8-2016

Molecular Dynamics Simulations of Enzymes with Quantum Mechanical/Molecular Mechanical Potentials

Yufei Yue
University of Tennessee, Knoxville, yyue@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Yufei Yue entitled "Molecular Dynamics Simulations of Enzymes with Quantum Mechanical/Molecular Mechanical Potentials." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biochemistry and Cellular and Molecular Biology.

Hong Guo, Major Professor

We have read this dissertation and recommend its acceptance:

Albrecht von Arnim, Jerome Baudry, Chunlei Su

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Molecular Dynamics Simulations of Enzymes with Quantum Mechanical/Molecular Mechanical Potentials

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Yufei Yue

August 2016

DEDICATION

To my parents

Xiangyu Zhang and Rulin Yue

To my grandparents

Tianju Lei, Fenghe Yue and Qifeng Long

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my advisor, Dr. Hong Guo, who has the attitude and the substance of a genius, for his professional advice, continuous encouragement and support for my academic study and career planning in the past 5 years. Beyond that, I am feeling very fortunate to have worked with someone who could be your lifelong mentor.

I would like to sincerely thank my committee members, Dr. Albrecht von Arnim, Dr. Jerome Baudry and Dr. Chunlei Su for their valuable suggestions during my course of study and completion of the dissertation. I also want to thank Dr. Cynthia Peterson and Dr. Feng Cheng who once served in my committee.

I want to thank all the members of Dr. Hong Guo's lab, including Dr. Qin Xu, Dr. Jianzhuang Yao, Dr. Yuzhuo Chu, Dr. Haobo Guo, Dr. Peng Lian and Dr. Ping Qian for their technical help and meaningful discussions, and our collaborator Dr. Xiaohan Yang from Oak Ridge National Lab. Completion of this dissertation is impossible without their support.

Finally, I would like to send my gratitude to my parents Xiangyu Zhang and Rulin Yue. They are always encouraging and supporting me with their best wishes.

ABSTRACT

S-adenosyl methionine (SAM) dependent methylation process is universally found in all branches of life. It has important implications in mammalian pathogenesis and plant metabolism. The methyl transfer is normally catalyzed by SAM-dependent methyltransferases(MTases). Two MTases are studied in this dissertation: the 1,7-dimethylxanthine methyltransferase (DXMT) which involve in plant caffeine biosynthesis, and the protein arginine methyltransferase 5(PRMT5) that participates in eukaryotic posttranslational modification. The late phase of caffeine biosynthesis starts from the substrate xanthosine and ends with the product caffeine, with theobromine as an intermediate product. DXMT is a key enzyme in this process and catalyzes two methylation steps: 1)methylation of 7-methylxanthine to form theobromine; 2)methylation of theobromine to form caffeine. The catalytic mechanism and product promiscuity of DXMT is intriguing. In Chapter 1, the quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) and free energy simulations were performed to explain the dual catalytic roles of DXMT. Simulation results show that a histidine residue may act as a general base catalyst during methylations. PRMTs can work as modifiers for histones and methylate the substrate arginine, thus interfering with histone code orchestration. The product specificity of PRMTs refers to their ability to produce either symmetric di-methylarginine(SDMA), asymmetric di-methylarginine(ADMA) or mono-methylarginine(MMA). Understanding the product specificity of PRMTs is important since different methylations may cause distinctive, even inverse

biological consequences. PRMT5 produces SDMA, as compared to PRMT1 and PRMT3 that produce ADMA. In Chapter 2, simulations of PRMT5 have drawn a theoretical insight into the catalytic difference between SDMA and ADMA. Neddylation is a type of eukaryotic Ubiquitin-like (UBL) protein modification that is essential in cell division and development. Unlike ubiquitin and other small ubiquitin-like modifiers which target variety of protein substrates, the UBL NEDD8 is highly selective on modifying cullin proteins and contributes to 10% ~20% of all cellular ubiquitination and ubiquitination-like modification. In Chapter 3, the crystal structure of a trapped E3-E2~NEDD8-CUL1 intermediate was used for modeling, and simulations were applied to investigate the catalytic mechanism of NEDD8 transfer from E2 to the substrate. Some important insights were observed that may be used to understand the functional properties of the enzyme.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Caffeine Biosynthesis and Dimethylxanthine Methyltransferase

As a well-known plant alkaloid, caffeine is synthesized in the seeds, leaves and fruits of a variety of flowering plants, including three eudicots of great economical value: coffea (*Coffea arabica* and *Coffea canephora*), cacao (*Theobroma cacao*) and tea plant (*Camellia sinensis*).[1] It is a plant secondary metabolite originally working as an insect repellent[2,3] in leaves, and as inhibitor to seed germination of competing species in fruits and seeds[4]. But it is more recognized as a human central nervous system stimulant, widely used in in the beverage and pharmaceutical industries. The late stage of caffeine biosynthesis involves three *S*-adenosyl-L-methionine (SAM) dependent N-methyl transfer processes and one ribose removal step, converting the initial substrate of xanthosine (XR) to the final product of caffeine, with 7-methylxanthine (7mX) and theobromine (Tb) as two intermediate successive products.[1] Eleven enzymes participating in this pathway have been isolated and biochemically identified as *N*-methyltransferase (NMT).[5,6,7,8,9] Comparative genomic analyses suggests NMTs are highly conserved in eudicots.[10]

Only two NMTs, xanthosine methyltransferases (XMT) and 1,7-dimethylxanthine methyltransferase(DXMT) from *Coffea canephora* ‘robusta’, were structurally characterized[11]. NMTs share a conserved domain, from the wider B’ motif methyltransferase family, which functions in the SAM-dependent methylation.[9] A

structurally characterized salicylic acid *O*-methyltransferase, SAMT, also belongs to the same family[12]. The overall structure of DXMT is similar to SAMT[11]. Yao et al. performed hybrid quantum mechanical/molecular mechanical (QM/MM) simulations to study the methyl transfer process catalyzed by SAMT and gained some key insights into the catalysis mechanism not experimentally observable[13]. Moreover, computational approaches appear to be well suited for the investigations of enzyme-catalyzed methyl transfer processes, based on a good correlation between enzymes' activities(k_{cat}) observed experimentally and the results (free energy barriers) of the computer simulations in previous research[14,15,16].

The substrate recognition and catalytic efficiency of DXMT appears to have subtle and promiscuous characteristics. DXMT shows dual activity toward 7mX and Tb, but not toward XR[11]. One histidine residue, His160, is conserved in the theobromine synthases and caffeine synthases, and is thought as having a catalytic role[11,17]. It is thus of great interest to apply computer simulations to study the trivial catalytic mechanism of DXMT. In Chapter 1, quantum mechanical /molecular mechanical (QM/MM) MD and free energy simulations were performed to illustrate the catalytic mechanism of DXMT and to explain experimental observations concerning the activity of this enzyme. It is proposed that His160 may act as a general base/acid catalyst during the enzyme-catalyzed methyl transfer to both 7mX and Tb.

Histone Code and Protein Arginine Methyltransferase

Methylation of histone on its arginine side chain during post-translation is an epigenetic mark that plays a vital role in variety of cell function.[18] Arginine is unique with its guanidino group which contains five potential hydrogen bond donors that if well positioned may interact with biological hydrogen bond acceptors. In protein-DNA and protein-RNA complexes, arginine residues frequently work as hydrogen bond donors to backbone phosphate groups and to A, T, G bases[19,20]. And the salt bridge interactions between arginine and the two H-bonds of aspartate/glutamate is especially stable in proteins.[21,22] Methyl addition to an arginine residue not only changes its shape, but also removes a potential hydrogen bond donor.[18] Formation of some protein quaternary structures could thus be inhibited because of the disturbance of arginine methylation.[23,24] Methylation of arginine residues might also change their affinity to aromatic rings in cation-pi interactions[25,26] Therefore methylation on protein arginine(s) residues could modulate protein binding interactions, resulting in regulation of physiological functions.

Protein arginine methyltransferases (PRMTs) catalyze the transfer of the methyl group from S-adenosyl-L-methionine (AdoMet) to the guanidino nitrogen on arginine side chain. Histone and many other substrate proteins with target arginine could thus be modified.[18] Depending on types of PRMTs (I, II or III), there are three known methylation products, asymmetric dimethylarginine (ADMA, Type I), symmetric dimethylarginine (SDMA, Type II) and monomethylarginine (MMA, Type

III). The Type II PRMT5 catalyzes the transfers of methyl groups to the two ω-guanidino nitrogen atoms of arginine in the target protein, resulting in the ω-N^G, N'^G symmetric dimethylation of arginine (SDMA).[27] Symmetric dimethylation of those proteins profoundly impacts many biological processes including epigenetic control of gene expression, splicing regulation, DNA damage response and germ cell development and pluripotency.[28] Interestingly, the Type II PRMT5s often share the common recognition sequences with the Type I PRMTs such as PRMT1 and PRMT3 which add two methyl groups to the same ω-guanidino nitrogen atom (ADMA).[28] And the same target arginine can be either symmetrically or asymmetrically dimethylated. However, SDMA and ADMA are isomeric protein posttranslational modifications with distinct and sometimes inverse biological effects. One common case is the methylation of Arginine 3 on histone H4 (H4R3): symmetric dimethylation of H4R3 could repress gene expression[29,30,31], while asymmetric dimethylation of H4R3 is associated with gene activation[32,33].

The crystal structures of several PRMTs have been identified, which include PRMT1, PRMT3, PRMT6 of Type I PRMTs, PRMT5 of Type II PRMTs and PRMT7 of Type III PRMTs. And computational simulations have investigated the catalytic mechanism of PRMT1 and PRMT3.[34,35] Important questions concerning the product specificity of Type I ADMA were addressed. In Chapter 2, the methylation reaction catalyzed by Type II PRMT5 was investigated by use of quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) and free

energy simulations. The simulations have identified some important interactions and proton transfers involving the active site residues. Similar to Type I PRMT1 and PRMT3, the conserved Glu435 residue in PRMT5 is suggested to function as general base catalyst during the catalysis. And the methyl transfer and proton transfer seem to be somehow concerted. However, unlike Type I PRMTs which energetically favor $N_{\eta 2}$ atom of arginine as target for both the 1st and 2nd methylations[34,35], PRMT5 shows an energetic preference of $N_{\eta 1}$ over $N_{\eta 2}$ as the target for the first methylation and subsequently targets $N_{\eta 2}$ for the second methylation. My research supplies a better mechanistic understanding of symmetric dimethylation and suggests a different mechanism from asymmetric dimethylation.

CULLIN-RING E3 Catalyzed Transfer of Ubiquitin-Like Protein NEDD8

Neddylation is a type of ubiquitination-like post-translational modification through which a ubiquitin-like(UBL) protein called NEDD8 is transferred to its highly selective substrate, cullin proteins(CULs) on a conserved lysine.[36] Neddylation takes approximately 10% - 20% of ubiquitination and ubiquitination-like modifications in eukaryotic proteome.[37,38,39] It participates in regulation of important biological processes such as signal transduction[40], cell division[41,42] and development[43].

Like ubiquitin(UB) ligation, neddylation involves three collaborative enzymes, E1, E2 (UBC12) and RING E3(RBX1) enzymes. E1 activates NEDD8 and loads it onto a conserved cysteine from E2, making a thioester bond to covalently link E2 and NEDD8.[44] RING E3 then recruits substrate CULs and the E2~NEDD8 intermediate and completes the transfer of NEDD8 to the substrate CULs.[36] Recently the active configuration mimicking NEDD8 ligation to the substrate CUL1 was achieved through crystallization of the RBX1(E3)-UBC12(E2)~NEDD8-CUL1(Substrate)-DCN1(Co-E3) complex.[36] Similar to other processes of UB and UBL ligation to substrate proteins[45,46], the RING-box protein 1(RBX1) utilizes its N-terminal domain(NTD) to heterodimerize with the substrate CUL1, and then binds and activates the UBC12~NEDD8 intermediate[36]. Different from some other UB and UBL ligation, RBX1 also needs assistance of a Co-E3, DCN1, which binds CUL1 and UBC12's NTD, to form a catalytic quaternary complex.[47,48] Nevertheless, RBX1 and DCN1 are away from the acceptor lysine (K720) on substrate CUL1 where ligation takes place.[36]

To obtain a crystal complex of E3-E2~NEDD8-CUL1-DCN1, three residue mutations are introduced: K720R from CUL1, C111S from E2 and N103S from E2.[36] The substitution of K720 by an arginine makes the receptor residue less active. The substitution of C111 by serine makes a more stable oxyester bond between the hydroxyl side chain of serine and the backbone carbonyl group of G76 from C-terminal of NEDD8 (as compared to a highly activated thioester bond

formed between the thiol side chain of cysteine and backbone carbonyl group of G76). N103 is a highly conserved residue that is believed to play a catalytic role of stabilizing the oxyanion intermediate formed during formation of isopeptide bond between K720 and G76.[49,50] Altogether, those three mutations help avoid NEDD8 transfer, making crystallization of the NEDD8 ligation intermediate possible. Although the structural analysis combined with mutagenesis have provided important insights into the NEDD8 ligation[36,48,51], the dynamic motions of the active site residues and their catalytic roles in the detailed ligation process are still not well understood, and these can be further studied by QM/MM MD and free energy simulations.

Previous computational study of ubiquitination process involving E3-E2~UB-SUMO2 has led to the proposal of two important events during the ubiquitin transfer from E2 to substrate SUMO2, including deprotonation of substrate lysine and the nucleophilic attack from the receptor lysine to the thioester bond.[52] According to this study, the proton transfer process during the deprotonation of substrate lysine by a general base, D117 on E2, has almost no energy barrier. The deprotonated lysine then approaches the thioester bond via a conformational change, and then initiates a nucleophilic attack on the thioester bond between UB and E2 to form an oxyanion intermediate. Finally, the nucleophilic elimination results in the cleavage of the thioester bond and subsequently formation of an isopeptide bond between UB and substrate SUMO2. In Chapter 3 we examined

the neddylation catalysis based on the crystal structure of E3-E2~NEDD8-CUL1-CoE3 complex[36] by use of QM/MM MD and free energy simulations. The results show that either E117 or D143 from E2 could be the general base catalyst for the deprotonation of substrate K720. Furthermore, the formation of the isopeptide between K720 side chain and G76, and the breaking of thioester bond seem to be concerted without forming an oxyanion intermediate. Moreover, it is shown that K720 could potentially work as a general acid catalyst to promote the departure of C111 from E2.

Computational study of enzymology through QM/MM Methods

Enzymes are essential for life due to their extraordinary catalytic efficiency and high specificity for biochemical reactions. Understanding working mechanisms of enzymes could not only provide important insights into fundamental problems in biology such as evolution and biosynthesis, but also contribute to bioengineering and rational drug design.[53] Experimental studies including enzyme kinetics, mutagenesis and structural biology have provided rich information on enzymes and reactions they catalyze. Nevertheless, enzymatic reactions usually occur in a transient manner and might be difficult to study with experimental techniques[54,55]. Computer simulations have demonstrated the capability of studying enzymatic dynamics and reactions at atomic detail and are very useful in interpreting, complementing and expanding experimental results.[56] One important case is that the simulations can be applied to determine the activation energy barriers for the

enzyme-catalyzed reactions and geometries of transition state structures, which are the keys to understand the reactivity but normally very difficult to obtain through experiments.

Enzyme's behaviors could be described by models generated through fundamental laws of quantum mechanics (QM) and classical mechanics integrated with statistical mechanics.[55,57] The QM methods, including density function theory(DFT),[58,59] now have the capacity to study chemical reactions with a decent accuracy for systems made up of with hundreds of atoms. However, treating the whole biological macromolecules containing thousands of atoms or more by QM would be extremely time consuming. The MM approaches based on classical mechanics have been well established and can be used to determine conformational energetics and non-bonded interactions in systems with up to hundreds of thousands of atoms.[57,60,61,62,63] The QM/MM[64,65,66,67,68] methods combine the accuracy of QM and speediness of MM, and they are widely used to model enzymatic reactions involving breaking and making of covalent bonds. The active sites of enzyme complexes are often treated with QM to encompass any important electronic structure changes such as the redistribution of electron clouds, and the remainder of the system are modeled with MM approaches which employ appropriate structural constraints, electrostatic and van der Waals interactions.[69]

Molecular Dynamic Simulations

Enzymes work like a dynamic machine. They are in motions at multiple timescales, which play important roles for their biological functions.[70,71] Molecular dynamics(MD) simulations are used to provide theoretical understanding of enzyme motions. A few force fields have been well developed in the past thirty years, including AMBER (Assisted Model Building and Energy Refinement)[72], CHARMM (Chemistry at HARvard Macromolecular Mechanics)[61], GROMOS (GROningen MOlecular Simulation)[62], and OPLS (Optimized Potential for Liquid Simulations)[73]. These empirical approaches have been applied to simulate biological macromolecules such as proteins and nucleic acids. Here in this dissertation, the all-atom force field CHARMM27[61] was adopted for MM and the modified TIP3P model[74] was used for solvent water molecules. For the efficiency of calculation on the solvents, three point charges were employed to describe water molecules while ignoring the internal interactions in this approach.[74] All atoms were treated explicitly.

QM approaches involve the solvation of the Schrodinger equation.[75] They are critical for determining electronic structure changes and electron redistributions caused by the bond-breaking and bond-making events. High level *ab initio* approaches such as MP2 and certain methods based on density functional theory (DFT) have been developed for the QM calculations.[76] B3LYP (a popular hybrid-function of DFT), for instance, works well for a variety of chemical systems.[77]

Nonetheless, the *ab initio* methods generally are very computationally expensive. Therefore, some semi-empirical methods, such as MNDO (Modified Neglect of Diatomic Overlap), AM1 (Austin Model 1), PM3 (Parameterized Model 3) and SCC-DFTB (Self-Consistent Charge Density Functional Tight Binding), have been widely used as cost-efficient solutions for QM simulations.[78,79] Among those approaches, SCC-DFTB method is derived as a second-order expansion from the DFT total energy function with respect to charge density variation relative to a chosen reference.[79] The energies calculated by SCC-DFTB can sometimes be comparable to the high-level B3LYP for systems with or without MM potentials. SCC-DFTB appears to yield reasonably good results for biological systems.[80] The integration of SCC-DFTB with CHARMM has led to the studies of a number of enzymes.[13,14,15,16,81] As such, the hybrid SCC-DFTB/CHARMM approaches are employed for our simulations.

Simulations carried out in this dissertation have to deal with the issue where the QM/MM boundary falls onto a covalent bond. For instance, we want to have the side chain of a catalytic residue being located in the QM region and its backbone in the MM region. The problem is how to partition a MM atom and a QM atom linked by a covalent bond. Methods such as link atom[65], delocalized Gaussian MM charges(DGMM) and generalized hybrid orbital (GHO) were mainly developed to deal with this problem. The link atom method, which utilizes a hydrogen to

saturate both ends of QM and MM regions, has been used in the SCC-DFTB/CHARMM simulations of this dissertation.

For enzymes, the catalytic activity may be linked mainly to the structural and dynamic features in relatively localized regions including, for instance, the active sites plus nearby residues/solvent molecules and substrates[82]. Thus the standard MD simulations of the entire solvated systems might be rather inefficient in studying enzymes. Stochastic boundary conditions (SBC) have been developed for taking the advantage of the spatial localization by replacing atoms beyond a given distance by thermal bath model[83]. A system with SBC is divided into two parts: the reaction zone and the reservoir zone (thermal bath model). The reaction zone is further partitioned by the reaction region and the buffer region, where the former is treated by molecular dynamics and the latter is treated by Langevin dynamics. The atoms that belong to the reservoir region are eliminated and those in buffer region work to equilibrate the forces from the reservoir region. In this dissertation, modeling with SBC chose an atom in the active site as reference center and set radii for reaction region, buffer region and reservoir zone, respectively.

QM/MM Free Energy Simulations

With the methodology adopted for MD simulations as discussed above, the free energy simulations further take the advantage of umbrella sampling[84] method along with the Weighted Histogram Analysis Method (WHAM)[85] to determine the

potential of mean force (PMF) as a function of predefined reaction coordinate(s). The reaction coordinate(s) were chosen based on how the catalytic process is described. For instance, they can be a structural parameter that changes during the chemical transformation, such as distance, bond length and dihedral angle. They also can be a linear combination of those structural parameters.

The transition state configurations are difficult to obtain during regular simulations. The umbrella sampling utilizes a harmonic potential function to expand the exploration for those rare configurations on the coordinate space. The function is described as,

$$U = \frac{\mathbf{1}}{E} \times K \times (r - r_{\mathbf{0}})^{E} \ ,$$

Where K is the force constant, r is the reaction coordinate, r_0 is the origin of harmonic function on reaction coordinate, and E is set as 2 for default. The WHAM program calculates the probability density function of the configurations along the reaction coordinate(s) in a histogram and then estimates the free energy change on the coordinate(s).[85,86] Except for regular WHAM that was applied to estimate the free energy change on one reaction coordinate, two-dimensional(2D) WHAM was employed in this dissertation to estimate the free energy change when two dimensional reaction coordinates were considered. The 2D free energy profiles computed by 2D-WHAM is good at implicating the synchronicity of multiple chemical steps.

References

1. Ashihara, H.; Sano, H.; Crozier, A., Caffeine and related purine alkaloids: Biosynthesis, catabolism, function and genetic engineering. *Phytochemistry* **2008,** *69* (4), 841-856.

2. Uefuji, H.; Tatsumi, Y.; Morimoto, M.; Kaothien-Nakayama, P.; Ogita, S.; Sano, H., Caffeine production in tobacco plants by simultaneous expression of three coffee N-methyltrasferases and its potential as a pest repellant. *Plant Mol Biol* **2005,** *59* (2), 221-227.

3. Nathanson, J. A., Caffeine and Related Methylxanthines - Possible Naturally-Occurring Pesticides. *Science* **1984,** *226* (4671), 184-187.

4. Pacheco, A.; Pohlan, J.; Schulz, M., Allelopathic effects of aromatic species intercropped with coffee: Investigation of their growth stimulation capacity and potential of caffeine uptake in Puebla, Mexico. *Allelopathy J* **2008,** *21* (1), 39-56.

5. Kato, M.; Mizuno, K.; Crozier, A.; Fujimura, T.; Ashihara, H., Plant biotechnology - Caffeine synthase gene from tea leaves. *Nature* **2000,** *406* (6799), 956-957.

6. Ogawa, M.; Herai, Y.; Koizumi, N.; Kusano, T.; Sano, H., 7-Methylxanthine methyltransferase of coffee plants - Gene isolation and enzymatic properties. *J Biol Chem* **2001,** *276* (11), 8213-8218.

7. Mizuno, K.; Kato, M.; Irino, F.; Yoneyama, N.; Fujimura, T.; Ashihara, H., The first committed step reaction of caffeine biosynthesis: 7-methylxanthosine

synthase is closely homologous to caffeine synthases in coffee (Coffea arabica L.). *Febs Lett* **2003,** *547* (1-3), 56-60.

8. Mizuno, K.; Okuda, A.; Kato, M.; Yoneyama, N.; Tanaka, H.; Ashihara, H.; Fujimura, T., Isolation of a new dual-functional caffeine synthase gene encoding an enzyme for the conversion of 7-methylxanthine to caffeine from coffee (Coffea arabica L.). *Febs Lett* **2003,** *534* (1-3), 75-81.

9. Kato, M.; Mizuno, K., Caffeine synthase and related methyltransferases in plants. *Front Biosci* **2004,** *9*, 1833-1842.

