a contribution towards closing the 7 Å gap between two reacting centers. Both α - and β -subunit diphosphate oxygen atoms involve in the Mg²⁺ coordination in both models, consisting with the model proposed by Fierke et al¹⁵¹. Furthermore, the involvement of Mg²⁺ in FPP³⁻ binding pocket provides extra stabilization to the big negative charge on diphosphate moiety in addition

to the hydrogen bond interactions between diphosphate oxygen atoms and side chains of Lys164 α , His248 β , and Tyr300 β , presenting great potential of decreasing the free energy barrier of chemical step by stabilizing the developing 4 units of negative charge of the leaving group. In summary, it is reasonable to regard the Mg²⁺-Asp352 β interaction observed in both WT1 and WT2 models as our best estimate of the Mg²⁺ binding motifs in the FTase/CVIM/FPP³⁻/Mg²⁺ complex.

Similarities have been discovered in both models. The active site conformations from both WT1 and WT2 share great similitude. The key hydrogen bonds between diphosphate moiety and surrounding functionally important AA residues including Lys164 α , His248 β and Tyr300 β are well preserved, and a key hydrogen bond is identified between S γ of Cys1p and a Mg²⁺ bound water molecule in both models. In addition, Lys353 β , one of the neighbouring residues of Asp352 β , is found to move into FPP³⁻ binding pocket, localizing itself within hydrogen bond interaction range with β -subunit of diphosphate moiety and carboxylate group of Asp352 β in both cases (as well as one Mg²⁺ bound water in WT2 model). Such a movement brings an extra stabilization to the Mg²⁺ binding site. Another metal site, the tetrahedral coordination of Zn²⁺ maintains fairly well in both models. The bound of Mg²⁺ introduces strong electrostatic interactions with diphosphate moiety, disrupting the salt bridge interactions between the diphosphate moiety and Arg291 β and Lys294 β temporarily. However, in the extend MD simulations with WT2 model, these two salt bridges have been restored (Figure 3-5). An estimation of approximate 5 kcal/mol $\cdot\text{\AA}^2$ free energy penalty associated with breaking these interactions has been made by Cui and Merz¹⁴¹, while experimental results revealed a binding free energy of approximate 3-4 kcal/mol provided by these interactions. Thus, it is necessary to identify the exact roles these two positively charged residues played in the chemical step.

Based on results from mutagenesis experiments, Tyr300 β is functionally important to the catalysis process by forming a hydrogen bond interaction with α -subunit of diphosphate. In previous FTase/CVIM/FPPH²⁻ and FTase/CVLM/FPPH²⁻ studies, such a hydrogen bond was clearly identified. However, in the 6 ns MD simulation of both models, this hydrogen bond is not found. Instead, Tyr300 β moves a little bit away from the diphosphate and forms a hydrogen bond with the acetyl capping group. Another positively charged residue that acts as a hydrogen bond donor to the α -subunit of diphosphate, Lys164 α , maintains a stable salt bridge interaction with a non-bridging α -subunit oxygen atom throughout entire simulation for each model. It is interesting to note that these two residues change roles from previously FTase/CVIM/FPPH²⁻ study, in which Lys164 α forms a hydrogen bond with acetyl cap while Tyr300 β forms a hydrogen bond with α -subunit of diphosphate moiety. With such a roles exchange, the total contribution from these two residues to the system is maintained, as each of the peptide and the α -subunit of diphosphate moiety are stabilized by one of two residues.

Comparing WT1 and WT2, obviously the main difference is the function of zinc bound Asp297 β . WT1 model also adopts it as a Mg²⁺ binding ligand, which resembles the Mg²⁺ binding scheme proposed by Fierke and coworkers, while in WT2 model it functioning as only a zinc bound ligand. In WT1 model, the two metal centers are 3.3 Å apart, and the equilibrium d_{RC} is 5.8 Å, which suggests this scheme might represent the energetically more favored FTase/CVIM complex than either FTase/CVIM/FPP²⁻ or FTase/CVIM/FPP³⁻ studied previously. With two 2+ charged metal ions on each side, the bridging Asp297 β might constrain the system, making the active site less flexible. Such an assumption is supported by the evidence that d_{RC} only fluctuates in a fairly small range in WT1 simulation. In this sense, despite the short equilibrium d_{RC} , the

bridged bimetallic site may result in increased difficulty in performing required conformational transition to make two reacting atoms further approaching each other.

In WT2 model, there is no such an aspartic acid bridge. The two metal ions are around 5.1 Å separated in space. A water molecule takes the place of Asp297β to fill up the octahedral coordination of Mg²⁺. The equilibrium distance of C₁-S_Y is fluctuating around 6.1 Å, which is 0.3 Å longer than that found in WT1 model. A magnesium bound water molecule has been found to form a stable hydrogen bond with Cys1p in the peptide, resembling a substitute bridge connecting two metal sites. It is rational to suppose with this water ‘bridge’, the strain between two metal binding sites is less than that in WT1 model thus a more flexible active site in WT2 model can be expected. As expected, a larger fluctuation in d_{RC} is observed in the simulation of WT2 (Figure 3-4), making it easier for two reacting centers to further approach each other and reach a pre-organized state prior to the chemical step.

In contrast to the model proposed by Fierke and coworkers, either WT1 or WT2 model does not adopt a double Mg²⁺-Asp352β interaction. Instead, the aspartic acid side chain only donates a single carboxylate oxygen atom to Mg²⁺ coordination. Attempts were made to test the possibility of Mg²⁺ interacting with both carboxylate oxygen atoms of Asp352β. Nevertheless, a stable structure cannot be obtained.

In the 10 ns extended MD simulations carried out with WT2 model, the interactions of Arg291β and Lys294β to the β-subunit of diphosphate moiety are restored (Figure 3-5). The Lys353β, on the other hand, is pushed away accompanied by Lys294β moving back into the FPP³⁻ binding pocket. On the other hand, side chain residues of Lys164α, His248β and Tyr300β are all back into range to form hydrogen bond interactions with diphosphate. The distance between Zn²⁺ and Mg²⁺ is greatly relaxed to approximate 7.6 Å, allowing higher degree of

flexibility in the active site. The Asp352 β is less than 3 Å away from Asp297 β , while the backbone nitrogen of Asp352 β is within hydrogen bond range to either zinc bound Cys1p or Asp297 β , indicating the close relationship between two metal sites. Finally, the equilibrium d_{RC} is maintained as 6.1 Å, in support of Mg²⁺ playing active role in conformational transition step.

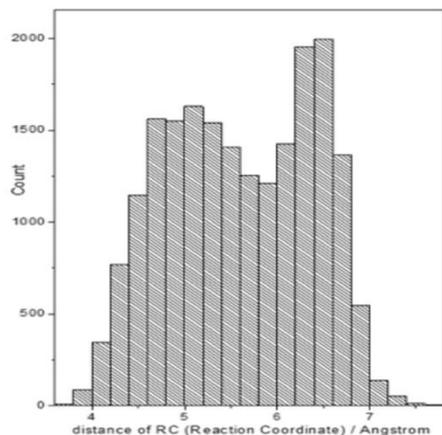


Figure 3-4. Distribution of d_{RC} in the original 6 ns MD simulation of WT2 model.

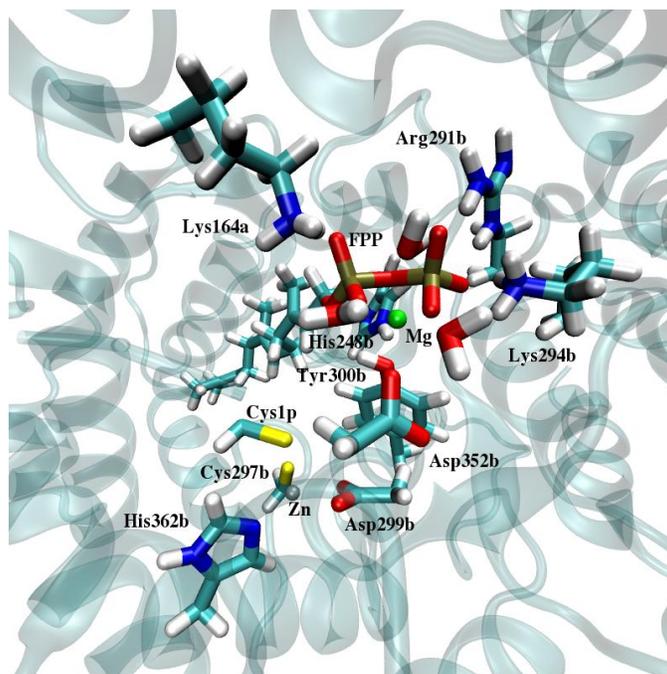


Figure 3-5. Snapshot of active site from the extend 10 ns MD simulation of WT2 model.

3.3.2 D β 352 Mutant and Mg²⁺/H₂O Swap Simulation

An approximate 30-fold difference in the rate of product formation has been observed associated with D β 352A mutation in mutagenesis experiments. In the mutant system, the most desired binding coordinating candidate, Asp352 β , is missing, thus another aspartate ligand, Asp297 β that is also a zinc bound ligand, becomes the substitute.

MD simulations are also conducted to study the impact of this mutation. The initial structure is taken from and modified based on a snapshot from the simulation of WT2 model. Through the new series of simulations, a substitute Mg²⁺ binding motif is observed and validated. Because of the involvement of Asp297 β , a bridging bimetallic site is restored, similar to that in WT1 model. However, different to WT1, a water molecule picks up the role of Asp352 β , filling up the octahedral coordination. A distance of 3.4 Å between two metal centers is observed, marginally longer than that observed in WT1 but significantly shorter than that in WT2. It is worthy noting that Asp297 β , as a substitute of sequence conserved Asp352 β , is also sequence conserved.

Several important hydrogen bond interactions observed in WT2 are well preserved in this MUT model, including the α -diphosphate to Lys164 α , β -diphosphate to Lys353 β and Tyr300 β to the capping acetyl group. Water molecules play important roles in stabilizing both substrates. Several solvent water molecules form hydrogen bonds with oxygen atoms on the β -diphosphate moiety, while a couple of Mg²⁺ bound water molecules interacting with Cys1p via hydrogen bonds. In the equilibrium state, His248 β is found to forming hydrogen bond interaction with Tyr300 β instead of α -diphosphate, although it is still within a range to form hydrogen bond with the nearest α -diphosphate oxygen. On the other hand, Lys294 β has moved back forming a salt bridge with β -diphosphate oxygen, while Arg291 β is also back to within 4 Å away from the closest oxygen atom of FPP³⁻.

The equilibrium d_{RC} is approximate 6.5 Å, slightly longer than that of both WT1 and WT2 model. However, it is still clearly shorter than that observed in FTase/CVIM/FPPH²⁻ complex in the absence of Mg²⁺. The value of d_{RC} somehow can be interpreted as a measure of difficulty level to undergo a conformational transition required by the system prior to the chemical step. In this sense, the predicted order of difficulty is: FTase/CVIM/FPPH²⁻ > MUT > WT1/WT2. Nevertheless, this trend is not obtained based on free energy, thus it is necessary to carry out PMF studies to determine the real trend.

Another series of MD simulations are carried out to study the SWAP model, in which the Mg²⁺ has been switched with an outside water molecule. Such a switch mimics the effect of removal Mg²⁺ from the enzyme complex, resulting in the system relaxing itself. Ideally, such a relaxation should give the same conformation of FTase/CVIM/FPP³⁻ without Mg²⁺ from previous study. Through the MD simulation, a similar FPP³⁻ binding pocket as that observed in the resting state of FTase/CVIM/FPP³⁻ complex is obtained. More importantly, key interactions are all restored as expected. In the resulting structure, Lys164 α , His248 β , Arg291 β and Lys294 β all form salt bridge or hydrogen bond interactions with diphosphate while Tyr300 β is also within the hydrogen bond range to the nearest diphosphate oxygen. Asp352 β , on the other hand, moves away from the binding pocket accompanied with the disappearance of Mg²⁺. Also moved away is Cys1p, leaving the equilibrium d_{RC} back to approximate 7.4 Å, well represent its status without Mg²⁺ binding.

3.3.3 PMF Studies

Another good approach to validate these Mg²⁺ binding models is the PMF study. However, it is not a straightforward task to study free energy in MD simulations due to the difficulty to sample certain phase spaces, in particular those associated with high energies, along the RC. In order to achieve this goal, special techniques must be implemented in order to obtain distribution

information over all phase spaces required by free energy calculation. In this study, USP method is adopted. For each model, a set of at least 15 simulations is carried out with a harmonic bias potential placed between 4.5 and 8.0 Å. For WT1 model, in order to achieve better overlap between adjacent windows, a much stronger bias potential is employed and trajectories are propagated at 29 points along the same range of RC with an interval of 0.125 Å.

The most important observation is that the inclusion of Mg^{2+} clearly alters the free energy profile (Figure 3-6). For instance, the global minimum, which represents the resting state, shifts from 7.4 Å in FTase/CVIM/FPPH²⁻ complex to 5.8 Å in WT1 model. The 1.8 Å decrease in the equilibrium d_{RC} implies WT1 model energetically more favored than FTase/CVIM/FPPH²⁻ system, since less distance is required to bear a conformational transition. However, the free energy profile suggests that it might not be a correct Mg^{2+} binding model.

In fact, the free energy curve of WT1 model represents steep increases on both side of the global minimum, in accordance with the small fluctuations of d_{RC} observed in the MD simulation. The system, as a result, seems locked into the conformation associated with this resting state. A reasonable explanation is that the bridged bimetallic Zn^{2+} -Asp297 β - Mg^{2+} site is too rigid for the two substrates to get closer. Another possibility is that the zinc parameters based on a bonded approach are not good enough. In fact, the bond parameters to describe Zn^{2+} -Asp297 β coordination are 180.94 kcal/mol for force constant and 1.90 Å for equilibrium bond length. However, a set of newer fashioned parameters built with MTK++¹⁵⁸ that is based on bonded approach and Seminario method gives a 51.96 kcal/mol force constant and a 2.01 Å equilibrium bond length. Obviously, the latter set of parameters represents more flexibility. However, this discussion is not relevant to this study thus will not be further expanded here.

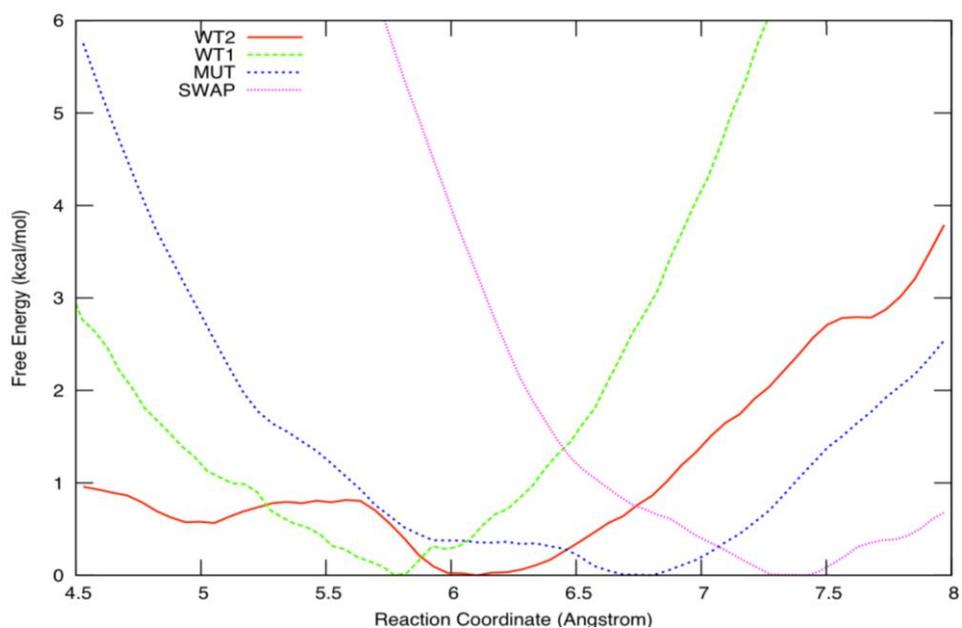


Figure 3-6. Free energy profiles of the modeled systems.

WT2 model, on the other hand, shows a global minimum at 6.1 Å along the RC and a second minimum at around 5.0 Å corresponding to an intermediate state during conformation transition. Although there is an about 0.3 Å increase in equilibrium d_{RC} the free energy curve appears more reasonable than WT1 model. The highest barrier of the transition is approximate 0.8 kcal/mol at around 5.6 Å, and the intermediate on the other hand is about 0.5 kcal/mol higher in free energy than the resting state. Furthermore, the curve on the left side of the global minimum is fairly flat, indicating it is easy for the system to undergo a conformational transition to close the gap between two reacting atoms and bring them into reacting range. During the transition, FPP^{3+} approaches the peptide substrate via the rotation of its first and second isoprene unit (Figure 3-7), which is also in agreement with the isoprene-rotation mechanism. Comparing to the free energy profile of FTase/CVIM/ $FPPH^{2-}$ complex that has a global minimum at around 7.4 Å and an 0.5 kcal/mol transition intermediate at around 5.0 Å separated by a barrier of about 1.0 kcal/mol at around 5.8 Å, clearly WT2 possesses a much shorter d_{RC} at equilibrium and a

lower transition barrier. In addition, the free energy at 4.5 Å is lower than 1.0 kcal/mol for WT2 model, which is nearly only a half of that observed for FTase/CVIM/FPPH²⁻ complex. However, due to the strong VDW clashes possibly at distances shorter than 4.5 Å, a QM/MM level PMF study is required to further validate the last point.

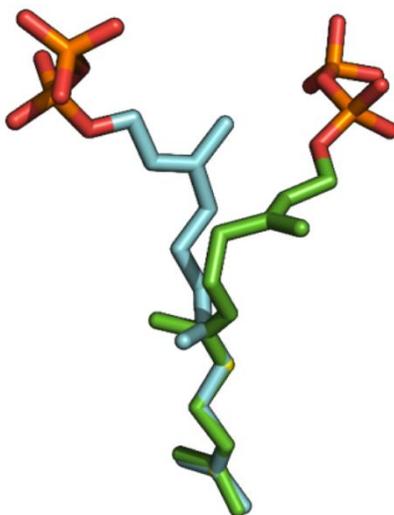


Figure 3-7. Comparison of FPP³⁻ orientation found in WT2 MD simulation and from previous FTase/CVIM/FPPH²⁻ (without Mg²⁺) MD simulation.

The free energy profile computed for MUT model possesses a global minimum at around 6.6 Å, approximately 0.5 Å longer than that before mutation. No second minimum has been found for this mutant, while the free energy penalty to bring two reacting centers from 6.6 Å to 4.5 Å is approximate 6 kcal/mol. From mutagenesis study, such a mutation results in a loss of reactivity by nearly 30-fold. In terms of free energy, the difference is approximate 2 kcal/mol. Apparently this value is much lower than the 5 kcal/mol difference between WT2 and MUT found from PMF study. However, by breaking down the profile of MUT into two parts, clearly a large portion of the free energy penalty comes with the 4.5 – 5.5 Å part, while from 5.5 Å to the resting state the transition requires only less than 1.0 kcal/mol free energy. This interesting fact

once again lifts the discussion on the quality of currently employed zinc parameters and the limitation of MM force field at the edge of potential VDW clashes. Nevertheless, it appears that this D β 352A mutation significantly impacted the physical step thus it is of great interest to further determine its impact in the following chemical step.

The PMF of SWAP model finds great similarity with the profile obtained from previously studied FTase/CVIM/FPP³⁻ complex in the absence of Mg²⁺. Both models find the global minimum at around 7.5 Å, and both systems require an approximate 3 kcal/mol•Å free energy penalty to close the big gap between two reacting centers. The excellent agreement between these two models further confirms the validity of our Mg²⁺ binding models complexed with FTase, in particular the WT2 model.

3.3.4 Modelling and Simulation of GGTase I

The independency of GGTase I reactivity on Mg²⁺ is interesting if one considers the great homologous shared by GGTase I and FTase. In the absence of Mg²⁺, the most important question is whether there is an AA side chain can play a substitute role. Experimental results reveal that Lys311 β is functionally important to the catalysis of GGTase I²¹. However, from the simulations carried out with both GGTase/CVIL/GGPP³⁻ and GGTase/CVIL/GGPPH²⁻, it is revealed that this residue will approach the β -diphosphate after around 5 ns of simulation and forms salt bridge interaction to stabilize that moiety together with Arg263 β and Lys266 β . Such an approach associates with the decrease of $d_{C1-S\gamma}$ (see Figure 3-8). This correlation is particularly evident in GGTase/CVIL/GGPPH²⁻ complex. Another question of the same importance is whether GGTase I adopts a monoprotonated GGPPH²⁻ like the FTase does in the absence of Mg²⁺ or it possesses a fully deprotonated GGPP³⁻. This question cannot be answered based on current results from regular MD simulations at present. Although the presence of positively charged Lys311 β near the β -diphosphate may prefer a more negatively charged diphosphate moiety, the possibility of

GGPPH²⁻ cannot yet be totally excluded. Hence, the best solution to these two questions is the PMF study, which is still ongoing currently.

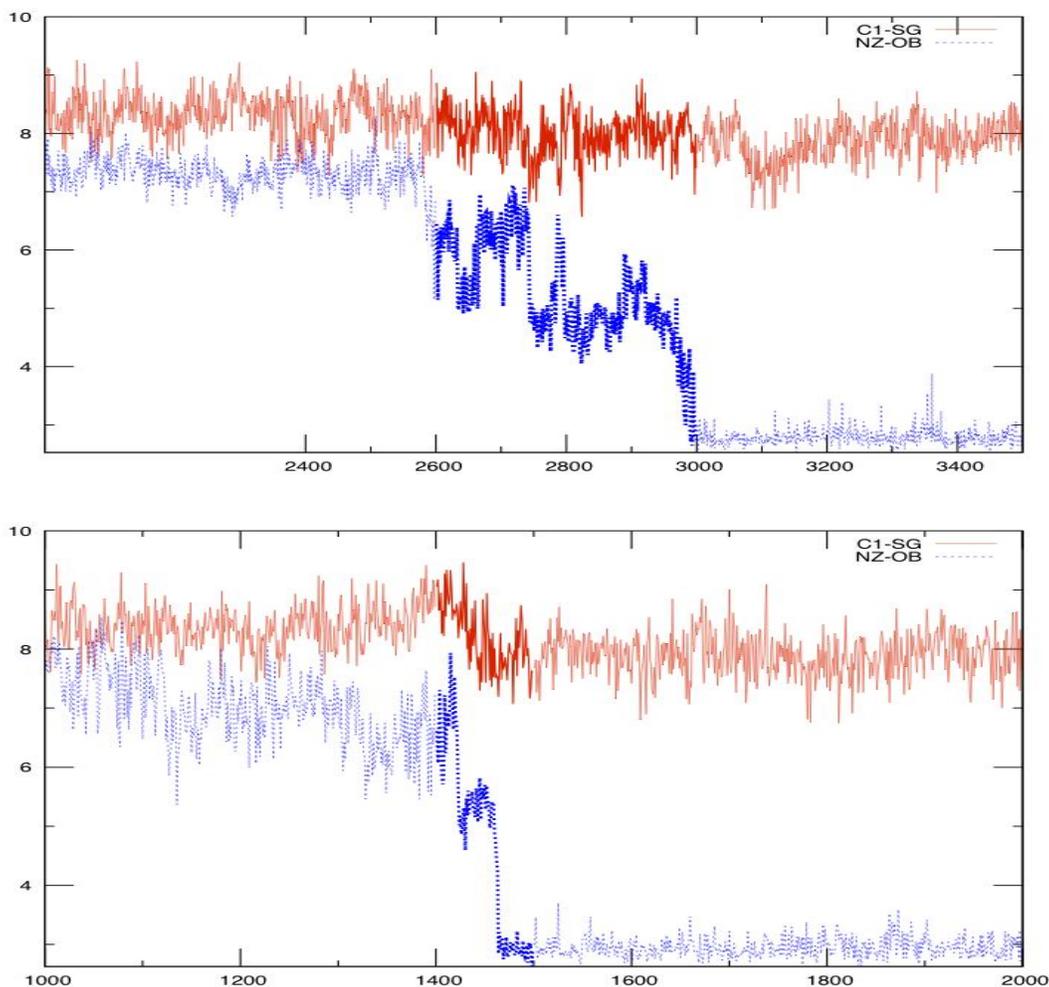


Figure 3-8. Distances variation in the simulation of GGTase/CVIL/GGPP³⁻ (top) and GGTase/CVIL/GGPPH²⁻ (bottom). X-axis represents frames. Y-axis represents distance in Å. Blue line represents the distance between Lys311 β and diphosphate while red line represents $d_{C1-S\gamma}$. Highlighted are the processes that Lys311 β approach the diphosphate.

On the other hand, MD simulations provide useful information to evaluate important ligand-substrate interactions. GGPP shares very similar binding pocket with FPP in FTase, the salt bridges between diphosphate and Lys164 α , Arg263 β , Lys266 β , as well as the hydrogen bonds between diphosphate and His219 β and Tyr272 β , are all well preserved during the

simulations for both models, making it impossible to determine a more favorable model based on these interactions. Tyr272 β corresponds to Tyr300 β in FTase, which is also found interacting with β -diphosphate in the resting state simulation but switch to stabilize α -diphosphate via hydrogen bond interaction after the conformational transition. Thus it is reasonable to assume Tyr272 β shall move away from β -diphosphate to α -diphosphate accompanied by system bearing conformational transition.

The 8.2 Å equilibrium d_{RC} represents more challenge for the conformation transition preformed by GGTase I than that conducted by FTase. A reasonable explanation of this increasing in d_{RC} lies on GGTase I possessing a binding pocket that ‘locks’ the terminal isoprene unit of GGPP thus making the rotation of this substrate to peptide cysteine more difficult. Kinetic experiments reveal that the k_{cat} for GGTase I is $(0.18 \pm 0.02) \text{ s}^{-1}$, corresponding to 18.6 kcal/mol in Gibbs free energy. This value is lower than FTase complexes without Mg^{2+} but higher than the FTase/ Mg^{2+} complexes. It is yet clear to say whether the difference attributes to the longer distance of RC or the less stabilization Lys311 β provides. Furthermore, GGTase I can also accept FPP and catalyze the farnesylation. A QM/MM study is needed to elucidate the nature of GGTase I chemistry.

3.4 Summary

Mg^{2+} binding schemes are modeled and validated through regular MD simulations and PMF studies with umbrella sampling. Among the three possible binding motifs, only WT2 is found to match all the existing experimental observations. Asp352 β is coordinating with Mg^{2+} , as proposed based on mutagenesis results, however, another proposed Mg^{2+} bound ligand, Asp297 β is found not directly involved in Mg^{2+} binding. The Zn^{2+} -Asp297 β - Mg^{2+} bridged bimetallic site appears incapable of lowering the free energy barrier associated with conformational transition, thus making itself an unreasonable choice for Mg^{2+} coordination. A

hydrogen bond formed between Cys1p and Mg^{2+} bound water molecule is detected and proposed to contribute to reduce the d_{RC} .

D β 352A mutant is also studied. In this system, the Asp297 β is detected to bridge the tetrahedral Zn^{2+} site and octahedral Mg^{2+} site. However, difficulty in closing the distance gap between two reacting atoms is found in this model, although the bridge seems increasing the connection between two binding sites thus shortened the d_{RC} .

USP method is employed to study the PMF of both models. WT2 possesses a flat free energy curve with less than 1.0 kcal/mol difference between the resting state and 4.5 Å. However, an approximate 6 kcal/mol free energy penalty is observed for the MUT system, making it much more difficult for FPP^{3-} to approach Cys1p in this case.

Introducing Mg^{2+} into the FPP^{3-} binding pocket enables the fully deprotonated FPP to the zinc bound peptide cysteine residue via isoprene rotation mechanism. It also disturbs the active site structure. For example, diphosphate moiety is obviously pulled away from Arg291 β and Lys294 β , at least temporarily. The loss of interactions between negatively charged diphosphate and these two positively charged residues is compensated by Mg^{2+} itself via coordinating to both a α - and a β -diphosphate oxygen atom. However, with longer relaxation time, the salt bridges between diphosphate and Arg291 β and Lys294 β are restored, along with the increasing distance between Mg^{2+} and Zn^{2+} .

After removing Mg^{2+} from the active site, the system relaxes itself back to a configuration represents the model of FTase/CVIM/ FPP^{3-} without Mg^{2+} . The equilibrium d_{RC} is also reverted to 7.4 Å. Such a complete restoration strengthens our conclusion that Mg^{2+} binding is the solo reason for the conformational change.

GGTase I, a highly homologous system to FTase, has also been modeled and simulated. Comparisons are made between two systems. Clearly, in GGTase I, Lys311 β , which takes the place of Asp352 β in FTase, reduces the Mg²⁺ affinity and at least partially plays a substituting role in GGTase I chemistry. However, during MD simulations of GGTase/CVIL complexed with both GGPP³⁻ and GGPPH²⁻, the distance between this lysine and its nearest diphosphate oxygen is more than 5 Å. Thus it either reaches in during the conformation transition or stabilizes the diphosphate indirectly via interaction with water molecule in between. Both protonation states give stable MD simulations, thus PMF study seems the only way to identify the substrate preference of GGTase I.

CHAPTER 4 SYSTEMATICALLY MECHANISTIC STUDY OF FTASE COMPLEXED WITH DIFFERENT PEPTIDE SUBSTRATES

4.1 Background

Conformational transition step is required by FTase complexes to close the big distance gap between two reacting atoms. However, this step does not account for the highest free energy barrier in FTase chemistry. Kinetic studies reveal that the rate-limiting step for FTase catalysis is the product release step^{159,160}. With single turnover kinetics, it is feasible to measure the free energy barrier for the product formation that covers both the conformational transition and the chemical reaction step^{139,161}. It is understandable the chemical step draws more attention, since reaction is the real essence of chemistry. In addition, as an important pharmaceutical target enzyme, more efficient FTase inhibitors are always of great interest to researchers^{23,162-176}. In drug discovery, more than often the enzyme inhibitor is designed to adopt or mimic the pose of transition state or intermediate state of enzyme catalyzed reactions. In this sense, it is of great importance to study the chemical step of FTase catalyzed farnesylation.

Many useful information and interesting facts have been uncovered by experimental approaches. Long, Casey and Beese have resolved the crystal structures of FTase complex at several key steps of its catalysis process. By analyzing the active site configuration of reactant complex and product complex, they proposed a transition state configuration which resembles a S_N2 reaction TS¹⁸. This assumption is supported by various experimental evidence, such as the observance of 'inversion' configuration in FTase catalysis¹⁷⁷ and the nearly unity measurement of α -SKIE by Weller and Distefano¹⁷⁸. However, not only S_N2 characteristics were found for FTase chemistry. Harris, Derdowski and Poulter have proposed a zinc dissociation mechanism featured the break of Cys1p and Zn^{2+} coordination before the chemical step¹⁷⁹. Nevertheless, this is inconsistent to the product complex structure resolved by Long, Casey and Beese¹⁸, thus

making itself irrelevant to the discussion. Also, several experimental groups have proposed a carbocation invoked dissociative mechanism for farnesylation^{139,180,181}. To explain the existence of both nucleophilic and electrophilic characters identified in FTase chemistry^{139,144,182-184}, another mechanism has been proposed. This new mechanism possesses an S_N2-like TS with dissociative characters. More evidence in support of this combined mechanism came from a computational approach conducted by Klein and coworkers. QM/MM MD simulations coupled with PMF study was employed and as a result a clear S_N2-like TS with S_N1-like dissociative characteristics was identified¹⁸⁵. Unfortunately, the energetic parameters obtained from their simulation are fairly off track, partially attributed to the use of FPP³⁻ instead of FPPH²⁻ as isoprenoid substrate in the absence of Mg²⁺. In addition, the QM theory they adopted in the study, BLYP, has been proved problematic with transition metals especially zinc¹⁸⁶.

Kinetic studies under single turnover condition provide valuable information regarding to the rate of product formation. One such example is the approximate up to 700-fold difference in rate of product formation observed in the presence or absence of Mg²⁺. FTase complexed with Dns-GCVLS peptide gives a k_{chem} of $8.1 \pm 0.3 \text{ s}^{-1}$ at saturating concentration of MgCl₂ and a k_{chem} of $0.025 \pm 0.001 \text{ s}^{-1}$ in the absence of MgCl₂¹⁵¹. Such k_{chem} values correspond to 16.3 kcal/mol and 19.8 kcal/mol Gibbs free energy with Mg²⁺ present or absent, respectively. This kind of energetic information, coupled with TST, could efficiently guide theoretical studies to illustrate the nature of FTase catalysis.

Some recent interesting experiment results made FTase catalytic mechanism more fascinating. Fierke and coworkers carried out ³H α -SKIE experiments to examine the reaction mechanism for FTase complexed with GCVLS and TKCVIF peptide separately. The measurement for TKCVIF peptide reads 1.154 ± 0.004 , while the GCVLS peptide gives a $k_{\text{H}}/k_{\text{T}}$

ratio of 1.00 ± 0.04 ¹⁸⁴. Based on the α -SKIE measurement for analogues enzymatic system, the first value falls well into the so-called associative mechanism with dissociative characteristics¹⁸⁷. However, the second value, which is nearly unity, is not straightforward to interpret. Fierke and coworkers proposed that such a value indicates the switch of the RLS from the chemical step to the conformational transition step. Later, the same group has also systematically analyzed the recognition of different $\text{Ca}_1\text{a}_2\text{X}$ motifs by FTase³¹ and proposed an interesting context-dependent a_2 and X selectivity of the substrate recognition. The impact of different a_2 and X could occur either at or before the chemical step. Combining these two studies a very interesting scheme with regard to the nature of FTase chemistry is given.

Despite the large amount of studies carried out on FTase, it is unfortunate to say that the study of farnesylation mechanism, in particular the transition state configuration, has not gone far away from proposals and hypotheses on paper. This is partially resulted from the limitation of current experimental techniques. However, computational study at QM/MM level can provide valuable insight into the catalytic mechanism of this important enzyme.

4.2 Methods and Simulation Details

4.2.1 FTase/CVIM/FPPH²⁻ and Y β 300F Mutant

FTase/CVIM/FPPH²⁻ system has been well studied in our lab¹⁴¹. This complex will be referred to as FPPH model in this chapter. It is reasonable to adopt one of the configurations extracted from the production simulation performed previously as the initial structure for this study. Because the system has been well equilibrated, any more classical simulation with this model is redundant.

The structure is partitioned into two parts: QM region and MM region. In order to limit the QM region to a moderate size, only those residues, ligands, or parts of them that functionally important to the chemical reaction are included. In this sense, the diphosphate group, the first

isoprene unit, the side chain of Cys1p, Asp297 β , Cys299 β and His362 β , all of which are zinc coordinating ligands, along with Zn²⁺ itself are selected as QM atoms (Figure 4-1). All the side chains are cut at the covalent bond between C $_{\alpha}$ and C $_{\beta}$, thus there is no polarized bond cut.

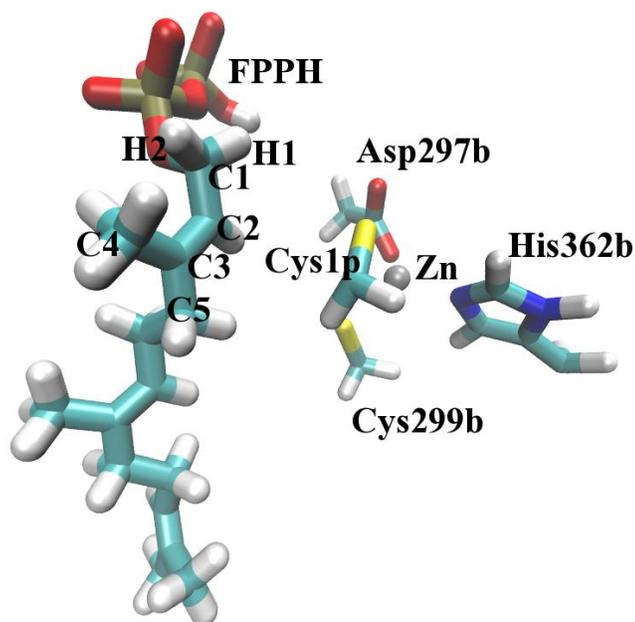


Figure 4-1. QM region selection: side chains of Asp297 β , Cys299 β , His362 β , Cys1p, Zn²⁺ and diphosphate and the first isoprene unit of FPPH. Note only labeled carbon and their hydrogen atoms in the farnesyl group are considered in QM.

Due to the existence of diphosphate, in particular the phosphor, traditional semiempirical methods such as PM3 and AM1 cannot handle the simulation properly¹⁸⁸. Therefore, SCC-DFTB^{189,190} becomes the best, if not the only, choice.

The system is first minimized using steepest descent method for 1000 steps to fully relax the system with hybrid SCC-DFTB/ff99SB potential¹⁹¹. A 100 ps Langevin dynamic with a 0.5 fs time step is employed to heat the system temperature back to 300 K in the canonical ensemble. Followed is an equilibration of 500 ps in NPT ensemble with a 1 fs time step. A cutoff of 8 Å is selected for both QM and MM portion of the system. SHAKE is applied to MM part to constrain

the motion of bond with hydrogen involved. Long-range electrostatic interactions in MM region are computed using PME method with the default settings. In QM region, SHAKE is turned off so that the free motion of hydrogen involved covalent bond is allowed, and a modified PME implementation designed for QM region is employed.

A couple of SMD simulations with a 5000 kcal/mol•Å force constant are conducted to propagate the trajectory along the RC to both 1.8 Å (product state) and 8.0 Å (reactant state). The RC is still defined as the d_{C1-SY} . Although at present a RC in the form of $d_{RC} = d_{C1-SY} - d_{C1-O1}$ is more popular, our preliminary results showed this selection will lead to a very unreasonable dissociative pathway, in accordance with the claim given by Klein and coworkers¹⁸⁵. 63 structures are extracted with an interval of 0.1 Å along the RC in order for the subsequent US. A set of 63 windows of USP is carried out to obtain the free energy profile. Each window is sampled in NTP ensemble for 150 ps wherein the first 50 ps is for equilibration purpose while the latter 100 ps is used for data collection. A harmonic force constant of 100 kcal/mol•Å is applied to bias all the windows. An approximate 10 ns total simulation is performed for US. Grossfield's WHAM code is employed to unbias the distribution and construct the PMF.

In order to study the impact of Yβ300F mutation, a single modification of replacing the hydroxyl group with a hydrogen atom at Tyr300β is made. The resulting system, denoted as Yβ300F model in this study, is partitioned and treated in the same way as FPPH model, except the QM/MM study for this model does not cover the 4.5 Å to 8.0 Å range that has already been studied classically. As a result, 28 windows, instead of 63, are simulated with USP method.

4.2.2 FTase/CVLM/FPPH²⁻

FTase/CVLM/FPPH²⁻ system (denoted as CVLM model in this chapter) is another complex that has been well studied by Cui and Merz¹⁴¹. The original purpose of this study is to testify the possibility of a switch of RLS from chemical step to physical step, as observed in

TKCVIF and GCVLS complexes. The computational results show an increase of conformational transition barrier, which is, however, too small to cause a RLS switch. Considering the context-dependent selectivity of both a_2 and X position in the Ca_1a_2X motif, it is not surprising to see this result, since only one position of the motif is changed in the simulation thus it is insufficient to reproduce the impact of double change studied in the experiment.

On the other hand, a complete mechanism of both the chemical and physical step will help to understand the substrate dependency at a_2 position of this four-unit peptide and provide useful information to illustrate the more complicated context-dependent substrate selectivity.

In this sense, a QM/MM PMF study is carried out to energetically study the chemical step. In the structure preparation step, isoleucine in the peptide is modified to leucine within the FPPH model. Thereafter all the procedures and treatments are exactly identical to that applied on $Y\beta 300F$ model.

4.2.3 FTase/CVLS/FPPH²⁻

The most interesting mutational effect associated with FTase chemistry seems related to the FTase/CVLS complex. In nature, GCVLS motif represents the C-terminus of human Ras enzyme, making this complex more attractive.

A crystal structure (PDB code: 1TN8)²² obtained at 2.25 Å resolution is adopted to prepare the initial structure. This structure represents FTase enzyme complexed with GCVLS peptide. In order to focus our study on the context-dependent of a_2 and X position, the glycine is modified to an acetyl capping group, in order to be consistent with our previous study. The result acetyl- Ca_1a_2X motif is consistent with all the models investigated in this chapter. The modified system is then solvated in an octahedral box with a distance of 8 Å to separate the edge of box and closest solute atom. The resulting model is denoted as CVLS in this study.

CVLS model is first treated classically. A cutoff of 8 Å is applied, while SHAKE and PME method are employed to improve the simulation. A series of minimizations with a combination of steepest descent method and conjugated gradient method is performed in the first stage to remove possible close contacts in the system. A weak positional restraint is applied to all of the non-water components. In the next stage, the system is slowly heating to 300 K with Langevin dynamics in NVT ensemble. During this state the restraint potential gradually removed. Then the model is allowed to fully relax in the NPT ensemble followed with a 10 ns production run.

A snapshot extracted from the last 1 ns of production simulation is chosen as the starting structure for the following MM PMF study and QM/MM simulations. The MM PMF study adopts the same strategy as described for WT2 model in Chapter 3.2.5. The QM region and MM region are selected following the same procedure described before. The subsequent treatment is identical to that applied to FPPH model.

4.2.4 QM/MM Validation of FTase/CVIM/FPP³⁻/Mg²⁺ Model

Mg²⁺ plays an important catalytic role in the FTase chemistry. Based on the available evidence, this divalent metal cation does not cause the approximate 3.5 kcal/mol difference in free energy during the conformational transition. Thus, it is rational to suppose the difference in rate of product formation only comes from the chemical step.

An approach is made to study the chemical step of our WT2 model from chapter 3 with QM/MM potential. Due to the lack of Zn-Mg parameter, Mg coordinating ligands with the exception of FPP³⁻ are not modeled with SCC-DFTB. This leaves the QM selection of WT2 model identical to that of FPPH except one less hydrogen from diphosphate group.

The starting structure of WT2 model is extracted from the trajectory of the extra 10 ns relaxation instead of the original 6 ns simulation. This is a reasonable move since it is found that during the extra 10 ns, the system relaxed itself more deeply. Thus the strain associated with the

inclusion of Mg^{2+} will be minimal. The Mg^{2+} and Zn^{2+} are more than 6 Å apart, adding less extra trouble to the troublesome SCC-DFTB convergence procedure.

The WT2 model is partitioned and treated with QM/MM potential in the same way as applied to FPPH and other systems.

4.3 Results and Discussions

4.3.1 A Glimpse of Entire Product Formation Process of FTase/CVIM/FPPH²⁻

In FTase complex, two reacting atoms, C_1 of FPP and S_γ of Cys1p are more than 7 Å away in the resting state¹⁸. In order to complete the catalysis, both a physical step that closes the distance gap and a chemical step that attaches the farnesyl moiety to the target cysteine are required. Thus, a comprehensive understanding of the catalytic mechanism requires a PMF study to cover both steps.

To accomplish this goal, we carried out a couple of SMD simulations to force the RC moving from the resting state to the product state and reactant state, respectively. The obtained free energy profile reveals a resting state at approximate 7.0 Å, an intermediate state at about 5.3 Å, a second intermediate state at approximate 3.9 Å that is very likely represent a pre-organized state of the complex, and a TS at approximate 2.6 Å as well as a product state at 1.8 Å. Although the barrier predicted by SMD simulation is frequently less accurate than that computed with USP, it usually provides a free energy curve with the ‘right’ shape. In this sense, an $\text{S}_{\text{N}}2$ -like TS is found, however, with a unusual long C-S bond observed that strongly implies ‘dissociative’ characteristics.

4.3.1.1 Transition state and free energy barrier

With the structures extracted from SMD trajectories, a set of 63 windows USP is then conducted in order to refine the results. The generated free energy profile (Figure 4-2) is similar to that gathered from SMD simulations. The resting state is located at 6.90 Å, while the first

intermediate state is identified at 5.32 Å, separated by an approximate 1.0 kcal/mol barrier. These results are in excellent agreement with the MM PMF results obtained by Cui and Merz¹⁴¹. In the region that has not been covered in their previous study, a second intermediate is found at 3.90 Å, associated with an approximate 6.5 kcal/mol free energy. It is worthy noting that at this point, O₁ and C₁ of FPPH²⁻ and S_Y of Cys1p start to align linearly. Such evidence can be well explained by the pre-organized state theory. The TS is located at 2.63 Å and the barrier height is approximate 20.6 kcal/mol. The experimental measurements are 21.1 kcal/mol for TKCVIF and 20.0 kcal/mol for GCVLS¹⁸⁴. Obviously, our result perfectly falls into this range. The product state is detected at 1.82 Å, associated with a free energy of approximate 4.5 kcal/mol higher than the resting state.

Upon reaching the transition state (see Figure 4-3), the C₁-O₁ bond starts to break and the distance between them starts to explode. The S_Y of Cys1p, C₁ of FPPH²⁻ and O₁ of FPPH²⁻ still stay linearly, however, with C₁ slightly out of plane. The H₁-H₂-C₁-C₂ (see Figure 4-1 for atom specification) dihedral is giving the highest value at this point, reading 168 ° that is very close to the ideal value of 180 ° for a typical S_N2-like TS. Moreover, such a dihedral changes sign after this point, meaning the two hydrogens bend out of plane in a different direction. The distance of C₁-S_Y is 2.63 Å, approximately 0.2 Å longer than d_{C-S} observed in a S_N2 reaction calculation carried out at MP2/6-31+G**//MP2/6-31+G* level¹⁹². The difference in TS C-S bond length might be a good indicator of the dissociate characteristic. Also, the distance of Zn-S_Y elongates to about 2.45 Å from its equilibrium value 2.38 Å, implying a weaker coordination of zinc.

After carefully examining the structure of each window along the RC, it is conspicuous that the distance gap between two reactants is closed by FPP rotating its first and second isoprenoid units, which has been proposed by several experimental groups. AA's in the FPP

binding pocket play important roles in the catalysis process. Lys164 α and Tyr300 β help stabilizing the α -diphosphate, His248 β forms hydrogen bond with the bridging oxygen atom, while the β -diphosphate is stabilized via salt bridges formed between its non-protonated oxygen atoms and the side chains of Arg291 β and Lys294 β .