10. Denoeud, F.; Carretero-Paulet, L.; Dereeper, A.; Droc, G.; Guyot, R.; Pietrella, M.; Zheng, C. F.; Alberti, A.; Anthony, F.; Aprea, G.; Aury, J. M.; Bento, P.; Bernard, M.; Bocs, S.; Campa, C.; Cenci, A.; Combes, M. C.; Crouzillat, D.; Da Silva, C.; Daddiego, L.; De Bellis, F.; Dussert, S.; Garsmeur, O.; Gayraud, T.; Guignon, V.; Jahn, K.; Jamilloux, V.; Joet, T.; Labadie, K.; Lan, T. Y.; Leclercq, J.; Lepelley, M.; Leroy, T.; Li, L. T.; Librado, P.; Lopez, L.; Munoz, A.; Noel, B.; Pallavicini, A.; Perrotta, G.; Poncet, V.; Pot, D.; Priyono; Rigoreau, M.; Rouard, M.; Rozas, J.; Tranchant-Dubreuil, C.; VanBuren, R.; Zhang, Q.; Andrade, A. C.; Argout, X.; Bertrand, B.; de Kochko, A.; Graziosi, G.; Henry, R. J.; Jayarama; Ming, R.; Nagai, C.; Rounsley, S.; Sankoff, D.; Giuliano, G.; Albert, V. A.; Wincker, P.; Lashermes, P., The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* **2014,** *345* (6201), 1181-1184.

11. McCarthy, A. A.; McCarthy, J. G., The structure of two N-methyltransferases from the caffeine biosynthetic pathway. *Plant Physiol* **2007,** *144* (2), 879-889.

12. Zubieta, C.; Ross, J. R.; Koscheski, P.; Yang, Y.; Pichersky, E.; Noel, J. P., Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family. *Plant Cell* **2003,** *15* (8), 1704-1716.

13. Yao, J. Z.; Xu, Q.; Chen, F.; Guo, H., QM/MM Free Energy Simulations of Salicylic Acid Methyltransferase: Effects of Stabilization of TS-like Structures on Substrate Specificity. *J Phys Chem B* **2011,** *115* (2), 389-396.

14. Guo, H. B.; Guo, H., Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity. *P Natl Acad Sci USA* **2007,** *104* (21), 8797-8802.

15. Xu, Q.; Chu, Y. Z.; Guo, H. B.; Smith, J. C.; Guo, H., Energy Triplets for Writing Epigenetic Marks: Insights from QM/MM Free-Energy Simulations of Protein Lysine Methyltransferases. *Chem-Eur J* **2009,** *15* (46), 12596-12599.

16. Chu, Y. Z.; Xu, Q.; Guo, H., Understanding Energetic Origins of Product Specificity of SET8 from QM/MM Free Energy Simulations: What Causes the Stop of Methyl Addition during Histone Lysine Methylation? *J Chem Theory Comput* **2010,** *6* (4), 1380-1389.

17. Mizuno, K.; Kurosawa, S.; Yoshizawa, Y.; Kato, M., Essential Region for 3-N Methylation in N-Methyltransferases Involved in Caffeine Biosynthesis. *Z Naturforsch C* **2010,** *65* (3-4), 257-265.

18. Bedford, M. T.; Clarke, S. G., Protein Arginine Methylation in Mammals: Who, What, and Why. *Mol Cell* **2009,** *33* (1), 1-13.

19. Calnan, B. J.; Tidor, B.; Biancalana, S.; Hudson, D.; Frankel, A. D., Arginine-Mediated Rna Recognition - the Arginine Fork. *Science* **1991,** *252* (5009), 1167-1171.

20. Luscombe, N. M.; Laskowski, R. A.; Thornton, J. M., Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res* **2001,** *29* (13), 2860-2874.

21. Mitchell, J. B. O.; Thornton, J. M.; Singh, J.; Price, S. L., Towards an Understanding of the Arginine Aspartate Interaction. *J Mol Biol* **1992,** *226* (1), 251-262.

22. Melo, A.; Ramos, M. J.; Floriano, W. B.; Gomes, J. A. N. F.; Leao, J. F. R.; Magalhaes, A. L.; Maigret, B.; Nascimento, M. C.; Reuter, N., Theoretical study of arginine-carboxylate interactions. *J Mol Struc-Theochem* **1999,** *463* (1-2), 81-90.

23. Bedford, M. T.; Richard, S., Arginine methylation: An emerging regulator of protein function. *Mol Cell* **2005,** *18* (3), 263-272.

24. Bedford, M. T., Arginine methylation at a glance. *J Cell Sci* **2007,** *120* (24), 4243-4246.

25. Sprangers, R.; Groves, M. R.; Sinning, L.; Sattler, M., High-resolution X-ray and NMR structures of the SMN tudor domain: Conformational variation in

the binding site for symmetrically dimethylated arginine residues. *J Mol Biol* **2003,** *327* (2), 507-520.

26. Hughes, R. M.; Waters, M. L., Arginine methylation in a beta-hairpin peptide: Implications for Arg-pi interactions, Delta Cp degrees, and the cold denatured state. *J Am Chem Soc* **2006,** *128* (39), 12735-12742.

27. Wang, M.; Xu, R. M.; Thompson, P. R., Substrate Specificity, Processivity, and Kinetic Mechanism of Protein Arginine Methyltransferase 5. *Biochemistry-Us* **2013,** *52* (32), 5430-5440.

28. Sun, L. T.; Wang, M. Z.; Lv, Z. Y.; Yang, N.; Liu, Y. F.; Bao, S. L.; Gong, W. M.; Xu, R. M., Structural insights into protein arginine symmetric dimethylation by PRMT5. *P Natl Acad Sci USA* **2011,** *108* (51), 20538-20543.

29. Pal, S.; Vishwanath, S. N.; Erdjument-Bromage, H.; Tempst, P.; Sif, S., Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes. *Mol Cell Biol* **2004,** *24* (21), 9630-9645.

30. Wang, X.; Zhang, Y.; Ma, Q. B.; Zhang, Z. L.; Xue, Y. B.; Bao, S. L.; Chong, K., SKB1-mediated symmetric dimethylation of histone H4R3 controls flowering time in Arabidopsis. *Embo J* **2007,** *26* (7), 1934-1941.

31. Zhao, Q.; Rank, G.; Tan, Y. T.; Li, H. T.; Moritz, R. L.; Simpson, R. J.; Cerruti, L.; Curtis, D. J.; Patel, D. J.; Allis, C. D.; Cunningham, J. M.; Jane, S. M., PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling

histone and DNA methylation in gene silencing. *Nat Struct Mol Biol* **2009,** *16* (3), 304-311.

32. Wang, H. B.; Huang, Z. Q.; Xia, L.; Feng, Q.; Erdjument-Bromage, H.; Strahl, B. D.; Briggs, S. D.; Allis, C. D.; Wong, J. M.; Tempst, P.; Zhang, Y., Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Science* **2001,** *293* (5531), 853-857.

33. Fuchs, S. M.; Krajewski, K.; Baker, R. W.; Miller, V. L.; Strahl, B. D., Influence of Combinatorial Histone Modifications on Antibody and Effector Protein Recognition. *Curr Biol* **2011,** *21* (1), 53-58.

34. Chu, Y. Z.; Li, G. H.; Guo, H., QM/MM MD and free energy simulations of the methylation reactions catalyzed by protein arginine methyltransferase PRMT3. *Can J Chem* **2013,** *91* (7), 605-612.

35. Zhang, R. H.; Li, X.; Liang, Z. J.; Zhu, K. K.; Lu, J. Y.; Kong, X. Q.; Ouyang, S. S.; Li, L.; Zheng, Y. G.; Luo, C., Theoretical Insights into Catalytic Mechanism of Protein Arginine Methyltransferase 1. *Plos One* **2013,** *8* (8).

36. Scott, D. C.; Sviderskiy, V. O.; Monda, J. K.; Lydeard, J. R.; Cho, S. E.; Harper, J. W.; Schulman, B. A., Structure of a RING E3 trapped in action reveals ligation mechanism for the ubiquitin-like protein NEDD8. *Cell* **2014,** *157* (7), 1671-84.

37. Soucy, T. A.; Smith, P. G.; Milhollen, M. A.; Berger, A. J.; Gavin, J. M.; Adhikari, S.; Brownell, J. E.; Burke, K. E.; Cardin, D. P.; Critchley, S.; Cullis, C. A.; Doucette, A.; Garnsey, J. J.; Gaulin, J. L.; Gershman, R. E.; Lublinsky, A. R.;

McDonald, A.; Mizutani, H.; Narayanan, U.; Olhava, E. J.; Peluso, S.; Rezaei, M.; Sintchak, M. D.; Talreja, T.; Thomas, M. P.; Traore, T.; Vyskocil, S.; Weatherhead, G. S.; Yu, J.; Zhang, J.; Dick, L. R.; Claiborne, C. F.; Rolfe, M.; Bolen, J. B.; Langston, S. P., An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. *Nature* **2009,** *458* (7239), 732-6.

38. Osaka, F.; Saeki, M.; Katayama, S.; Aida, N.; Toh-e, A.; Kominami, K.; Toda, T.; Suzuki, T.; Chiba, T.; Tanaka, K.; Kato, S., Covalent modifier NEDD8 is essential for SCF ubiquitin-ligase in fission yeast. *Embo J* **2000,** *19* (13), 3475-3484.

39. Tateishi, K.; Omata, M.; Tanaka, K.; Chiba, T., The NEDD8 system is essential for cell cycle progression and morphogenetic pathway in mice. *The Journal of cell biology* **2001,** *155* (4), 571-9.

40. Read, M. A.; Brownell, J. E.; Gladysheva, T. B.; Hottelet, M.; Parent, L. A.; Coggins, M. B.; Pierce, J. W.; Podust, V. N.; Luo, R. S.; Chau, V.; Palombella, V. J., Nedd8 modification of cul-1 activates SCF(beta(TrCP))-dependent ubiquitination of IkappaBalpha. *Molecular and cellular biology* **2000,** *20* (7), 2326-33.

41. Liakopoulos, D.; Doenges, G.; Matuschewski, K.; Jentsch, S., A novel protein modification pathway related to the ubiquitin system. *Embo J* **1998,** *17* (8), 2208-2214.

42. Kurz, T.; Pintard, L.; Willis, J. H.; Hamill, D. R.; Gonczy, P.; Peter, M.; Bowerman, B., Cytoskeletal regulation by the Nedd8 ubiquitin-like protein modification pathway. *Science* **2002,** *295* (5558), 1294-8.

43. Pozo, J. C.; Timpte, C.; Tan, S.; Callis, J.; Estelle, M., The ubiquitin-related protein RUB1 and auxin response in Arabidopsis. *Science* **1998,** *280* (5370), 1760-3.

44. Huang, D. T.; Hunt, H. W.; Zhuang, M.; Ohi, M. D.; Holton, J. M.; Schulman, B. A., Basis for a ubiquitin-like protein thioester switch toggling E1-E2 affinity. *Nature* **2007,** *445* (7126), 394-8.

45. Gareau, J. R.; Reverter, D.; Lima, C. D., Determinants of small ubiquitin-like modifier 1 (SUMO1) protein specificity, E3 ligase, and SUMO-RanGAP1 binding activities of nucleoporin RanBP2. *The Journal of biological chemistry* **2012,** *287* (7), 4740-51.

46. Plechanovova, A.; Jaffray, E. G.; Tatham, M. H.; Naismith, J. H.; Hay, R. T., Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **2012,** *489* (7414), 115-20.

47. Kim, A. Y.; Bommelje, C. C.; Lee, B. E.; Yonekawa, Y.; Choi, L.; Morris, L. G.; Huang, G.; Kaufman, A.; Ryan, R. J.; Hao, B.; Ramanathan, Y.; Singh, B., SCCRO (DCUN1D1) is an essential component of the E3 complex for neddylation. *The Journal of biological chemistry* **2008,** *283* (48), 33211-20.

48. Huang, G.; Kaufman, A. J.; Ramanathan, Y.; Singh, B., SCCRO (DCUN1D1) promotes nuclear translocation and assembly of the neddylation E3 complex. *The Journal of biological chemistry* **2011,** *286* (12), 10297-304.

49. Wu, P. Y.; Hanlon, M.; Eddins, M.; Tsui, C.; Rogers, R. S.; Jensen, J. P.; Matunis, M. J.; Weissman, A. M.; Wolberger, C. P.; Pickart, C. M., A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *Embo J* **2003,** *22* (19), 5241-5250.

50. Berndsen, C. E.; Wiener, R.; Yu, I. W.; Ringel, A. E.; Wolberger, C., A conserved asparagine has a structural role in ubiquitin-conjugating enzymes. *Nat Chem Biol* **2013,** *9* (3), 154-156.

51. Calabrese, M. F.; Scott, D. C.; Duda, D. M.; Grace, C. R. R.; Kurinov, I.; Kriwacki, R. W.; Schulman, B. A., A RING E3-substrate complex poised for ubiquitin-like protein transfer: structural insights into cullin-RING ligases. *Nature structural & molecular biology* **2011,** *18* (8), 947-949.

52. Zhen, Y.; Qin, G.; Luo, C.; Jiang, H.; Yu, K.; Chen, G., Exploring the RING-catalyzed ubiquitin transfer mechanism by MD and QM/MM calculations. *PloS one* **2014,** *9* (7), e101663.

53. van der Kamp, M. W.; Mulholland, A. J., Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry-Us* **2013,** *52* (16), 2708-2728.

54. Yang, W.; Gao, Y. Q.; Cui, Q.; Ma, J.; Karplus, M., The missing link between thermodynamics and structure in F1-ATPase. *P Natl Acad Sci USA* **2003,** *100* (3), 874-879.

55. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The Biomolecular Simulation Program. *J Comput Chem* **2009,** *30* (10), 1545-1614.

56. van der Kamp, M. W.; Shaw, K. E.; Woods, C. J.; Mulholland, A. J., Biomolecular simulation and modelling: status, progress and prospects. *J R Soc Interface* **2008,** *5*, S173-S190.

57. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J Comput Chem* **1983,** *4* (2), 187-217.

58. Becke, A. D., Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys Rev A* **1988,** *38* (6), 3098-3100.

59. Parr, R. G.; Yang, W., *Density-functional theory of atoms and molecules*. Oxford University Press ;

Clarendon Press: New York

Oxford England, 1989; p x, 333 p.

60. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput Phys Commun* **1995,** *91* (1-3), 1-41.

61. 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., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **1998,** *102* (18), 3586-3616.

62. 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., The GROMOS biomolecular simulation program package. *J Phys Chem A* **1999,** *103* (19), 3596-3607.

63. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* **2001,** *105* (28), 6474-6487.

64. Warshel, A.; Levitt, M., Theoretical Studies of Enzymic Reactions - Dielectric, Electrostatic and Steric Stabilization of Carbonium-Ion in Reaction of Lysozyme. *J Mol Biol* **1976,** *103* (2), 227-249.

65. Field, M. J.; Bash, P. A.; Karplus, M., A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations. *J Comput Chem* **1990,** *11* (6), 700-733.

66. Gao, J., Combined Qm/Mm Simulation Study of the Claisen Rearrangement of Allyl Vinyl Ether in Aqueous-Solution. *J Am Chem Soc* **1994,** *116* (4), 1563-1564.

67. Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K., ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)(3))(2)+H-2 oxidative addition. *J Phys Chem-Us* **1996,** *100* (50), 19357-19363.

68. Zhang, Y. K.; Lee, T. S.; Yang, W. T., A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J Chem Phys* **1999,** *110* (1), 46-54.

69. Friesner, R. A.; Guallar, V., Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu Rev Phys Chem* **2005,** *56*, 389-427.

70. Henzler-Wildman, K.; Kern, D., Dynamic personalities of proteins. *Nature* **2007,** *450* (7172), 964-72.

71. Zwier, M. C.; Chong, L. T., Reaching biological timescales with all-atom molecular dynamics simulations. *Current opinion in pharmacology* **2010,** *10* (6), 745-52.

72. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *J Am Chem Soc* **1996,** *118* (9), 2309-2309.

73. Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* **1996,** *118* (45), 11225-11236.

74. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **1983,** *79* (2), 926-935.

75. Alhaidari, A. D.; Ismail, M. E. H., Quantum mechanics without potential function. *J Math Phys* **2015,** *56* (7).

76. Gao, J. L.; Truhlar, D. G., Quantum mechanical methods for enzyme kinetics. *Annu Rev Phys Chem* **2002,** *53*, 467-505.

77. Kim, K.; Jordan, K. D., Comparison of Density-Functional and Mp2 Calculations on the Water Monomer and Dimer. *J Phys Chem-Us* **1994,** *98* (40), 10089-10094.

78. Stewart, J. J. P., Special Issue - Mopac - a Semiempirical Molecular-Orbital Program. *J Comput Aid Mol Des* **1990,** *4* (1), 1-45.

79. Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G., Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys Rev B* **1998,** *58* (11), 7260-7268.

80. Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M., A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J Phys Chem B* **2001,** *105* (2), 569-585.

81. Guo, H.; Cui, Q.; Lipscomb, W. N.; Karplus, M., Understanding the role of active-site residues in chorismate mutase catalysis from molecular-dynamics simulations. *Angew Chem Int Edit* **2003,** *42* (13), 1508-1511.

82. Brooks, C. L.; Brunger, A.; Karplus, M., Active-Site Dynamics in Protein Molecules - a Stochastic Boundary Molecular-Dynamics Approach. *Biopolymers* **1985,** *24* (5), 843-865.

83. Brunger, A.; Brooks, C. L.; Karplus, M., Stochastic Boundary-Conditions for Molecular-Dynamics Simulations of St2 Water. *Chem Phys Lett* **1984,** *105* (5), 495-500.

84. Torrie, G. M.; Valleau, J. P., Monte-Carlo Free-Energy Estimates Using Non-Boltzmann Sampling - Application to Subcritical Lennard-Jones Fluid. *Chem Phys Lett* **1974,** *28* (4), 578-581.

85. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J Comput Chem* **1992,** *13* (8), 1011-1021.

86. Kobrak, M. N., Systematic and statistical error in histogram-based free energy calculations. *J Comput Chem* **2003,** *24* (12), 1437-46.

CHAPTER 1 CATALYTIC MECHANISM AND THE ROLE OF KEY RESIDUES IN CAFFEINE BIOSYNTHESIS CATALYZED BY DIMETHYLXANTHINE METHYLTRANSFERASE

A version of this chapter was originally published by Yufei Yue and Hong Guo: Yufei Yue, Hong Guo. “Quantum Mechanical/Molecular Mechanical (QM/MM) Study of Catalytic Mechanism and the Role of Key Residues in Caffeine Biosynthesis Catalyzed by Dimethylxanthine Methyltransferase” *J. Chem. Inf. Model.,* 54 (2014): 593-600. My contributions to this study include: 1) design of experiments; 2) building models and running simulations; 3) analyzing results and preparing the manuscript. Dr. Hong Guo supervised in the experimental design and reviewed the manuscript.

Abstract

The caffeine biosynthetic pathway is of considerable importance for beverage and pharmaceutical industries which produces two blockbuster products, theobromine and caffeine. The major biochemistry in caffeine biosynthesis starts from the initial substrate of xanthosine and ends with the final product caffeine, with theobromine serving as an intermediate. The key enzyme, S-adenosyl-L-methionine (SAM) dependent 1,7-dimethyl-xanthine methyltransferase (DXMT), catalyzes two important methyl transfer steps in caffeine biosynthesis: 1) methylation of N3 on 7-methylxanthine (7mX) to form theobromine (Tb); 2) methylation of N1 on theobromine to form caffeine (Cf). Although DXMT has been structurally characterized recently, our understanding of the detailed catalytic mechanism and role of key catalytic residues are still lacking. In this work, the quantum mechanical/ molecular mechanical (QM/MM) molecular dynamics (MD) and free energy

simulations are performed to elucidate the catalytic mechanism of the enzyme-catalyzed reactions and to explain experimental observations concerning the activity of this enzyme. The roles of certain active-site residues are studied, and it is proposed that based on the results of computer simulations that a histidine residue (His160) at the active site of DXMT may act as a general base/acid catalyst during the methyl transfer process.

Introduction

As one of the best known plant alkaloids, caffeine (1,3,7-trimethylxanthine) is produced in seeds, leaves and fruits of many higher plants, including three very important crops: coffee (*Coffea arabica* and *Coffea canephora*), cacao (*Theobroma cacao*) and tea (*Camellia sinensis*).[1] It is a secondary metabolite in plants and can work as a native insect repellent[2] and inhibitor to the germination of other seeds[3]. However, it is probably better known as a human central nervous system stimulant, widely used in beverage and pharmaceutical industries. The main pathway of caffeine biosynthesis starts from the initial substrate xanthosine(XR) and ends with the final product of caffeine, involving three *S*-adenosyl-L-methionine(SAM) dependent N-methyl transfer processes and one ribose removal step(Figure 1.1).[1] In this biosynthetic pathway, the ribose is removed and 7-N, 3-N and 1-N on the purine alkaloid are methylated(Figure 1.1), producing 7-methylxanthosine(7mXR),7-methylxanthine(7mX) and theobromine (3,7-dimethy-lxanthine, Tb) as intermediates.

Eleven enzymes that participate in this pathway have been isolated so far and identified as *N*-methyltransferase.[4,5,6,7,8,9] These enzymes catalyze the N-methylation of nitrogen-containing compounds and belong to the wider motif B' methyltransferase family, which are plants specific and involve in SAM-dependent substrate methylation.[10] Only two of these enzymes in the pathway, xanthosine methyltransferases (XMT) and 3,7-dimethylxanthine methyltransferase (DXMT), both from *Coffea canephora* 'robusta', have been structurally characterized.[11] XMT is capable of catalyzing the methyl transfer to and, probably, nucleotide cleavage of XR.[11] DXMT has dual activities toward both 7mX(3-N) and Tb(1-N).[11]

Previous experimental studies have indicated that the substrate recognition and catalytic efficiency of DXMT and other members of the family appear to have subtle biochemical characteristics, and their dependences on the active site and other residuals may be complicated.[9,11,12] It is thus of great interest to assess the subtle structure features in the active site of DXMT that may lead to a better understanding of substrate recognition. Moreover, a detailed study of the substrate specificity for DXMT could help to identify the key residues in the active site that could be genetically modified by site-directed mutagenesis to generate decaffeinated *Coffea* species. The overall structure of DXMT was also found to resemble that of salicylic acid *O*-methyltransferase, SAMT,[11] which is another SAM-dependent motif B' methyltransferase that has been structurally identified.[13] Yao et al. performed computer simulations on the SAMT catalyzed methyl transfer

from SAM to salicylic acid(SA), by use of quantum mechanics and molecular mechanics hybrid(QM/MM) molecular dynamic(MD) simulations.[14] They suggested based on a good correlation between the activities of the enzymes observed experimentally and the results of the computer simulations that the computational approaches seem to be well suited for the investigations of enzyme-catalyzed methyl transfer processes.[15,16,17]

Although the X-ray structures of DXMT have been determined, the catalytic mechanism of the enzyme is still not clear. For instance, Figure 1.1 shows that a general base seems to be required to remove the proton of the hydroxyl group (e.g., at the 2 position) in order to have the methyl group transferred to the nitrogen (e.g., either the 3 or 1 position) in each case. However, the general base or general bases have not been identified so far. In this work, the quantum mechanical /molecular mechanical (QM/MM) MD and free energy simulations are performed to determine the catalytic mechanism of the enzyme-catalyzed reactions involving DXMT and to explain some experimental observations. It is proposed that a conserved residue, His160, may act as a general base catalyst for the conversion of 7-methylxanthine to caffeine with theobromine as an intermediate product.