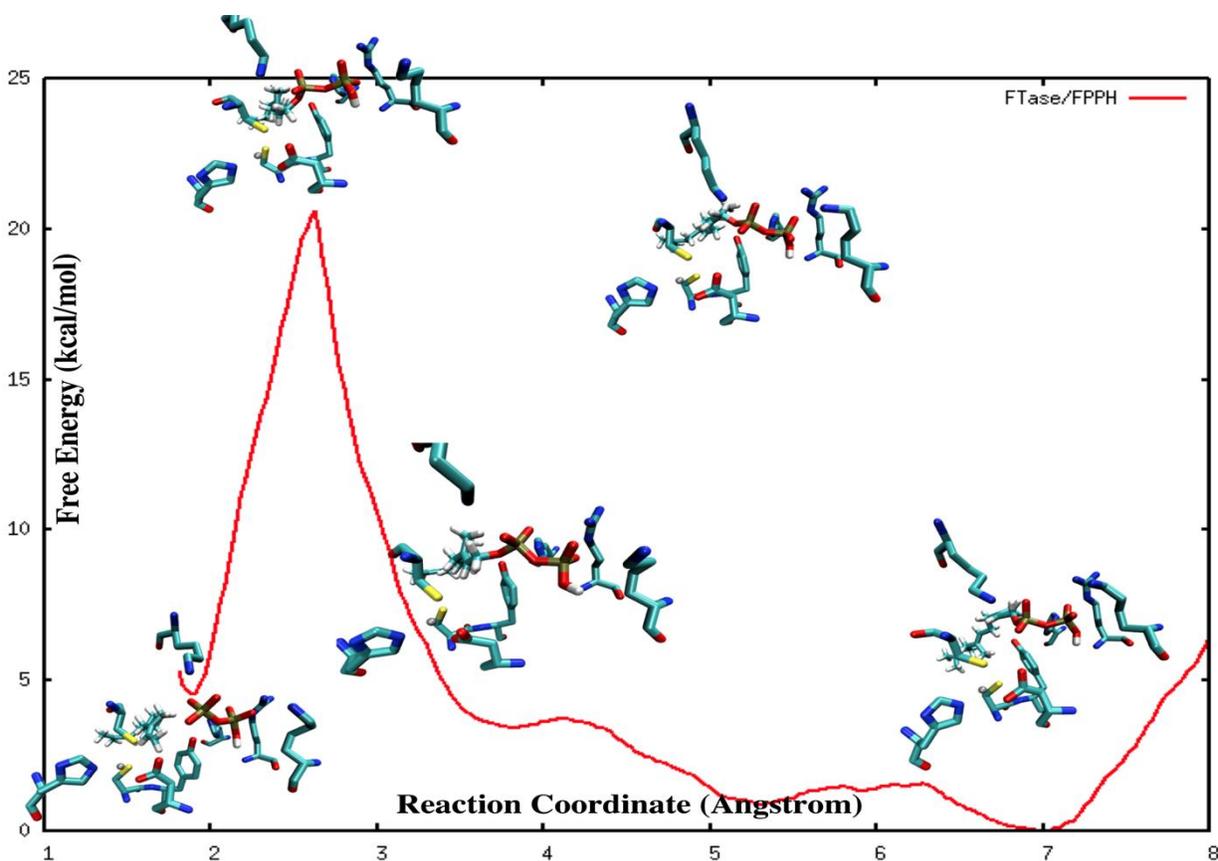


Figure 4-2. Free energy profile of FPPH model obtained from a set of 63 windows of umbrella sampling. Active site snapshots are given at key points. From right to left, (1) resting state; (2) first transition intermediate; (3) second transition intermediate; (4) transition state and (5) product state.

Zinc coordinates the peptide cysteine throughout the entire process. An increase of approximately 0.3 \AA of $d_{\text{Zn-S}_Y}$ from the FTase reactant complex (2.4 \AA) to the product complex (2.7 \AA) has been observed though comparing the crystallographic structures¹⁸. Such a difference results in a weaker Zn-S_Y coordination, which has been proposed to increase the nucleophilicity

at S_{γ} thus improving the reactivity. From our simulations, a change of about 0.2 \AA has been detected. Furthermore, the $Zn-S_{\gamma}$ starts elongating when the C_1-O_1 bond is about to break. Such observations seems in support to the hypothesis of weak zinc-sulfur interaction improve the nucleophilicity at S_{γ}^{139} .

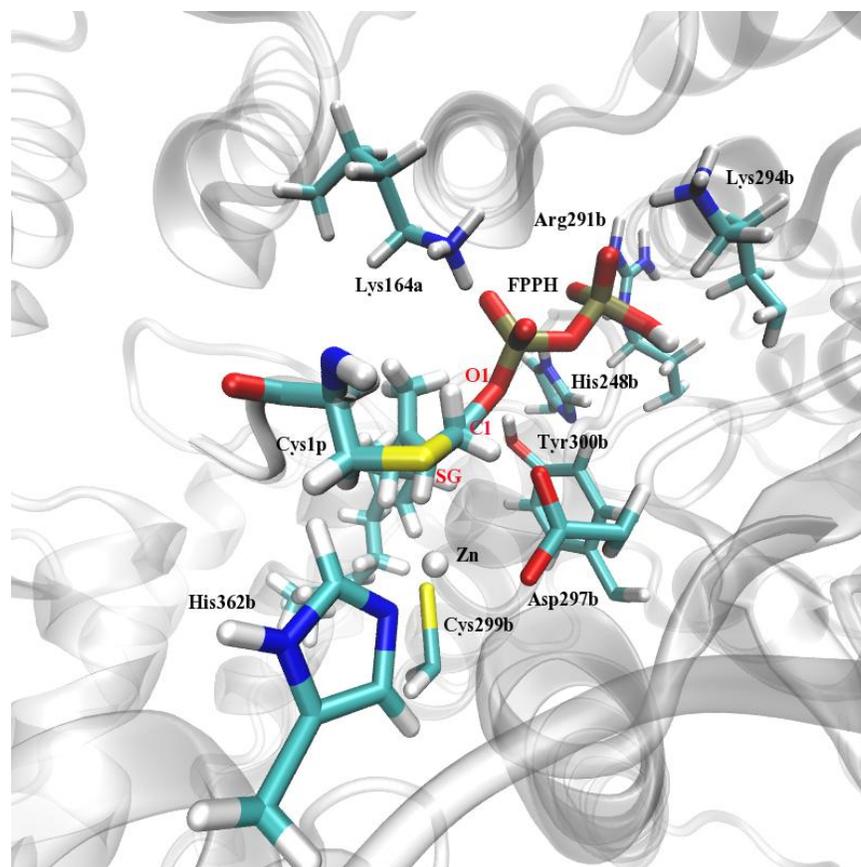


Figure 4-3. Snapshot of the active site of FPPH at the transition state. O_1 and C_1 of FPPH, and S_{γ} of Cys1p are labeled in red.

4.3.1.2 Comparison of QM/MM and MM PMF of conformational transition step

The difference between our QM/MM PMF results and MM PMF results of Cui and Merz in regard to the conformational transition step is fairly marginal, except at close to 4.5 \AA of RC where the potential of VDW clashes becomes significant (due to the sum of carbon and sulfur VDW radius is about 4 \AA). The little discrepancy at the resting state location, 6.9 \AA of QM/MM

prediction and 7.2 Å of MM results, as well as at the intermediate state, 5.3 Å from QM/MM and 5.0 Å from MM, are acceptable, since QM and MM are totally different approaches. In fact, the energy barrier separating the two states predicted by QM/MM and MM matches fairly well.

The biggest difference is the free energy value at 4.5 Å, which QM/MM gives around 2.7 kcal/mol while MM shows only 1.8 kcal/mol. As mentioned before, at this point, results from MM is suspicious due to the potential of VDW clashes. Whereas QM method does not have this limitation, therefore the difference is generated. In our following studies of Yβ300F, CVLM and WT2, the QM/MM PMF studies do not cover the 4.5 Å to 8.0 Å region, instead, MM results are used to predict the total reaction barrier based on the difference between corresponding value at 4.5 Å and that of CVIM at 4.5 Å. This is a reasonable approach, since the MM error at this point for each model will very likely cancel. Hence, a great deal of computational cost is saved.

4.3.2 Effect of Mutations

Mutagenesis study plays a very important role in the mechanistic study of enzymatic catalysis. For example, it helps to determine the function of key AA residues. Hence, it is of great importance to reproduce the results from such experiments in computational simulations.

Tyr300β is determined to stabilize the α-diphosphate in the chemical step of FTase catalysis. The k_{cat} measured for the product formation changes from $0.016 \pm 0.008 \text{ s}^{-1}$ of wild type (WT) FTase/GCVLS complex to $0.0002 \pm 0.00004 \text{ s}^{-1}$ of its Yβ300F mutant (both in the absence of Mg^{2+}). A nearly 80-fold loss of activity accompanies by the mutation, which corresponds to 2.6 kcal/mol difference in Gibbs free energy¹⁸².

The mutation in the peptide substrate plays more important role. It not only alters the free energy barrier height, but also appears to disturb the reaction mechanism. One good example is the α-SKIE experiment of FTase/TKCVIF complex and FTase/GCVLS complex¹⁸⁴. The $k_{\text{H}}/k_{\text{T}}$ measurement of the former system is 1.154 in the absence of Mg^{2+} while for the latter is nearly

unity, 1.00 ± 0.04 , under same condition. Such a phenomenon has been explained as the RLS switches from the chemical step (for TKCVIF) to the physical step (for GCVLS). However, consider the free energy barrier of FTase/GCVLS complex in the absence of magnesium ion is approximate 20.0 kcal/mol, it is rather unusual to see such high a energy barrier accompanies with the conformational transition.

4.3.2.1 Impact of Y β 300F mutation

Y β 300F mutant is prepared by manual changing Tyr300 β residue to Phe300 β . The resulting system is decently minimized and then simulated in the same manner as applied to CVIM model. However, PMF study has not been applied to the 4.5 Å – 8.0 Å range of RC due to the reason explained before.

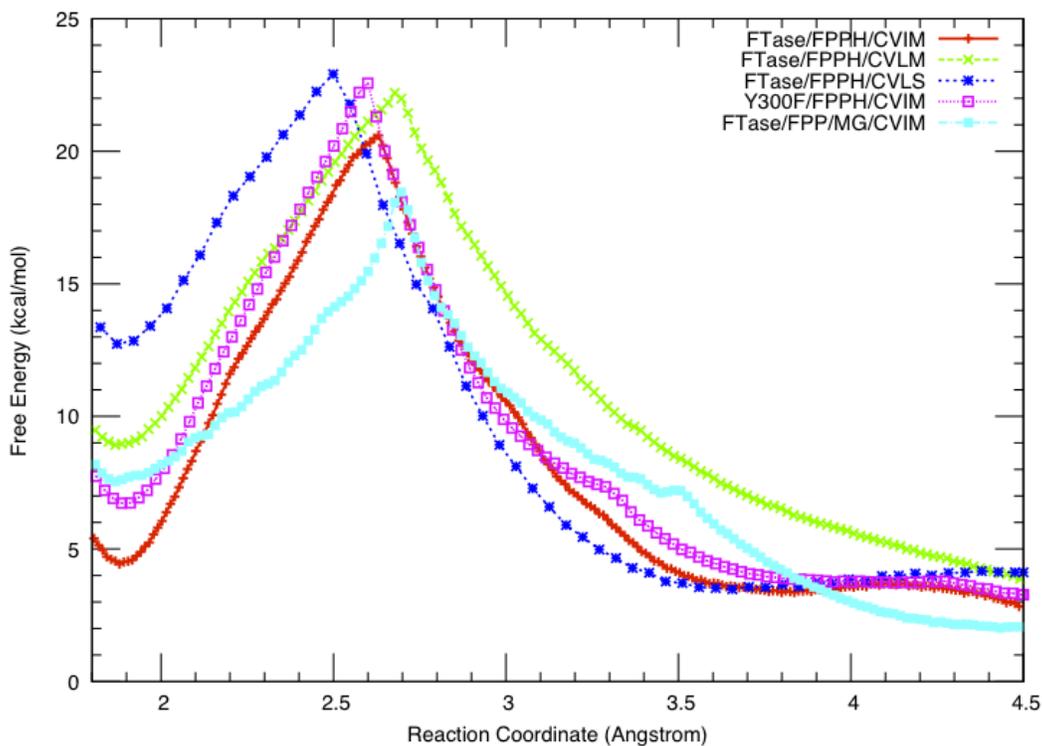


Figure 4-4. Free energy profiles of all five models included in this study.

The result free energy profile (see Figure 4-4) shows a TS at 2.61 Å associated with a free energy of 22.5 kcal/mol (see Table 4-1). Comparing to the 20.6 kcal/mol free energy barrier of CVIM model covering the same range of RC, the 0.7 kcal/mol discrepancy between calculated difference and experiment observation is almost within the chemical accuracy (1.0 kcal/mol) to the experimental observed difference.

Table 4-1. Free energy parameters obtained from QM/MM and MM PMF studies.

	$\Delta G_{\text{TS-CONF}}$ (kcal/mol) ¹	$\Delta G_{4.5\text{\AA-MM}}$ (kcal/mol) ²	$\Delta G_{4.5\text{\AA-QMMM}}$ (kcal/mol) ³	$\Delta G_{\text{TS-QMMM}}$ (kcal/mol) ⁴	$\Delta G_{\text{TS-PROD}}$ (kcal/mol) ⁵
FTase/FPPH/CVIM	1.0 (1.0 ⁶)	1.8	2.7 ⁶	17.9	20.6 ⁶
FTase/FPPH/CVLM	2.7	3.0	3.9 ⁷	19.3	22.7
FTase/FPPH/CVLS	2.5	2.8	3.7 ⁷	18.3	22.0
Y300F/FPPH/CVIM	1.5	2.3	3.2 ⁷	19.3	22.5
FTase/FPP/Mg/CVIM	0.8	1.0	1.9 ⁷	16.5	18.4

¹ $\Delta G_{\text{TS-CONF}}$ refers to the free energy barrier associated with the conformational transition taken place in 4.5-8.0Å region. It's obtained from MM PMF calculations.

² $\Delta G_{4.5\text{\AA-MM}}$ refers to the free energy at 4.5Å obtained from MM PMF calculations.

³ $\Delta G_{4.5\text{\AA-QMMM}}$ refers to the free energy at 4.5Å obtained from QM/MM PMF calculations.

⁴ $\Delta G_{\text{TS-QMMM}}$ refers to the free energy barrier of the chemical step, obtained from QM/MM calculations in 4.5-8.0Å region.

⁵ $\Delta G_{\text{TS-PROD}}$ refers to the free energy barrier of the entire product formation, covering reaction coordinate of 1.8-8.0Å.

⁶ These values are directly produced from QM/MM PMF in 1.8-8.0Å region.

⁷ These values are calculated as follows: $\Delta G_{4.5\text{\AA-QMMM}} = \Delta G_{4.5\text{\AA-MM}} + \Delta\Delta G_{4.5\text{\AA-FTase/FPPH/CVIM}}$, where $\Delta\Delta G_{4.5\text{\AA-FTase/FPPH/CVIM}} = \Delta G_{4.5\text{\AA-QMMM}} - \Delta G_{4.5\text{\AA-MM}} = 2.7 - 1.8 = 0.9$ kcal/mol.

Analyzing the TS structure (see Figure 4-5), it is not surprise to see similarities to that of CVIM model. The biggest difference is the loss of hydrogen bond interaction between α -diphosphate and the side chain of 300 β residue. This change allows more flexibility of FPPH²⁻ and later developed diphosphate leaving group (PPi). The shorter length of the forming C₁-S_Y bond at TS seems resulting from this extra flexibility, so does the longer distance between C₁ and O₁ after the bond broken. Phe300 β might also play a role in the catalysis by stabilizing the

farnesyl group via cation- π interaction, since shortest distance between the C₁, C₂ or C₃ atom of farnesyl (where based on the dissociative characteristic the delocalization of positive charge should take place) and the aromatic ring of Phe300 β is always within 4.0 Å throughout the transition from TS to the product state. However, more analysis such as higher level QM calculation is needed to validate this assumption.

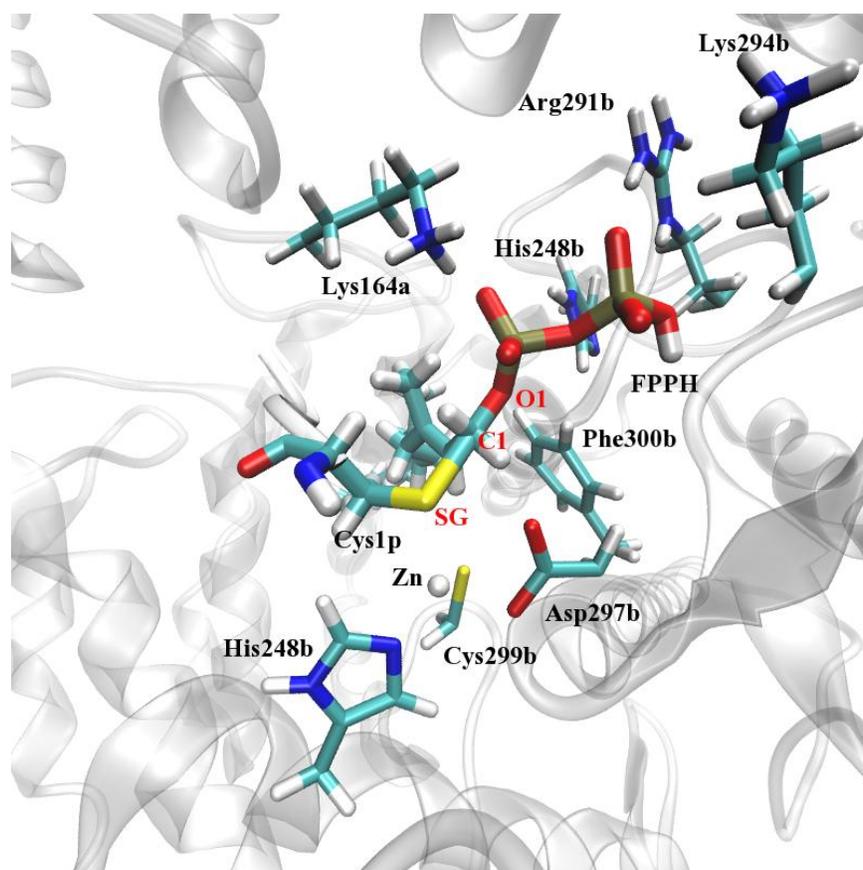


Figure 4-5. Snapshot of the active site of Y300F mutant at the transition state. O₁ and C₁ of FPPH, and S _{γ} of Cys1p are labeled in red. Note the position and orientation of mutant Phe300 β makes it possibly able to interact with partial positively charged farnesyl moiety via cation- π interaction.

It seems that the loss of hydrogen bond interaction between the α -diphosphate and β 300 residue cause the increase of free energy barrier. The 1.9 kcal/mol difference between the WT and mutant is divided into two parts, approximate 30% (0.5 kcal/mol) in the physical step and

70% (1.4 kcal/mol) in the chemical step. With experimental measurements of other enzyme residue mutation available, it will be interesting to find out what is the impact associated with those residue mutations.

4.3.2.2 Interesting context-dependent effect

The most interesting mutational effect of FTase complex might be the so-called context dependent selectivity of the a_2 and X in the Ca_1a_2X motif³¹. Such an effect does not only impact the free energy barrier height but also alters the reaction pathway¹⁸⁴. The most solid evidence came from the α -SKIE experiments of FTase/TKCVIF and FTase/GCVLS complexes.

The most straightforward effort based on our previous study would be to simply substitute CVIM with CVLM. CVLM model is prepared and studied with the same strategy as applied to Y β 300F model. The result free energy profile features a TS at 2.68 Å and a barrier height of 22.0 kcal/mol. The experimental value with regard to this motif in the absence of Mg^{2+} is yet available, while a peptide with this motif in the presence of Mg^{2+} gives a Gibbs free energy of 16.6 kcal/mol, 0.3 kcal/mol larger than that measured for the CVIM motif (16.3 kcal/mol) under identical experimental condition. Thus the difference of calculated free energy difference between CVLM and CVIM model, 1.4 kcal/mol (22.0 – 20.6), is again in excellent agreement with the experimental evidence and almost reaches the chemical accuracy requirement.

The TS active site configuration (see Figure 4-6) once again represents great similarity to the CVIM model. The biggest difference this time is the longer C_1-S_Y bond. The 2.68 Å bond length corresponds to a 0.05 Å increase, possibly resulting from the increased hydrophobic interaction between the side chain of a_2 residue and the farnesyl group of $FPPH^{2-}$. In addition, the mutation of Ile to Leu results in bulkier side chain terminal, thus possibly also increasing the steric effect of the residue to the surrounding enzyme AA side chains.

In spite of increased barrier height and $d_{C1-S\gamma}$ at the TS, the RLS remains unchanged. Hence, it seems a single change at a_2 position is not sufficient to result in mechanism change. In order to further illustrate this fascinating context-dependent effect, another mutant, CVLS model, is then prepared and studied.

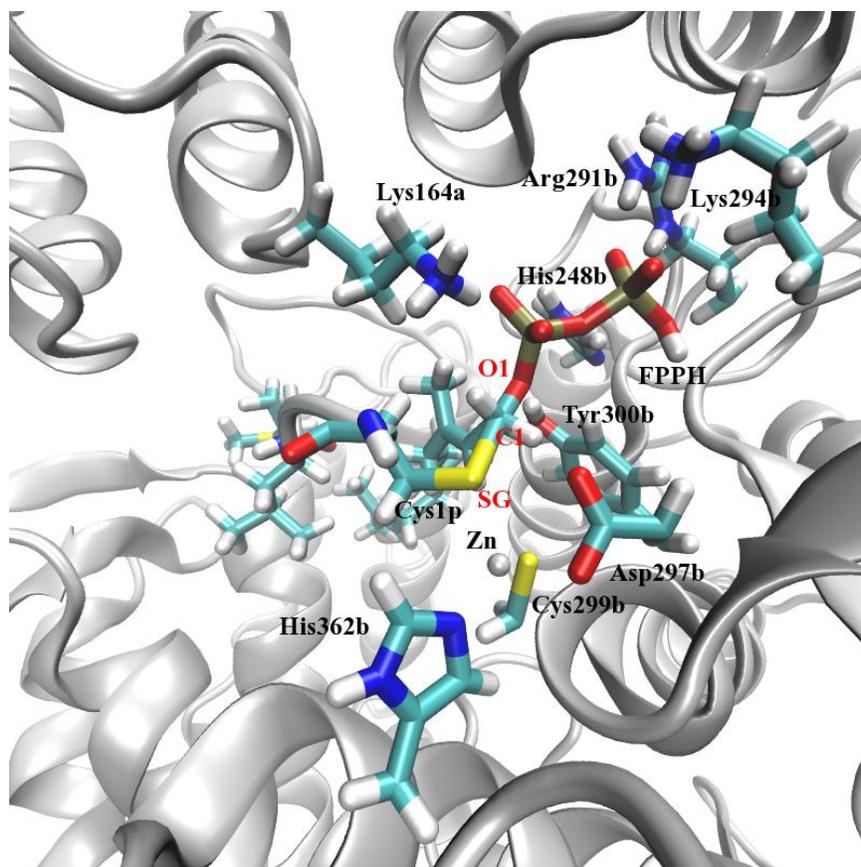


Figure 4-6. Snapshot of the active site of CVLM mutant at the transition state. O_1 and C_1 of FPPH, and S_γ of Cys1p are labeled in red.

4.3.2.3 Interpretation of α -SKIE results of FTase/CVLS complex

The double mutation at both a_2 and X position of the peptide substrate accompanies bigger change in peptide conformation, therefore simply modifying the existing CVIM model might not be as good as utilizing a crystal structure of FTase/GCVLS complex. The preparation and simulation detail has already been described in 4.2.3. The free energy curve of MM PMF study

clearly locates the resting state at approximate 7.65 Å (Figure 4-7), which is significantly longer than the WT of either CVIM or CVLM, and a transition intermediate state at around 5.0 Å. The free energy barrier associated with the transition is around 2.75 kcal/mol. Although this value is inarguably higher than either WT CVIM or WT CVLM, it is definitely unable to cause the RLS switch. This finding leads to the necessity of another interpretation of the nearly unity α -SKIE result observed on this system. An explanation of such observations might be the chemical reaction mechanism alters from an associative mechanism with dissociative characteristics to a more ‘independent’ S_N2 mechanism.