Methods

The initial coordinates for the reactant complexes of the methyl transfers to N3 or N1 of the purine ring are based on the crystallographic complexes (PDB ID: 2EFJ) of DXMT containing AdoHcy and theobromine with two different orientations. The

QM/MM MD and free energy (potential of mean force) simulations were performed in the determination of the free energy profiles for the methylation processes with the CHARMM program.[18] Water spheres based on a modified TIP3P water model[19] with radius of 30 Å, centered at either N3 or N1 of the purine ring, were used to solvate the system. Stochastic boundary with a Poisson-Boltzmann charge-scaling scheme[20] was applied. The reservoir region has r > 22 Å, and the buffer region has radius(r) equal to 20 Å ≤ r ≤ 22 Å. The reaction region has r ≤ 20 Å. The methyl group of AdoMet, the substrate (7mX or Tb) and the side chain of His160 were treated by QM and the rest of the system by MM. All-hydrogen potential function (PARAM27)[21] was used for the MM region, and the self-consistence charge density functional tight binding (SCC-DFTB)[22,23] method was used for the QM region. The link-atom approach[24] was used to separate the QM and MM regions. A small system containing the side chain of His160, a part of AdoMet, and the complete 7mX molecule was studied by using the SCC-DFTB and high level MP2/6–31G** methods with energy minimization to determine the accuracy of the SCC-DFTB method (Figure S1.1), suggesting that the SCC-DFTB method performs well for this system.

To build the initial reactant complex for the N3 methylation on 7mX from the X-ray structure, the substrate analog was modified to 7mX by deleting the methyl group at its N3 position, and a methyl group was added to S_δ of AdoHcy to generate the methyl-donor AdoMet. Similarly, to build the initial reactant complex for the N1

methylation on Tb, AdoHcy was modified to AdoMet by adding a methyl group. The initial reactant complex for the N1 methylation on 7mX is built by deleting the methyl group on N3 of the original Tb substrate in the crystal structure. The initial reactant complex containing paraxanthine (Px) is established by adding a methyl group to and deleting a methyl group on N1 of the 7Mx-like Tb in the crystal structure. The reference center for partitioning the system was chosen to be either N3 or N1 of the substrate, depending on which atom is the methyl acceptor. The resulting systems contain about 5800 atoms, including about 600 water molecules. The initial structures for the entire stochastic boundary systems were optimized using the steepest descent (SD) and adopted-basis Newton-Raphson (ABNR) methods. The systems were gradually heated from 50.0 to 298.15 K in 50 ps. A 1-fs time step was applied for the integration of equation of motion, and the coordinates were saved every 50 fs. 1.5 ns QM/MM MD simulations were executed for each of the reactant complexes initially. Similar to the methyl transfer from AdoMet to salicylic acid catalyzed by SAMT[14], the S_N2 methyl transfer reaction from AdoMet to $N_{1/3}$ on the DXMT's substrate is presumably more efficient if the S_δ-CH_3 group of AdoMet is well aligned with the lone pair of electrons of $N_{1/3}$ in the reactant complexes. The reaction coordinate was defined as a linear combination of $r(C_M \cdots N_{1/3})$ and $r(S_\delta \cdots C_M)$, which is $R = r(S_\delta \cdots C_M) - r(C_M \cdots N_{1/3})$. The umbrella sampling method[25] implemented in the CHARMM program along with the Weighted Histogram Analysis Method (WHAM) [26] was applied to determine the free energy (potential of mean force, PMF) as a function of the reaction

coordinate(s) for the methyl transfer process. For each methyl transfer process, twenty simulation windows were saved, and for each window 50 ps production runs were performed after 50 ps equilibration. The force constants of the harmonic biasing potentials used in the PMF simulations were 50 to 500 kcal mol^{-1} $Å^{-2}$.

The simulations were also performed with the O-H bond on the substrate fixed by SHAKE[27] algorithm of CHARMM to avoid proton transfer from the substrate to NE2 position of His160 in order to estimate the effect of the proton transfer on the methyl transfer. The distance between the proton donated from the substrate and its acceptor, NE2 on His160, was monitored in the whole simulations as a function of the reaction coordinate R= $r(S_{\delta}\cdots C_M)$ - $r(C_M\cdots N_{1/3})$.

The 2D free energy map (PMF) for the conversion of 7mX to Tb was determined with the umbrella sampling method [25] and two-dimensional WHAM[26]. The time step for QM/MM MD simulation is 1 fs. One reaction coordinate is Rx = $r(S_{\delta}\cdots C_M)$-$r(C_M\cdots N_3)$, describing the methyl transfer process. The other reaction coordinate is the distance between H on the hydroxyl group of 7mX and NE2 on His160, Ry = $r(H\cdots NE2)$, describing the proton transfer process. 209 windows were used in the calculation of 2D PMF, and 50 ps production runs were preformed after 50 ps equilibration for each window. The force constants for each window were in the range of 100-400 kcal·Mol^{-1}·$Å^{-2}$ for both Rx and Ry.

Results and Discussion

Reactant complexes for N3 and N1 methylation

The X-ray structures of the DXMT complexes have been available.[11] As shown in Figure 1.2A, Tb has its N3 methyl group pointing to S_δ of AdoHcy in the X-ray structure. This is in contrast with the case of another orientation of Tb in the structure in which N1 of Tb instead of N3 is close to S_δ (Figure 1.2B). Both orientations are presumably stabilized in the active site by hydrogen binding interactions, including the interactions of O2 of Tb with the side chains of His160 and Trp161 and the hydrogen bond involving the hydroxyl group of Ser237 and O6 in Fig.2A (N9 in Figure 1.2B). Moreover, water mediated hydrogen bonds are also formed involving the backbone N-H group of Ser237 and O6 in Figure 1. 2A (or N9 in Figure 1.2B) and the hydroxyl group of Tyr18 and N9 in Figure 1.2A (Try18 and O6 in Figure 1.2B). In addition to the hydrogen bonding interactions, the existence of hydrophobic environment can also be seen in Figure 1.2A and B. For instance, the purine ring forms hydrophobic contacts with Ile332 and Tyr157, while Phe27, Leu26, Tyr368 and Val328 produce a hydrophobic core that may also contribute the binding.

The average structures of the reactant complexes for the conversion of 7mX to Tb and for the conversion of Tb to Cf obtained after 1ns QM/MM MD simulations are shown in Figure 1.2C and D, respectively. The hydrogen bonding and hydrophobic interactions discussed earlier for the original crystal structures (Figure 1.2A and B)

generally remain (see Figure 1.2C and D). However, there are some significant deviations, as can be seen by comparison of Figure 1. 2A with C, and Figure 1. 2B with D, respectively. Those differences might be explained in part by the fact that, unlike the crystal structure, the reactant complex contains AdoMet rather than AdoHcy. The existence of a positively charged methyl group on AdoMet is expected to produce some changes in the interactions, leading to modifications of the active-site structure. For instance, the distance from S_δ to N1 increased from 3.6Å in Figure 1.2B to 4.7Å in Figure 1.2D. For the reactant complexes of other methyltransferases including PKMT[15,31,32], PRMT[33] and SAMT,[14] the methyl transfer processes appear to be favorable if the distance between the methyl group from methyl donor and acceptor nitrogen (or oxygen) is around 3 Å. This seems to be the case as shown in Figure 1.2C and D. Figure 1.2C and D also show that the methyl group of AdoMet is well aligned with the lone-pair electrons of N3 from 7mX and that of N1 from Tb in the both reactant complexes. It has been indicated from previous studies on PKMTs[14] and SAMT[15,16,17] that a good alignment between the transferable methyl group and the lone pair of electrons of the its acceptor, such as nitrogen and oxygen, is of importance for an efficient methyl transfer. The simulation results here imply that the DMXT reactant complex containing either 7mX or Tb could meet the requirement for a good alignment of the reactive groups participating in the methyl transfer, which is consistent with the substrate specificity[9,11] of this enzyme. As discussed below, free energy simulations from the reactant complex to the transition-like state indicate that a proper balance of

hydrogen bonding and hydrophobic interactions appears to be responsible for the good alignment of the reactive groups in the process of methyl transfer.

His160 as general acid/base catalyst for N3 and N1 methylation

7mX may exist in one of the three tautomers in solution (see Figure 1.1).[28] Methyl transfer from AdoMet to N3 on 7mX is presumably a S_N2 reaction in which the lone pair of electrons from the nucleophile N3 attacks C_M of the methyl group. Although the normal pKa of histidine imidazole is around 6, different prediction methods seem to indicate that His160 in the active site of DXMT may have a lower pKa (3.21 by ProPka[29]; <0 by H++[30]). Therefore His160 may be neutral, making it more likely to act as a general base. The neighboring O2 of 7mX and the side chain of Tyr157 might contribute to lowering the pKa of imidazole of His160.[29] In the crystal structure, the atom NE2 of His160 is within 2.5 Å of the atom O2 on the substrate (Figure 1.2A). After 1 ns MD simulation, the distance between NE2 and O2 is slightly increased to be 2.8 Å (Figure 1.2C). Thus, it is likely that NE2 on His160 might help to stabilize the hydroxyl O2 on 7mX through strong hydrogen bonding in the reactant complex and act as the general base during the methyl transfer. Tb also has different tautomers, with either deprotonated N1 or deprotonated O2 (Figure 1.1). After 1 ns MD simulation, NE2 of His160 is well kept at 2.9 Å to O2 of the substrate (Figure 1.2B), the same as in the crystal structure (Figure 1.2D). It's thus likely that, similar to the case involving 7mX, NE2 on His160 might help to stabilize the hydroxyl O2 on Tb through strong hydrogen bonding and act as a general base as well. Consistent with the suggestion above, the distance between

Trp161 and O2 are reasonably maintained during the simulations (see Figure 1.3 and 1.4), possibly contributing as another stabilizing force of O2 on the substrate.

The hydrophobic interactions in all those simulations (see Figure 1.3 and 1.4) remain basically the same as that observed in the X-ray structure; for clarity they are not shown in Figure 1.3 and 1.4. It is of interest to note that the active-site water molecule, TIP85, moves away from the active site. Nevertheless, the distance between the nitrogen on the side chain of Trp161 and O2 of substrates is well kept at round 3.2Å as shown in Figure 1.3 and 1.4. The hydrogen bond between Trp161 and the hydroxyl O2 on the substrate appears to help to maintain the substrate in the reactive configuration for the methyl transfer. The hydrogen bonding interactions of the backbone nitrogen from Ser237 with either O6 of 7mX or N9 of Tb as well as those mediated by the water molecule are generally stable in all simulations; they may play a role in generating the reactive configuration for the reactant complex as well.

The role of His160 as a general base is further demonstrated in the free energy simulations of methyl transfer processes. As shown in Figure 1.3 and 1.4, a spontaneous proton transfer from O2 to neutral His160 occurred during methylation of N3 (7mX->Tb) and N1 (Tb->Cf), respectively. In Figure 1.3 where the average structure near TS is given, window10 shows that the average distance between the proton and NE2 of His160 is 1.4 Å, window11 shows that the proton

transfer has basically completed (H···NE2 average distance is 1.1 Å). In Figure 1.4, window10 show that the average distance between proton and NE2 of His160 is 1.9 Å, and window11 show that the proton transfer has basically completed (H···NE2 average distance is 1.1 Å).

To monitor the breakage of O-H bond of the substrate and the formation of N-H bond between H and NE2 of His160, the distance between H and NE2 as a function of the reaction coordinate, R= $r(S_{\delta}\cdots C_M)$- $r(C_M\cdots N_{1/3})$, was plotted in Figure 1.5. The N-H covalent bond is considered being formed when the distance between H and NE2 is less than 1.1 Å. The reaction coordinate for the transition state (R_{TS}) is estimated to correspond to the highest free energy point in the free energy profile shown in Figure 1.5. It shows in Figure 1.5A and B that the N-H bond is formed near the transition state (R_{TS}=0.17 Å for N3, R_{TS}=0.43 Å for N1) of methylation. Furthermore, Figure 1.3 and 1.4 show that the proton transfer and the breakage of S_{δ}-C_M bond are not concerted, which is consistent with the two dimensional free energy map (see below in Figure 1.7).

The free energy profiles of stepwise methyl transfer processes

The free energy profiles (potential of mean force) for the methyl transfer from AdoMet to 7mX and Tb are plotted in Figure 1.6A and B, respectively, as a function of the reaction coordinate (R = $r(S_{\delta}\cdots C_M)$ - $r(C_M\cdots N_3)$ in Figure 1.6A and R=$r(S_{\delta}\cdots C_M)$- $r(C_M\cdots N_1)$ in Fig. 6B). As can be seen in Figure 1.6A, the free energy barrier from the initial reactant complex to the transition state is calculated to be

about 16-17 kcal/mol for the first methyl transfer to 7mX. And this free energy barrier is estimated to be approximately 18 kcal/mol for the second methyl transfer to Tb as shown in Figure 1.6B. Moreover, preventing the proton transfer could greatly increase the free energy barrier of methylation. Indeed, as shown in Figure 1.6A and B, preventing the proton transfer (by SHAKE algorithm in CHARMM[27]) increases the energy barrier for the 7mX methylation by 3.7 kcal/mol and for the Tb methylation by 12.7 kcal/mol, respectively.

By comparing the two free energy profiles of methylation of 7mX in Figure 1.6A, it is interesting to note that the profile with the proton fixed overlaps with the profile without fixing the proton until the two profiles approaches near transition state(i.e., around R=0.17Å). Without the proton transfer, the transition state of methylation is much lagged behind and is reached at around R=0.64 Å. This is consistent with the fact that the average structure of each window in Figure 1.3 before transition state is quite similar to the corresponding one in Figure S1.2. By contrast, the free energy profile of methyl transfer to N1 (Figure 1.6B) shows that preventing the proton transfer increases the free energy change at the beginning of methylation of Tb already. The free energy difference is also reflected in Figure S1.3. It seems in Figure S1.3 that the inability of the formation of a stable hydrogen bonding between NE2 of His160 and the fixed proton of Tb may be responsible for the increasement of the free energy barrier. In summary, we suggest that His160 is

likely to act as the general acid/base catalyst and mediate the proton transfer for the methylation of N3 and N1.

It should be pointed out that the stepwise bond-breaking and -making events might not well be reflected in the one-dimension (1D) free energy simulation and therefore two dimension (2D) simulation involving both methylation and proton transfer might offer a clearer picture on those events. The 2D free-energy contour map with reaction coordinates for both methylation (Rx= $r(S_{\delta}\cdots C_M)$ - $r(C_M\cdots N_3)$) and proton transfer (Ry=r(H···NE2)) is shown in Figure 1.7. As demonstrated in Figure 1.7, the progress of the proton transfer seems to be ahead of the methylation process. Indeed, the proton transfer is basically completed at around the transition state of N3 methylation (point B in Figure 1.7). The completion of proton transfer could thereafter promote the final attachment of the positive charged methyl group to the substrate.

The major pathway for DXMT catalyzed caffeine synthesis is thought by a step-by-step methylation of N3 (7mX->Tb) and then N1 (Tb->Cf).[11] Uefuji et al. detected the activity of DXMT toward paraxanthine(Px).[34] Px is an isomer of Tb which has two methyl groups on N1 and N7 of its purine ring(Figure 1.1), while Tb has two methyl groups on N3 and N7. The enzymatic activity of DXMT on Px is about 10 times higher than that for Tb.[34] The MD simulations of the step-by-step methylation was carried out for N1(7mX->Px) and then N3(Px->Cf). The free energy profiles

for the methyl transfer from AdoMet to 7mX and Px are as a function of the reaction coordinate, respectively. The free energy barrier for the N1 methylation on 7mX is 17.4 kcal/mol, which is 1.1 kcal/mol higher than that of N3 methylation on 7mX. This difference might suggest that DMXT prefer to methylate N3 other than N1 of 7mX. Furthermore, the free energy barrier for the conversion of Px to Cf is estimated to be 15.8 kcal/mol, which is 2.2 kcal/mol lower than that for the conversion of Tb to Cf. DXMT is thus estimated to be much more efficient for converting paraxanthine to caffeine than for converting theobromine to caffeine. This is consistent with the experimental result that the conversion efficiency of Tb to Cf is only 3.8% of that of Px to Cf by DXMT from *Coffee arabica.*[34] As an intermediate product in the caffeine biosynthesis, paraxanthine is more quickly to be converted to caffeine than theobromine. This might explain why theobromine is detectable in the coffee and tea plants while paraxanthine is rarely detected in the same species.

Conclusions

In this study, a histidine residue, His160, in the active site of DXMT was identified by the QM/MM MD and free energy simulations as a general base/acid catalyst that accepts the proton from the hydroxyl group of both 7-methylxanthine and theobromine. The computational simulations provided a clearer picture for the catalytic mechanism of DXMT catalyzed methyl-transfer process.

Acknowledgement

We thank Professor Martin Karplus for a gift of the CHARMM program. This work was supported by the National Science Foundation (Grant 0817940 to H.G.). We are also grateful for the computer resources (Newton) from University of Tennessee, Knoxville.

References

1. Ashihara, H.; Sano, H.; Crozier, A. Caffeine and related purine alkaloids: Biosynthesis, catabolism, function and genetic engineering. Phytochemistry 2008, 69, 841−856.

2. UefujI, H.; Tatsumi, Y.; Morimoto, M.; Kaothien-Nakayama, P.; Ogita, S.; Sano, H. Caffeine production in tobacco plants by simultaneous expression of three coffee N-methyltransferases and its potential as a pest repellant. Plant Mol. Biol. 2005, 59, 221−227.

3. Waller, G. Biochemical frontiers of allelopathy. Biol. Plant. 1989, 31, 418−447.

4. Kato, M.; Mizuno, K.; Crozier, A.; Fujimura, T.; Ashihara, H. Caffeine synthase gene from tea leaves. Nature 2000, 406, 956−957.

5. Ogawa, M.; Herai, Y.; Koizumi, N.; Kusano, T.; Sano, H. 7- Methylxanthine Methyltransferase of Coffee Plants. J. Biol. Chem. 2001, 276, 8213−8218.

6. Mizuno, K.; Kato, M.; Irino, F.; Yoneyama, N.; Fujimura, T.; Ashihara, H. The first committed step reaction of caffeine biosynthesis: 7-methylxanthosine synthase is closely homologous to caffeine synthases in coffee (Coffea arabica L.). FEBS Lett. 2003, 547, 56−60.

7. Mizuno, K.; Okuda, A.; Kato, M.; Yoneyama, N.; Tanaka, H.; Ashihara, H.; Fujimura, T. Isolation of a new dual-functional caffeine synthase gene encoding an enzyme for the conversion of 7-methylxanthine to caffeine from coffee(Coffea arabica L.). FEBS Lett. 2003, 534, 75−81.

8. Uefuji, H.; Ogita, S.; Yamaguchi, Y.; Koizumi, N.; Sano, H. Molecular Cloning and Functional Characterization of Three Distinct N-Methyltransferases Involved in the Caffeine Biosynthetic Pathway in Coffee Plants. Plant Physiol. 2003, 132, 372−380.

9. Yoneyama, N.; Morimoto, H.; Ye, C. X.; Ashihara, H.; Mizuno, K.; Kato, M. Substrate specificity of N-methyltransferase involved in purine alkaloids synthesis is dependent upon one amino acid residue of the enzyme. Mol. Genet. Genomics 2006, 275, 125−135.

10. Kato, M.; Mizuno, K. Caffeine synthase and related methyltransferases in plants. Front. Biosci. 2004, 9, 1833−1842.

11. McCarthy, A.; McCarthy, J. The Structure of Two N-Methyltransferases from the Caffeine Biosynthetic Pathway. Plant Physiol. 2007, 144, 879−889.

12. Mizuno, K.; Kurosawa, S.; Yoshizawa, Y.; Kato, M. Essential region for 3-N methylation in N-methyltransferases involved in caffeine biosynthesis. Z. Naturforsch. C 2010, 65, 257−265.

13. Zubieta, C.; Ross, J. R.; Koscheski, P.; Yang, Y.; Pichersky, E.; Noel, J. Structural Basis for Substrate Recognition in the Salicylic Acid Carboxyl Methyltransferase Family. Plant Cell 2003, 15, 1704−1716.

14. Yao, J.; Xu, Q.; Chen, F.; Guo, H. QM/MM Free Energy Simulations of Salicylic Acid Methyltransferase: Effects of Stabilization of TS-like Structures on Substrate Specificity. J. Phys. Chem. B 2011, 115, 389−396.

15. Guo, H.; Guo, H. Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity. Proc. Natl. Acad. Sci. U.S.A. 2007, 104, 8797−8802.

16. Xu, Q.; Chu, Y. Z.; Guo, H. B.; Smith, J. C.; Guo, H. Energy Triplets for Writing Epigenetic Marks: Insights from QM/MM Free Energy Simulations of Protein Lysine Methyltransferases. Chem. Eur. J. 2009, 15, 12596−12599.

17. Chu, Y.; Xu, Q.; Guo, H. Understanding Energetic Origins of Product Specificity of SET8 from QM/MM Free Energy Simulations: What Causes the Stop of Methyl Addition during Histone Lysine Methylation? J. Chem. Theory Comput. 2010, 6, 1380−1389.

18. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 1983, 4, 187−217.

19. Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 1983, 79, 926−935.

20. Brooks, C. L.; Brunger, A.; Karplus, M. Active site dynamics in protein molecules: a stochastic boundary molecular-dynamics approach. Biopolymers 1985, 24, 843−865.

21. 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. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B 1998, 102, 3586−3616.

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

23. Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. A QM/MM Implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) Method. J. Phys. Chem. B 2001, 105, 569−585.

24. Field, M. J.; Bash, P. A.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. J. Comput. Chem. 1990, 11, 700−733.

25. Torrie, G. M.; Valleau, J. P. Monte-Carlo free energy estimates using non-Boltzmann sampling. Application to the subcritical LennardJones fluid. Chem. Phys. Lett. 1974, 28, 578−581.

26. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The weighted histogram analysis method for free energy calculations on biomolecules. I. The method. J. Comput. Chem. 1992, 13, 1011−1021.

27. Ryckaert, J.; Ciccotti, G.; Berendsen, H. Numerical integration of the Cartesian equation of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. 1977, 23, 327−341.

28. Callahan, M. P.; Gengeliczki, Z.; Svadlenak, N.; Valdes, H.; Hobza, P.; de Vries, M. S. Non-standard base pairing and stacked structures in methyl xanthine clusters. Phys. Chem. Chem. Phys. 2008, 10, 2819−2816.

29. Rostkowski, M.; Olsson, M.; Søndergaard, C. R.; Jensen, J. H. Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. BMC Struct. Biol. 2011, 11, 6.

30. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Res. 2012, 40, W537−W541.

31. Zhang, X.; Bruice, T. C. Catalytic Mechanism and Product Specificity of Rubisco Large Subunit Methyltransferase: QM/MM and MD Investigations. Biochemistry 2007, 45, 5505−5514.

32. Hu, P.; Wang, S.; Zhang, Y. How Do SET-Domain Protein Lysine Methyltransferases Achieve the Methylation State Specificity? Revisited by Ab Initio QM/MM Molecular Dynamics Simulations. J. Am. Chem. Soc. 2008, 130, 3806−3813.

33. Yao, J.; Chu, Y.; An, R.; Guo, H. Understanding Product Specificity of Protein Lysine Methyltransferases from QM/MM Molecular Dynamics and Free

Energy Simulations: The Effects of Mutation on SET7/9 beyond the Tyr/Phe Switch. J. Chem. Inf. Model. 2012, 52, 449−456.

34. Chu, Y.; Li, G.; Guo, H. QM/MM MD and free energy simulations of the methylation reactions catalyzed by protein arginine methyltransferase PRMT3. Can. J. Chem. 2013, 91, 605−612.

35. Roberts, M. F.; Waller, G. R. N-methyltransferases and 7- methyl-N-nucleoside hydrolase activity in Coffea arabica and the biosynthesis of caffeine. Phytochemistry 1979, 18, 451−455.

Appendix

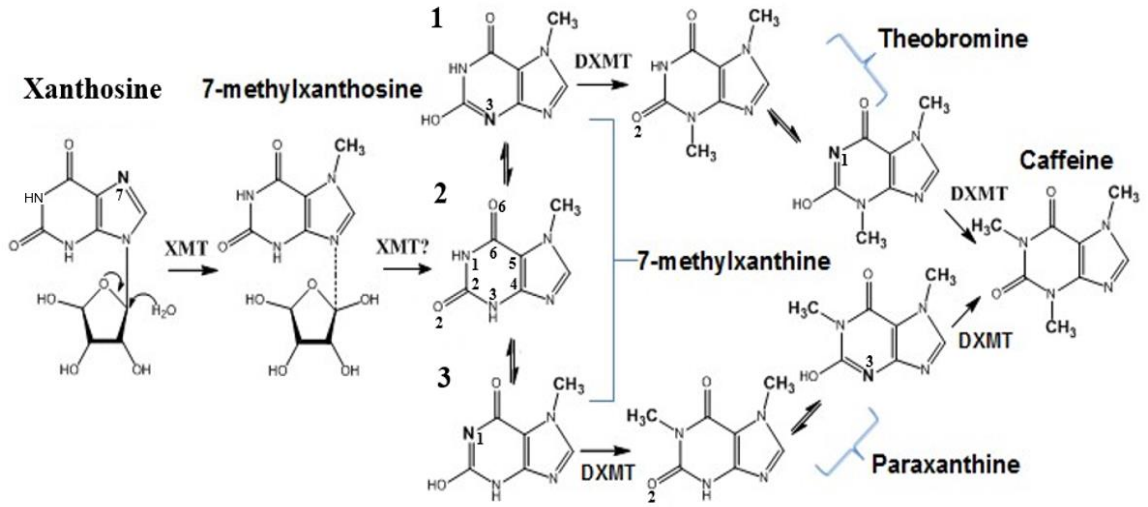


Figure 1.1 Caffeine biosynthesis pathway in *Coffea canephora* 'robusta'

The conversion of xanthosine to caffeine involves three N-methylation steps and the removal of ribose moiety. The major pathway is through an intermediate product, theobromine. However, the pathway through the intermediate paraxanthine is also possible. The nitrogen undergoing methylation and transferred methyl groups are highlighted in bold. Note 7-methylxanthine, theobromine and paraxanthine may exist as different tautomers. For instance, 7-methylxanthine may exist as one of the three tautomers (1, 2 or 3).

Figure 1.2 Key structures of DXMT complex

(A) The active site in the X-ray structure of DXMT (PDB ID: 2EFJ) complexed with AdoHcy and Tb. DXMT is shown in sticks, and AdoHcy and Tb are in balls and sticks. Only the three atoms, C5', CG and $S_{\delta,}$ from AdoHcy and the residues that are close to Tb are shown for clarity. Possible hydrogen bonds are shown in blue dotted lines. Some distances are given with unit of angstrom. (B) The active site in the X-ray structure of DXMT complexed with AdoHcy and Tb with different orientation. (C) The average active-site structure of the reactant complex containing AdoMet and 7mX from simulations. (D) The average active-site structure of the reactant complex containing AdoMet and Tb from simulations.

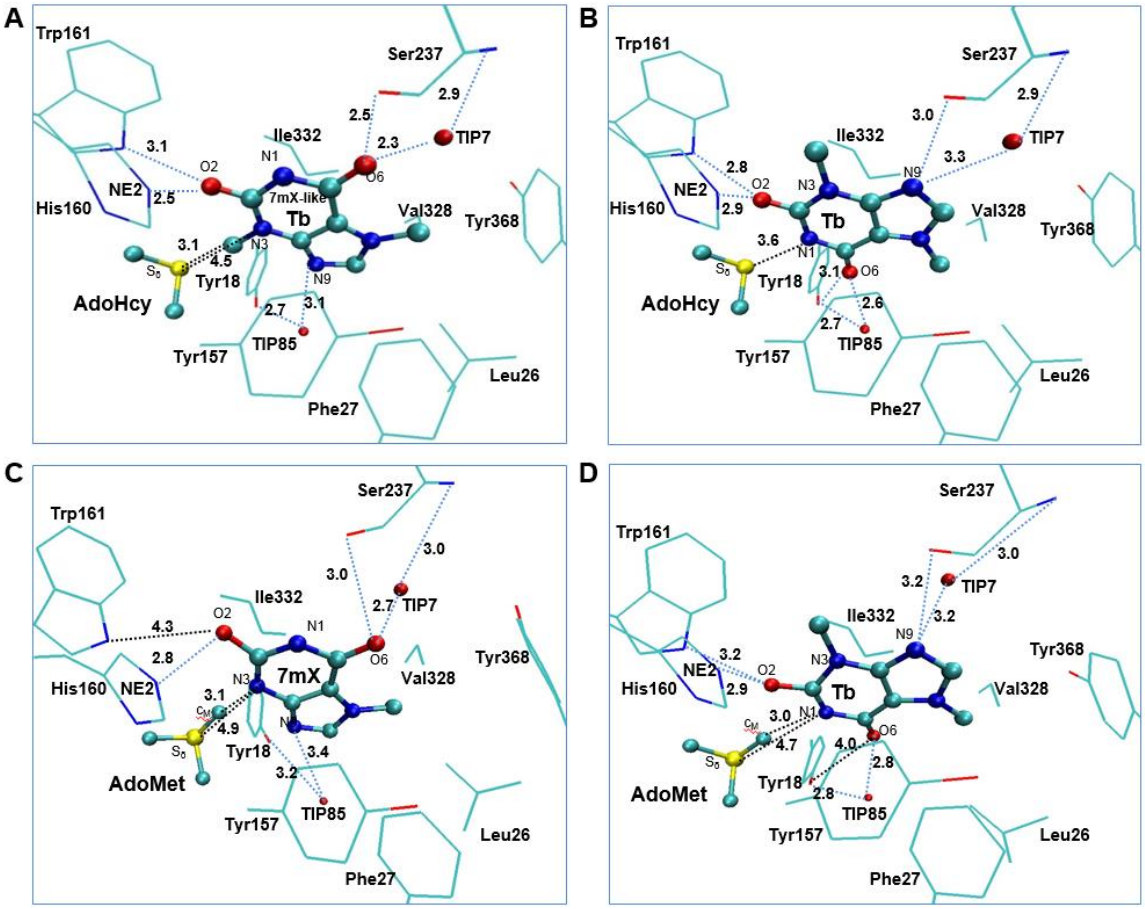


Figure 1.2 continued

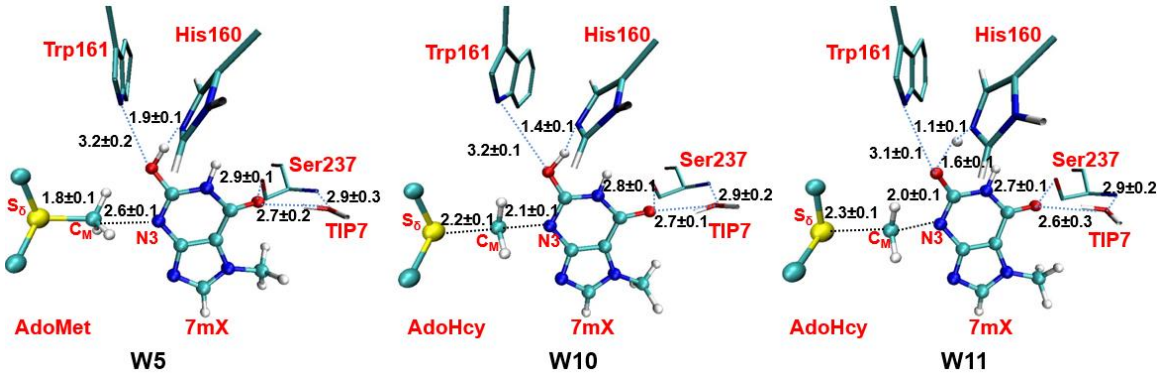


Figure 1.3 The active-site structure during the free energy simulation of N3 methylation

The average structures in window 5(W5, before the transition state, TS), window 10(W10, approaching near TS), window 11(W11, passed near TS) of N3 methylation are shown. The possible hydrogen bonding between 7mX and nearby residues are shown in blue dotted line. Some average distances are given with the unit of angstrom.

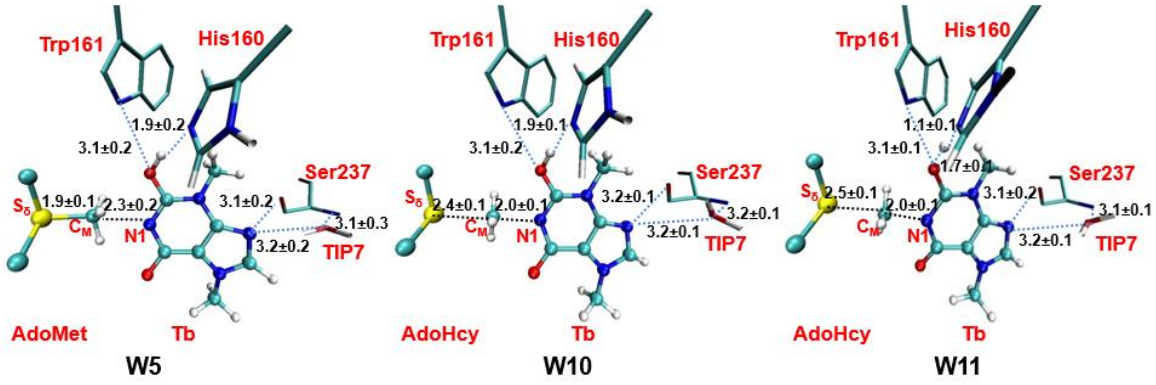


Figure 1.4 The active-site structure during the free energy simulation of N1 methylation

The average structures in window 5(W5, before the transition state (TS)), window 10(W10, approaching near TS), window 11(W11, passed near TS) of N1 methylation are shown. The possible hydrogen bonding between Tb and nearby residues are shown in blue dotted line. Some average distances are given with the unit of angstrom.

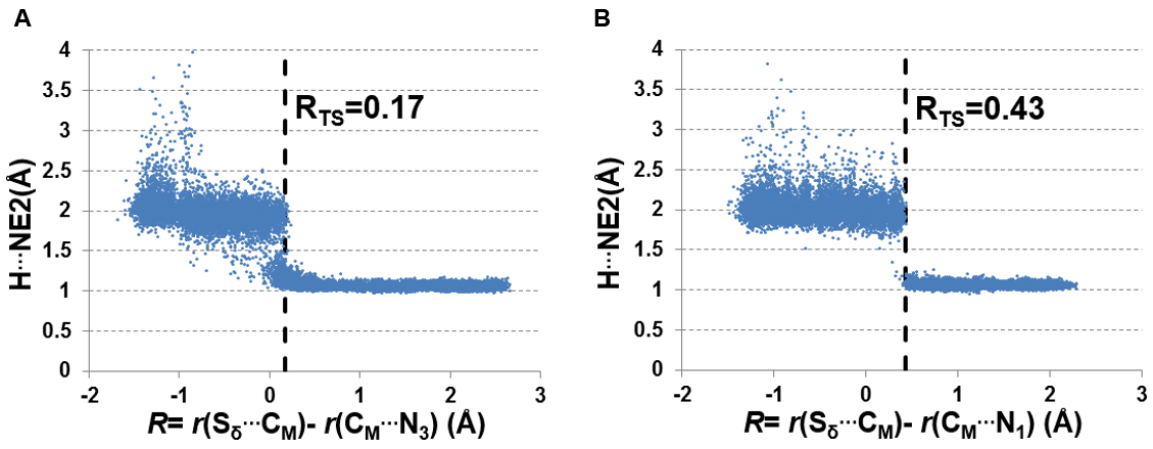


Figure 1.5 The distance between H on the substrate and NE2 on His160 as a function of the reaction coordinate

(A) The methylation of N_3(7mX->Tb); (B) The methylation of N_1(Tb->Cf). Note that O···H is broken and NE2···H is formed when the distance between H and NE2 is about 1.1 Å. The reaction coordinate for transition state is shown as R_{TS} in a dotted line.

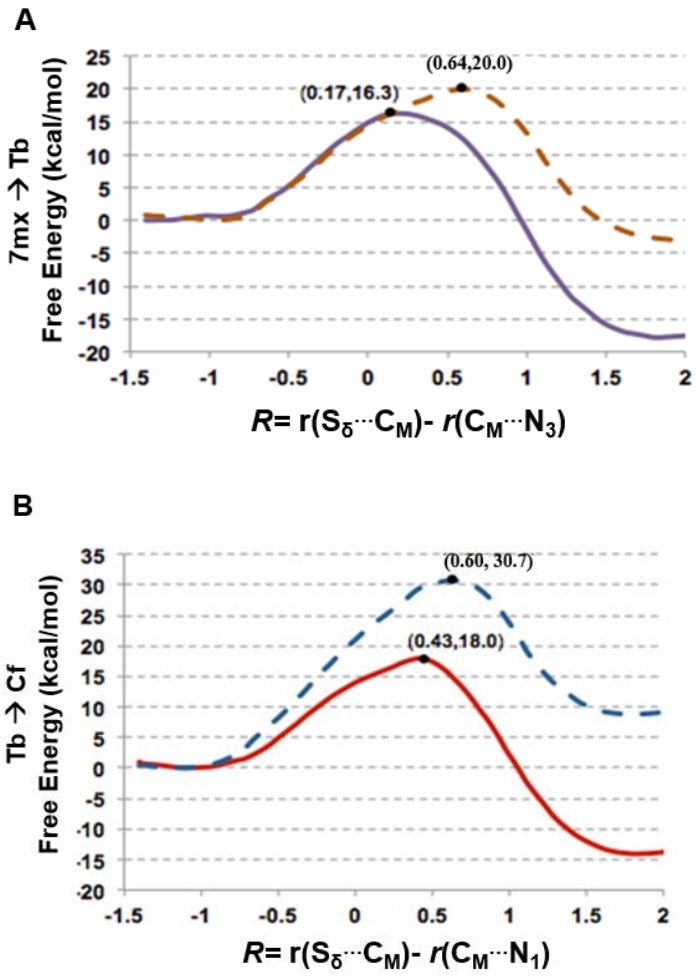


Figure 1.6 Free energy profiles of methyl transfer to N3 and to N1 of the purine ring

The free energy is calculated as the result of potential mean force and respectively plotted against the reaction coordinate R= r(C_M···S_δ) - r(C_M··· N_3) for methylation of N_3, and R = r(C_M···S_δ) - r(C_M··· N_1) for methylation of N_1. The solid line shows the energy profile of simulation without O-H bond fixed; the dotted line shows the energy profile of simulation with O-H fixed. The relative free energy at the transition state for each reaction is shown in bracket.

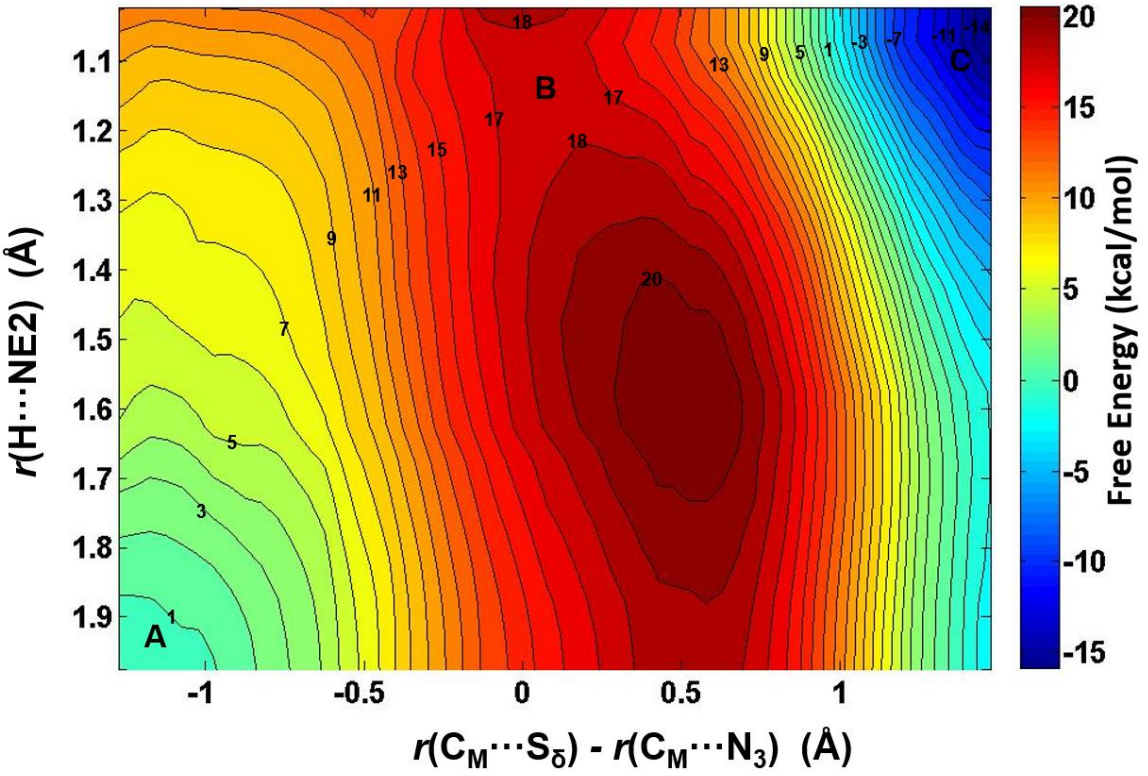


Figure 1.7 2D free energy contour map of N3 methylation on 7mX

The horizontal axis represents the reaction coordinate of methyl transfer and the vertical axis represents the proton transfer. Points A, B, C designate the reactant, transition state, and product complexes, respectively. Some contour lines are shown with respective energy value. The energy bar in shown on the right.

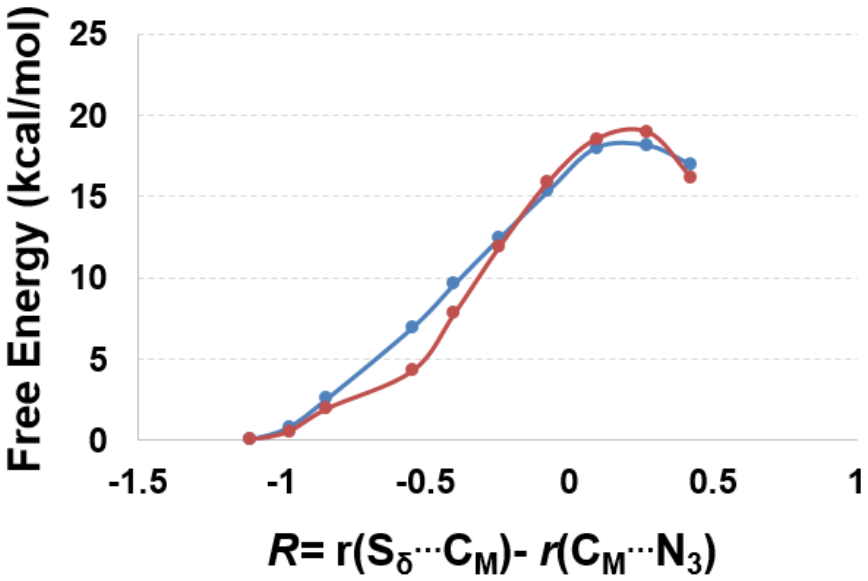


Figure S1.1 Comparison between SCC-DFTB and high level MP2/6–31G method**

The potential energy curve calculated by the self-consistent charge density functional tight-binding method(SCC-DFTB, in blue) and the single-point energy curve calculated by *ab initio* MP2/6–31G** method(in red) were applied in a small model.

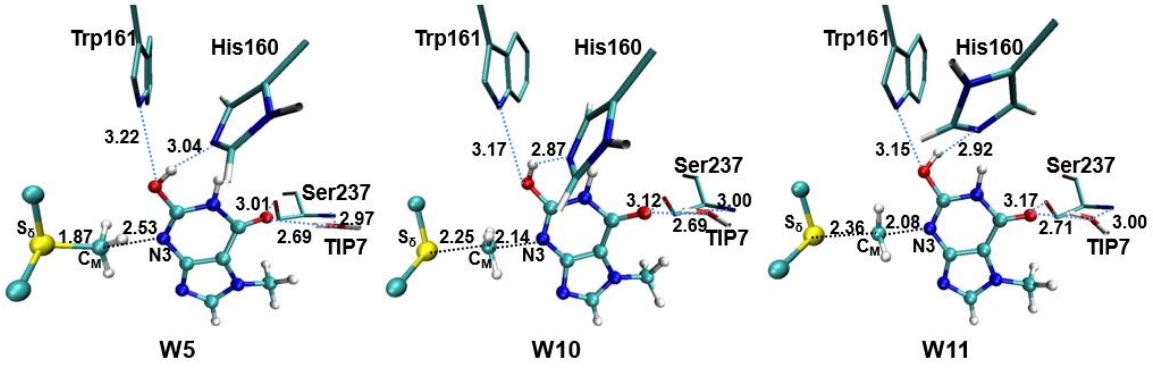


Figure S1.2 The dynamics of active site during the free energy simulation of N3 methylation with the fixed covalent hydrogen bond

The average structures in window 5(W5, before the transition state, TS), window 10(W10, approaching near TS), window 11(W11, passed near TS) of N3 methylation are shown.

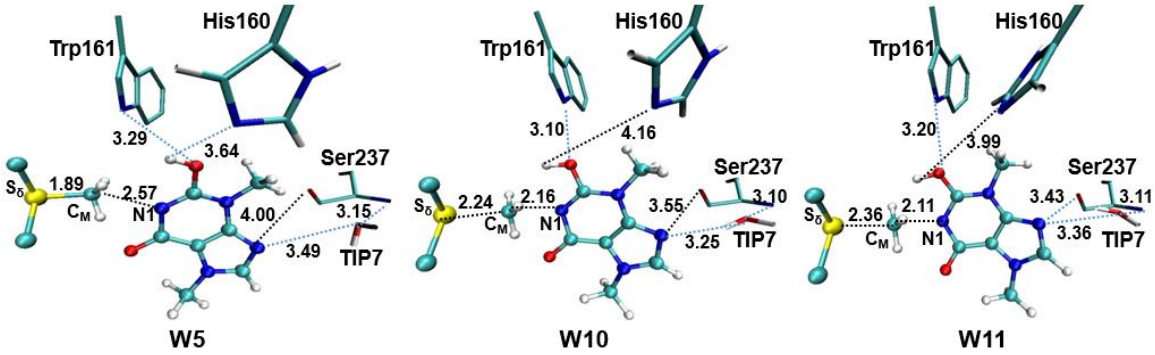


Figure S1.3 The dynamics of active site during the free energy simulation of N1 methylation with the fixed covalent hydrogen bond

The average structures in window 5(W5, before the transition state(TS)), window 10(W10, approaching near TS), window 11(W11, passed near TS) of N1 methylation are shown.

CHAPTER 2 COMPUTATIONAL STUDY OF SYMMETRIC METHYLATION ON HISTONE ARGININE CATALYZED BY PROTEIN ARGININE METHYLTRANSFERASE PRMT5

A version of this chapter was originally published by Yufei Yue, Yuzhuo Chu and Hong Guo: Yufei Yue, Yuzhuo Chu, Hong Guo. “Computational Study of Symmetric Methylation on Histone Arginine Catalyzed by Protein Arginine Methyltransferase PRMT5 through QM/MM MD and Free Energy Simulations.” *Molecules,* 20 (2015): 10032-10046. My contributions to this study include: 1) design of experiments; 2) building models and running simulations; 3) analyzing results and preparing the manuscript. Dr. Yuzhuo Chu contributed with some preliminary results regarding PRMT3. Dr. Hong Guo supervised in the experimental design, participated in discussion and reviewed the manuscript.