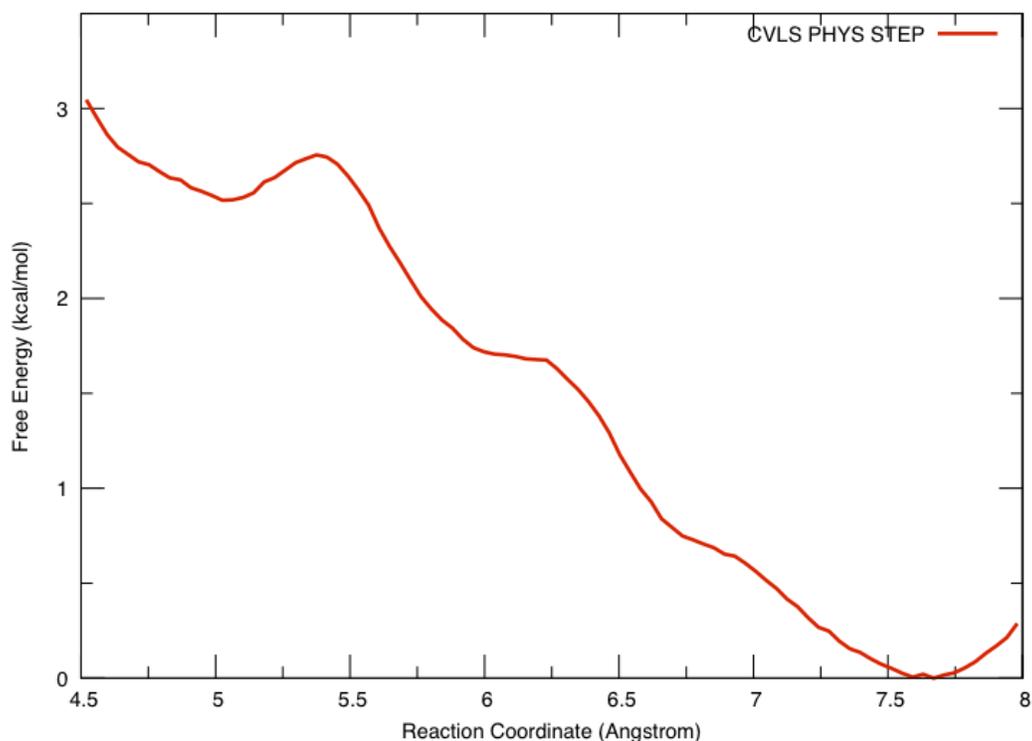


Figure 4-7. Free energy profile associated with the conformational transition step of CVLS model.

The QM/MM PMF study results provide more support to this conclusion. The free energy profile represents a 2.50 Å of TS C_1-S_Y distance and a 22.7 kcal/mol free energy barrier.

Comparing to CVIM model, markedly the $d_{C_1-S_Y}$ is 0.13 Å shorter and more close to the results

of S_N2 reaction computationally studied at MP2/6-31+G**//MP2/6-31+G* level¹⁹². Moreover, the S_Y of Cys1p, C_1 of FPPH²⁻ and O_1 of FPPH²⁻ are almost linearly aligned, despite C_1 is slightly out of the plane (see Figure 4-8). Unlike in CVIM model, where the $H_1-H_2-C_1-C_2$ dihedral remains fluctuating around 160° for a period of time corresponding to the RC change from 2.65 Å to 2.50 Å, the same dihedral in CVLS model reaches a peak at TS, which reads 178.6° (see Figure 4-9), and changes rapidly on both sides of the peak. All of this structural information strongly pushes toward the conclusion of a more independent S_N2 reaction mechanism is adopted by FTase/CVLS complex.

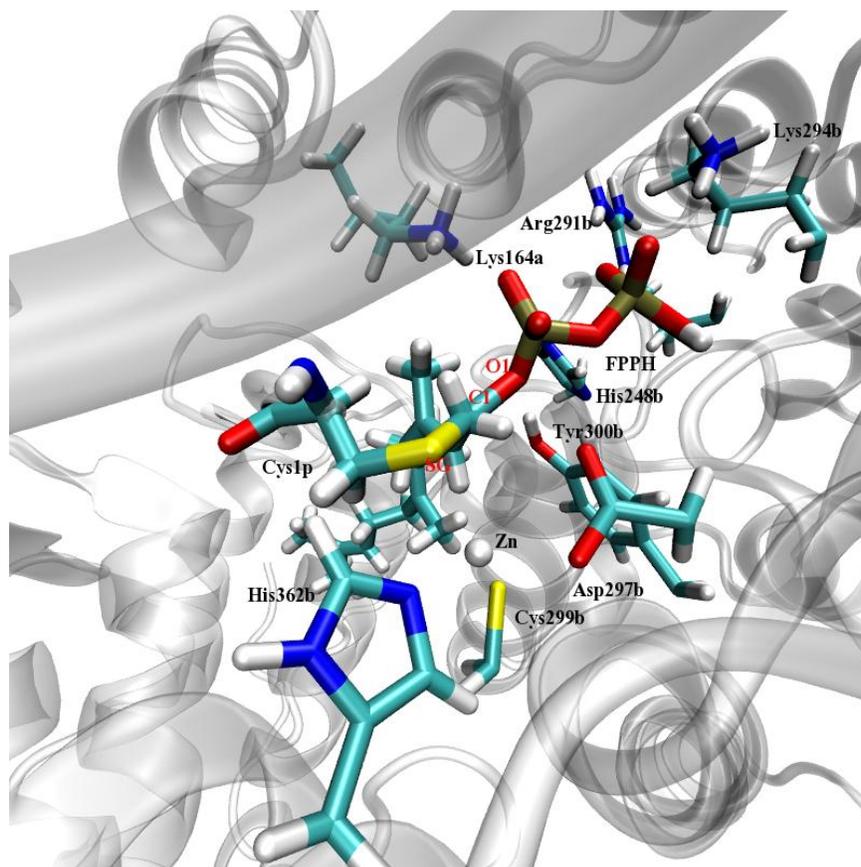


Figure 4-8. Active site snapshot of CVLS model at the transition state. O_1 and C_1 of FPPH, and S_γ of Cys1p are labeled in red.

On the other hand, the experimentally observed free energy difference for FTase complexed with CVLS (16.8 kcal/mol) and CVIM (16.3 kcal/mol) in the presence of Mg^{2+} is 0.5 kcal/mol, while the barrier for FTase/GCVLS without Mg^{2+} is about 20.0 kcal/mol, both of which are in excellent agreement with our results.

Before reaching the TS, there is another intermediate state with regard to the conformational transition is identified at approximate 3.7 Å. Like in CVIM model, the three atoms involved in bond formation and broken, namely O_1 of $FPPH^{2-}$, C_1 of $FPPH^{2-}$ and S_Y of Cys1p, appear to align linearly upon this point, suggesting this intermediate state can be referred to as the pre-organized state.

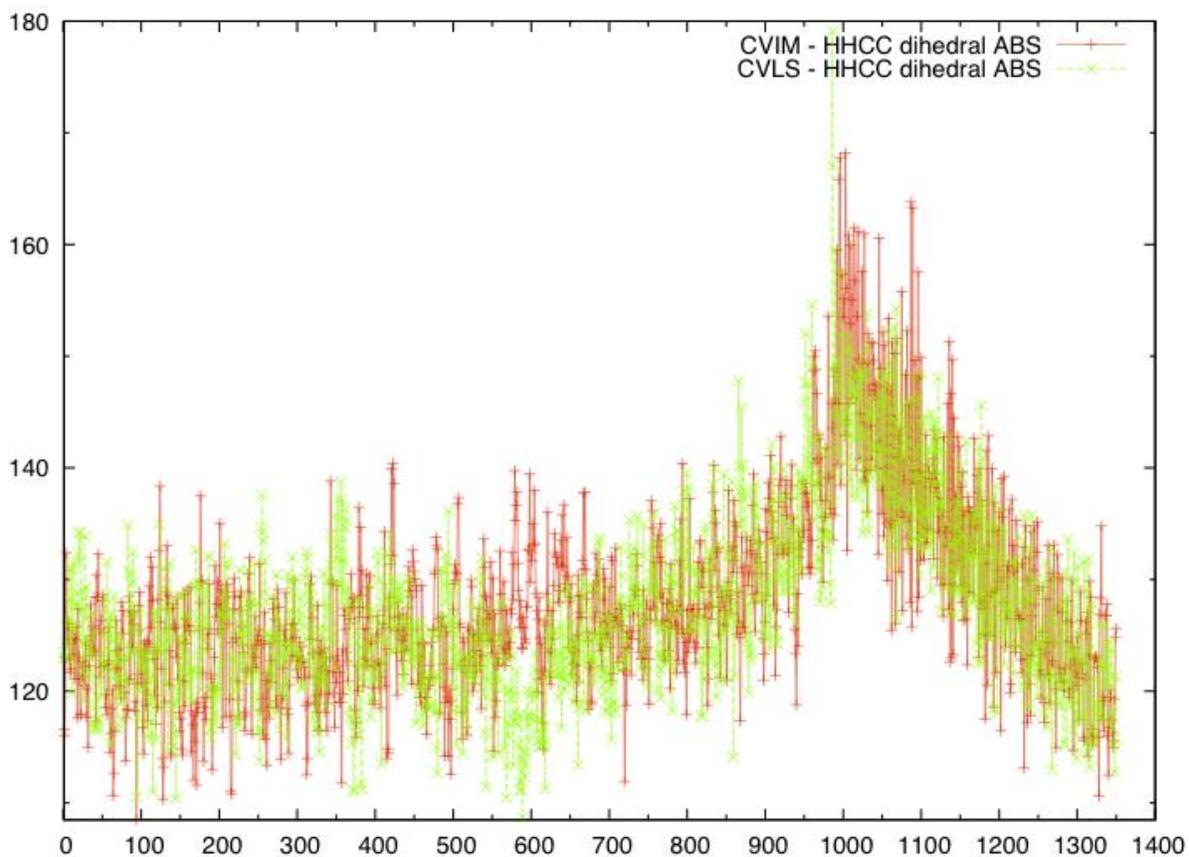


Figure 4-9. $H_1-H_2-C_1-C_2$ dihedral variation in the farnesylation of FTase/FPPH/CVIM (red) and FTase/FPPH/CVLS (green) complexes. X-axis of the figure represents number of frames, Y-axis of the figure represents the dihedral in degree.

Comparison between CVLS and CVIM reveals structural information of interactions between peptide substrate and enzyme active site residues. In CVIM, the sulfur of methionine side chain forms stable hydrogen bond interaction with side chains of S99 β and W102 β . In CVLS, the side chain of terminal serine only forms hydrogen bond with W102 β while the backbone of serine forms water mediate hydrogen bond interactions with A98 β and S99 β . These observations are supported by crystal structures of FTase/KKKSKTKCVIM complex (PDB 1D8D)¹⁹³ and FTase/CVLS complex (PDB 1TN8). Thus, despite less bulky in terminal residue size, the peptide of CVLS model shows the same level of stabilization as CVIM model.

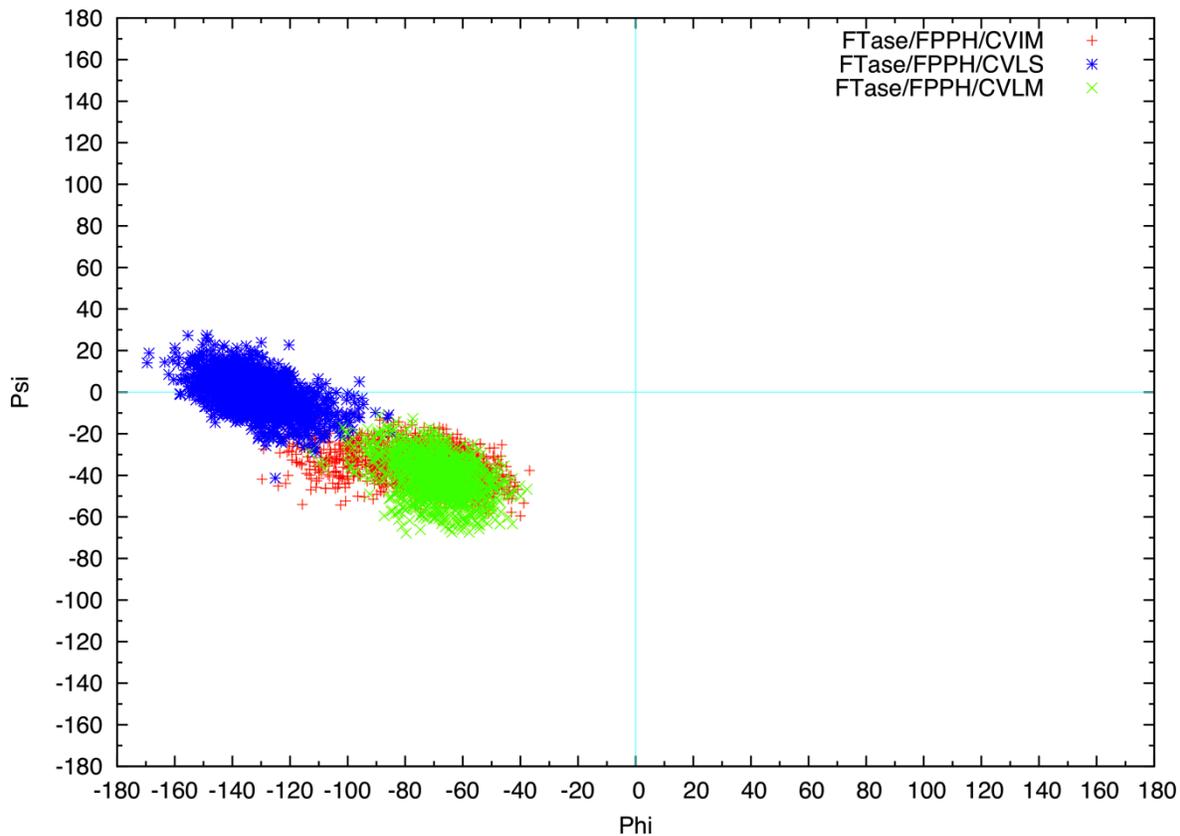


Figure 4-10. Ramachandran plot of a_2 residue of the Ca_1a_2X motif throughout the farnesylation of FTase/FPPH/CVIM (red), FTase/FPPH/CVLS (blue) and FTase/FPPH/CVLM (green) complexes.

The Ramachandran plots¹⁹⁴ of the peptide a₂ residue of CVIM, CVLM and CVLS model during the chemical step (see Figure 4-10) provide more interesting facts about the context-dependent a₂-X selectivity. In CVIM or CVLM model, either the isoleucine or the leucine stays in the α region throughout the entire process, while the leucine in the CVLS model appears fluctuating in the transition area connecting the α and β region. This implies that the peptide CVLS might sacrifice stability to help two reactants approaching. Furthermore, the closest contact between leucine side chain and farnesyl moiety in CVLM model is around 3.5 Å while nearly remains over 3.9 Å in CVLS. Such a difference might result in an extra flexibility for FPPH²⁻ to rotate its farnesyl group due to the less steric strain from peptide leucine side chain.

To further validate our assumption about unity α -SKIE measurement of FTase/GCVLS complex, calculations to address the ZPE difference at TS and resting state of both CVLS and CVIM model are undergoing, since ZPE difference dominates the free energy difference resulted from isotope substitution. Moreover, what role that the residue before Ca₁a₂X plays in the catalysis and what impact the length of the peptide does to the reaction pathway are of great interests to us and needs to be clarified in future studies.

4.3.3 Mg²⁺ in Chemical Step

In Chapter 3, a reasonable Mg²⁺ binding model of FTase/CVIM complex has been proposed. Although results from MM level simulations match experimental observations fairly well, it is necessary to carry out QM/MM level studies, in particular the PMF study, to further testify this binding scheme and illustrate how this metal ion alters the free energy barrier.

The preparation and simulation strategy is identical to that applied to CVIM model. The original simulation fails because of the difficulty of convergence possibly resulting from the complex electrostatic interactions associated with the close distance between two metal ions. However, freezing water molecules at least 10 Å away from the active site ligands seems

improving the electrostatic interaction calculation and thus solving the convergence problems. This approximation is not unusual to see in many other successful QM/MM simulations. Thus, the restraint has been applied throughout all simulations with regard to this model.

The QM/MM PMF study of WT2 model gives a free energy profile featuring an approximate 18.3 kcal/mol at approximate 2.69 Å along the RC. It is interesting to notice the TS d_{C1-SY} increases about 0.06 Å from an identical complex without Mg^{2+} present. Consider the Yβ300F simulation reveals a -0.02 Å from the same complex with a single mutation at β300 residue, it seems the inclusion of Mg^{2+} coordination introduces extra degree of inflexibility as well as extra degree of stabilization to the diphosphate moiety.

The TS active site configuration looks somewhat different to that of the CVIM TS (Figure 4-11). The major difference is of course the Mg^{2+} binding throughout the reaction process. Mg^{2+} coordinates to two diphosphate oxygen atoms from each subunit, a carboxylate oxygen atom of Asp352β and three water molecules. This coordination appears strongly limiting the flexibility of diphosphate moiety. Tyr300β has been found interacting with the capping acetyl group in MM PMF study, however, in this study it appears to form a hydrogen bond with the peptide cysteine stabilizing the peptide substrate. His248β does not interact with diphosphate directly but helps stabilizing the -4 charged PPI leaving group via water mediated hydrogen bond interaction. Lys164α, Arg291β and Lys294β all function as hydrogen bond donors to the diphosphate. The Mg^{2+} and Zn^{2+} are more than 7 Å apart throughout the process, however, the two metal sites are correlated via a hydrogen bond interaction between the side chain of zinc bound Asp297β and the backbone of magnesium bound Asp352β.

The computed free energy barrier is 18.4 kcal/mol. It is approximately 2.2 kcal/mol lower than the barrier height of CVIM model but about 2.1 kcal/mol higher than the experimental

measurement. This discrepancy is significantly bigger than any other model studied in this series. A rational explanation to this increased discrepancy might be the implementation of freezing water 10 Å away from the substrates. However, unless an improved set of SCC-DFTB Mg^{2+} parameters including the most important Zn-Mg interaction parameters is introduced, to simulate this system with SCC-DFTB/AMBER level QM/MM theory would remain challenging thus the freeing outside water approach would still be necessary. Other approaches, such as *ab initio* level of QM/MM simulation, should also be able to contribute to the mechanistic study of FTase chemistry in the presence of Mg^{2+} , nevertheless, the computational cost almost prohibits the implementation of such methods coupled with USP approach.

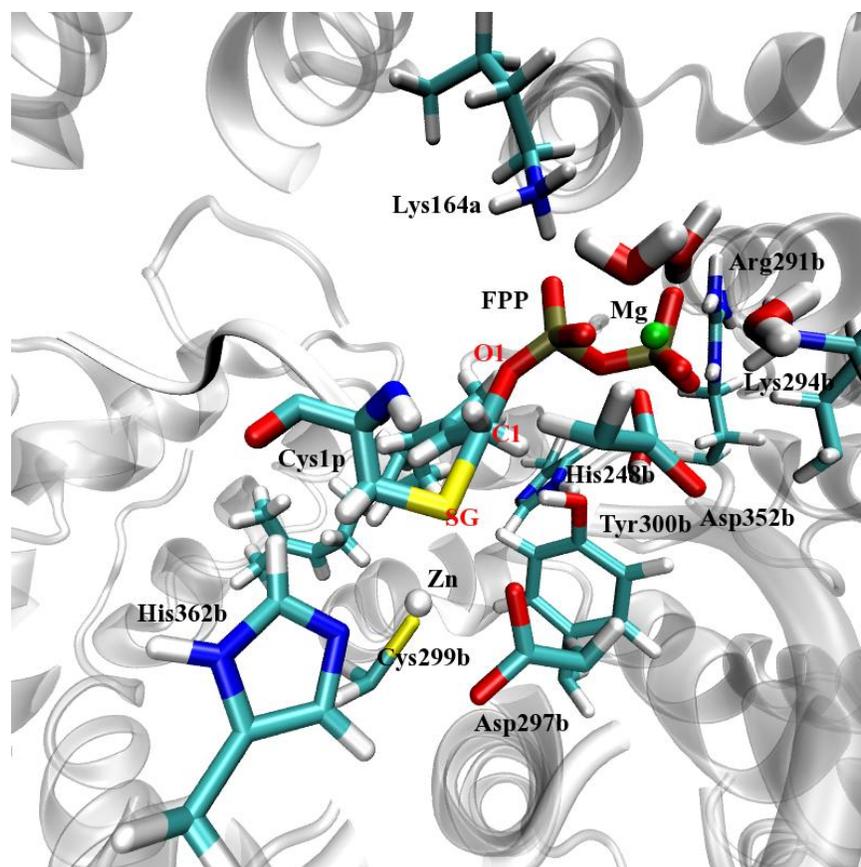


Figure 4-11. The active site snapshot of WT2 at the transition state. C_1 and O_1 of FPP, and S_γ of Cys1p are labeled in red.

4.4 Summary

In this series of studies, the chemical step of FTase catalyzed farnesylation has been systematically studied. The resulting TS active configuration of CVIM model with Mg^{2+} absent is in perfect agreement with experimental evidence, both structurally and energetically. In particular, the 20.6 kcal/mol free energy barrier is impressively falling into the range of 20.0 kcal/mol for FTase/GCVLS complex and 21.1 kcal/mol for FTase/TKCVIF complex with zero concentration of MgCl_2 . On the other hand, the 2.63 Å TS $d_{\text{C}_1\text{-S}_\gamma}$ is significantly longer than the computed TS $d_{\text{C-S}}$ of a $\text{S}_\text{N}2$ reaction in gas phase at MP2/6-31+G**//MP2/6-31+G* level, indicating this reaction adopts an associative mechanism with dissociative characteristics. In addition, the QM/MM PMF study of the conformational transition step finds its results and the results from MM PMF study match excellently.

The impact of key residue mutation is then studied. The PMF study of Y β 300F, CVLM and CVLS model in the absence of Mg^{2+} all give excellent free energy predictions. Additionally, the difference in TS distance of $\text{C}_1\text{-S}_\gamma$ varies among these models. Although the exact reason remains puzzle, the trend seems related to the level of flexibility possessed by FPPH²⁻ and/or the peptide substrate. The recognition of peptide residue as a_2 and X position in the Ca_1a_2X motif exhibits fascinating context-dependent selectivity. Amongst, the FTase/CVLS model is particularly important since it seems making impact not only on the free energy barrier height, but also on disturbing the reaction mechanism. The α -SKIE experiments reveal a nearly unity measurement for FTase/GCVLS complex while a value of 1.154 for FTase/TKCVIF complex. The latter obviously falls into the range of an associative mechanism with dissociative characteristics, while the former has been explained as the RLS switches from the chemical step to the physical step. However, our results lead to an alternative interpretation of the nearly unity ³H α -SKIE result of CVLS model that the chemical step remains the RLS but the mechanism

changes to a more ‘independent’ or ‘pure’ associative (or S_N2) mechanism. This assumption is being further tested with ZPE and α -SKIE calculations at MP2/6-31+G** level. The Ramachandran plot of a_2 residue shows great difference between CVLS and CVIM/CVLM models. The leucine in the CVLS model flexibly fluctuates in the transition area connecting α and β region, while either the isoleucine in CVIM or the leucine in CVLM stably remains in the same α region.

Finally, the WT2 model obtained from studies described in Chapter 3 is validated through QM/MM PMF simulations. The result free energy barrier features a free energy barrier of 18.4 kcal/mol corresponding a TS d_{C1-SY} of 2.69 Å, which is the longest through this study. Tyr300 β and His248 β are discovered to play different roles in the presence of Mg^{2+} than in the absence of such a metal ion. With an approximation of freezing water molecules 10 Å outside any substrates applied, the resulted free energy barrier shows the biggest discrepancy among all the models discussed in this chapter. Therefore, an improved simulation of this model is still required.

CHAPTER 5
COMPUTATIONAL APPROACH TO UNDERSTAND AROMATIC
PRENYLTRANSFERASE CATALYTIC MECHANISMS

5.1 Background

Prenyltransferases include not only three members of protein prenyltransferases that are famous for their important roles in G-proteins regulated signal transduction pathway but also aromatic prenyltransferase such as bacterial prenyltransferase NphB³ (or Orf2 before rename) purified from *Streptomyces* sp. Strain CL190 and fungal indole prenyltransferase FtmPT1⁴ extracted from *Aspergillus fumigatus*.