Abstract

Protein arginine methyltransferases (PRMTs) catalyze the transfer of the methyl group from S-adenosyl-L-methionine (AdoMet) to arginine residues. There are three types of PRMTs (I, II and III) that produce different methylation products, including asymmetric dimethylarginine (ADMA), symmetric dimethylarginine (SDMA) and monomethylarginine (MMA). Since these different methylations can lead to different biological consequences, understanding the origin of product specificity of PRMTs is of considerable interest. In this article, the quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) and free energy simulations are performed to study SDMA catalyzed by the Type II PRMT5 on the basis of experimental observation that the dimethylated product is

generated through a distributive fashion. The simulations have identified some important interactions and proton transfers during the catalysis. Similar to the cases involving Type I PRMTs, a conserved Glu residue (Glu435) in PRMT5 is suggested to function as general base catalyst based on the result of the simulations. Moreover, our results show that PRMT5 has an energetic preference for the first methylation on $N_{\eta 1}$ followed by the second methylation on a different ω-guanidino nitrogen of arginine ($N_{\eta 2}$).The first and second methyl transfers are estimated to have free energy barriers of 19–20 and 18–19 kcal/mol respectively. The computer simulations suggest a distinctive catalytic mechanism of symmetric dimethylation that seems to be different from asymmetric dimethylation.

Introduction

Posttranslational methylation of histone proteins on their arginine residues is an epigenetic mark that plays a vital role in cell function and is related with cell disorders and diseases.[1] Protein arginine methyltransferases (PRMTs) catalyze the transfer of methyl group(s) from S-adenosyl-L-methionine (AdoMet) to the guanidine nitrogen of arginine residue, resulting in the reaction products containing methylarginine along with S-adenosyl homocysteine (SAH).[2] In addition to the methylation of histone proteins, PRMTs could also modify a variety of other proteins.[3] Depending on types of PRMTs (I, II or III), the methylation products may contain asymmetric dimethylarginine (ADMA), symmetric dimethylarginine (SDMA) or monomethylarginine (MMA) (shown in Figure 2.1)[1].

The Type II PRMT5 to be investigated in this work catalyzes the transfer of methyl groups to the two different ω-guanidino nitrogen atoms on the arginine residue of the target protein, producing the ω-N^G, N'^G symmetrically dimethylated arginine (SDMA)[4]. PRMT5 has a variety of substrates that include histones, transcription factors, splicesomal proteins and piRNA biogenesis related proteins, and this enzyme functions in both nucleus and cytoplasm[3]. SDMA may profoundly impact many biological processes including epigenetic control of gene expression[5], circadian rhythms[6,7], splicing regulation [8,9], germ cell development and pluripotency[10,11], and DNA damage response[12,13]. The Type II PRMT5, however, often share the common recognition sequence with the Type I PRMTs which add two methyl groups to the same ω-guanidino nitrogen atom (ADMA).[14] Thus, the same target arginine may be either symmetrically or asymmetrically dimethylated. As such, SDMA and ADMA are isomeric protein posttranslational modifications with distinct and sometimes opposite biological effects.[14] One example is the methylation of arginine 3 on histone H4 (H4R3): symmetric dimethylation of H4R3 could repress gene expression[15,16], while asymmetric dimethylation of H4R3 is correlated with gene activation[17,18]. Recent studies have focused on understanding the enzymatic mechanisms that differentiate the two chemically isomeric but functionally antagonistic posttranslational modifications[14].

The crystal structures have been determined for several PRMTs, including PRMT1[19], PRMT3[20], PRMT6 [21], PRMT10 [22] of Type I, PRMT5[23] of Type II and

PRMT7[24] of Type III. Computer simulations have been applied to investigate the catalytic mechanism of PRMT1[25] and PRMT3[26], and some important questions concerning the product specificity of Type I ADMA have been addressed. The computational approaches used in these earlier studies include molecular dynamics and free energy simulations (potential of mean force) with the hybrid quantum mechanical and molecular mechanical (QM/MM) potential that seem to be suitable to investigate the enzyme-catalyzed methyl transfer process and have also been widely used for some other methyltransferases [26–32].

In this study, the methylation reaction catalyzed by Type II PRMT5 is investigated by use of QM/MM MD and free energy simulations on the basis of experimental observation that the dimethylated product is generated through a distributive fashion. The simulations have identified some important interactions and proton transfers involving the active site residues. Similar to the cases of Type I PRMT1[25] and PRMT3 [26], a homologous Glu residue (Glu435) in PRMT5 seems to function as a general base catalyst during the catalysis, and the corresponding proton transfer is found to be somehow concerted with the methyl transfer process. However, unlike Type I PRMTs which energetically favors a single ω-guanidino nitrogen ($N_{\eta 2}$) of arginine as the target for the both 1st and 2nd methylations [25,26], PRMT5 is found to have an energetic preference of targeting $N_{\eta 1}$ for the first methylation and then targeting a different ω-guanidino nitrogen ($N_{\eta 2}$) for the second methylation. The first and second methyl transfers are estimated to have

free energy barriers of 19–20 and 18–19 kcal/mol, respectively, from the simulations. These results are consistent with the existing experimental data [23]. Our computational study provides a better understanding of the symmetric di-methylation mechanism that is different from the asymmetric di-methylation.

Methods

The simulation coordinates for the reactant complexes of methyl transfers were based on the crystallographic complex (PDB ID: 4GQB, 2.06 Å) of PRMT5[23] that contains AdoMet analog and the H4 peptide with the methylation targeting substrate, Arg3. The coordinates of PRMT3 for comparison is from the crystal structure (PDB ID:1F3L, 2.06 Å) and modified based on the previous work[26]. The QM/MM MD and free energy (potential of mean force, PMF) simulations were applied for monitoring the methylation processes and determining the free energy profiles with the CHARMM program [35]. A water sphere based on a modified TIP3P water model [36] with radius(r) of 30 Å, centered at CZ of Arg3, was pre-equilibrated to the system. A stochastic boundary with a Poisson-Boltzmann charge-scaling scheme[37] was applied for the model. The reservoir region had r > 22 Å, and the buffer region had r equal to 20 Å ≤ r ≤ 22 Å. The reaction region had r ≤ 20 Å. The -CH_2-CH_2-S^+(Methyl)-CH_2- part of AdoMet, the side chain of substrate Arg3/monomethylated Arg3, the side chains of E435 and E444 were treated by QM and the rest of the system by MM. The resulting systems contained around 5500 atoms with about 800 water molecules. The all-hydrogen potential function

(PARAM27)[38] was used for the MM region, the self-consistence charge density functional tight binding (SCC-DFTB) [39,40] method was used for the QM region. The link-atom approach [41] was applied to separate the QM and MM regions.

The initial structures for the entire resulting system were optimized by use of the steepest descent (SD) and adopted-basis Newton-Raphson (ABNR) methods. The systems were gradually heated from 50 to 310.15 K in 50 ps. A 1 fs time step was applied to integrate equation of motion. 1.5 ns QM/MM MD simulations were initially executed for each of the reactant complex. The reaction coordinate was defined as a linear combination of $r(C_M \cdots S_\delta)$ and $r(C_M \cdots N_{\eta 1/2})$, which is $R = r(C_M \cdots S_\delta) - r(C_M \cdots N_{\eta 1/2})$. The umbrella sampling method[42] in the CHARMM program along with the weighted histogram analysis method (WHAM) [43] was used to determine the free energy(PMF) change as a function of the reaction coordinate(s). The Monte Carlo Simulation bootstrapping integrated with WHAM [43] was applied 100 times to estimate the errors (shown in Figure 2.3(b) and indicated in the legend of Figure 2.7) of the PMF profile. 20–22 simulation windows were saved for each methyl transfer process. And for each window 50 ps production runs were performed after 50 ps equilibration. The force constants of the harmonic biasing potentials used in the PMF simulations were 50–500 kcal mol^{-1} $Å^{-2}$. During the 50-ps production run of each window, structural coordinates were saved every 50 steps (1 fs/step), and consequently there are 1000 frames saved for each window. The distances shown in Figures 2.4, 2.5, 2.8 and 2.9 are the means of

corresponding distances with stand deviation (S.D.) collected from those 1000 frames. The 2D free energy (PMF) map were also determined with umbrella sampling method and two-dimensional-WHAM. The time step for simulations is 1 fs. The horizontal reaction coordinate for the 1st methylation on $N_{\eta 1}$, $R_x = r(C_M \cdots S_\delta) - r(C_M \cdots N_{\eta 1})$, was used to describe the methyl transfer process. And the vertical reaction coordinate, $R_y = r(HH_{12} \cdots N_{\eta 1}) - r(HH_{12} \cdots OE_1)$, was used to describe the proton transfer process. Similarly, for the 2nd methylations, the methyl transfer was explained by $R_x = r(C_M \cdots S_\delta) - r(C_M \cdots N_{\eta 2})$ and the proton transfer by $R_y = r(HH_{22} \cdots N_{\eta 2}) - r(HH_{22} \cdots OE_1)$. Approximated 300–400 windows were used in the construction of 2D free energy map, with 50 ps production run following 50 ps equilibration for each window. The force constants for each window were in the range of 100–800 kcal mol^{-1} $Å^{-2}$.

Results and Discussion

Comparison of the Active site Structures of Type II PRMT5 and Type I PRMT3

The two invariant glutamate residues from the “double-E” loop found in all PRMTs [33] are structurally conserved in both Type I PRMT3 and Type II PRMT5 active sites (E435 and E444 in Figure 2.2A; E144 and E153 in Figure 2.2B). As shown in Figure 2.2, the carboxylate side chain of E444 in PRMT5 (E153 in PRMT3) forms a stable salt bridge with the N_ε and $N_{\eta 1}$ atoms of the substrate arginine. The side

chain of E435 in PRMT5 (E144 PRMT3) interacts with $N_{\eta 1}$ and $N_{\eta 2}$ of the substrate. These two Glu residues are required for the enzymatic activities; the mutation of either of them could greatly decrease the enzymatic activity [14].

Previous computational studies have identified E144 as the general base to accept proton from the arginine during the methyl transfer catalyzed by Type I PRMTs [25,26]. One interesting questions is whether the corresponding E435 from PRMT5 would play a similar role during the catalysis. Four other residues that are conserved in the active site of PRMT5 are F327, K333, S578 and S439; the corresponding residues in PRMT3 are M48, R54, Y148 and H293, respectively. Although the mutation of S439 and S578 of PRMT5 diminishes the enzymatic activity significantly [14], the exact role of these residues is not clear; both S439 and S578 seem to be far away from the methyl donating AdoMet (Figure 2.2A). Interestingly, the F327M mutant of PRMT5 could produce both ADMA and SDMA [14], indicating that F327 may occupy a key position in PRMT5 and its properties may be important in determining the product specificity. Moreover, K333 is in the vicinity of F327 and could form hydrogen bonds with both the carboxylate group of AdoMet and E435.

The previous computational studies have shown that for PRMT1 and PRMT3 the 1st and 2nd methyl transfers would be energetically more favorable with relatively lower barriers if $N_{\eta 2}$ (see Figure 2.2B) is the methyl acceptor in each of the cases

[25,26]. This is consistent with the fact that they are both Type I PRMTs and the methylation products contain asymmetric dimethylarginine (ADMA). In the active site of the PRMT5 structure (Figure 2.2A), $N_{\eta 1}$ appears to be in a much better position for acting as the methyl acceptor compared to $N_{\eta 2}$; this is in contrast with the case of PRMT3 where $N_{\eta 2}$ seems to be in a better position for accepting the methyl group [26]. Indeed, the distance between $N_{\eta 1}$ and C_M is 3.2 Å in the active site of PRMT5 as compared to 4.9 Å between $N_{\eta 2}$ and C_M.

The First Methylation Catalyzed by PRMT5

The acceptor site for the first methylation process catalyzed by PRMT5 is studied by comparing the energetics of the methyl addition to $N_{\eta 1}$ and $N_{\eta 2}$, respectively, on the arginine via QM/MM MD and free energy simulations. As shown in Figure 2.3 (b), the free energy barrier for the methyl transfer to $N_{\eta 1}$ is 20.4 kcal/mol, which is 9 kcal/mol lower than that for the methyl transfer to $N_{\eta 2}$. This suggests that the 1st methyl group is likely to be transferred to the $N_{\eta 1}$ atom, as the corresponding process is energetically more favorable. This is in contrast with the 1st methylation catalyzed by PRMT3 which favors $N_{\eta 2}$ as the methyl acceptor [26].

The average active-site structures of the reactant complex and near transition state for the 1st methyl transfer are shown in Figures 2.4 and 2.5 for the methyl transfer to $N_{\eta 1}$ and $N_{\eta 2}$, respectively. The interactions obtained from the simulations for the reactant complex generally resemble the interactions observed in the crystal structure. This seems to indicate that the conformation of the active site in the

crystal structure is well maintained after initial 1.5 ns MD simulations. In the reactant complex for the methyl transfer to $N_{\eta 1}$ (Figure 2.4A), $N_{\eta 1}$ is well aligned with the methyl group of AdoMet, and this good alignment presumably leads to a relatively low barrier for the methyl transfer. The arginine on the substrate appears to be well stabilized through the salt bridge with E444 as well as hydrogen bonding interaction with E435. K333 may also help to adjust the orientation of E435 through a hydrogen bonding interaction. F327 seems to be involved in the π-cation interaction with the arginine in the reactant state. Moreover, as mentioned earlier the bulk size of this residue may prevent the formation of a good alignment between the transferable methyl group and $N_{\eta 2}$ and interfere with the methyl transfer to $N_{\eta 2}$. Indeed, the structures in Figures 2.2A shows that the distance between the transferable methyl group (C_M) and $N_{\eta 2}$ can be as much as 4.9 Å in the reactant complex. By contrast, the corresponding distance in PRMT3, which has M48 at the location rather than a Phe residue, is much smaller (e.g., 3.9 Å in Figure 2.2B). It is of interest to note from Figure 4B that the distance between F327 and the arginine increases significantly near the transition state during the methyl transfer to $N_{\eta 2}$.

The 1D free energy simulations only used the reaction coordinate for the methyl transfer. Nevertheless, the proton transfer occurred near the transition state of methylation (Figures 2.4B and 2.5B). To better understand the relationship between the methyl transfer and the proton transfer process, the 2D free energy simulations were performed (Figure 2.6). As is shown in Figure 2.6, the reaction

path goes from point A (designated as the reactant complex) to point B (the transition state), and finally reaches point D as the product complex. Before reaching the transition state at point B, the proton has been basically transferred from $N_{\eta 1}$ of the substrate arginine to the carboxyl oxygen of E435. And the free energy barrier is estimated to be 19–20 kcal/mol from the 2D free energy simulations, which is almost the same as that obtained from the 1D free energy simulations (20.4 kcal/mol). Moreover, point D in Figure 2.6 is about 6 kcal/mol lower than point C, suggesting that the deprotonated product is more stable in the active site than the protonated product and that the proton transfer is coupled with the methyl transfer during methylation.

The Second Methylation Catalyzed by PRMT5

With the first methyl group being transferred already, the 2nd methylation reaction can now be examined to determine the mechanism for the formation of the symmetrically dimethylated product. Previous experimental study on *Caenorhabditis elegans* PRMT5 (cPRMT5) has shown that the dimethylated product is generated through a distributive fashion [4] in which the peptide is released prior to rebinding to facilitate a second round of methylation. The distributive mechanism was further confirmed from the study of the human hPRMT5•MEP50 complex [34]. This is in contrast with the cases of some protein lysine methyltransferases for which the multiple rounds of methylation are believed to proceed processively without the release of the intermediates from the active sites. The kinetic data (K_m and k_{cat}) for cPRMT5, hPRMT5 and hPRMT5•MEP50

complex have also been obtained for a variety of un-methylated and mono-methylated substrates. For the hPRMT5•MEP50 complex for which the current investigation is based on, the differences in K_m and k_{cat} for the first and second methyl transfers are rather small and are beyond the accuracy of most of quantum mechanical approaches used in QM/MM studies. Although several experimental structures for PRMTs are available, the structure for the reactant complex of the second methyl transfer has not been determined. In our earlier investigations of protein lysine methyltransferases[27–30], we developed an approach based on the free energy simulations of the methyl transfer processes that allows us to determine whether or not the enzymes would be able to add methyl groups. This approach has been successfully applied to study the product specificity for a number of protein lysine methyltransferases, and its usefulness has also been confirmed on some other methyltransferases2[6,31,32], including PRMT1 [26]. This approach is especially suitable for the current investigation because the K_m values for the first and second methyl transfers are rather similar (see above), and the hypothetical non-existent methyl transfers can be presumably eliminated using this approach in the similar manner as the product specificity is determined [26–30]. In Figure 2.7, the three different configurations for the mono-methyl arginine without destroying the salt bridge with E444 are given. The nitrogen atom already connecting to the methyl group is designated as $N_{\eta 1}$ and that without the methyl group (*i.e.*, NH_2) designated as $N_{\eta 2}$. As is demonstrated in Figure 2.7, for each configuration there are two possible di-methylation products (SDMA and ADMA).

The QM/MM free energy simulations have been performed for each of the six methylation processes. Figure 2.7(I) shows that the free energy barriers were calculated to be 32.3 and 36 kcal/mol for the 2nd methyl transfer to the $N_{\eta 1}$ and $N_{\eta 2}$ atoms, respectively. Since either of these energy barriers seems to be too high, the 2nd methylation may not start from the configuration in Figure 2.7(I). The similar argument can be made for the configuration in Table Figure 2.7(III). Figure 2.7(II) shows that the second methyl transfer has the lowest free energy barrier (20.1 kcal/mol) if it is transferred to $N_{\eta 2}$ (*i.e.*, the SDMA product), while the free energy barrier for the second methyl transfer to $N_{\eta 1}$ (*i.e.*, the ADMA product) based on the configuration in Figure 2.7(II) is 11.2 kcal/mol higher.

The average structures of the reactant complex and near transition state for the 2nd methyl transfer to $N_{\eta 2}$ and $N_{\eta 1}$ are exhibited in Figures 2.8 and 2.9, respectively. The active-site structures of the reactant complex (Figure 2.8A) show that the $N_{\eta 2}$ atom is well aligned with the methyl group of AdoMet. The relatively high barrier for the 2nd methyl transfer to $N_{\eta 1}$ appears due in part to the steric hindrance of F327 (as in the case of the 1st methyl transfer), although other factors may be involved as well. One interesting observation for the second methyl transfer is that the proton has not been transferred to the general base E435 at the transition state (see Figures 2.8B and 2.9B). This is in contrast with the case of the first methyl transfer (shown in Figures 2.4B and 2.5B) where the proton transfer occurs before the methylation reaches the transition state. This observation is also

reflected in the 2D free energy map involving both the 2nd methyl transfer and the proton transfer (Figure 2.10). As is shown in the reaction path illustrated in Figure 2.10, the reaction reaches the transition state at point B without the deprotonation of the monomethylarginine.

Another interesting observation from Figures 2.8B and 2.9B is that E435 forms a salt bridge with both $N_{\eta 1}$ and $N_{\eta 2}$ near the transition state. This salt bridge may be relatively strong compared to that in the reactant complex and help to stabilize the transition state; it may also play a role for the delayed proton transfer mentioned earlier. The free energy barrier is estimated to be 18–19 kcal/mol by the 2D free energy simulation (Figure 2.10), consistent with 20.1 kcal/mol from the 1D free energy simulations (Figure 2.7).

Conclusions

In this study, the Type II PRMT5 was investigated by the QM/MM MD and free energy simulations. E435 was found to function as a general base catalyst for accepting a proton from the substrate arginine during the 1st and the 2nd methylation. This residue seems to play a similar role as E144 in Type I PRMT3. The simulations provide the detailed mechanism for the symmetric di-methylation by the enzyme and suggest the possible role for some other key residues as well during the catalysis. The k_{cat} for the arginine methylation catalyzed by PRMT5 is measured to be between 10 to 50 h^{-1}, indicating that the activation energy barrier

is about 18 kcal/mol based on transition state theory [23]. Thus, our estimates of the free energy barriers for the methyl transfers (19–20 and 18–19 kcal/mol for the first and the second methylation, respectively) are reasonable compared with the experimental observations. Furthermore, the simulations suggest that the symmetric di-methylation by PRMT5 seems to be energetically favorable, in agreement with the fact that PRMT5 is Type II PRMT. The proposed mechanism here based on the simulation results is different from the asymmetric dimethylation [25,26].

Acknowledgement

We thank Martin Karplus for a gift of the CHARMM program. This work was supported by the National Science Foundation (Grant 0817940 to H.G.). We are also grateful for the computer resources (Newton) from University of Tennessee, Knoxville and from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575.

References

1. Bedford, M.T.; Clarke, S.G. Protein arginine methylation in mammals: Who, what, and why. *Mol. Cell* **2009**, *33*, 1–13.

2. Di Lorenzo, A.; Bedford, M.T. Histone arginine methylation. *FEBS Lett.* **2011**, *585*, 2024–2031.

3. Biggar, K.K.; Li, S.S. Non-histone protein methylation as a regulator of cellular signalling and function. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 5–17.

4. Wang, M.; Xu, R.M.; Thompson, P.R. Substrate specificity, processivity, and kinetic mechanism of protein arginine methyltransferase 5. *Biochemistry* **2013**, *52*, 5430–5440.

5. Wysocka, J.; Allis, C.D.; Coonrod, S. Histone arginine methylation and its dynamic regulation. *Front. Biosci.* **2006**, *11*, 344–355.

6. Sanchez, S.E.; Petrillo, E.; Beckwith, E.J.; Zhang, X.; Rugnone, M.L.; Hernando, C.E.; Cuevas, J.C.; Godoy Herz, M.A.; Depetris-Chauvin, A.; Simpson, C.G.; *et al.* A methyl transferase links the circadian clock to the regulation of alternative splicing. *Nature* **2010**, *468*, 112–116.

7. Hong, S.; Song, H.R.; Lutz, K.; Kerstetter, R.A.; Michael, T.P.; McClung, C.R. Type II protein arginine methyltransferase 5 (prmt5) is required for circadian period determination in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21211–21216.

8. Meister, G.; Eggert, C.; Buhler, D.; Brahms, H.; Kambach, C.; Fischer, U. Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. *Curr. Biol.* **2001**, *11*, 1990–1994.

9. Deng, X.; Gu, L.; Liu, C.; Lu, T.; Lu, F.; Lu, Z.; Cui, P.; Pei, Y.; Wang, B.; Hu, S.; *et al.* Arginine methylation mediated by the Arabidopsis homolog of prmt5 is essential for proper pre-mRNA splicing. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 19114–19119.

10. Nishida, K.M.; Okada, T.N.; Kawamura, T.; Mituyama, T.; Kawamura, Y.; Inagaki, S.; Huang, H.; Chen, D.; Kodama, T.; Siomi, H.; *et al.* Functional involvement of tudor and dprmt5 in the pirna processing pathway in drosophila germlines. *EMBO J.* **2009**, *28*, 3820–3831.

11. Gonsalvez, G.B.; Rajendra, T.K.; Tian, L.; Matera, A.G. The sm-protein methyltransferase, dart5, is essential for germ-cell specification and maintenance. *Curr. Biol.* **2006**, *16*, 1077–1089.

12. Jansson, M.; Durant, S.T.; Cho, E.C.; Sheahan, S.; Edelmann, M.; Kessler, B.; la Thangue, N.B. Arginine methylation regulates the p53 response. *Nat. Cell Biol.* **2008**, *10*, 1431–1439.

13. Yang, M.; Sun, J.; Sun, X.; Shen, Q.; Gao, Z.; Yang, C. Caenorhabditis elegans protein arginine methyltransferase prmt-5 negatively regulates DNA damage-induced apoptosis. *PLoS Genet* **2009**, *5*, e1000514.