This kind of aromatic prenyltransferase makes great contribution to the biosynthesis of terpenoid in addition to a variety of terpenoid synthase¹⁹⁵. The products of such a prenylation, the terpenoids, are very widely distributed in a large variety of natural products, such as in fungi, animals and plants. Amongst these terpenoids products, plant isoprenoids have been well known for their application into traditional herbal remedies because of their aromatic qualities⁶. Some isoprenoids and their derivatives have been revealed to possess certain pharmaceutically desired characteristics such as antiviral, antioxidant as well as anticancer effect¹⁹⁶⁻²⁰⁵. In addition, the low cellular toxicity and excellent ability to penetrate cell membrane associated with this kind of compounds have made them desired drug templates. The biosynthesis of these terpenoids has been of great interest and importance to biochemistry and medical chemistry researches¹⁹⁵.

NphB, identified recently from *Streptomyces*, possesses a rather interesting 3-D structure that features a novel α/β barrel fold termed PT-barrel. Inside the barrel a spacious binding pocket is provided to two reacting substrates, the isoprenoid substrate GPP and the aromatic substrate 1,6-DHN. GPP is stabilized via interaction between negatively charged diphosphate and several AA side chains, including Lys119, Lys169, Arg228, Tyr216 and Lys284, in addition to a Mg²⁺.

The magnesium ion is found coordinating the side chain of Asp62, a α -diphosphate oxygen atom and four water molecules, in spite of the missing of a DDXDX motif usually found in Mg^{2+} metalloenzymes. Asp110 situates within 3.5 Å to a Mg^{2+} bound water molecule, provide extra stabilization to the metal site.

NphB exhibits interesting substrate selectivity and product regioselectivity^{3,153,206}. In the catalysis of geranylation of 1,6-DHN, at least 3 products have been identified. The major product, a trans-5-geranyl-1,6-DHN, and the minor product, a 2-geranyl-1,6-DHN, are originally characterized, while recently a extra-minor product has been purified and identified as a 4-geranyl-1,6-DHN. Previous computational study has revealed that the different orientation of 1,6-DHN causes the product diversity. The major product is defined as S1 state, the minor product is defined as S3 state, while an intermediate state referred to as S2 state connects them in the FES. Unfortunately, no binding orientation has been found responsible for the extra-minor state. We will call this extra-minor state the S4 state in this study.

Indole prenyltransferase FtmPT1 has been only recently identified from *Aspergillus fumigatus*. This family of indole prenyltransferases catalyzes the attachment of isoprenoid moiety to tryptophan and its derivatives. The product of FtmPT1 complexed with Brevianamide F, the Tryprostatin B, shows high cytotoxicity to a number of cancer cells, making this enzyme a promising anticancer target^{56-58,207-210}.

The 3-D structure of FtmPT1 also exhibits a α/β barrel fold composed of 10 anti-parallel β -strands. Outside the barrel situated 10 solvent accessible α -helices. However, this barrel is slightly different to the one observed in NphB at one of the $\alpha\beta$ repeat. FtmPT1 prefers DMAPP as isoprenoids substrate. The active site consists of a highly positive charged binding pocket for DMAPP and a tyrosine-rich environment. The diphosphate is anchored by the salt bridge

interactions between itself and two positively charged residues, namely Arg113 and Lys294. The side chains of Tyr203 and Tyr450 help stabilize the diphosphate via hydrogen bond interactions. These two tyrosine residues, along with Tyr296 and Tyr382, circle a ring-shaped ‘shield’. The DMAPP binding site and brevianamide F binding site appear to be separated by such a ‘shield’. Hence, this ‘shield’ has been proposed to be essential for the catalysis since it seems able to provide stabilization to the dimethylallylic cation via cation- π interactions. Additionally, the reaction between DMAPP and brevianamide F very likely includes a rotation of the dimethylallylic cation due to the large separate between two reacting atoms. Therefore, the tyrosine ‘shield’ might also serve as a steer to engineer such a rotation.

FtmPT1 also exhibits interesting regioselectivity. However, it is activated by a single mutation at Gly115. Mutating it to Thr115 results in the dimethylallylic cation connecting to the C₃ (C_{3A}) instead of C₂ (C_{2A}) position of the brevianamide F via its own C₃ (C_{3P}) atom instead of C₁ (C_{1P}) atom. However, G115A mutant does not show the same regioselectivity.

The reaction mechanism of NphB has been proposed to be an associative pathway featured an S_N2-like TS or a dissociative mechanism featured a carbocation invoked electrophilic capture. Although the dissociative mechanism is more preferred, the possibility of an associative mechanism cannot be totally excluded without further study. On the other hand, the catalysis of FtmPT1 has been assumed to adopt a dissociative pathway. Three steps mechanism has been proposed: in the first step the hydrolysis takes place to form a PPi leaving group and a carbocation, next in the second step the carbocation rotates then attaches either C_{1P} or C_{3P} (in G115T mutant) to the indole nucleus forming a new C-C bond, consequently the proton transfer occurs in the third step to restore aromatic quality of the indole substrate. KIE experiment has determined that both the hydrolysis and bond forming steps are partially rate limiting. Consider

the rate constant of product formation (k_{cat}) is 5.57 s^{-1} under single turnover condition, the KIE result is reasonable since the time scale for such a rotation is usually not in this order of magnitude.

5.2 Methods and Simulation Details

5.2.1 QM/MM Study of NphB Prenylation

The initial structure of NphB is obtained from the QM/MM X-ray refinement conducted in the previous study¹⁵³. The S1 state structure is selected and then solvated in a truncated octahedron box filled with TIP3P water models. The edge of the box is at least 8 Å away from the nearest solute atom. The enzyme is modeled using AMBER force field ff99SB, while the substrates are modeled using GAFF. The atomic charges of two substrates from previous study are adopted here. PME is applied to describe the long-range electrostatic interactions. An 8 Å cutoff of the real-space nonbond interaction and an approximately 1 Å for the reciprocal space grid spacing are used. All hydrogen-involved bonds are constrained with SHAKE algorithm.

The resulting system is first fully minimized with a weak, harmonic restraint applied to remove any possible close contacts. The temperature of the system is slowly heated to 300 K using Langevin dynamics during a 100 ps time period. During this heating process, the positional restraints are gradually removed. Consequently, the system is simulated in the NPT ensemble for 900 ps with a 1 fs time step to further equilibrate the density. After that, a 4 ns NPT simulation with a 2 fs time step is performed to obtain a descent starting point for QM/MM simulations.

A snapshot of the final 1 ns simulation is taken to serve as the initial structure for QM/MM simulation. The system is then partitioned into QM region and MM region. Only two substrates are included in the QM region since preliminary results showed this selection produces the most stable simulation while does not sacrifice accuracy (Figure 5-1). The system is quickly minimized with the SCC-DFTB/ff99SB hybrid potential followed by re-heating to 300 K in

NVT ensemble during a 50 ps period with a 0.5 fs time step. The pressure regulation is then turned on to have the system equilibrated in the NPT ensemble for 450 ps with a 1 fs time step.

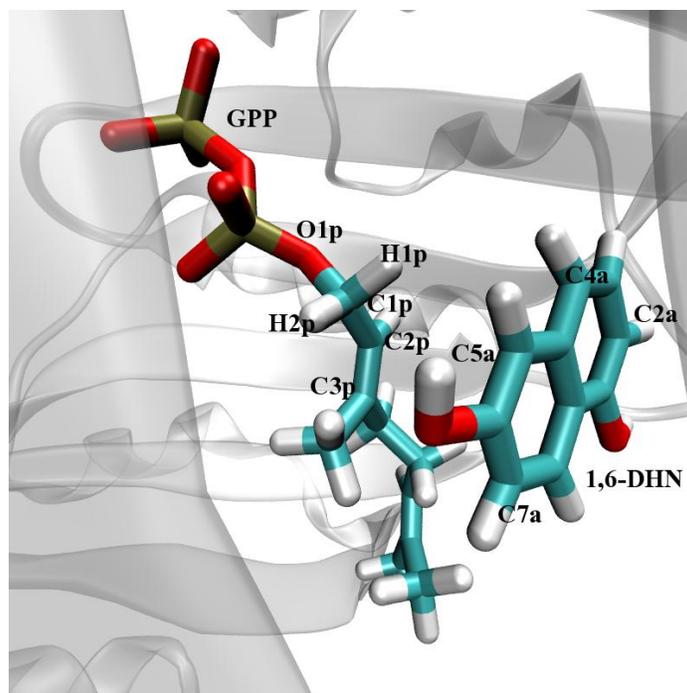


Figure 5-1. QM partition of NphB: GPP and 1,6-DHN. Important atom specifications are given. O1p in figure denotes O_{1P} in text (which means O_1 from diphosphate substrate), and C5a denotes C_{5A} in text (which means C_5 in aromatic substrate).

The resulting system is then used for a SMD scan to access a quick view of the FES shape. The RC is defined as the distance between C_1 (C_{1P}) of GPP and C_5 (C_{5A}) of 1,6-DHN. The linear combination of RC, which defined as $d_{RC} = d_{C_{1P}-O_{1P}} - d_{C_{1P}-C_{5A}}$, appears bringing instability into the PMF study, possibly due to the existence of multiple reacting-capable atoms in 1,6-DHN. While a 2-D PMF study, where the RC's defined as $d_{RC}^1 = d_{C_{1P}-O_{1P}}$ and $d_{RC}^2 = d_{C_{1P}-C_{5A}}$, increases the computational cost significantly but does not provide additional accuracy. A set of 26 starting structures are extracted from the trajectory of SMD covering the RC from 1.5 Å (product state) to 4.0 Å (resting state), with a 0.1 Å interval. Another 14 windows are prepared to cover the middle points of two originally generated windows for the reaction important area, namely 1.6 Å

to 3.0 Å. A set of resulting 40 windows of USP is then carried out with each window simulated for 250 ps, thus a total of 10 ns QM/MM simulation is accomplished for this study. The last 150 ps simulation of each window is used for data collecting. The results are then processed with Dr. Grossfield's WHAM code to reconstruct the free energy profile after unbias the distribution.

The S3 state is reached by applying the following NMR restraints ($C_{1P}-C_{5A}$: 5.9 Å, 10 kcal/mol•Å; $C_{1P}-C_{2A}$: 4.0 Å, 10 kcal/mol•Å) to an extended QM/MM simulation after the already-mentioned 4 ns simulation. A SMD scan is then carried out with the new generated S3 state structure with the new RC defined as $d_{C_{1P}-C_{2A}}$. The same strategy as described for the S1 state is adopted into the QM/MM PMF study of S3 state coupled with USP method. WHAM code is utilized again to generate the free energy profile.

The S4 state is also studied. The initial structure is obtained from a snapshot of QM/MM equilibration of S3 state. The reason of such a selection will be explained later. A SMD scan is carried out with the resulting structure to propagate the trajectory over the RC, where RC is defined as $d_{RC} = d_{C_{1P}-C_{3A}}$. A following QM/MM PMF study coupled with USP technique, however, has not yet been performed.

5.2.2 Modelling and Simulation of Proton Transfer Step in NphB Catalysis

After the prenylation, a proton transfer step is required to remove the redundant hydrogen from the prenylated 1,6-DHN, thus restoring its aromatic quality. The proton transfer is usually much more complicated than other reactions due to the highly possible quantum tunneling effect resulting from the small atomic mass of hydrogen atom. In this study, we still adopt the SMD-US strategy to study this reaction with QM/MM simulation. However, *ab initio* level calculation will be conducted later to study the tunneling effect.

The initial structure (see Figure 5-2) is a snapshot taken from the S1 QM/MM USP at $d_{RC} = 1.5$ Å that corresponds to the product state. After carefully examine the resulted structure, two

candidates to perform such a proton abstraction have isolated. One of them is the Tyr216 and another one is a crystal water molecule.

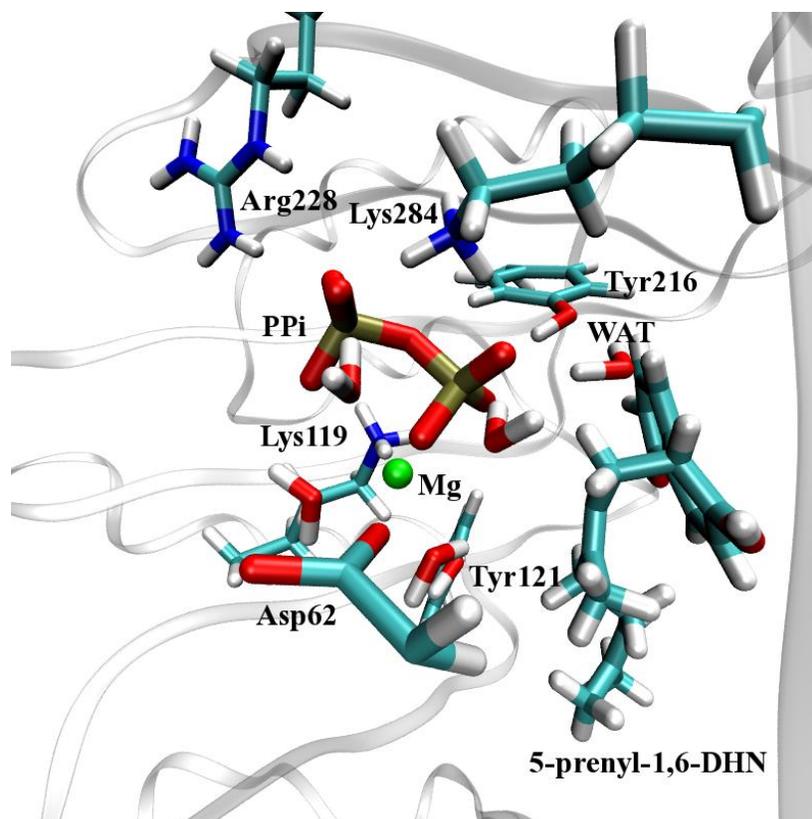


Figure 5-2. Active site configuration of the initial structure selected for proton transfer step simulation. WAT refers to the water molecule that is included in QM region at this stage.

Tyr216 stabilize the PPi by donating a hydrogen bond to the α -diphosphate, and it situates within 4.5 Å away from C_{5A} of 1,6-DHN. A reasonable tyrosine mediate proton transfer process would be the diphosphate first abstracts the hydroxyl oxygen and then consequently the tyrosine compensates itself by capturing the extra proton from geranylated 1,6-DHN. However, our preliminary results show the first step associates with an unreasonably large free energy barrier (45+ kcal/mol from SMD scan). Therefore, this hypothesis has not been further studied.

On the other hand, a crystal water molecule has been observed approaching 1,6-DHN when the new C-C bond starts to form. This water molecule interacts with the hydroxyl group at C_{6A}, thus staying close to the redundant proton at C_{5A}. An equilibration of 250 ps is first performed to validate that this water molecule remains its position during longer simulation. Within this equilibration, this water molecule is added into QM region. Then a snapshot is extracted from the trajectory resulted from the last 50 ps of such equilibration. The resulting structure is scanned with a SMD simulation with RC defined as $d_{RC} = d_{OW-H5A}$. A following QM/MM USP study has not been performed yet.

5.2.3 Classical and QM/MM study of FtmPT1

The FtmPT1 model is prepared based on the structure of PDB 3O2K (Figure 5-3). The system is first gone through MolProbity (at <http://molprobity.biochem.duke.edu/>) to improve the structure. Several flip are taken place at a couple of glutamines and an asparagine residue. The resulting system is then prepared with LeAP in the AMBER suite of programs. AMBER force field ff99SB and GAFF are used to model the enzyme and substrates, respectively. The atomic charge of DMAPP and brevianamide F and derived following a two-stage RESP fitting procedure. After that, the system is solvated into a truncated octahedron water box with TIP3P water model applied. The edge of the box situates at least 8 Å away from the closest solute atom. Long-range electrostatic interactions are computed using PME method. SHAKE algorithm is implemented to apply constraints to all bonds with hydrogen atoms. This model is termed as DPP in the following text.

Another system is prepared in the same way for FtmPT1 complexed with monoprotonated DMAPP²⁻, in order to testify the substrate preference of this enzyme. This new model will be called DPH in this study.

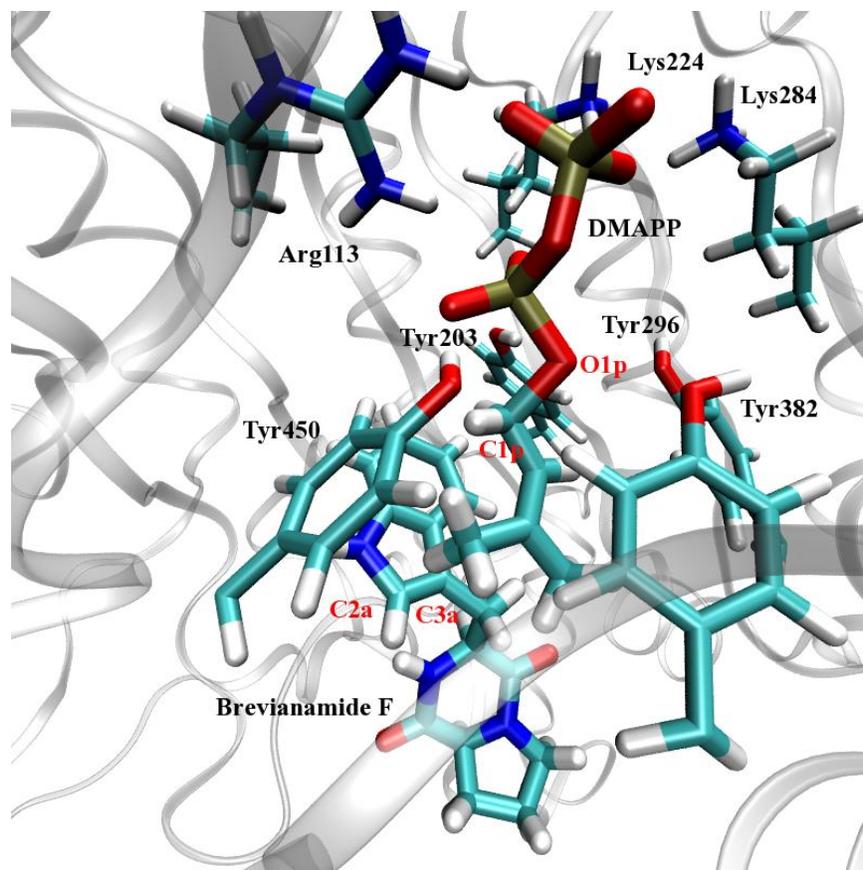


Figure 5-3. Active site configuration of PDB ID 3O2K. DMAPP and Brevianamide F are treated with QM in QM/MM simulation. Important atoms are labeled in red.

Two resulting systems are first fully minimized. A weak, harmonic force is applied to restrain the coordinates of protein and ligand atoms. Then the systems are heated to 300 K during a 100 ps period with a 1 fs time step in canonical ensemble. The positional restraints are gradually removed during the heating processes. The resulting systems are equilibrated in the NPT ensemble for 900 ps with a 1 fs time step, followed by a 4 ns simulation for each system.

A snapshot of the final 1 ns simulation of DPP is extracted for each system and served as the initial structure for the following QM/MM study. The QM/MM study of DPH has not yet been conducted. Similar to NphB, only two reactants are selected in the QM partition of the system, while all remaining system is modeled with MM potential. The resulting two systems are then minimized with hybrid QM/MM potential and re-heating to 300 K during a 50 ps time

period in NVT ensemble with a 0.5 fs time step. Then both systems are equilibrated with QM/MM potential for 450 ps with a 1 fs time step. A snapshot from the last 50 ps simulation for each system is chosen as the initial structure for the PMF study.

Due to the large separation between C_{1P} and C_{2A} , a 2-D PMF study would be very computationally consuming. Thus a two-stage 1-D PMF study is applied to each system. In the first-stage, the RC is defined as $d_{C_{1P}-O_{1P}}$ while in the second state $d_{C_{1P}-C_{2A}}$ denotes the RC. A first-stage SMD scan for each model is conducted to simulate the hydrolysis of DMAPP. The resulting structures are then undergoing another SMD scan at the second-stage to study the bond forming process. A two-stage study of USP method coupled PMF is also required. Such a study is currently undergoing in our lab.

After obtaining knowledge of the RC variation from two-stage 1-D PMF studies, it is possible to carry out 2-D PMF studies with a ‘reduced’ number of windows that only cover those important areas while leaving physically unfavorable areas out of study. In such a study, the two RC’s are defined as $d_{RC}^1 = d_{C_{1P}-C_{2A}}$ and $d_{RC}^2 = d_{C_{1P}-O_{1P}}$. This study would provide additional support to the conclusion drawn from 1-D PMF study.

5.3 Results and Discussions

5.3.1 Orientation of Substrate Orientation in QM/MM Equilibration of NphB

A QM/MM X-ray refined structure of NphB S1 state has been adopted as the initial structure and solvated into a water box in truncated octahedron shape. The resulting system is first fully relaxed through MM equilibration. During this procedure, $d_{C_{1P}-C_{2A}}$ and $d_{C_{1P}-C_{5A}}$ fluctuates as the system is transforming between the major product of S1 and the minor product state of S3. A snapshot from the last 1 ns MM simulation that best represents the S1 state is extracted to serve as the initial structure of the consequent QM/MM simulations.

During QM/MM simulation, GPP and its binding pocket are fairly stable. The salt bridge interactions between negatively charged side chains of Lys119, Arg228 and Lys284 help stabilizing the isoprenoids substrate. In addition, Tyr216 functions as a hydrogen bond donor to the α -diphosphate, while Lys169 reaches into hydrogen bond range with the β -diphosphate.

Tyr121 situates on the opposite side of geranyl group to the aromatic substrate thus it looks like the lipid group being sandwiched by the two aromatic units. Such a pose would be essential to the catalysis if the reaction adopts a dissociative pathway, since two aromatic groups could stabilize the developing carbocation through cation- π interactions²¹¹⁻²¹⁸. Carefully examining the relative positions of Tyr121 and 1,6-DHN, it is found that the aromatic ring of the tyrosine residue points to the C_{3P} position of the geranyl moiety while the aromatic group of 1,6-DHN (in the S1 state orientation) is in a perfect position to interact with C_{1P}. Considering the fact that the positive charge will be delocalized between C_{1P} and C_{3P}, this staggered arrangement of two aromatic units is particular interesting and important.

The orientation of 1,6-DHN fluctuates between S1 and S3 state through the QM/MM equilibration (see Figure 5-4). The side chains of Ser214 and Tyr288 play important role to help stabilize the S3 state orientation via hydrogen bond interactions. A configuration of 1,6-DHN that leads to the C_{4A} closer to C_{1P} of GPP³⁻ than either C_{5A} or C_{2A} has been observed during the orientation transition, however, is fairly unstable as it nearly instantly transforms to other orientations. That finding leads to an assumption that the 4-prenyl-1,6-DHN is an associated product of either S1 state or S3 state orientation. This assumption makes sense if one considers the fact that at the equilibrium conformation of S3 state C_{4A} is quite accessible to the C_{1P}, since the distance between C_{1P} and C_{4A} is fluctuating between approximately 1.0 Å shorter to 1.2 Å longer than that $d_{C1P-C2A}$.

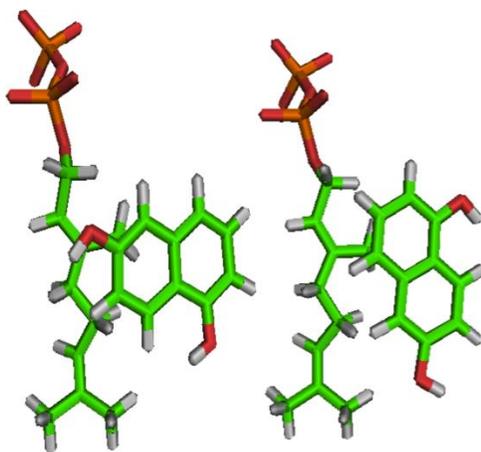


Figure 5-4. 1,6 DHN orientation in S1 (left) and S3 (right) State.

5.3.2 Free Energy Surface of NphB Catalyzed Prenylation

5.3.2.1 S1 state

QM/MM PMF studies are carried out to study the FES associated with different product formation processes of NphB catalysis. A snapshot from the final 50 ps QM/MM equilibration of S1 state that best represents the resting state of such an orientation ($d_{C1P-C5A} \approx 4.0 \text{ \AA}$; and $d_{C1P-C2A} \approx 7.0 \text{ \AA}$) is prepared as the initial structure for the future PMF study.

A SMD scan is performed with the resulting structure. 40 structures are consequently extracted from the SMD trajectory to serve as starting structures for the following US. After unbiasing the distribution of each windows using WHAM code, a free energy profile is generated. An intermediate is identified on such a FES at approximate 2.10 \AA along the RC. Two free energy barriers, at 2.55 \AA and 2.00 \AA respectively, separate the intermediate state to either the reactant state or the product state. The barrier at 2.55 \AA (TS1) is the highest barrier, corresponding to 12.35 kcal/mol free energy (see Figure 5-5), while the second barrier at 2.00 \AA (TS2) is fairly small (about 5.55 kcal/mol) comparing to the intermediate state (approximately 5.10 kcal/mol).

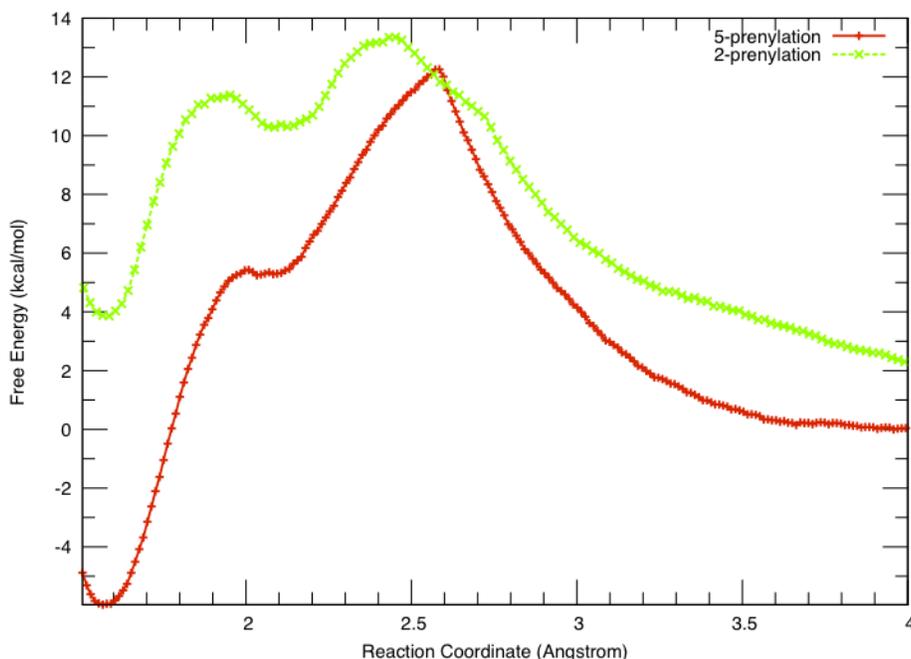


Figure 5-5. Free energy profile associated with prenylation step for both S1 and S3 model. The profile of S3 does not start from 0 kcal/mol, instead, it starts at 2.3 kcal/mol, which is the free energy difference between S1 and S3 orientation.

The intermediate features a carbocation (Figure 5-5). The $H_{1P}-H_{2P}-C_{1P}-C_{2P}$ (see Figure 5-1 for atom specifications) dihedral reads nearly 180° upon the forming of such a carbocation, which strongly indicating the generation an sp^2 carbon at C_{1P} . Moreover, this dihedral stays above 160° during the RC moving from approximately 2.6 \AA to 2.3 \AA and fluctuates above 150 \AA until the intermediate state is passed. Comparing to the fluctuation of such a dihedral in the farnesylation of FTase/FPPH/CVIM complex (see Figure 5-7), the characteristic of a carbocation in this geranylation is more evident. Therefore, we propose this NphB catalyzed geranylation adopts a typical S_N1 reaction mechanism.

Tyr121 sandwiches the developing carbocation along with 1,6-DHN via cation- π interactions. The staggered arrangement of two aromatic groups guarantees that the positive charge is stabilized by at least one cation-aromatic π interaction despite the delocalization of the charge. In fact, after the $C_{1P}-O_{1P}$ bond breaks, the generated carbocation immediately ‘slips’ into

the pocket ‘sandwiched’ by these two aromatic units. On the other hand, Tyr216, which forms stable hydrogen bond interaction to an oxygen atom from α -diphosphate, is within 4.0 to 4.8 Å away from the C_{1P} atom after the carbocation forms. In spite of the distance significantly longer than that between the carbocation and either Tyr121 or 1,6-DHN, it is still within the interaction range of such a cation-aromatic π interaction. Thus, an interesting T-shaped cation-aromatic π interaction pocket is formed and possibly steers the carbocation to electrophilic attack the C_{5A} of 1,6-DHN.

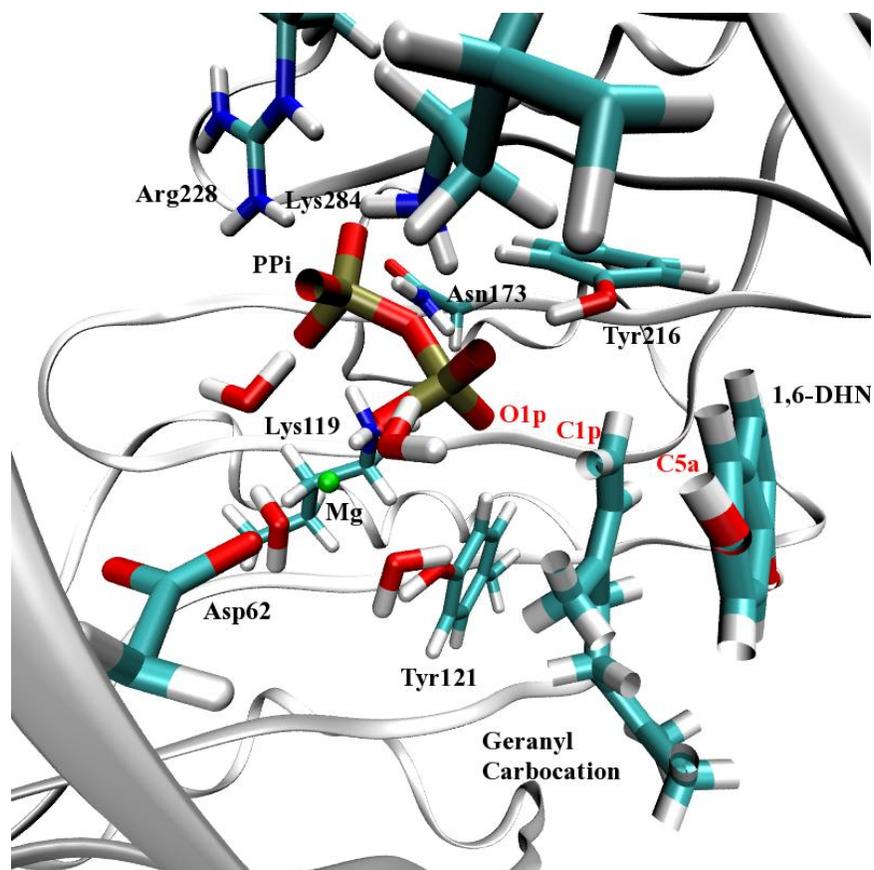


Figure 5-6. Active site configuration at the intermediate state of S1 model. O_{1P} of PPI, C_{1P} of carbocation and C_{5A} of 1,6-DHN are labeled in red. Note Tyr121 and 1,6-DHN are staggered ‘sandwiching’ the carbocation.

The diphosphate leaving group, with a big 4 units negative charge, is stabilized by Lys119, Asn173, Arg228, Lys284 and Tyr216 via hydrogen bond interactions. Lys169 might also

contribute to the stabilization via a water mediate hydrogen bond to the β -diphosphate moiety. Upon passing the TS2, the C_{1P} - C_{5A} bond starts forming. At the same time, a crystal water molecule reaches in to within reacting range to the 1,6-DHN. Such a water molecule might function to abstract the extra proton at C_{5A} .

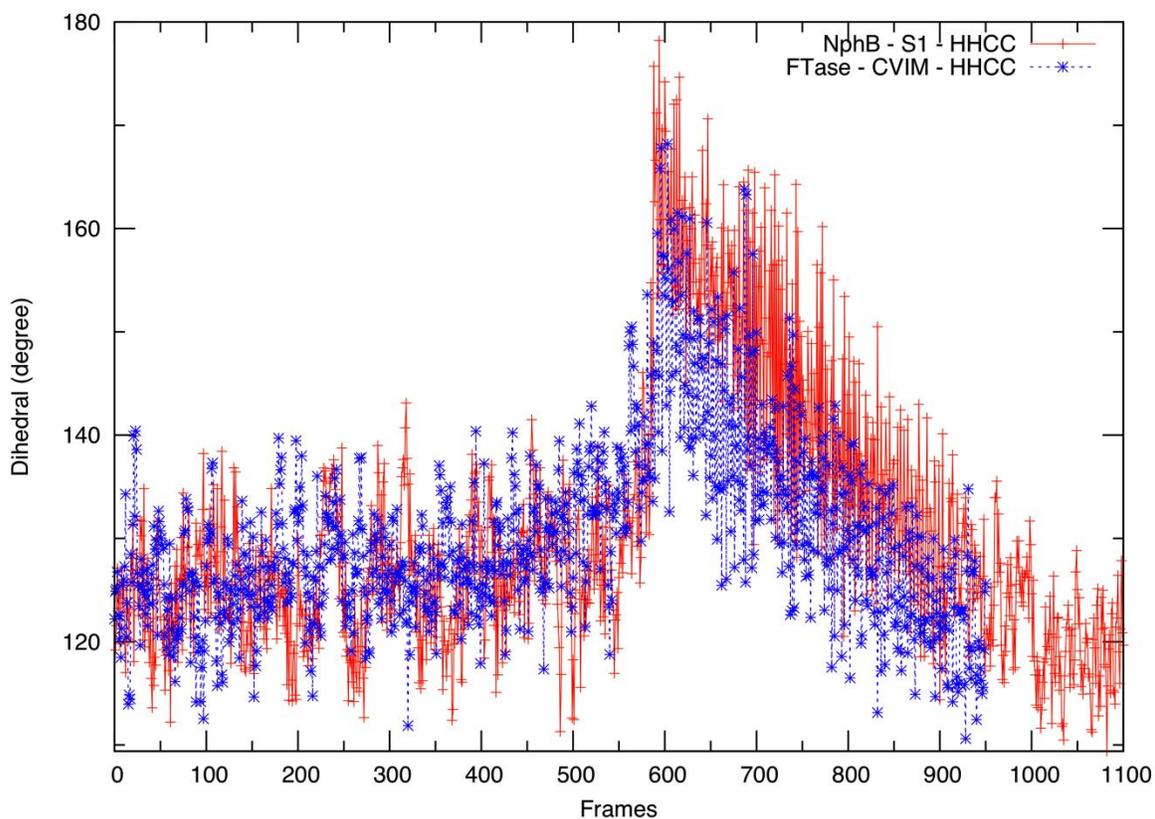


Figure 5-7. $H_{1(P)}-H_{2(P)}-C_{1(P)}-C_{2(P)}$ dihedral variation in the chemical reaction of NphB catalyzed 5-prenylation (red) and FTase/FPPH/CVIM catalyzed farnesylation (blue).

The computed free energy barrier is 12.35 kcal/mol, while the experimentally measured k_{cat} for this reaction is $4200 \pm 200 \text{ s}^{-1}$, corresponding to 12.6 kcal/mol in barrier height associated with the product formation. In addition to the excellent agreement between our calculation and experimental observation, this value suggests that the hydrolysis step is the RLS for this NphB catalyzed geranylation. Although the possibility of the proton abstraction step being the RLS cannot be totally excluded, it is very unlikely that the free energy barrier associated with such a

proton transfer will be larger than 12 kcal/mol, leave alone the highly possible accompanied quantum tunneling effect due to the small mass of hydrogen atom.

5.3.2.2 S3 state and S4 state

The initial structure of S3 state for the QM/MM simulation is obtained by artificially building such a configuration with NMR restraints applied. One of the snapshots during last 10 ps of such a restraint simulation is then extracted and prepared for the QM/MM PMF study.

Like S1 study, a SMD scan is first conducted to propagate the trajectory along the RC from 4.0 Å (reactant state) to 1.5 Å (product state). A set of 40 structures are then extracted from the trajectory and prepared for the consequent US. After 40 windows of simulations finish, data is collected from each window and processed with WHAM code to generate the unbiased free energy profile. The resulting profile once again implied a dissociative mechanism (Figure 5-5). Similar to S1, TS1 is identified at approximate 2.5 Å with a barrier height of about 13.50 kcal/mol, and TS2 is found at approximate 1.95 Å with a barrier height of around 11.35 kcal/mol. Two TS are connected by an intermediate state at approximately 2.10 Å with a free energy approximately 10.40 kcal/mol higher than the S1 resting state.

Comparing to the profile previously obtained for S1. An interesting note is the difference at the RLS barrier height. S1 gives a 12.35 kcal/mol barrier while S3 produces a barrier of 13.50 kcal/mol, which is 1.15 kcal/mol higher. Experimental results show an approximate 4 fold-difference in their corresponding k_{cat} that results in a 0.8 kcal/mol difference (S3 higher than S1). Such a value is perfectly within chemical accuracy comparing to the experimental measurement.

In previous MM PMF study, S1 state is found approximately 2.3 kcal/mol energetically more favorable than S3 state. Put this energetic difference into consideration, a 11.15 kcal/mol barrier height for S3 chemical step is obtained.

S3 clearly has a much higher free energy than S1 with regard to the second barrier and the intermediate state. Structural analysis provides usually information to help understanding this difference. In their corresponding resting state, 1,6-DHN seems orienting ‘horizontally’ in S1 model while ‘vertically’ in S3 state. Thus, the resting state of S3 provides more surface area for the hydrophobic interactions with geranyl group. The resulted dispersion might be the source of an easier hydrolysis, or C_{1P}-O_{1P} bond broken. However, the orientation of 1,6-DHN in the product state is more similar to the orientation in S1 resting state. One possible explanation of this orientation change is in its original orientation the repulsion between aromatic groups and the attaching geranyl group is too strong thus 1,6-DHN has to adopt a pose similar to S1. Such a repulsion force or the reorientation of 1,6-DHN possibly causes the energy increase associated with the intermediate state and TS2.

The active site configuration of the intermediate state (see Figure 5-8) represents a very similar look to S1 state with the exception of a different orientation of 1,6-DHN. The carbon atoms in the first isoprene unit almost situate in the same plane, while the H_{1P}-H_{2P}-C_{1P}-C_{2P} dihedral is nearly 180°. The carbocation is sandwiched between the aromatic moiety of Tyr121 and ‘vertically’ oriented 1,6-DHN. C_{1P}, C_{2P} and C_{3P}, where the charge delocalization could take place, all have close aromatic moiety available to provide cation-aromatic π interactions. The side chains of Lys119, Lys169, Asn173, Tyr216 Arg228 and Lys284 help stabilized the PPI via hydrogen bond interactions, in addition to the stable Mg²⁺ coordination. Tyr288 forms a hydrogen bond with the hydroxyl group at C_{6P} of 1,6-DHN, by serving as either a hydrogen bond donor or acceptor. When the C_{1P} and C_{2A} reach into the reacting range, a couple of crystal water molecules, one of which is exactly the possible proton acceptor in S1, approach to the hydroxyl group of 1,6-DHN at C_{1A} and possibly serve as the proton abstractor.

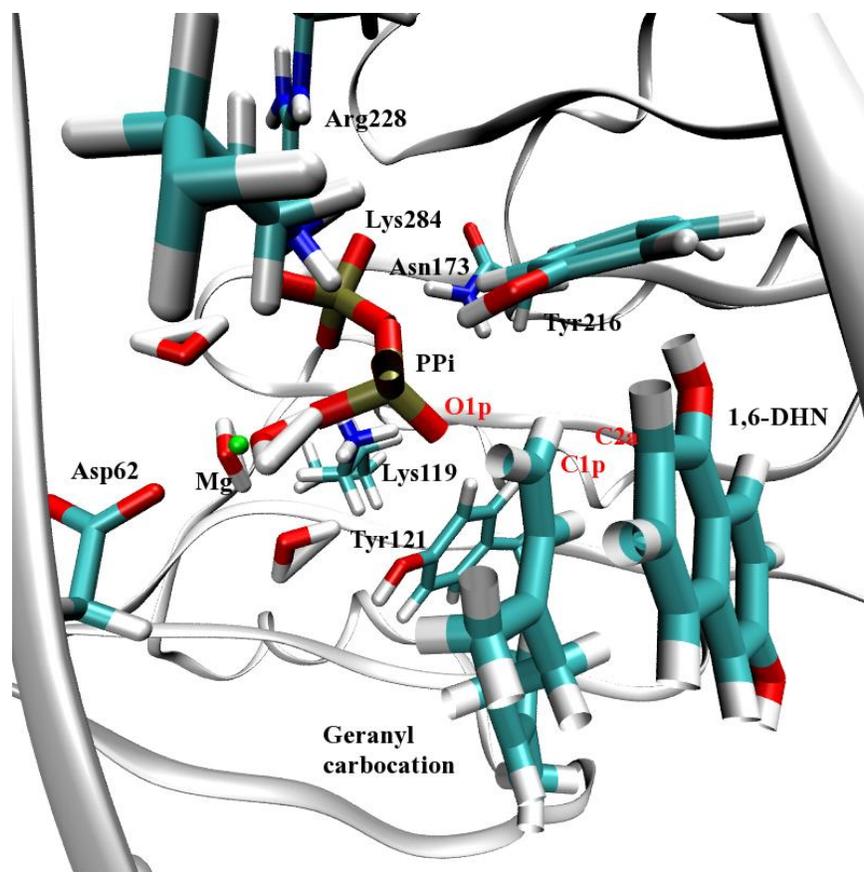


Figure 5-8. Active site conformation in S3 model at the intermediate state. O_{1P} of PPI, C_{1P} of carbocation, and C_{2A} of 1,6-DHN are labeled in red. Notice 1,6-DHN starts to rotate counter-clockwise in order to reduce steric repulsion for the following attachment.

The free energy profile obtained from SMD simulation (see Figure 5-9) also supports a dissociative mechanism for this model. The active site configuration after carbocation forming looks very similar to that of S3 state (Figure 5-10). 1,6-DHN is nearly ‘vertically’ oriented thus able to stabilize the carbocation via cation-aromatic π interactions that cover C_{1P} , C_{2P} and C_{3P} positions. Tyr121, although still cooperates with 1,6-DHN sandwiching the carbocation, seems moving away from C_{3P} . Thus it is unclear whether it still helps stabilizing the carbocation through cation- π interactions. The free energy barrier associated with this product is more than 15.0 kcal/mol. Qualitatively, this value is in agreement with the experiment evidence that S4 associated with a free energy higher than both S1 and S3. Quantitatively, however, this value is

not conclusive enough since several other important factors still need to be considered, such like SMD usually produces a free energy barrier than USP as well as the free energy difference associated with the orientation of 1,6-DHN between S1 or S3 and S4 remains unknown.

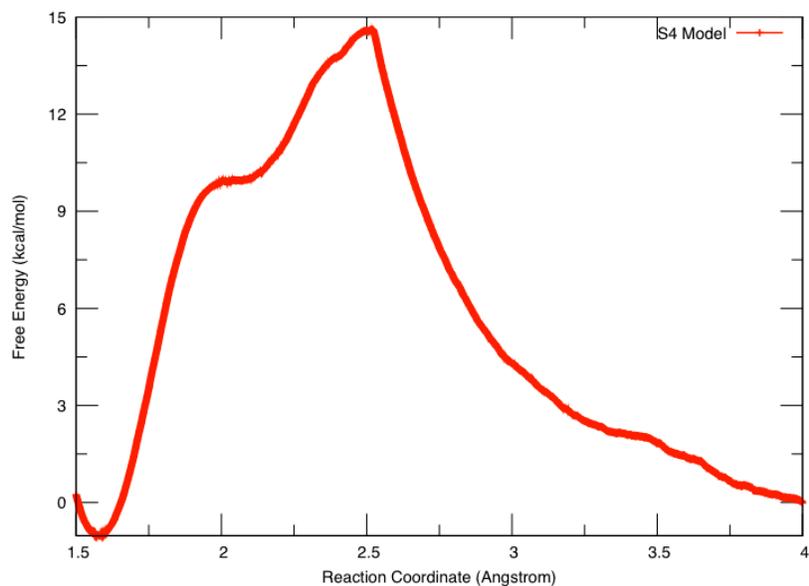


Figure 5-9. Free energy profile associated with S4 model prenylation. This profile starts at 2.3 kcal/mol, which is the free energy difference between S1 and any observed state orientation state of 1,6-DHN.

5.3.3 Proton Transfer Step of NphB Catalysis

The NphB chemistry includes both a prenylation step and a proton transfer step that restores aromatic quality of prenylated 1,6-DHN.

Proton transfer is usually a tricky reaction in computational simulation. The quantum tunneling effect is usually strong in this kind of reactions because hydrogen possesses such a small atomic mass. Therefore, QM calculation is required to fully understand such a scenario. In this study, we mainly focus on whether or not the deprotonation of geranylated 1,6-DHN is the RLS of NphB catalyzed prenylation reaction. Regarding this question, our previous analysis of the prenylation step has strongly implied that the deprotonation step is not the RLS. In addition, KIE experiment results of indole prenyltransferase system also disaffirm the deprotonation step

in a similar catalysis as the RLS. However, a more persuasive conclusion still requires QM/MM simulation coupled with PMF study.

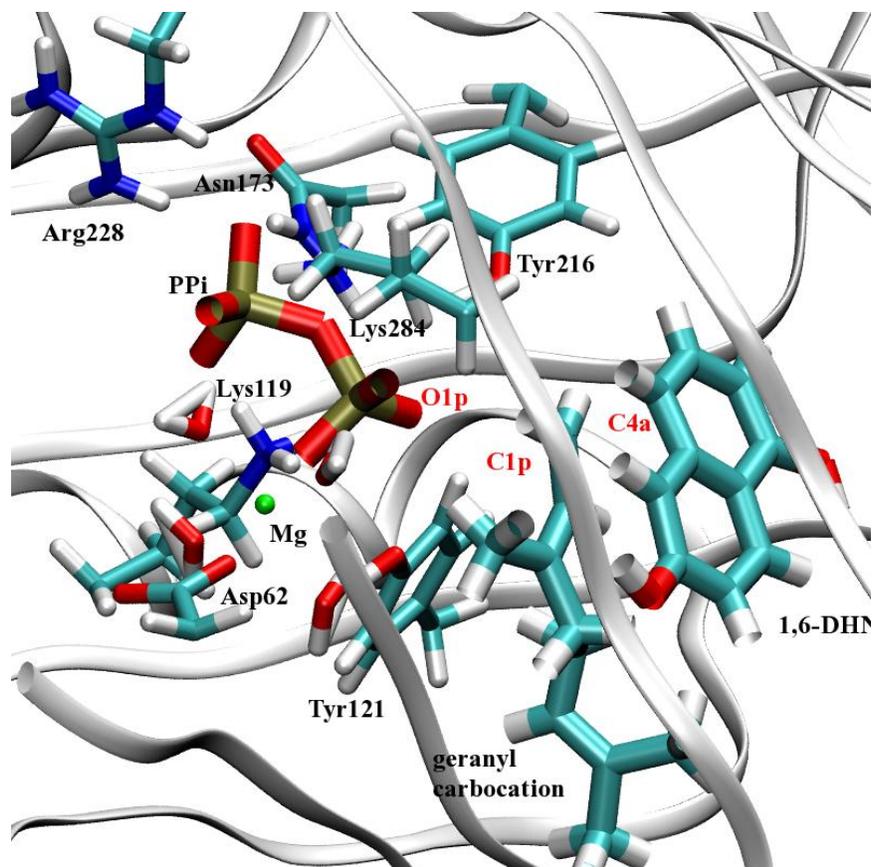


Figure 5-10. Active site snapshot of S4 prenylation. O_{1P} of PPI, C_{1P} of carbocation and C_{4A} of 1,6-DHN are red labeled.