14. Sun, L.; Wang, M.; Lv, Z.; Yang, N.; Liu, Y.; Bao, S.; Gong, W.; Xu, R.M. Structural insights into protein arginine symmetric dimethylation by prmt5. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20538–20543.

15. Wang, X.; Zhang, Y.; Ma, Q.; Zhang, Z.; Xue, Y.; Bao, S.; Chong, K. Skb1-mediated symmetric dimethylation of histone h4r3 controls flowering time in Arabidopsis. *EMBO J.* **2007**, *26*, 1934–1941.

16. Zhao, Q.; Rank, G.; Tan, Y.T.; Li, H.; Moritz, R.L.; Simpson, R.J.; Cerruti, L.; Curtis, D.J.; Patel, D.J.; Allis, C.D.; *et al.* Prmt5-mediated methylation of histone h4r3 recruits dnmt3a, coupling histone and DNA methylation in gene silencing. *Nat. Struct. Mol. Biol.* **2009**, *16*, 304–311.

17. Wang, H.; Huang, Z.Q.; Xia, L.; Feng, Q.; Erdjument-Bromage, H.; Strahl, B.D.; Briggs, S.D.; Allis, C.D.; Wong, J.; Tempst, P.; *et al.* Methylation of histone h4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Science* **2001**, *293*, 853–857.

18. Strahl, B.D.; Briggs, S.D.; Brame, C.J.; Caldwell, J.A.; Koh, S.S.; Ma, H.; Cook, R.G.; Shabanowitz, J.; Hunt, D.F.; Stallcup, M.R.; *et al.* Methylation of histone h4 at arginine 3 occurs *in vivo* and is mediated by the nuclear receptor coactivator prmt1. *Curr. Biol.* **2001**, *11*, 996–1000.

19. Zhang, X.; Cheng, X. Structure of the predominant protein arginine methyltransferase prmt1 and analysis of its binding to substrate peptides. *Structure* **2003**, *11*, 509–520.

20. Siarheyeva, A.; Senisterra, G.; Allali-Hassani, A.; Dong, A.; Dobrovetsky, E.; Wasney, G.A.; Chau, I.; Marcellus, R.; Hajian, T.; Liu, F.; *et al.* An allosteric inhibitor of protein arginine methyltransferase 3. *Structure* **2012**, *20*, 1425–1435.

21. Wang, C.; Zhu, Y.; Chen, J.; Li, X.; Peng, J.; Chen, J.; Zou, Y.; Zhang, Z.; Jin, H.; Yang, P.; *et al.* Crystal structure of arginine methyltransferase 6 from Trypanosoma brucei. *PLoS ONE* **2014**, *9*, e87267.

22. Cheng, Y.; Frazier, M.; Lu, F.; Cao, X.; Redinbo, M.R. Crystal structure of the plant epigenetic protein arginine methyltransferase 10. *J. Mol. Biol.* **2011**, *414*, 106–122.

23. Antonysamy, S.; Bonday, Z.; Campbell, R.M.; Doyle, B.; Druzina, Z.; Gheyi, T.; Han, B.; Jungheim, L.N.; Qian, Y.; Rauch, C.; *et al.* Crystal structure of the human prmt5:Mep50 complex. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 17960–17965.

24. Wang, C.; Zhu, Y.; Caceres, T.B.; Liu, L.; Peng, J.; Wang, J.; Chen, J.; Chen, X.; Zhang, Z.; Zuo, X.; *et al.* Structural determinants for the strict monomethylation activity by Trypanosoma brucei protein arginine methyltransferase 7. *Structure* **2014**, *22*, 756–768.

25. Zhang, R.; Li, X.; Liang, Z.; Zhu, K.; Lu, J.; Kong, X.; Ouyang, S.; Li, L.; Zheng, Y.G.; Luo, C. Theoretical insights into catalytic mechanism of protein arginine methyltransferase 1. *PLoS ONE* **2013**, *8*, e72424.

26. Chu, Y.; Li, G.; Guo, H. QM/MM md and free energy simulations of the methylation reactions catalyzed by protein arginine methyltransferase prmt3. *Can. J. Chem.* **2013**, *91*, 605–612.

27. Guo, H.B.; Guo, H. Mechanism of histone methylation catalyzed by protein lysine methyltransferase set7/9 and origin of product specificity. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 8797–8802.

28. Xu, Q.; Chu, Y.Z.; Guo, H.B.; Smith, J.C.; Guo, H. Energy triplets for writing epigenetic marks: Insights from qm/mm free-energy simulations of protein lysine methyltransferases. *Chemistry* **2009**, *15*, 12596–12599.

29. Chu, Y.; Yao, J.; Guo, H. QM/MM md and free energy simulations of g9a-like protein (glp) and its mutants: Understanding the factors that determine the product specificity. *PLoS ONE* **2012**, *7*, e37674.

30. Chu, Y.; Xu, Q.; Guo, H. Understanding energetic origins of product specificity of set8 from qm/mm free energy simulations: What causes the stop of methyl addition during histone lysine methylation? *J. Chem. Theory Comput.* **2010**, *6*, 1380–1389.

31. Yao, J.; Xu, Q.; Chen, F.; Guo, H. Qm/mm free energy simulations of salicylic acid methyltransferase: Effects of stabilization of ts-like structures on substrate specificity. *J. Phys. Chem. B* **2010**, *115*, 389–396.

32. Yue, Y.; Guo, H. Quantum mechanical/molecular mechanical study of catalytic mechanism and role of key residues in methylation reactions catalyzed

by dimethylxanthine methyltransferase in caffeine biosynthesis. *J. Chem. Inf. Model.* **2014**, *54*, 593–600.

33. Cheng, X.D.; Collins, R.E.; Zhang, X. Structural and sequence motifs of protein (histone) methylation enzymes. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 267–294.

34. Wang, M.; Fuhrmann, J.; Thompson, P.R. Protein arginine methyltransferase 5 catalyzes substrate dimethylation in a distributive fashion. *Biochemistry* **2014**, *53*, 7884–7892.

35. Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macro-molecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

36. Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

37. Brooks, C.L.; Brunger, A.; Karplus, M. Active site dynamics in protein molecules: A stochastic boundary molecular-dynamics approach. *Biopolymers* **1985**, *24*, 843–865.

38. MacKerell, A.D.; Bashford, D.; Bellott, M.; Dunbrack, R.L.; Evanseck, J.D.; Field,M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; *et al*. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

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

40. Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. A QM/MM Implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) Method. *J. Phys. Chem. B* **2001**, *105*, 569–585.

41. Field, M.J.; Bash, P.A.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **1990**, *11*, 700–733.

42. Torrie, G.M.; Valleau, J.P. Monte-Carlo free energy estimates using non-Boltzmann sampling. Application to the subcritical Lennard Jones fluid. *Chem. Phys. Lett.* **1974**, *28*, 578–581.

43. Kumar, S.; Bouzida, D.; Swendsen, R.H.; Kollman, P.A.; Rosenberg, J.M. The weighted histogram analysis method for free energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

Appendix

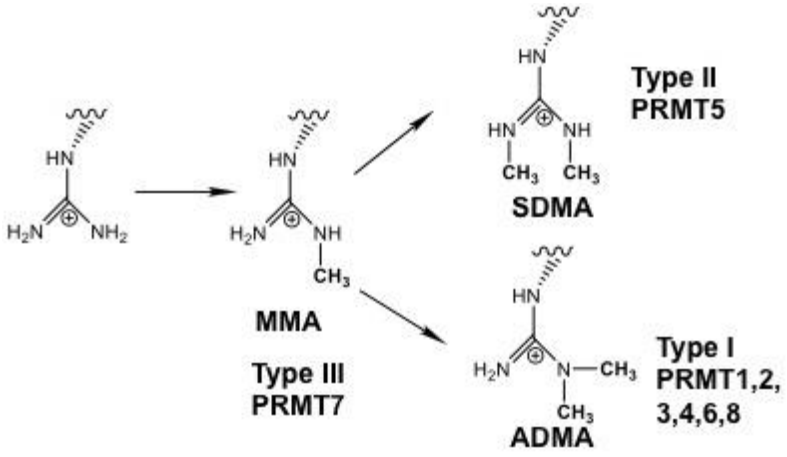


Figure 2.1 Methylation of arginine by different types of PRMTs

Type I PRMT can produce both monomethylarginine (MMA) and asymmetric dimethylarginine (ADMA). Type II PRMT can produce both MMA and symmetric dimethylarginine (SDMA). Type III PRMT can only produce MMA.

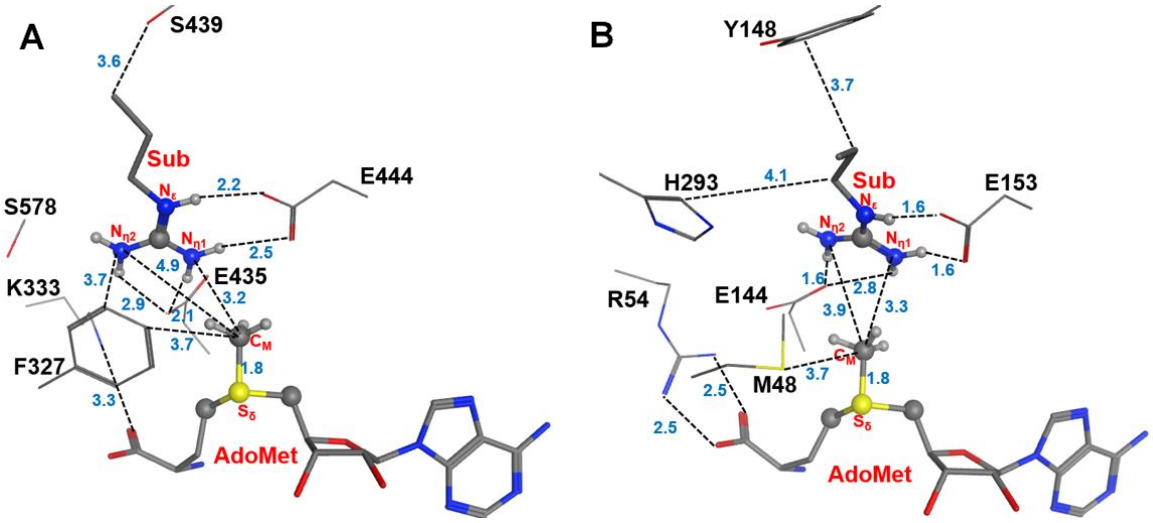
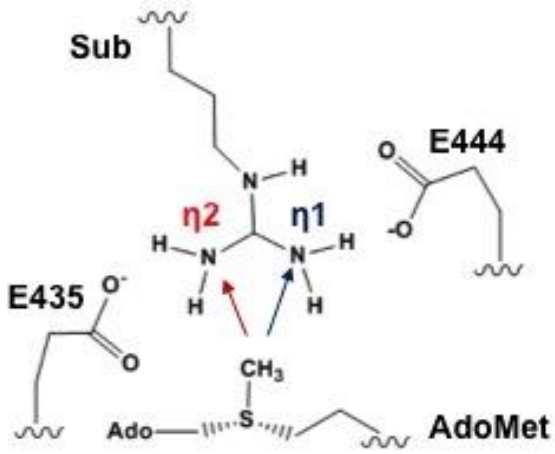
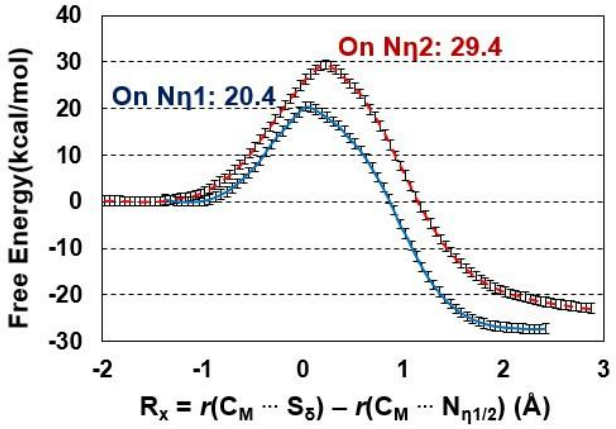


Figure 2.2 Comparison of the active sites of the crystal structures for Type-II PRMT5 and Type-I PRMT3

(A) The active site of PRMT5 (PDB ID: 4QGB); (**B**).The active site of PRMT3 (PDB ID: 1F3L). The two glutamate residues, E435 (E144) and E444 (E153) in PRMT5 (PRMT3), are conserved in the both types of PRMTs. F327, K333, S578 and S439 in PRMT5 are the conserved residues among Type-II PRMTs. The corresponding residues in Type-I PRMT3 are M48, R54, H293 and Y148, respectively. The substrate arginine is labeled as Sub. Some distances are shown in blue with the unit of angstrom.



(a)



(b)

Figure 2.3 Two possible pathways for the 1st methylation

(a) Two possible sites for the 1st methylation. The substrate arginine is labeled as Sub. **(b)** Free energy profiles of the 1st methylation on $N_{\eta1}$ and $N_{\eta2}$. To $N_{\eta1}$: blue line with a free energy barrier of 20.4 kcal/mol; to $N_{\eta2}$: red dotted line with a free energy barrier of 29.4 kcal/mol. Error bars are shown on the free energy profile (within ± 0.12 kcal/mol).

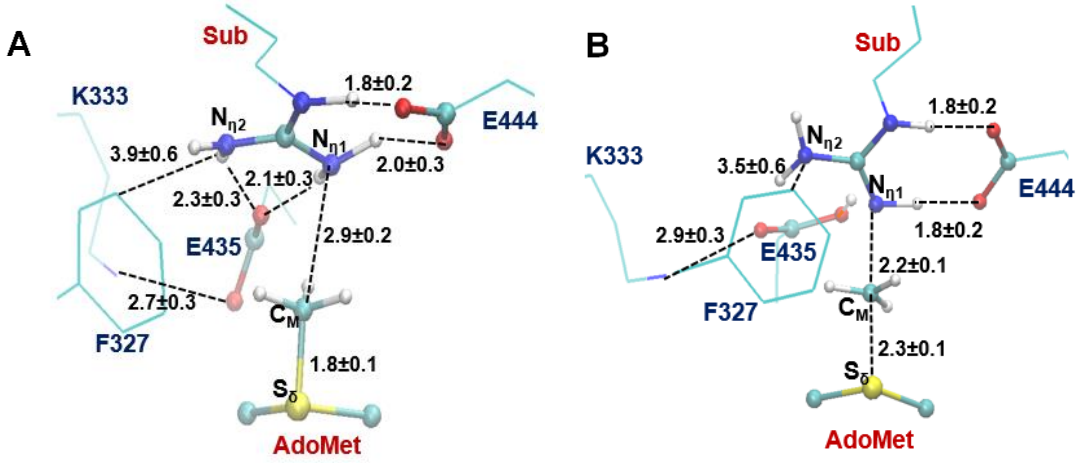


Figure 2.4 The average active-site structure obtained for the 1st methylation on $N_{\eta 1}$

(A) Reactant complex. **(B)** Near the transition state. Some average distances in active sites are shown in angstrom. Note that E435 works as the proton acceptor based on the simulations.

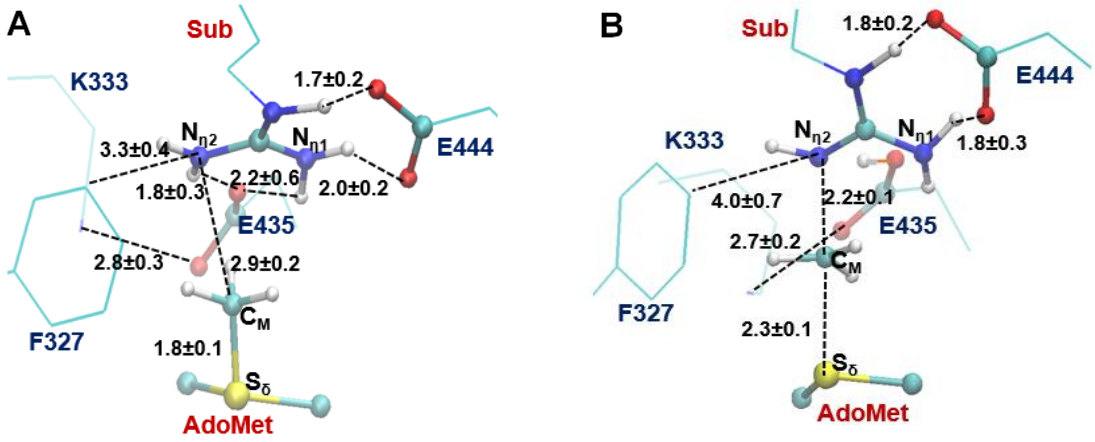


Figure 2.5 The average active-site structure obtained for the 1st methylation on $N_{\eta 2}$

(A) Reactant complex. **(B)** Near the transition state.

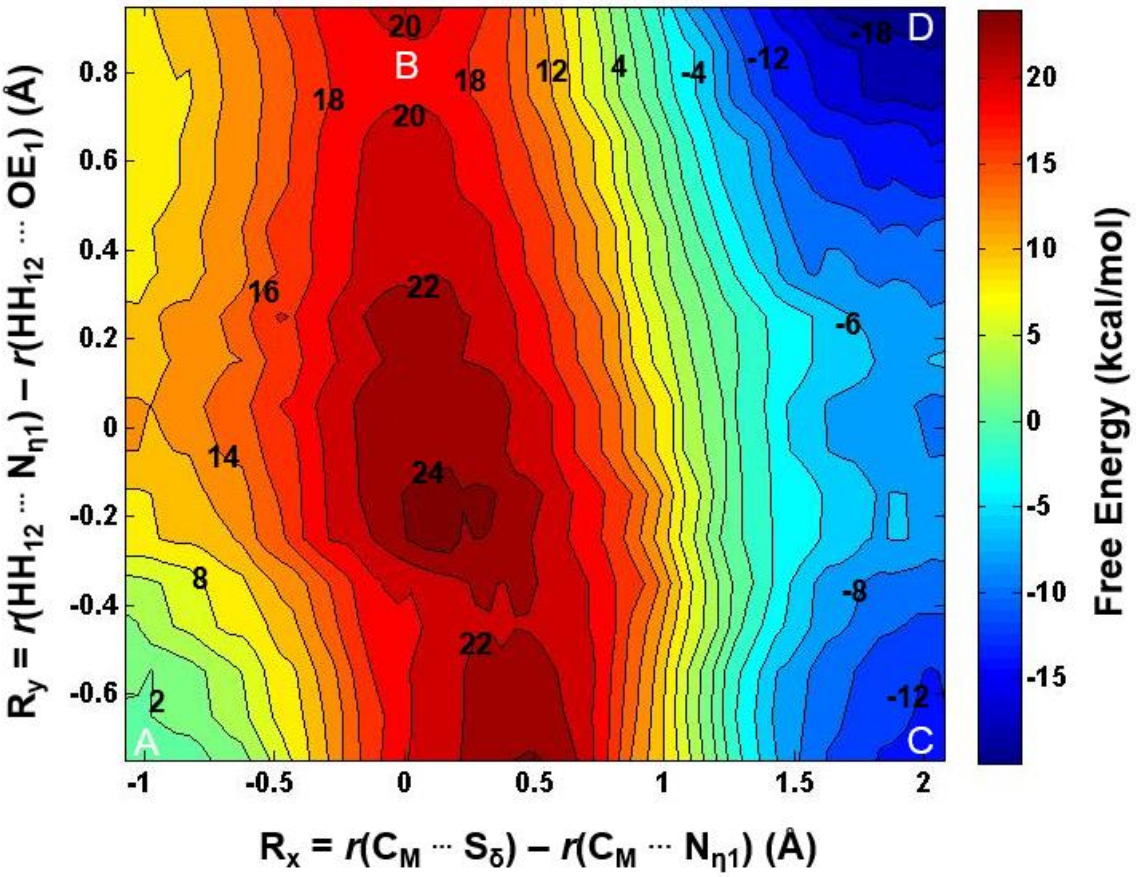
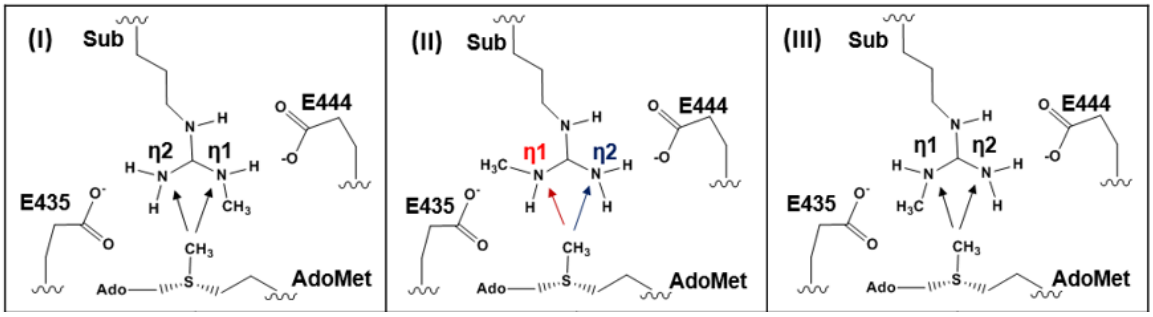


Figure 2.6 2D free-energy contour map for the 1st methylation on $N_{\eta 1}$

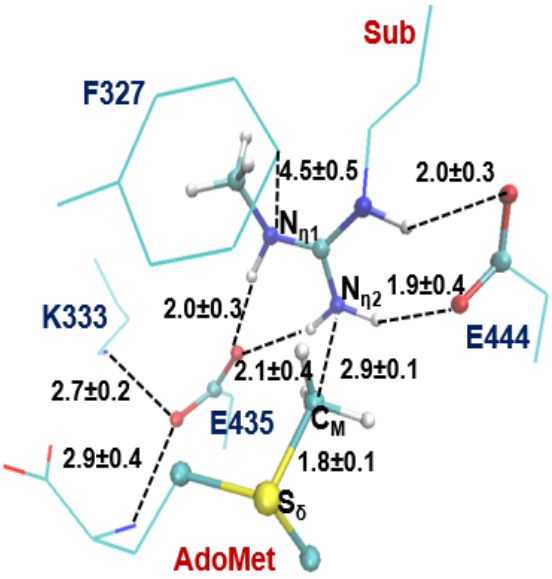
R_x: the reaction coordinate of the methyl transfer; R_y: the reaction coordinate for the proton transfer. Points A, B, and D designate the reactant, near transition state, and product complexes, respectively. Some contour lines are shown with respective energy values. The energy barrier is estimated to be around 19–20 kcal/mol.



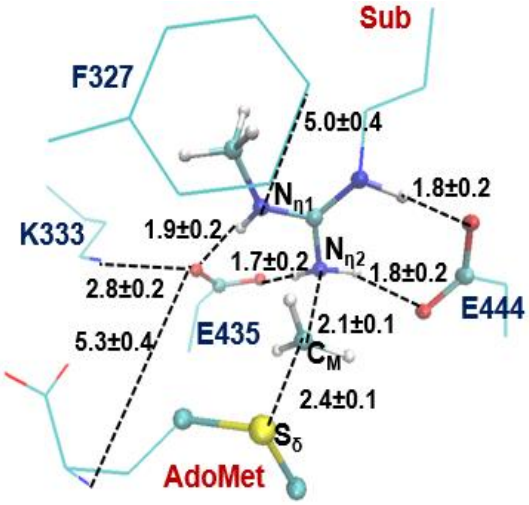
To $N_{\eta1}$	To $N_{\eta2}$	To $N_{\eta1}$	To $N_{\eta2}$	To $N_{\eta1}$	To $N_{\eta2}$
32.3 kcal/mol	36 kcal/mol	31.3 kcal/mol	20.1 kcal/mol	43.9 kcal/mol	32.5 kcal/mol

Figure 2.7 Free energy barriers for the 2nd methylation

The 2nd methylation can be on either the mono-methylated nitrogen atom (designated as $N_{\eta1}$) or un-methylated nitrogen atom (designated as $N_{\eta2}$) of the mono-methylated arginine (MMA) with three possible configurations for the reactant complex without destroying the salt bridge involving E444. The only difference between Configurations II and III is whether the methyl group is pointing to the AdoMet side (III) or away from the AdoMet side (II). The conformation in (II) has the lowest energy barrier of 20.1 kcal/mol when the 2nd methylation occurs to $N_{\eta2}$ which will eventually produce the symmetrically dimethylated arginine (SDMA). Error bars are also determined, similar to those given in Figure 2.3 (within ± 0.2 kcal/mol).