The initial structure is first equilibrated with QM/MM potential for 250 ps in NPT ensemble. During the last 50 ps simulation, the water oxygen (O_w) remains in the middle of diphosphate group and geranyl-1,6-DHN (Figure 5-2). The distance between O_w and H_{5A} fluctuates around 3.0 Å, while the closest distance between water hydrogen (H_w) and O_{1P} stays around 3.5 Å. In addition, this water molecules forms hydrogen bond with both the hydroxyl group of geranyl-1,6-DHN and a non-O_{1P} α-diphosphate oxygen atom, thus making itself able to maintain its position and serve as a proton capturer as well.

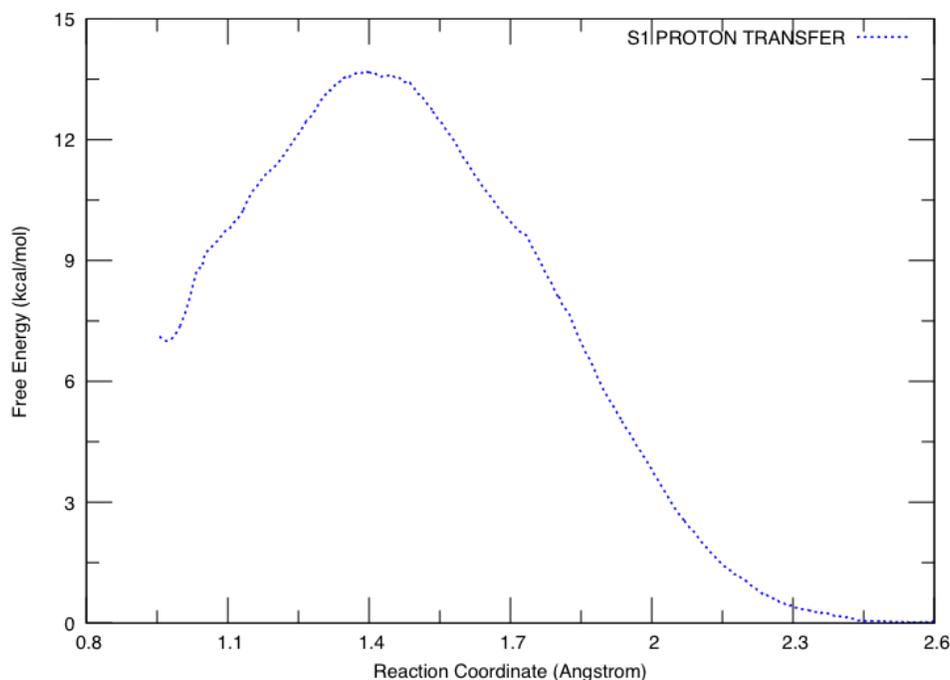


Figure 5-11. Free energy profile associated with the S1 proton transfer step. This profile is obtained from a SMD scan. Note the starting point (2.6 Å) corresponds to the end point (1.5 Å) in Figure 5-5.

A SMD scan is then taken to propagate the trajectory along the RC, $d_{\text{Ow-H5A}}$, from 3.0 Å to 0.95 Å, which is the default O-H bond length of TIP3P water model though this water is no longer a TIP3P water model. The resulting free energy profile gives an approximate 13.6 kcal/mol free energy barrier (see Figure 5-11). However, consider (1) the start point that is the end point of prenylation step (of S1 state) corresponds to a -6.0 kcal/mol free energy difference to the resting state and (2) the SMD free energy barrier height is usually higher than the USP result and (3) the existence of quantum tunneling effect will significantly drops the real barrier height, it seems the prenylation step is indeed the RLS of the NphB catalyzed geranylation.

The resulting product structure, 5-prenyl-1,6-DHN, displays a relaxed configuration (Figure 5-12). Tyr288 help stabilize this product by donating a hydrogen bond to one of the carboxyl groups of 5-prenyl-1,6-DHN. The H₃O resulted from the proton abstraction forms hydrogen bond with the hydroxyl group of Tyr216 as a hydrogen bond donor, while the tyrosine

serves as a hydrogen bond donor to interact with the PPi leaving group. Such a configuration implies there might be a context proton transfer scheme between these three residues, however, it is out of our discussion.

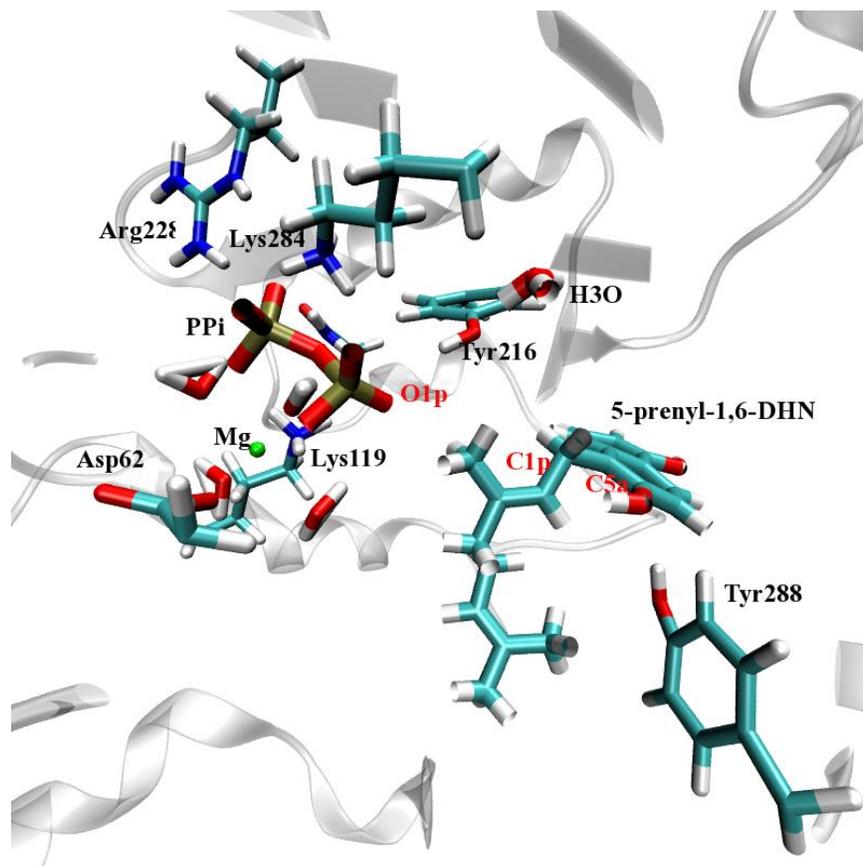


Figure 5-12. Active site configuration after proton abstraction from 5-prenyl-1,6-DHN to H3O (WAT in Figure 5-2). O_{IP} of PPi, C_{1P} of geranyl moiety and C_{5A} of aromatic product are labeled in red.

5.3.4 Substrate Protonation State Preference of FtmPT1

The MM simulation of FtmPT1 complexed with brevianamide F and DMAPP has been simulated with two models with two different protonation state of DMAPP adopted. In the first model, fully deprotonated DMAPP (DMAPP³⁻) is selected while in the second model the monoprotonated form of such an isoprenoid diphosphate (DMAPPH²⁻) is chosen.

Both models have been equilibrated with MM potential in NPT ensemble for several nano seconds. The resulting trajectories are examined and compared carefully. RMSD (root mean square deviation) and RMSF (root mean square fluctuation) information of two models in the last 2 ns of simulation are extracted. Both simulations appear stable over this period of time, though the fluctuation of DPP appears slightly bigger than DPH. The same trend has been observed in the RMSF plot, whereas the important areas appear stable in both models. The distances between two reacting atoms, C_{1P} of DMAPP and C_{2A} of brevianamide F, have also been pulled out. The result shows that DPH adopts a much shorter RC than DPP. In previous computational study of FTase¹⁴¹, the monoprotonated form the FPP seems the only form able to steer the conformational transition in the absence the magnesium. In FtmPT1 chemistry, however, Mg²⁺ dependency is not observed, thus whether the enzyme prefers DMAPPH²⁻ or its deprotonated form, DMAPP³⁻, is of great interest to us.

The active site configurations of two models are examined separately. In DPH model, side chains of positively charged residues such as Arg113, Lys201 and Lys294 anchor the negatively charged diphosphate moiety via salt bridges. In addition, the side chains of Tyr203, Gln380, Thr448 and Tyr450 help stabilize the diphosphate group via hydrogen bond interactions. Four tyrosine residues, including Tyr203, Tyr296, Tyr382 and Tyr450 (see Figure 5-3), surround the binding pocket of DMAPPH²⁻. These four aromatic AA's, along with Phe280 and brevianamide F itself, form an aromatic rich environment, which is proposed to function as a steer to engineer the rotation of developing carbocation required by the new bond forming. Two reacting atoms are approximately 4.5 Å apart, while C_{3P} of DMAPPH²⁻ is also within 4.5 Å to the C_{3A} atom of brevianamide F, where an alternative reaction will take place upon the G115T mutation.

On the other hand, in DPP model, the diphosphate moiety is also anchored by Arg113, Lys201 and Lys294 via salt bridge interactions, and the side chains of Tyr203, Gln308 and Tyr450 still form hydrogen bond interactions with diphosphate, providing additional support. The distance between C_{1P} and C_{2A} fluctuates around 5.0 Å while the C_{3P} - C_{3A} distance is approximate 4.0 Å. The tyrosine shield and aromatic environment are stable throughout the simulation as well.

Different to FTase, which requires a conformational transition step prior to the chemical step, no evidence shows that FtmPT1 requires such a physical step, although the two reacting atoms are at least 4.5 Å apart (in DPH model, more than 5 Å away in DPP model). Therefore, a PMF study at MM level might be unnecessary and useless. As a result, QM/MM PMF study is the only way to testify the substrate preference of this enzyme.

5.3.5 A Look at FtmPT1 Catalysis

The chemical reaction of FtmPT1 catalysis is studied by dividing the PMF into two parts. The first part is the hydrolysis step that adopts $d_{C_{1P}-O_{1P}}$ as the RC, while the second part is the bond formation step that a RC of $d_{C_{1P}-C_{2A}}$ is employed. Such a strategy is a compromise of a 1-D PMF study with a RC of either $d_{C_{1P}-O_{1P}}$ or $d_{RC} = d_{C_{1P}-O_{1P}} - d_{C_{1P}-C_{2A}}$ and a 2-D PMF study with $d_{RC}^1 = d_{C_{1P}-O_{1P}}$ and $d_{RC}^2 = d_{C_{1P}-C_{2A}}$. The former case fails to give a reasonable FES because the simulations tend to move the DMAPP close to brevianamide F before breaks the C_{1P} - O_{1P} bond thus the free energy barrier is mistakenly raised due to the huge steric effect associated with the movement of DMAPP. On the other hand, the 2-D PMF will be a very expensive and time-consuming approach without gaining any knowledge about the reaction pathway. The adopted strategy is rational because within each step the free energy is described as a function of the most important property.

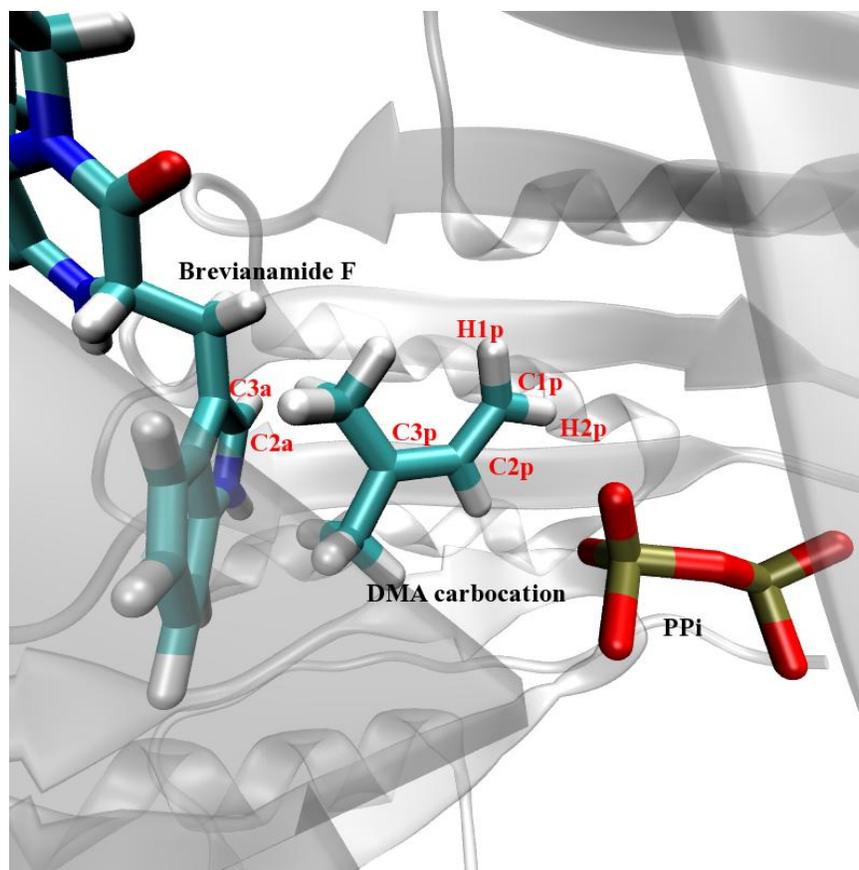


Figure 5-13. Carbocation formed in FtmPT1 catalysis. Important atoms are labeled in red. H1p corresponds to H_{1P} in text where p means ‘from diphosphate substrate’. C2a denotes C_{2A} in text where a means ‘from aromatic substrate’.

A QM/MM SMD scan is conducted to study the hydrolysis step first. The RC is moved from 1.5 Å (reactant state) to 3.5 Å where C_{1P} and O_{1P} are sufficiently apart thus a carbocation should have already formed. The resulted free energy profile features a steadily increase before $d_{RC} = 2.7$ Å and a flat curve thereafter. Such a FES implies the formation of a carbocation after the two bonding atoms are forced sufficiently apart. The active site configuration extracted from the trajectory provides addition support to the existence of a carbocation as five carbon atoms of dimethylallylic group nearly situate in the same plane, while the dihedral of H_{1P}-H_{2P}-C_{1P}-C_{2P} remains nearly 180° after 2.7 Å point (Figure 5-13). The resulted dimethylallylic cation is sandwiched by two aromatic units: brevianamide F and a tyrosine residue involved in the

tyrosine shield. Strong cation-aromatic π interactions are expected from both sides of the carbocation. In addition, the shape of the tyrosine shield enables the steer of the carbocation rotation in order to approach another reactant, since continuous cation-aromatic π is provided by at least one of four tyrosine residues regardless the orientation of dimethylallylic cation inside the pocket.

An initial structure of the second step is extracted from the SMD trajectory with $d_{RC} = 2.8$ Å upon where a carbocation is already generated. The resulted structure is then undergoing another QM/MM SMD scan with a RC of $d_{C_{1P}-C_{2A}}$. In the configuration of such an initial structure, C_{1P} and C_{2A} are 5.8 Å apart, thus the trajectory is propagated to move the RC from 5.8 Å to 1.5 Å. The resulted free energy curve (data not shown) indicates the existence of an intermediate state after the carbocation rotates its orientation slightly. Another free energy barrier is observed at $d_{RC} = 2.1$ Å, due to the strong repulsion generated in the approaching of two carbon atoms. The hydrogen at C_{2A} position distinctly bends out of the aromatic plane, indicating a proton removal step is required to restore the aromatic quality of the aromatic substrate.

Despite the qualitatively understanding of FtmPT1 catalyzed prenylation, the quantitative properties, such as free energy barrier height or the exact location of important state, obtained from SMD simulation are not sufficiently conclusive. In fact, it is not unusual to find distinguishable difference between SMD generated free energy profile and USP resulted free energy curve. Therefore, a set of QM/MM USP is still required to complete our understanding of this catalysis. In addition, effort should be also made to study the proton abstraction step and the impact of key residue mutation such as G115T. These works are currently undergoing in our lab.

5.4 Summary

In this study, effort is made to illustrate the reaction mechanism catalyzed by two aromatic prenyltransferases: NphB and FtmPT1.

NphB is determined to catalyze the geranylation via a dissociative pathway. A stable carbocation is observed for each of three models leading to three different products. The free energy parameters obtained from QM/MM USP (12.35 kcal/mol for S1 and 13.50 kcal/mol for S3) are in excellent agreement with results from kinetic experiments (12.6 kcal/mol for S1 and 13.3 kcal/mol for S3). Although the free energy barrier obtained for S4 through a QM/MM SMD simulation qualitatively matches the experimentally observed trend, quantitatively it is not convincing enough. The proton transfer step is studied with S1 model, the given free energy profile suggests a barrier height lower than the hydrolysis of GPP, therefore, the RLS is determined to be the hydrolysis step in the prenylation.

Indole prenyltransferase FtmPT1 is also studied. In the MM simulation of both FtmPT1 complexed with DMAPP²⁻ and FtmPT1 complexed DMAPP³⁻, stable simulations are reached. The distance between two reacting atoms is slightly smaller in DPH model than in DPP model. Except that, everything looks similar in two active site configurations. QM/MM PMF is required to identify which protonation state of DMAPP is more favorable to FtmPT1. However, due to the large separation between two reactants, traditional 1-D PMF study fails because the system prefers to move the reactants into a close range before forcing bond cleavage. A 2-D PMF without preliminary knowledge of reaction pathway requires a huge number of sampling windows, thus making itself not an ideal implementation. As a compromise, a two stage 1-D PMF strategy is adopted and gives exciting results. The 2-stage 1-D SMD scan reveals a stable carbocation and an intermediate state associated with the rotation of such a carbocation. Tyrosine rich environment in DMAPP binding pocket plays an important role to stabilize the lipid group and furthermore engineer its rotation via cation-aromatic π interactions.

The aromatic rich environment in the active sites of aromatic prenyltransferases makes huge impact on the mechanism selection of this type of enzymes. Their ability to stabilize the developing carbocation via cation-aromatic π interactions is essential to the S_N1 -like mechanism. Such findings make an interesting argument about the target-dependent mechanism selectivity of prenyltransferases, in addition to the mechanism switch associated with the fascinating context-dependent substrate selectivity across different FTase/peptide complexes.

LIST OF REFERENCES

- (1) Cox, A. D. *Drugs* **2001**, *61*, 723-32.
- (2) Zhang, F. L.; Casey, P. J. *Annu. Rev. Biochem.* **1996**, *65*, 241-69.
- (3) Kuzuyama, T.; Noel, J. P.; Richard, S. B. *Nature* **2005**, *435*, 983-7.
- (4) Jost, M.; Zocher, G.; Tarcz, S.; Matuschek, M.; Xie, X. L.; Li, S. M.; Stehle, T. *J. Am. Chem. Soc.* **2010**, *132*, 17849-58.
- (5) Pojer, F.; Wemakor, E.; Kammerer, B.; Chen, H. W.; Walsh, C. T.; Li, S. M.; Heide, L. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 2316-21.
- (6) Piironen, V.; Lindsay, D. G.; Miettinen, T. A.; Toivo, J.; Lampi, A. M. *J. Sci. Food Agric.* **2000**, *80*, 939-66.
- (7) Park, H. W.; Boduluri, S. R.; Moomaw, J. F.; Casey, P. J.; Beese, L. S. *Science* **1997**, *275*, 1800-4.
- (8) Casey, P. J.; Seabra, M. C. *J. Biol. Chem.* **1996**, *271*, 5289-92.
- (9) Clarke, S. *Annu. Rev. Biochem.* **1992**, *61*, 355-86.
- (10) Caldwell, G. A.; Naider, F.; Becker, J. M. *Microbiological Reviews* **1995**, *59*, 406-22.
- (11) Schafer, W. R.; Rine, J. *Annu. Rev. Genet.* **1992**, *26*, 209-37.
- (12) Omer, C. A.; Gibbs, J. B. *Mol. Microbiol.* **1994**, *11*, 219-25.
- (13) Glomset, J. A.; Farnsworth, C. C. *Annual Review of Cell Biology* **1994**, *10*, 181-205.
- (14) Newman, C. M. H.; Magee, A. I. *Biochim. Biophys. Acta* **1993**, *1155*, 79-96.
- (15) Park, H. W.; Beese, L. S. *Curr. Opin. Struct. Biol.* **1997**, *7*, 873-80.
- (16) Fu, H. W.; Beese, L. S.; Casey, P. J. *Biochemistry* **1998**, *37*, 4465-72.
- (17) Long, S. B.; Casey, P. J.; Beese, L. S. *Biochemistry* **1998**, *37*, 9612-8.
- (18) Long, S. B.; Casey, P. J.; Beese, L. S. *Nature* **2002**, *419*, 645-50.
- (19) Zhang, F. L.; Casey, P. J. *Biochem. J.* **1996**, *320*, 925-32.
- (20) Taylor, J. S.; Reid, T. S.; Terry, K. L.; Casey, P. J.; Beese, L. S. *EMBO J.* **2003**, *22*, 5963-74.

- (21) Hartman, H. L.; Bowers, K. E.; Fierke, C. A. *J. Biol. Chem.* **2004**, *279*, 30546-53.
- (22) Reid, T. S.; Terry, K. L.; Casey, P. J.; Beese, L. S. *J. Mol. Biol.* **2004**, *343*, 417-33.
- (23) Maurer-Stroh, S.; Washietl, S.; Eisenhaber, F. *Genome Biology* **2003**, *4*, -.
- (24) Pfeffer, S. R.; Diracsvejstrup, A. B.; Soldati, T. *J. Biol. Chem.* **1995**, *270*, 17057-9.
- (25) Nuoffer, C.; Balch, W. E. *Annu. Rev. Biochem.* **1994**, *63*, 949-90.
- (26) Reiss, Y.; Goldstein, J. L.; Seabra, M. C.; Casey, P. J.; Brown, M. S. *Cell* **1990**, *62*, 81-8.
- (27) Seabra, M. C.; Reiss, Y.; Casey, P. J.; Brown, M. S.; Goldstein, J. L. *Cell* **1991**, *65*, 429-34.
- (28) Seabra, M. C.; Goldstein, J. L.; Sudhof, T. C.; Brown, M. S. *J. Biol. Chem.* **1992**, *267*, 14497-503.
- (29) Reiss, Y.; Stradley, S. J.; Gierasch, L. M.; Brown, M. S.; Goldstein, J. L. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 732-6.
- (30) Troutman, J. M.; Andres, D. A.; Spielmann, H. P. *Biochemistry* **2007**, *46*, 11299-309.
- (31) Houglund, J. L.; Lamphear, C. L.; Scott, S. A.; Gibbs, R. A.; Fierke, C. A. *Biochemistry* **2009**, *48*, 1691-701.
- (32) Zhang, F. L.; Fu, H. W.; Casey, P. J.; Bishop, W. R. *Biochemistry* **1996**, *35*, 8166-71.
- (33) Reiss, Y.; Brown, M. S.; Goldstein, J. L. *J. Biol. Chem.* **1992**, *267*, 6403-8.
- (34) Myers, L. C.; Jackow, F.; Verdine, G. L. *J. Biol. Chem.* **1995**, *270*, 6664-70.
- (35) Edwards, D. J.; Gerwick, W. H. *J. Am. Chem. Soc.* **2004**, *126*, 11432-3.
- (36) Sacchettini, J. C.; Poulter, C. D. *Science* **1997**, *277*, 1788-9.
- (37) Bohlmann, J.; Meyer-Gauen, G.; Croteau, R. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 4126-33.
- (38) Ashby, M. N.; Kutsunai, S. Y.; Ackerman, S.; Tzagoloff, A.; Edwards, P. A. *J. Biol. Chem.* **1992**, *267*, 4128-36.
- (39) Suvarna, K.; Stevenson, D.; Meganathan, R.; Hudspeth, M. E. S. *J. Bacteriol.* **1998**, *180*, 2782-7.