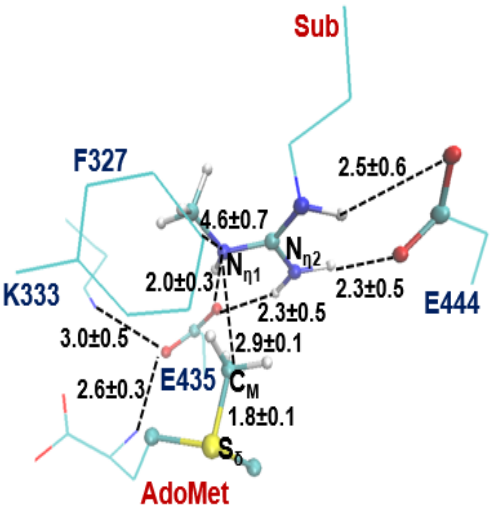
(A)



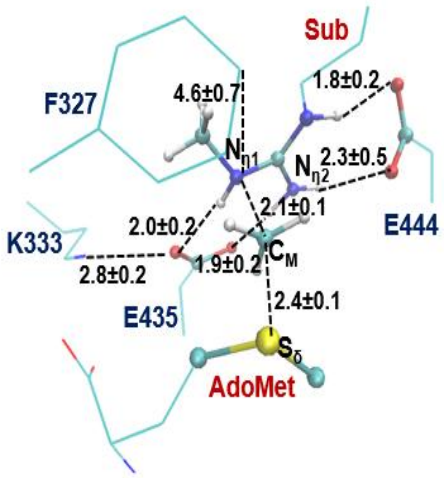
(B)

Figure 2.8 Results from the MD simulations for the 2nd methylation on $N_{\eta 2}$

(**A**) Active-site structure of the reactant complex; (**B**) Active-site structure near the transition state.



(A)



(B)

Figure 2.9 Results from the MD simulations for the 2nd methylation on $N_{\eta 1}$

(**A**) Active-site structure of the reactant complex; (**B**) Active-site structure near the transition state.

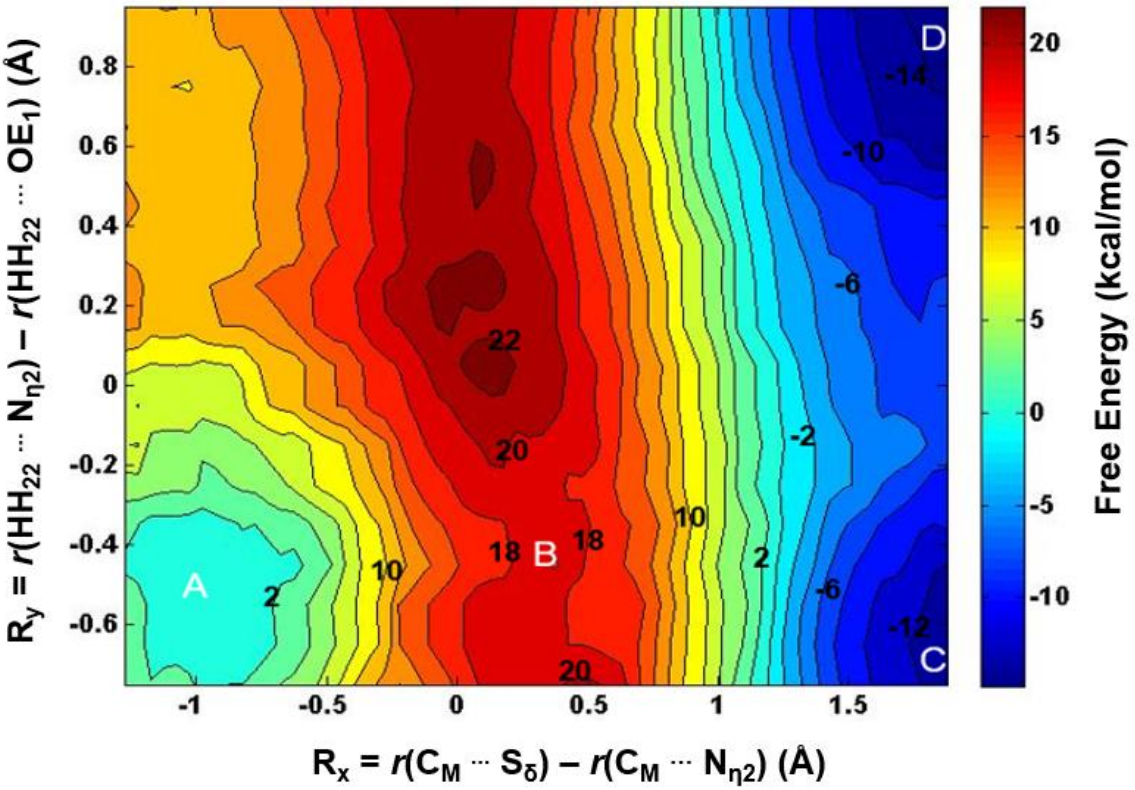


Figure 2.10 2D free-energy contour map for the 2nd methylation on $N_{\eta 2}$

Points A, B, and D designate the reactant, near transition state, and product complex, respectively. Point C corresponds to the hypothetical product complex with proton not being transferred, which is only 2 kcal/mol higher than Point D in the free-energy contour map. Some contour lines are shown with respective energy value in kcal/mol. The energy bar is shown on the right. The energy barrier is estimated to be around 18–19 kcal/mol.

CHAPTER 3 COMPUTATIONAL STUDY OF UBIQUITIN-LIKE NEDD8 TRANSFER IN RING E3-E2~NEDD8-TARGET COMPLEX

Abstract

Ubiquitin-like (UBL) protein modification of substrate proteins plays a key role in regulating protein function. Unlike ubiquitin (UB) and small ubiquitin-like modifier (SUMO) which are ligated to a massive segment of the proteome, the UBL NEDD8 is highly selective on modifying closely related cullin proteins (CULs) on a single lysine residue and contributes to 10% - 20% of all cellular ubiquitination. In this study, the X-ray structure of a trapped E3-E2 ~NEDD8-target intermediate (RBX1-UBC1~NEDD8-CUL1-DCN1) is used to build computer models, and combined quantum mechanics/molecular mechanics (QM/MM) and molecular dynamics (MD) simulations are performed to investigate the catalytic mechanism of the NEDD8 transfer from E2 to the substrate. The roles of the active site residues are examined. It is demonstrated that either E117 or D143 from E2 might work as a general base catalyst to deprotonate substrate K720. During the nucleophilic attack from substrate K720 to the thioester bond linking E2 and NEDD8, the formation of a new isopeptide bond between K720 and NEDD8 and the breaking of the thioester bond are found to be concerted. Furthermore, K720 is also found to act as a general acid catalyst to protonate the leaving group of C111 from E2. The free energy barrier for nucleophilic attack is estimated to be 14 -15 kcal/mol.

Introduction

Neddylation is the process by which the ubiquitin-like(UBL) protein NEDD8 is ligated to its highly selective substrate cullin proteins(CULs) on a single conserved lysine.[1] NEDD8 ligation to CULs are responsible for about 10% - 20% of all cellular ubiquitination and ubiquitination-like modification.[2] It is essential in eukaryotes[3,4] and regulates import biological processes such as signal transduction[5], cell division[6,7] and development[8].

Neddylation involves a catalytic cascades with E1, E2 (UBC12) and RING E3(RBX1) enzymes. E1 activates NEDD8 and loads it onto the catalytic cysteine from E2, forming a transient thioester-bound E2~NEDD8 intermediate.[9] RING E3 then recruits and loosely connects the substrate CULs with the E2~NEDD8 intermediate, promoting the transfer of NEDD8 at its C-terminal to the acceptor lysine on CULs.[1] The determination of the active configuration for NEDD8 ligation to the substrate CUL1 was recently achieved through crystallization of the RBX1(E3)-UBC12(E2)~NEDD8-CUL1(Substrate)-DCN1(Co-E3) complex(PDB ID: 4P5O,see Figure 3.1).[1] Similar to many processes of ubiquitin(UB) and UBL ligation to substrates[10,11] , RBX1 utilizes its N-terminal domain(NTD) as a substrate-binding domain to anchor in CUL1's C-terminal domain(CTD), and then binds and activates the UBC12~NEDD8 intermediate[1]. Different from some other UB and UBL ligation, RBX1 also needs a Co-E3, DCN1, which binds CUL1 and UBC12's NTD, to enhance the ligation.[12,13] However, RBX1 and its assistant DCN1

are both distal from the active site with a acceptor lysine (K720) on CUL1 where the ligation occurs (see Figure 3.1).[1]

The E3-E2~NEDD8-CUL1-DCN1 crystal structure was obtained by incorporating three mutations, K720R(CUL1), C111S(E2) and N103S(E2).[1] Those mutations were introduced to make the acceptor K720 less active and generate a more stable oxyester bond (as compared to the highly activated thioester bond) between E2 and NEDD8, thus preventing NEDD8 transfer and making crystallization of the NEDD8 ligation intermediate possible. K720 is the acceptor lysine from CUL1. C111 is the catalytic cysteine from E2 which forms the thioester bond with the backbone carbonyl carbon of C-terminal G76 from NEDD8. N103 of E2 is a highly conserved residue that is believed to play a catalytic role of stabilizing the oxyanion intermediate formed during the formation of the isopeptide bond between K720 and G76.[14,15] Although structural and activity analysis through mutagenesis have determined functions of some important residues in the active site during NEDD8 ligation[1,13, 16], the dynamic behaviors of the active site residues and their roles in determining the energy barriers on the ligation reaction pathway could be further understood by QM/MM molecular dynamics(MD) and free energy simulations.

Previous computational study of ubiquitination process involving E3-E2~UB-SUMO2 has highlighted two key events during the catalysis of ubiquitin transfer from E2 to substrate SUMO2, including the deprotonation of substrate lysine and

the following nucleophilic attack.[17] According to the study, the deprotonation of the substrate lysine by D117 on E2 occurs first with almost no energy barrier. The activated lysine then approaches the thioester bond via a conformational change and performs the nucleophilic attack on the thioester bond between UB and E2, leading to the formation of a tetrahedral intermediate. Finally, the nucleophilic elimination results in the cleavage of the thioester bond and formation of isopeptide bond between UB and substrate SUMO2. Here we examine the ubiquitin-like neddylation catalysis based on the structure of E3-E2~NEDD8-CUL1-DCN1 complex obtained in the previous investigations[1] by use of quantum mechanical/molecular mechanical (QM/MM) MD and free energy (potential of mean force) simulations. The results show that either E117 or D143 from E2 could be a general base catalyst for the deprotonation of K720 from the substrate. Furthermore, the bond making process between Nε of K720 and Cα of G76 on NEDD8, and the breaking of thioester bond seem to be concerted without forming a stable tetrahedral intermediate as previously proposed for SUMO2 ligation. Moreover, K720 could potentially work as a general acid catalyst for the leaving C111 group from E2.

Methods

The initial coordinates for the reactant complex is based on the crystallographic complex (PDB ID: 4P5O)[1] of RBX1-UBC12~NEDD8-CUL1-DCN1 containing five chains: RING E3 ligase RBX1, E2 conjugase UBC12, E2 bound ubiquitin-like

NEDD8, ubiquitination substrate protein Cullin-1(CUL1) and the assistant Co-E3 DCN-like protein 1(DCN1) as shown in Figure 3.1. The mutation residues R720 from CUL1, S103 and S111 from UBC12 of the crystal structure were manually mutated back to the wild type K720, N103 and C111 in the software MOE and the sidechain conformations of these residues were determined by using the rotamer explorer in the program. Structural preparing including optimizing protonation states of titratable residues and flips of Asn, Gln and His were done by MOE as well. Two initial models were built, one has K720 with protonated side chain entering the active site and the other has K720 with deprotonated side chain mimicking the reactant state of nucleophilic attack. The QM/MM MD and free energy (potential of mean force, PMF) simulations were performed to determine the free energy profiles as well as protein structures at different stages of the reaction with the CHARMM program[18]. Water spheres based on a modified TIP3P water model[19] with radius(r) of 30 Å and centered at Cβ were used for solvation. A stochastic boundary with a Poisson−Boltzmann charge-scaling scheme was applied[20]. The reservoir region has $r > 22$ Å, and the buffer region has $20 \leq r \leq 22$ Å. The reaction region has $r \leq 20$ Å. The side chain of C111(E2), N103(E2), K720(CUL1), E117(E2), D143(E2), the backbone carbonyl part of G75(NEDD8) and the full G76(NEDD8) were treated by QM and the rest of the system by MM. An all hydrogen potential function (PARAM27)[21] was used for the MM region, and the self-consistence charge density functional tight binding (SCC-DFTB)

method[22,23,24] was used for the QM region. The link-atom approach was used to separate the QM and MM regions.[25]

To build the initial structural model to simulate the deprotonation of substrate K720, K720 was chosen with its side chain in protonated state. To simulate the following nucleophilic attach, the model was modified by only neutralizing the lysyl side chain of K720. The resulting systems contain about 5500 atoms with approximately 700 water molecules. The initial structures for the entire stochastic boundary systems were optimized using the steepest descent (SD) and adopted-basis Newton−Raphson (ABNR) methods. The systems were gradually heated from 50.0 to 310.15 K in 50 ps. A 1-fs time step was applied for the integration of equation of motion, and the coordinates were saved every 50 fs. Initially, 5 ns QM/MM MD simulations were executed for each of the reactant complexes.

The reaction coordinate for the deportation process was defined as a linear combination of $d_{(N\varepsilon\ldots H)}$ and $d_{(O\ldots H)}$ as $RC = d_{(N\varepsilon\ldots H)} - d_{(O\ldots H)}$ where O is the side chain carboxyl oxygen of E117 or D143, H and Nε are from the lysyl side chain of K720. The umbrella sampling method[26], implemented in the CHARMM program along with the Weighted Histogram Analysis Method (WHAM)[27,28] was applied to determine the free energy (PMF) as a function of the reaction coordinate(s) for the proton transfer process. For each proton transfer process, twelve simulation windows were used, and for each window 50 ps production runs were performed

after 50 ps equilibration. The force constants of the harmonic biasing potentials used in the PMF simulations were 400 kcal mol^{-1} $Å^{-2}$. The 2D free energy map (PMF) for the nucleophilic attack was also determined with the umbrella sampling method and two-dimensional WHAM. The time step for QM/MM MD simulation is 1 fs. One reaction coordinate is RC1 = $d_{(C\alpha...S)}$, describing the bond breaking between Cα of G76(NEDD8) and Sγ of C111(E2). The other reaction coordinate is RC2 = $d_{(N\varepsilon \ldots C\alpha)}$ which describes the bond making between Cα of G76(NEDD8) and Nε of K720(substrate CUL1). 416 windows were used in the calculation of 2D PMF, and 100 ps production runs were preformed after 50 ps equilibration for each window. The structures from the second 50 ps out of the 100 ps production runs were saved for each window. The force constants for each window were in the range of 100−500 kcal mol^{-1} $Å^{-2}$ for both RC1 and RC2.

Results and Discussion

The E3-E2~NEDD8-CUL1-DCN1 complex after QM/MM MD simulations

The crystal structure of E3-E2~NEDD8-CUL1-DCN1 complex (PDB ID:4P5O) introduced three mutations, K720R (CUL1), N103S (E2) and C111S (E2), for the purpose of suppressing neddylation and obtaining a stable intermediate state that resembles RING E3 tapped in action of mediating NEDD8 ligation.[1] To investigate the molecular mechanism of NEDD8 transfer, the reverse mutations of these three

residues were employed to the initial configuration of the enzyme-substrate complex to build the simulation models.

The models with K720 in deprotonated state and protonated state were built respectively. The pK_a of lysine side chain in solution is 10.5 which makes it easy to be positively charged in physiological environment. The protonated state of K720 was further confirmed by MOE when preparing the structure for modeling. During the UB ligation process, it has been demonstrated both experimentally[11,29] and computationally[17] that an acidic residue, either glutamate or aspartate in the active site, could catalyze the deprotonation of the receptor lysine.

After 1 ns QM/MM MD simulations, the model with protonated K720 has two acidic residues in their deprotonated state that could potentially take the proton from K720. As shown in Figure 3.2, the ε-amino group of K720 is 3.9 angstrom away from the carboxyl group of D143 from E2, and 3.5 angstrom away from the carboxyl group of E117 from E2. The distance between Nε of K720 and the carbonyl C of thioester bond is 4.2 angstrom. However, the ε-amino group of K720 is not pointing to the C atom with a relatively flat angle, indicating it is not in a good near-attack conformation for the nucleophilic attack on the thioester bond. Moreover, a key interaction observed in UB ligation was recapped through our simulations on NEDD8 ligation: i.e., the side chain of N103 from E2 has one of the N-H bonds pointing with a distance of 2.2 angstrom to the carbonyl O by the reactive thioester

bond, establishing a hydrogen bonding interaction. The catalytic role of asparagine neighboring the thioester bond that helps to stabilize the oxyanion intermediate during UB ligation was discussed previously.[14,17,30,31] Y774 is shown in the crystal structure that aligns and directs CUL1's acceptor K720R.[1] Our simulations show there seems to be a strong pi-cation interaction between the positively charged ε-amino group of K720 and substrate Y774. As shown in Figure 3.2, the distance between Nε of K720 and aromatic side chain of Y774 is only 3.7 angstrom, and the lysyl side chain is parallel to the plane of aromatic side chain. Furthermore, the positively charged side chains of R74 from NEDD8 and R717 from CUL1 are all within 4 angstrom of the negatively charged carboxyl side chain of D143, indicating the position of D143 side chain could be potentially adjusted by those two arginines through salt bridge interaction(s).

After 1 ns QM/MM MD simulations(shown in Figure 3.3), the model with deprotonated K720 shows that the ε-amino group of K720 is approaching near-attack conformation where the distance between Nε of K720 and C of thioester bond is 3.5 angstrom and the lone-pair electrons of Nε points directly to the thioester bond. In the meantime, N103 is still in good position that helps to stabilize the carbonyl O bound to the thioester C through hydrogen bonding. The distance between N103 side chain H and the carbonyl O is 2.2 angstrom. Compared with the simulation results for the model with protonated K720, Y774 is further away from the neutral K720 side chain (5.2 angstrom vs 3.7 angstrom), possibly because

the pi-cation interaction no longer exists with a neutral ε-amino group. Furthermore, it also seems that the carboxyl side chain of D143 is moving away from K720 lysyl side chain and forms a stable salt bridge with R717 outside of the active site. However, E117 seems to still maintain a weak hydrogen bonding with the neutral ε-amino group with a distance of 3.4 angstrom.

Deprotonation of K720 could be catalyzed by either D143 or E117

As discussed before, there are two acidic residues, D143 and E117 in the active site that might potentially deprotonate K720. Interestingly the D143A mutation did not substantially affect neddylation involving UBC12 as E2[1], although the corresponding Asp is essential for SUMO ligation through UBC9 (E2)[30] and UB ligation through UBCH5(E2)[11]. It's also worth noting that when Asp does not exist on Ube2S (E2), a Glu from UB might instead play the catalytic role of Asp for deprotonation, and also help position the acceptor lysine in a good orientation for UB ligation.[32] Therefore, the question is whether D143 or E117 might be able to deprotonate K720.

QM/MM MD simulations showed that the D143 and K720 might form a stable salt bridge interaction during the initial substrate binding as shown in Figure 3.4A and 3.4C. The distance between carboxyl side chain of D143 and Nε of K720 is maintained at around 3 ~ 4 angstrom, and that between E117 and K720 is at around 6 ~ 8 angstrom. After 0.8 ns simulations (shown in Figure 3.4B, C), those two distances seem to converge at around 3 ~ 4 angstrom. Therefore, we would

propose that D143 might induce the initial binding of acceptor K720 to the active site while E117 might not.

Then free energy simulations were applied to compare the two potential deprotonation pathways. The proton transfer from K720 to E117 has almost no free energy barrier(1-2 kcal/mol)(Figure 3.5D). So does the proton transfer from K720 to D143 (Figure 3.6D). The average structures and free energies for the simulation windows of the reactant complex, near transition state and product state are shown in Figure 3.5A,B,C and Figure 3.6A,B,C for the deprotonation processes through E117 and D143 pathways, respectively. Altogether, the simulation results showed that either E117 or D143 could be potential general base catalyst for the deprotonation of K720. It seems that E117 might substitute the catalytic role of D143 when D143 is absent. This helps to explain why D143A mutation can't cause failure of neddylation in experiment observation.

Nucleophilic attack on the thioester bond by neutral K720 during neddylation

The nucleophilic attack in neddylation might involve three chemical steps, 1) breaking of the thioester bond between Cα and Sγ; 2) formation of isopeptide bond between Cα and Nε; 3) proton transfer from Nε to Sγ. However, these chemical steps still need to be confirmed, and the order of these steps and change of the active site structure along the reaction path have not been well understood.

The two-dimensional (2D) QM/MM MD and free energy simulations in this study led to the proposal that the breaking of the thioester bond is in concert with the making of the isopeptide bond during the nucleophilic attack. And the transition state occurs when the nucleophilic attack has already completed, but the proton has not been transferred from Nε to Sγ(see figure 3.7). The free energy path can be designated on the 2D free energy map following A -> B -> C -> D with an estimated energy barrier of 14 -15 kcal/mol(see Figure 3.8).

The average active site structure of the four key steps on the reaction path are exhibited in Figure 3.7. The reaction path starts from a free-energy favored pre-reactant complex (Figure 3.7A) which is similar to the structures obtained from MD simulations (Figure 3.3). In the pre-reactant complex, the distance between Nε of K720 and Cα of C111, $d_{(N\varepsilon\ldots C\alpha)}$, is 3.41 angstrom. The lone-pair electrons of Nε appears to point to the thioester bond, but not in good alignment with the electrophilic Cα. The amino group from the N103 side chain helps to stabilize the carbonyl O bound to Cα with a distance of 2.2 angstrom between H of the amino group and the carbonyl O. A weak hydrogen bond between Nε of K720 and oxygen of E117 side chain might help to position Nε for a better near attack orientation. After climbing up about 2 kcal/mol from point A to point B on the free energy map as shown in Figure 3.8, the reaction has reached a state of the reactant complex when the nucleophilic Nε is in a near attack position that points to Cα with a distance of 2.11 angstrom (Figure 3.7B). As shown in Figure 3.7B, the lone-pair

electrons of Nε also appears to be in good alignment with Cα. In the meantime, the carboxyl side chain of E117 seems not to be necessary for the nucleophilic attack and moves to a distance of 5.3 angstrom away from K720 side chain.