- (40) Yazaki, K.; Sasaki, K.; Tsurumaru, Y. *Phytochemistry* **2009**, *70*, 1739-45.
- (41) Koyama, T.; Tajima, M.; Sano, H.; Doi, T.; KoikeTakeshita, A.; Obata, S.; Nishino, T.; Ogura, K. *Biochemistry* **1996**, *35*, 9533-8.
- (42) Kharel, Y.; Koyama, T. *Nat. Prod. Rep.* **2003**, *20*, 111-8.
- (43) Liang, P. H.; Ko, T. P.; Wang, A. H. J. *Eur. J. Biochem.* **2002**, *269*, 3339-54.
- (44) Tarshis, L. C.; Yan, M. J.; Poulter, C. D.; Sacchettini, J. C. *Biochemistry* **1994**, *33*, 10871-7.
- (45) Cane, D. E. *Isoprenoids, Including Carotenoids and Steroids, in Comprehensive Natural Products Chemistry*; Elsevier: London, 1998.
- (46) Gerlt, J. A.; Raushel, F. M. *Curr. Opin. Chem. Biol.* **2003**, *7*, 252-64.
- (47) Sacchettini, J. C.; Gordon, J. I.; Banaszak, L. J. *J. Mol. Biol.* **1989**, *208*, 327-39.
- (48) Xu, Z. H.; Bernlohr, D. A.; Banaszak, L. J. *J. Biol. Chem.* **1993**, *268*, 7874-84.
- (49) Branden, C. I.; Tooze, J. *Introduction to Protein Structure*; 2nd ed. Garland, New York, 1999.
- (50) Shinya, K.; Imai, S.; Furihata, K.; Hayakawa, Y.; Kato, Y.; Vanduyne, G. D.; Clardy, J.; Seto, H. *J. Antibiot.* **1990**, *43*, 444-7.
- (51) Kazuo, S. Y.; Furihata, K.; Hayakawa, Y.; Seto, H. *Tetrahedron Lett.* **1990**, *31*, 6025-6.
- (52) Seto, H.; Watanabe, H.; Furihata, K. *Tetrahedron Lett.* **1996**, *37*, 7979-82.
- (53) Botta, B.; Vitali, A.; Menendez, P.; Misiti, D.; Delle Monache, G. *Curr. Med. Chem.* **2005**, *12*, 713-39.
- (54) Steffan, N.; Grundmann, A.; Yin, W. B.; Kremer, A.; Li, S. M. *Curr. Med. Chem.* **2009**, *16*, 218-31.
- (55) Li, S. M. *Nat. Prod. Rep.* **2010**, *27*, 57-78.
- (56) Zhao, S.; Smith, K. S.; Deveau, A. M.; Dieckhaus, C. M.; Johnson, M. A.; Macdonald, T. L.; Cook, J. M. *J. Med. Chem.* **2002**, *45*, 1559-62.
- (57) Woehlecke, H.; Osada, H.; Herrmann, A.; Lage, H. *Int. J. Cancer* **2003**, *107*, 721-8.
- (58) Jain, H. D.; Zhang, C. C.; Zhou, S.; Zhou, H.; Ma, J. X.; Liu, X. B.; Liao, X.; Deveau, A. M.; Dieckhaus, C. M.; Johnson, M. A.; Smith, K. S.; Macdonald, T. L.; Makeya, H.; Osada, H.; Cook, J. M. *Bioorg. Med. Chem.* **2008**, *16*, 4626-51.

- (59) Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P.; Moraes, M. A.; Piana, S.; Shan, Y.; Towles, B. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*; ACM: Portland, Oregon, 2009, p 1-11.
- (60) Nevins, N.; Chen, K. S.; Allinger, N. L. *Journal of Computational Chemistry* **1996**, *17*, 669-94.
- (61) Nevins, N.; Lii, J. H.; Allinger, N. L. *Journal of Computational Chemistry* **1996**, *17*, 695-729.
- (62) Nevins, N.; Allinger, N. L. *Journal of Computational Chemistry* **1996**, *17*, 730-46.
- (63) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551-66.
- (64) Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8566-75.
- (65) Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8576-82.
- (66) London, F. *Zeitschrift Fur Physik* **1930**, *63*, 245-79.
- (67) Halgren, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 7827-43.
- (68) Halgren, T. A. *Journal of Computational Chemistry* **1996**, *17*, 490-519.
- (69) Halgren, T. A. *Journal of Computational Chemistry* **1996**, *17*, 520-52.
- (70) Halgren, T. A. *Journal of Computational Chemistry* **1996**, *17*, 553-86.
- (71) Halgren, T. A.; Nachbar, R. B. *Journal of Computational Chemistry* **1996**, *17*, 587-615.
- (72) Halgren, T. A. *Journal of Computational Chemistry* **1996**, *17*, 616-41.
- (73) Claessens, M.; Ferrario, M.; Ryckaert, J. P. *Mol. Phys.* **1983**, *50*, 217-27.
- (74) Rick, S. W.; Berne, B. J. *J. Am. Chem. Soc.* **1996**, *118*, 672-9.
- (75) Sprik, M.; Klein, M. L. *J. Chem. Phys.* **1988**, *89*, 7556-60.
- (76) Dang, L. X.; Rice, J. E.; Caldwell, J.; Kollman, P. A. *J. Am. Chem. Soc.* **1991**, *113*, 2481-6.
- (77) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141-56.

- (78) Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. *J. Comput. Chem.* **1994**, *15*, 162-82.
- (79) Hwang, M. J.; Stockfisch, T. P.; Hagler, A. T. *J. Am. Chem. Soc.* **1994**, *116*, 2515-25.
- (80) Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Hagler, A. T. *Isr. J. Chem.* **1994**, *34*, 195-231.
- (81) Allinger, N. L.; Chen, K. S.; Lii, J. H. *Journal of Computational Chemistry* **1996**, *17*, 642-68.
- (82) Allinger, N. L.; Chen, K. S.; Katzenellenbogen, J. A.; Wilson, S. R.; Anstead, G. M. *Journal of Computational Chemistry* **1996**, *17*, 747-55.
- (83) Lifson, S.; Warshel, A. *J. Chem. Phys.* **1968**, *49*, 5116-&.
- (84) Hagler, A. T.; Huler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319-27.
- (85) Hagler, A. T.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5327-35.
- (86) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765-84.
- (87) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179-97.
- (88) Wang, J. M.; Cieplak, P.; Kollman, P. A. *Journal of Computational Chemistry* **2000**, *21*, 1049-74.
- (89) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins-Structure Function and Bioinformatics* **2006**, *65*, 712-25.
- (90) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586-616.
- (91) Mackerell, A. D.; Feig, M.; Brooks, C. L. *Journal of Computational Chemistry* **2004**, *25*, 1400-15.
- (92) MacKerell, A. D.; Banavali, N.; Foloppe, N. *Biopolymers* **2000**, *56*, 257-65.
- (93) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596-607.

- (94) Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; Van Gunsteren, W. F. *Journal of Computational Chemistry* **2005**, *26*, 1719-51.
- (95) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225-36.
- (96) Jorgensen, W. L.; Tiradorives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657-66.
- (97) Ren, P. Y.; Ponder, J. W. *Journal of Computational Chemistry* **2002**, *23*, 1497-506.
- (98) Ren, P. Y.; Ponder, J. W. *J. Phys. Chem. B* **2004**, *108*, 13427-37.
- (99) Ren, P. Y.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933-47.
- (100) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087-92.
- (101) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957**, *27*, 1208-9.
- (102) Verlet, L. *Physical Review* **1967**, *159*, 98-&.
- (103) Hockney, R. W. *Methods Computation Phys.* **1970**, *9*, 135.
- (104) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-41.
- (105) Woodcock, L. V. *Chem. Phys. Lett.* **1971**, *10*, 257-&.
- (106) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684-90.
- (107) Langevin, P. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences* **1908**, *146*, 530-3.
- (108) Leach, A. R. *Molecular Modelling: Principles and Applications*; 2nd ed.; Prentice Hall: New York, 2001.
- (109) Nicholls, A.; Honig, B. *Journal of Computational Chemistry* **1991**, *12*, 435-45.
- (110) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824-39.
- (111) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122-9.
- (112) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129-52.
- (113) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712-20.

- (114) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins-Structure Function and Bioinformatics* **2004**, *55*, 383-94.
- (115) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-35.
- (116) Stilling, Fh; Rahman, A. *J. Chem. Phys.* **1974**, *60*, 1545-57.
- (117) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910-22.
- (118) Ferguson, D. M. *Journal of Computational Chemistry* **1995**, *16*, 501-11.
- (119) Barnes, P.; Finney, J. L.; Nicholas, J. D.; Quinn, J. E. *Nature* **1979**, *282*, 459-64.
- (120) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Clarendon Press ; Oxford University Press: Oxford [England]: New York, 1987.
- (121) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269-71.
- (122) Bernal, J. D.; Fowler, R. H. *J. Chem. Phys.* **1933**, *1*, 515-48.
- (123) Ewald, P. P. *Annalen Der Physik* **1921**, *64*, 253-87.
- (124) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089-92.
- (125) Luty, B. A.; Davis, M. E.; Tironi, I. G.; Vangunsteren, W. F. *Mol. Simul.* **1994**, *14*, 11-20.
- (126) Luty, B. A.; Tironi, I. G.; Vangunsteren, W. F. *J. Chem. Phys.* **1995**, *103*, 3014-21.
- (127) Crespo, A.; Marti, M. A.; Estrin, D. A.; Roitberg, A. E. *J. Am. Chem. Soc.* **2005**, *127*, 6940-1.
- (128) Xiong, H.; Crespo, A.; Marti, M.; Estrin, D.; Roitberg, A. E. *Theor. Chem. Acc.* **2006**, *116*, 338-46.
- (129) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187-99.
- (130) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011-21.
- (131) Ryckaert, J. P.; Bellemans, A. *Faraday Discuss.* **1978**, *95*-106.
- (132) Jorgensen, W. L.; Gao, J.; Ravimohan, C. *J. Phys. Chem.* **1985**, *89*, 3470-3.
- (133) Goringe, C. M.; Bowler, D. R.; Hernandez, E. *Reports on Progress in Physics* **1997**, *60*, 1447-512.

- (134) Roskoski, R. *Biochem. Biophys. Res. Commun.* **2003**, *303*, 1-7.
- (135) Resh, M. D. *Cell. Signal.* **1996**, *8*, 403-12.
- (136) Sinensky, M. *Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids* **2000**, *1484*, 93-106.
- (137) Seabra, M. C. *Cell. Signal.* **1998**, *10*, 167-72.
- (138) Adjei, A. A. *Journal of the National Cancer Institute* **2001**, *93*, 1062-74.
- (139) Huang, C. C.; Hightower, K. E.; Fierke, C. A. *Biochemistry* **2000**, *39*, 2593-602.
- (140) Cui, G. L.; Wang, B.; Merz, K. M. *Biochemistry* **2005**, *44*, 16513-23.
- (141) Cui, G.; Merz, K. M., . *Biochemistry* **2007**, *46*, 12375-81.
- (142) Hightower, K. E.; Huang, C. C.; Casey, P. J.; Fierke, C. A. *Biochemistry* **1998**, *37*, 15555-62.
- (143) Merz, K. M.; Murcko, M. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1991**, *113*, 4484-90.
- (144) Saderholm, M. J.; Hightower, K. E.; Fierke, C. A. *Biochemistry* **2000**, *39*, 12398-405.
- (145) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-74.
- (146) Meagher, K. L.; Redman, L. T.; Carlson, H. A. *J. Comput. Chem.* **2003**, *24*, 1016-25.
- (147) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-80.
- (148) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357-77.

- (149) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc.: Wallingford CT, 2004.
- (150) Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; B.Wang; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossv áry, I.; K.F.Wong; Paesani, F.; Vanicek, J.; X.Wu; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A.; University of California, San Francisco: 2008.
- (151) Pickett, J. S.; Bowers, K. E.; Fierke, C. A. *J. Biol. Chem.* **2003**, *278*, 51243-50.
- (152) Aqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021-4.
- (153) Cui, G. L.; Li, X.; Merz, K. M. *Biochemistry* **2007**, *46*, 1303-11.
- (154) Grossfield, A.; 1.6d ed.
- (155) Ulitsky, A.; Elber, R. *J. Chem. Phys.* **1993**, *98*, 3380-8.
- (156) Simmerling, C.; Fox, T.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 5771-82.
- (157) Wang, X. S.; Roitberg, A. E.; Richards, N. G. J. *Biochemistry* **2009**, *48*, 12272-82.
- (158) Peters, M. B.; Yang, Y.; Wang, B.; Fusti-Molnar, L.; Weaver, M. N.; Merz, K. M. *J. Chem. Theory Comput.* **2010**, *6*, 2935-47.
- (159) Pais, J. E.; Bowers, K. E.; Stoddard, A. K.; Fierke, C. A. *Anal. Biochem.* **2005**, *345*, 302-11.
- (160) Furfine, E. S.; Leban, J. J.; Landavazo, A.; Moomaw, J. F.; Casey, P. J. *Biochemistry* **1995**, *34*, 6857-62.
- (161) Mathis, J. R.; Poulter, C. D. *Biochemistry* **1997**, *36*, 6367-76.

- (162) Gibbs, J. B.; Oliff, A.; Kohl, N. E. *Cell* **1994**, *77*, 175-8.
- (163) Graham, S. L.; Desolms, S. J.; Giuliani, E. A.; Kohl, N. E.; Mosser, S. D.; Oliff, A. I.; Pompliano, D. L.; Rands, E.; Breslin, M. J.; Deana, A. A.; Garsky, V. M.; Scholz, T. H.; Gibbs, J. B.; Smith, R. L. *J. Med. Chem.* **1994**, *37*, 725-32.
- (164) Garcia, A. M.; Rowell, C.; Ackermann, K.; Kowalczyk, J. J.; Lewis, M. D. *J. Biol. Chem.* **1993**, *268*, 18415-8.
- (165) Nigam, M.; Seong, C. M.; Qian, Y. M.; Hamilton, A. D.; Sebti, S. M. *J. Biol. Chem.* **1993**, *268*, 20695-8.
- (166) James, G. L.; Goldstein, J. L.; Brown, M. S.; Rawson, T. E.; Somers, T. C.; McDowell, R. S.; Crowley, C. W.; Lucas, B. K.; Levinson, A. D.; Marsters, J. C. *Science* **1993**, *260*, 1937-42.
- (167) Patel, D. V.; Schmidt, R. J.; Biller, S. A.; Gordon, E. M.; Robinson, S. S.; Manne, V. *J. Med. Chem.* **1995**, *38*, 2906-21.
- (168) Manne, V.; Yan, N.; Carboni, J. M.; Tuomari, A. V.; Ricca, C. S.; Brown, J. G.; Andahazy, M. L.; Schmidt, R. J.; Patel, D.; Zahler, R.; Weinmann, R.; Der, C. J.; Cox, A. D.; Hunt, J. T.; Gordon, E. M.; Barbacid, M.; Seizinger, B. R. *Oncogene* **1995**, *10*, 1763-79.
- (169) Manne, V.; Ricca, C. S.; Brown, J. G.; Tuomari, A. V.; Yan, N.; Patel, D.; Schmidt, R.; Lynch, M. J.; Ciosek, C. P.; Carboni, J. M.; Robinson, S.; Gordon, E. M.; Barbacid, M.; Seizinger, B. R.; Biller, S. A. *Drug Dev. Res.* **1995**, *34*, 121-37.
- (170) Patel, D. V.; Gordon, E. M.; Schmidt, R. J.; Weller, H. N.; Young, M. G.; Zahler, R.; Barbacid, M.; Carboni, J. M.; Gullobrown, J. L.; Hunihan, L.; Ricca, C.; Robinson, S.; Seizinger, B. R.; Tuomari, A. V.; Manne, V. *J. Med. Chem.* **1995**, *38*, 435-42.
- (171) Tamanoi, F. *Trends Biochem. Sci.* **1993**, *18*, 349-53.
- (172) Tamanoi, F.; Esson, K.; Gomez, R.; Wood, D.; Uh, M.; Hara, M.; Akasaka, K.; Okabe, M.; Nakano, H. *FASEB J.* **1993**, *7*, A1049-A.
- (173) Bishop, W. R.; Bond, R.; Petrin, J.; Wang, L.; Patton, R.; Doll, R.; Njoroge, G.; Catino, J.; Schwartz, J.; Windsor, W.; Syto, R.; Schwartz, J.; Carr, D.; James, L.; Kirschmeier, P. *J. Biol. Chem.* **1995**, *270*, 30611-8.
- (174) Chen, Z.; Sun, J. Z.; Pradines, A.; Favre, G.; Adnane, J.; Sebti, S. M. *J. Biol. Chem.* **2000**, *275*, 17974-8.
- (175) Maurer-Stroh, S.; Washietl, S.; Eisenhaber, F. *Biol. Chem.* **2003**, *384*, 977-89.

- (176) Gelb, M. H.; Van Voorhis, W. C.; Buckner, F. S.; Yokoyama, K.; Eastman, R.; Carpenter, E. P.; Panethymitaki, C.; Brown, K. A.; Smith, D. F. *Mol. Biochem. Parasitol.* **2003**, *126*, 155-63.
- (177) Mu, Y. Q.; Omer, C. A.; Gibbs, R. A. *J. Am. Chem. Soc.* **1996**, *118*, 1817-23.
- (178) Weller, V. A.; Distefano, M. D. *J. Am. Chem. Soc.* **1998**, *120*, 7975-6.
- (179) Harris, C. M.; Derdowski, A. M.; Poulter, C. D. *Biochemistry* **2002**, *41*, 10554-62.
- (180) Dolence, J. M.; Poulter, C. D. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 5008-11.
- (181) Cassidy, P. B.; Poulter, C. D. *J. Am. Chem. Soc.* **1996**, *118*, 8761-2.
- (182) Pickett, J. S.; Bowers, K. E.; Hartman, H. L.; Fu, H. W.; Embry, A. C.; Casey, P. J.; Fierke, C. A. *Biochemistry* **2003**, *42*, 9741-8.
- (183) Wu, Z.; Demma, M.; Strickland, C. L.; Radisky, E. S.; Poulter, C. D.; Le, H. V.; Windsor, W. T. *Biochemistry* **1999**, *38*, 11239-49.
- (184) Pais, J. E.; Bowers, K. E.; Fierke, C. A. *J. Am. Chem. Soc.* **2006**, *128*, 15086-7.
- (185) Ho, M. H.; De Vivo, M.; Dal Peraro, M.; Klein, M. L. *J. Chem. Theory Comput.* **2009**, *5*, 1657-66.
- (186) Yang, Y.; Weaver, M. N.; Merz, K. M. *J. Phys. Chem. A* **2009**, *113*, 9843-51.
- (187) Scheuring, J.; Berti, P. J.; Schramm, V. L. *Biochemistry* **1998**, *37*, 2748-58.
- (188) Yang, Y.; Yu, H. B.; York, D.; Elstner, M.; Cui, Q. *J. Chem. Theory Comput.* **2008**, *4*, 2067-84.
- (189) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Physical Review B* **1998**, *58*, 7260-8.
- (190) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, *24*, 565-81.
- (191) Seabra, G. D.; Walker, R. C.; Elstner, M.; Case, D. A.; Roitberg, A. E. *J. Phys. Chem. A* **2007**, *111*, 5655-64.
- (192) Gronert, S.; Lee, J. M. *J. Org. Chem.* **1995**, *60*, 6731-6.
- (193) Long, S. B.; Casey, P. J.; Beese, L. S. *Structure* **2000**, *8*, 209-22.
- (194) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95-&.
- (195) Christianson, D. W. *Chem. Rev. (Washington, DC, U. S.)* **2006**, *106*, 3412-42.

- (196) Santos, F. A.; Rao, V. S. N. *Phytomedicine* **1998**, *5*, 115-9.
- (197) Blanco-Colio, L. M.; Tunon, J.; Martin-Ventura, J. L.; Egido, J. *Kidney Int.* **2003**, *63*, 12-23.
- (198) Erdmann, K.; Berndt, G.; Grosser, N.; Schroder, H. *Naunyn-Schmiedebergs Archives of Pharmacology* **2004**, *369*, R14-R.
- (199) Grosser, N.; Oberle, S.; Berndt, G.; Erdmann, K.; Hemmerle, A.; Schroder, H. *Biochem. Biophys. Res. Commun.* **2004**, *314*, 351-5.
- (200) Chowdhury, S. A.; Kishino, K.; Satoh, R.; Hashimoto, K.; Kikuchi, H.; Nishikawa, H.; Shirataka, Y.; Sakagami, H. *Anticancer Res.* **2005**, *25*, 2055-63.
- (201) Jahangir, T.; Khan, T. H.; Prasad, L.; Sultana, S. *Redox Rep.* **2005**, *10*, 303-10.
- (202) Soria-Mercado, I. E.; Prieto-Davo, A.; Jensen, P. R.; Fenical, W. *J. Nat. Prod.* **2005**, *68*, 904-10.
- (203) Zhou, Y. D.; Kim, Y. P.; Mohammed, K. A.; Jones, D. K.; Muhammad, I.; Dunbar, D. C.; Nagle, D. G. *J. Nat. Prod.* **2005**, *68*, 947-50.
- (204) Boucher, K.; Chad, S. S.; Sharma, P.; Hauschka, P. V.; Solomon, K. R. *Microvasc. Res.* **2006**, *71*, 91-102.
- (205) Jahangir, T.; Khan, T. H.; Prasad, L.; Sultana, S. *Hum. Exp. Toxicol.* **2006**, *25*, 235-42.
- (206) Kumano, T.; Richard, S. B.; Noel, J. P.; Nishiyama, M.; Kuzuyama, T. *Bioorg. Med. Chem.* **2008**, *16*, 8117-26.
- (207) Rabindran, S. K.; He, H. Y.; Singh, M.; Brown, E.; Collins, K. I.; Annable, T.; Greenberger, L. M. *Cancer Res.* **1998**, *58*, 5850-8.
- (208) Rabindran, S. K.; Ross, D. D.; Doyle, L. A.; Yang, W. D.; Greenberger, L. M. *Cancer Res.* **2000**, *60*, 47-50.
- (209) van Loevezijn, A.; Allen, J. D.; Schinkel, A. H.; Koomen, G. J. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 29-32.
- (210) Allen, J. D.; van Loevezijn, A.; Lakhai, J. M.; van der Valk, M.; van Tellingen, O.; Reid, G.; Schellens, J. H. M.; Koomen, G. J.; Schinkel, A. H. *Mol. Cancer Ther.* **2002**, *1*, 417-25.
- (211) Ma, J. C.; Dougherty, D. A. *Chem. Rev. (Washington, DC, U. S.)* **1997**, *97*, 1303-24.
- (212) Dougherty, D. A. *J. Nutr.* **2007**, *137*, 1504S-8S.

- (213) Zacharias, N.; Dougherty, D. A. *Trends Pharmacol. Sci.* **2002**, *23*, 281-7.
- (214) Gallivan, J. P.; Dougherty, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 870-4.
- (215) Gallivan, J. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9459-64.
- (216) Dougherty, D. A. *Science* **1996**, *271*, 163-8.
- (217) Kumpf, R. A.; Dougherty, D. A. *Science* **1993**, *261*, 1708-10.
- (218) Mccurdy, A.; Jimenez, L.; Stauffer, D. A.; Dougherty, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 10314-21.

BIOGRAPHICAL SKETCH

Yang, Yue was born in Beijing, China in the end of 1979. He graduated from University of Science and Technology of China in 2003 with a Bachelor of Science degree major in chemical physics. After college, he worked in a lab in the Technique Institute of Physics and Chemistry, the Chinese Academy of Science of a couple of years before coming to University of Florida and joining the Kenneth Merz group. During his five years Ph.D. career, he has been focusing on the modeling and simulation of important enzymes and biological systems.