The nucleophilic attack reaches its transition state with the distance between Cα and Sγ, $d_{(C\alpha\ldots S\gamma)} \approx 2.7$ angstrom, and $d_{(N\varepsilon\ldots C\alpha)} \approx 1.6$ angstrom, as indicated in point C(Figure 3.8). The average structure of the active site near transition state is shown in Figure 3.7C. It shows that the thioester bond ($d_{(C\alpha\ldots S\gamma)} = 2.67$ angstrom) is broken and the isopeptide bond ($d_{(N\varepsilon\ldots C\alpha)} = 1.55$ angstrom) is formed in the near-transition configuration. In the meantime, Sγ from C111 is not bound to Cα and has become negatively charged in the current state. After the transition state, the reaction proceeds to the product state where a proton has transferred to the negative charged Sγ of C111(see Figure 3.7D and point D in Figure 3.8). In the process, K720 is proposed to act as a general acid catalyst to protonate the leaving group of C111.

The process of nucleophilic attack was proposed for UB ligation in a recent computational study.[17] Based on its QM/MM calculations on the potential energy surface (PES) of UB ligation, an oxyanion intermediate was found to exist on the reaction path in which Cα and Nε had formed a covalent bond while the thioester bond was still unbroken. This would suggest that the breaking of the thioester bond between Cα and Sγ occurs after the formation of isopeptide between Nε and Cα.

However, our study showed that the breaking of the thioester bond and the formation of the isopeptide bond during the NEDD8 ligation process would adopt a concerted way in which those two chemical events are highly synchronous. As shown in the 2D free energy map (Figure 3.8), point E corresponding to the oxyanion intermediate (where $d_{(N\varepsilon\ldots C\alpha)}$ = 1.7 angstrom and $d_{(C\alpha\ldots S\gamma)}$ = 2.0 angstrom) apparently deviates from the proposed reaction path of A->B->C->D. The difference between our study and the recent published computational study is possibly due to the fact that nucleophilic attack during NEDD8 ligation is slightly different from that during UB ligation, i.e., E117 might help adjust Nε of K720 with a better near-attack orientation attack through a weak hydrogen bonding interaction while such interaction is absent during the UB ligation. Another explanation might be that the reaction path for UB ligation was estimated based on mapping the potential energy surface which only considers potential energy changes, while the reaction path for NEDD8 ligation was discovered on a free energy map.

Conclusions

In this study, QM/MM MD and free energy simulations were employed to refine the mechanistic insights of RING E3 catalyzed NEDD8 transfer from E2 to substrate CUL1(see Figure 9). Either one of the acidic residues, E117 and D143, might work as the general base catalyst to deprotonate K720 with almost no free energy barrier. During the nucleophilic attack from the substrate K720 to the thioester bond between NEDD8 and E2, the formation of new isopeptide bond between

K720 and G76, and the breaking of thioester bond were found to be highly concerted. Moreover, K720 was found to transfer a proton to the leaving group of C111 after reaching the transition state, leading to the release of the NEDD8-bound substrate CUL1.

Acknowledgement

We thank Professor Martin Karplus for a gift of the CHARMM program. We are also grateful for the computer resources (Newton) from University of Tennessee, Knoxville.

References

1. Scott, D. C.; Sviderskiy, V. O.; Monda, J. K.; Lydeard, J. R.; Cho, S. E.; Harper, J. W.; Schulman, B. A., Structure of a RING E3 trapped in action reveals ligation mechanism for the ubiquitin-like protein NEDD8. *Cell* **2014,** *157* (7), 1671-84.

2. Soucy, T. A.; Smith, P. G.; Milhollen, M. A.; Berger, A. J.; Gavin, J. M.; Adhikari, S.; Brownell, J. E.; Burke, K. E.; Cardin, D. P.; Critchley, S.; Cullis, C. A.; Doucette, A.; Garnsey, J. J.; Gaulin, J. L.; Gershman, R. E.; Lublinsky, A. R.; McDonald, A.; Mizutani, H.; Narayanan, U.; Olhava, E. J.; Peluso, S.; Rezaei, M.; Sintchak, M. D.; Talreja, T.; Thomas, M. P.; Traore, T.; Vyskocil, S.; Weatherhead, G. S.; Yu, J.; Zhang, J.; Dick, L. R.; Claiborne, C. F.; Rolfe, M.; Bolen, J. B.; Langston, S. P., An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. *Nature* **2009,** *458* (7239), 732-6.

3. Osaka, F.; Saeki, M.; Katayama, S.; Aida, N.; Toh-e, A.; Kominami, K.; Toda, T.; Suzuki, T.; Chiba, T.; Tanaka, K.; Kato, S., Covalent modifier NEDD8 is essential for SCF ubiquitin-ligase in fission yeast. *Embo J* **2000,** *19* (13), 3475-3484.

4. Tateishi, K.; Omata, M.; Tanaka, K.; Chiba, T., The NEDD8 system is essential for cell cycle progression and morphogenetic pathway in mice. *The Journal of cell biology* **2001,** *155* (4), 571-9.

5. Read, M. A.; Brownell, J. E.; Gladysheva, T. B.; Hottelet, M.; Parent, L. A.; Coggins, M. B.; Pierce, J. W.; Podust, V. N.; Luo, R. S.; Chau, V.; Palombella, V.

J., Nedd8 modification of cul-1 activates SCF(beta(TrCP))-dependent ubiquitination of IkappaBalpha. *Molecular and cellular biology* **2000,** *20* (7), 2326-33.

6. Liakopoulos, D.; Doenges, G.; Matuschewski, K.; Jentsch, S., A novel protein modification pathway related to the ubiquitin system. *Embo J* **1998,** *17* (8), 2208-2214.

7. Kurz, T.; Pintard, L.; Willis, J. H.; Hamill, D. R.; Gonczy, P.; Peter, M.; Bowerman, B., Cytoskeletal regulation by the Nedd8 ubiquitin-like protein modification pathway. *Science* **2002,** *295* (5558), 1294-8.

8. Pozo, J. C.; Timpte, C.; Tan, S.; Callis, J.; Estelle, M., The ubiquitin-related protein RUB1 and auxin response in Arabidopsis. *Science* **1998,** *280* (5370), 1760-3.

9. Huang, D. T.; Hunt, H. W.; Zhuang, M.; Ohi, M. D.; Holton, J. M.; Schulman, B. A., Basis for a ubiquitin-like protein thioester switch toggling E1-E2 affinity. *Nature* **2007,** *445* (7126), 394-8.

10. Gareau, J. R.; Reverter, D.; Lima, C. D., Determinants of small ubiquitin-like modifier 1 (SUMO1) protein specificity, E3 ligase, and SUMO-RanGAP1 binding activities of nucleoporin RanBP2. *The Journal of biological chemistry* **2012,** *287* (7), 4740-51.

11. Plechanovova, A.; Jaffray, E. G.; Tatham, M. H.; Naismith, J. H.; Hay, R. T., Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **2012,** *489* (7414), 115-20.

12. Kim, A. Y.; Bommelje, C. C.; Lee, B. E.; Yonekawa, Y.; Choi, L.; Morris, L. G.; Huang, G.; Kaufman, A.; Ryan, R. J.; Hao, B.; Ramanathan, Y.; Singh, B., SCCRO (DCUN1D1) is an essential component of the E3 complex for neddylation. *The Journal of biological chemistry* **2008,** *283* (48), 33211-20.

13. Huang, G.; Kaufman, A. J.; Ramanathan, Y.; Singh, B., SCCRO (DCUN1D1) promotes nuclear translocation and assembly of the neddylation E3 complex. *The Journal of biological chemistry* **2011,** *286* (12), 10297-304.

14. Wu, P. Y.; Hanlon, M.; Eddins, M.; Tsui, C.; Rogers, R. S.; Jensen, J. P.; Matunis, M. J.; Weissman, A. M.; Wolberger, C. P.; Pickart, C. M., A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *Embo J* **2003,** *22* (19), 5241-5250.

15. Berndsen, C. E.; Wiener, R.; Yu, I. W.; Ringel, A. E.; Wolberger, C., A conserved asparagine has a structural role in ubiquitin-conjugating enzymes. *Nat Chem Biol* **2013,** *9* (3), 154-156.

16. Calabrese, M. F.; Scott, D. C.; Duda, D. M.; Grace, C. R. R.; Kurinov, I.; Kriwacki, R. W.; Schulman, B. A., A RING E3-substrate complex poised for ubiquitin-like protein transfer: structural insights into cullin-RING ligases. *Nature structural & molecular biology* **2011,** *18* (8), 947-949.

17. Zhen, Y.; Qin, G.; Luo, C.; Jiang, H.; Yu, K.; Chen, G., Exploring the RING-catalyzed ubiquitin transfer mechanism by MD and QM/MM calculations. *PloS one* **2014,** *9* (7), e101663.

18. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J Comput Chem* **1983,** *4* (2), 187-217.

19. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **1983,** *79* (2), 926-935.

20. Brooks, C. L.; Brunger, A.; Karplus, M., Active-Site Dynamics in Protein Molecules - a Stochastic Boundary Molecular-Dynamics Approach. *Biopolymers* **1985,** *24* (5), 843-865.

21. 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., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **1998,** *102* (18), 3586-3616.

22. Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G., Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys Rev B* **1998,** *58* (11), 7260-7268.

23. Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M., A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J Phys Chem B* **2001,** *105* (2), 569-585.

24. Gaus, M.; Cui, Q. A.; Elstner, M., DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J Chem Theory Comput* **2011,** *7* (4), 931-948.

25. Field, M. J.; Bash, P. A.; Karplus, M., A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations. *J Comput Chem* **1990,** *11* (6), 700-733.

26. Torrie, G. M.; Valleau, J. P., Monte-Carlo Free-Energy Estimates Using Non-Boltzmann Sampling - Application to Subcritical Lennard-Jones Fluid. *Chem Phys Lett* **1974,** *28* (4), 578-581.

27. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J Comput Chem* **1992,** *13* (8), 1011-1021.

28. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J Comput Chem* **1995,** *16* (11), 1339-1350.

29. Bernier-Villamor, V.; Sampson, D. A.; Matunis, M. J.; Lima, C. D., Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell* **2002,** *108* (3), 345-356.

30. Yunus, A. A.; Lima, C. D., Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. *Nature structural & molecular biology* **2006,** *13* (6), 491-499.

31. Reverter, D.; Lima, C. D., Insights into E3 ligase activity revealed by a SUMO-RanGAP1-Ubc9-Nup358 complex. *Nature* **2005,** *435* (7042), 687-692.

32. Wickliffe, K. E.; Lorenz, S.; Wemmer, D. E.; Kuriyan, J.; Rape, M., The mechanism of linkage-specific ubiquitin chain elongation by a single-subunit E2. *Cell* **2011,** *144* (5), 769-81.

Appendix

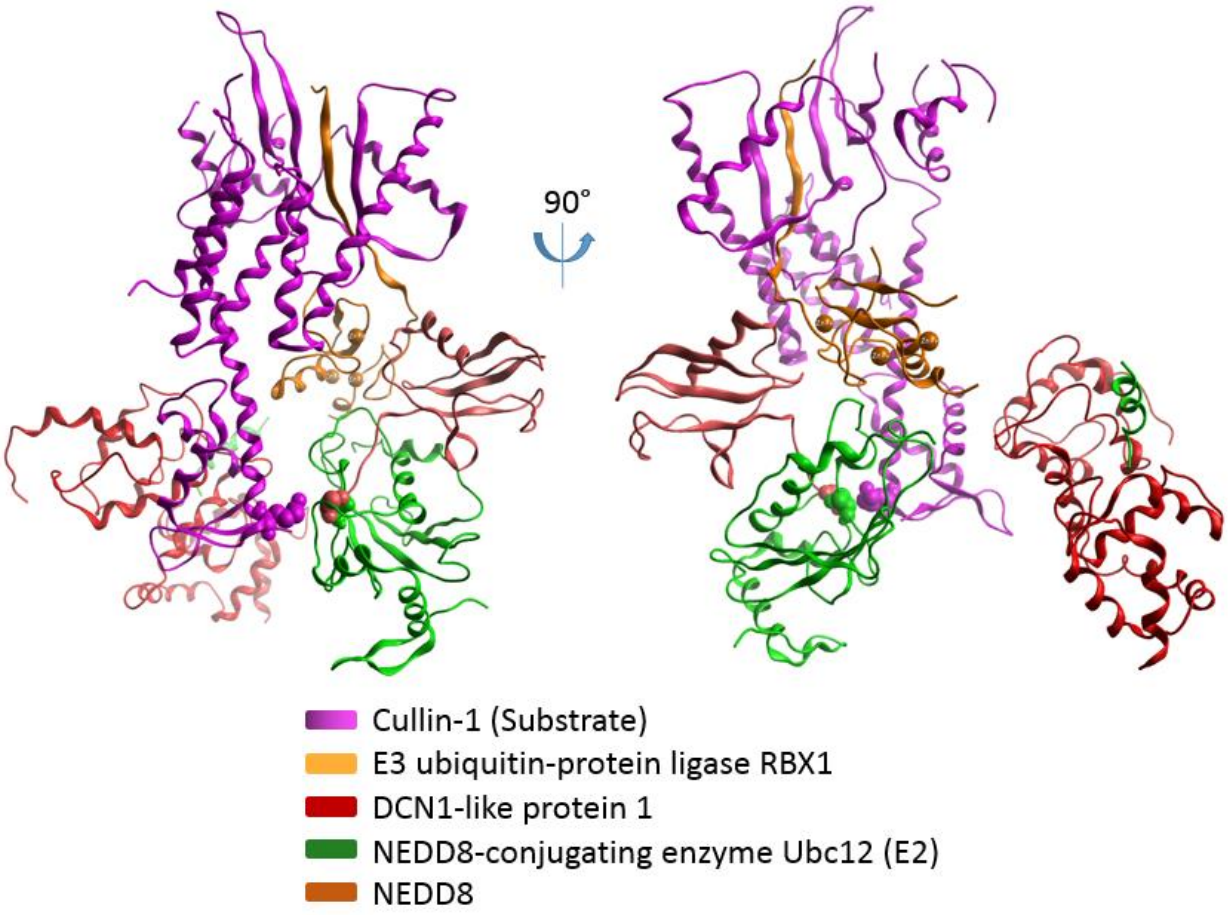


Figure 3.1 Crystal Structure of RBX1(E3)-UBC12(E2)~NEDD8(UBL)-CUL1(Substrate)-DCN1(CoE3) Complex

Modeled K720 from the substrate and C111 from E2 as well as G76 from UBL are shown in space fillings. The five chains are represented by ribbons with different colors.

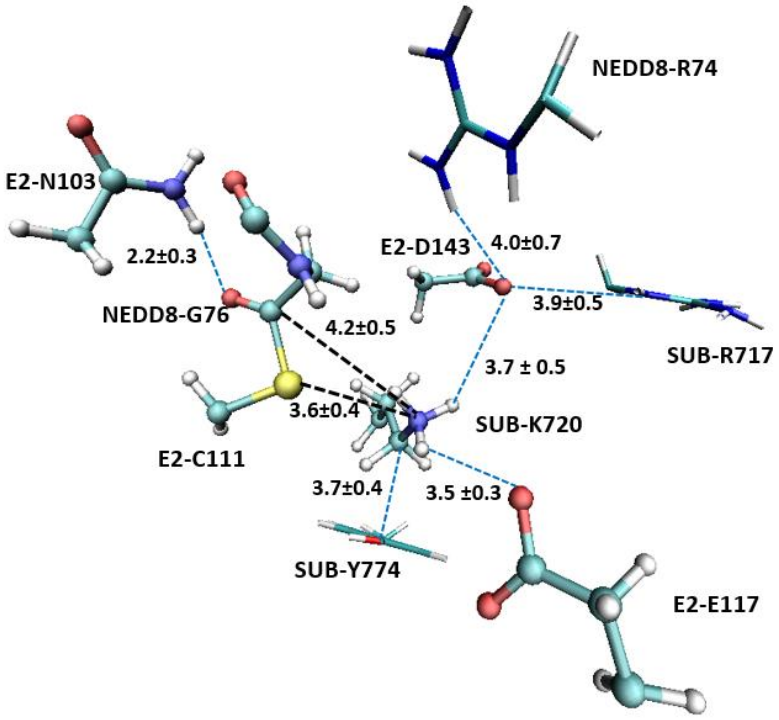


Figure 3.2 MD results of active-site structure of E3-E2~NEDD8-Substrate-CoE3 Complex

The results are generated after 1 ns simulations. The substrate K720 is in protonated state. The side chains of N103, C111, E117 and D143 from E2, the whole G76 and part of G75 from NEDD8, and substrate K720 are shown in ball and stick. Potential van der Waal's interactions and salt bridges are shown in blue dotted lines. And the related distances are given with standard deviations in angstrom. The average distances between Nε of K720 and Sγ OF C111, and between Nε of K720 and Cα of G76 are shown.

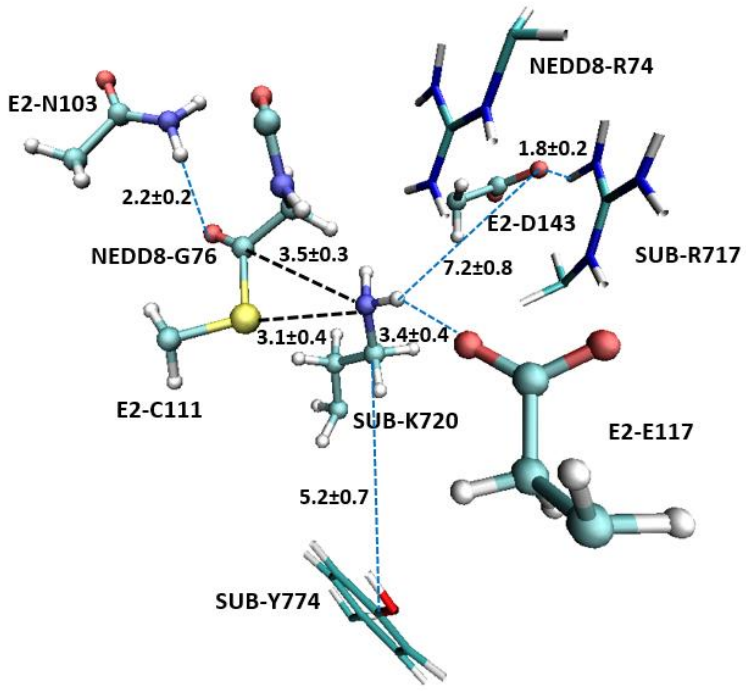


Figure 3.3 Average active-site structure of E3-E2~NEDD8-Substrate-CoE3 Complex with deprotonated K720

The side chains of N103, C111, E117 and D143 from E2, the whole G76 and part of G75 from NEDD8, and substrate K720 are shown in ball and stick. Potential van der Waal's interactions and salt bridges are shown in blue dotted lines. And the related distances are given with standard deviations in angstrom. The average distances between Nε of K720 and Sγ OF C111, and between Nε of K720 and Cα of G76 are shown.

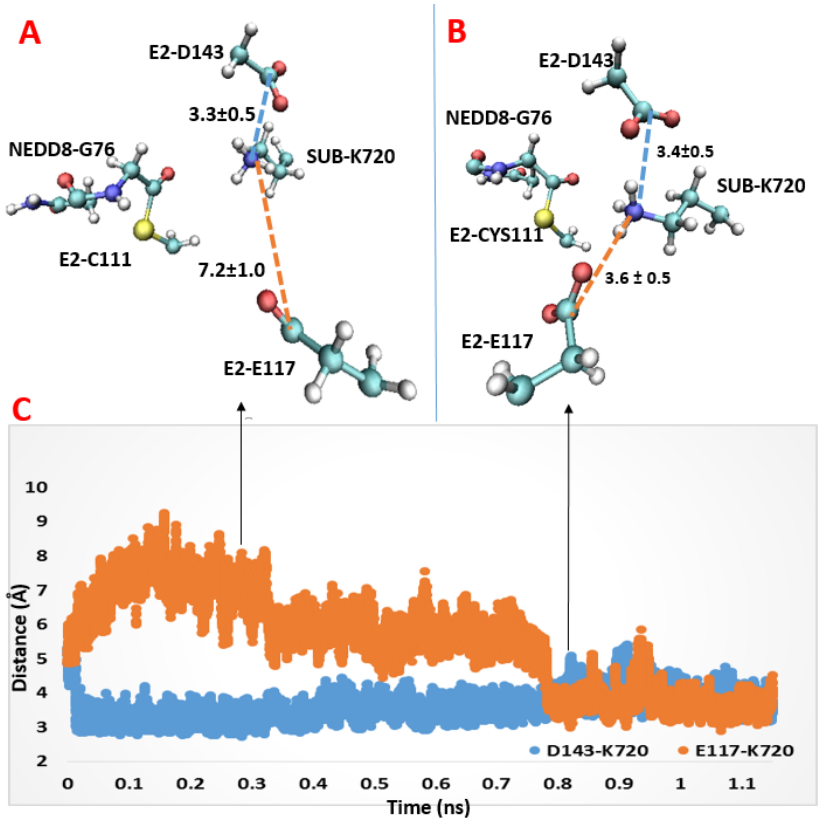


Figure 3.4 Two possible deprotonation pathways of substrate K720

(A) Average structure of the active site at around 0.3 ns MD simulation. (B) Average structure of the active site at around 0.8 ns MD simulation. The distance between the carboxyl carbon of D143 and side-chain amino nitrogen of K720 is shown with blue dotted lines in (A) and (B). The distance between carboxyl group of E117 and Nε of K720 is shown with orange dotted lines in (A) and (B). (C) Monitoring the distance between the carboxyl group of D143(in blue)/E117(in orange) and the amino group of K720 during MD simulations. The two arrows indicate when the screenshot of average structure in (A) and average structure in (B) are taken respectively.

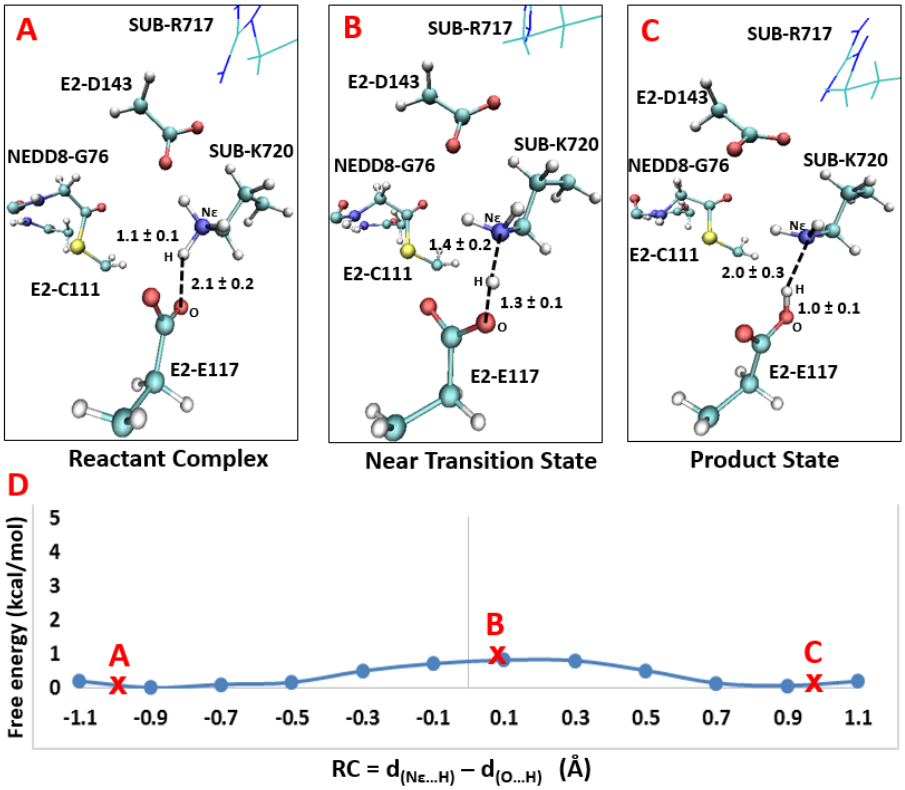


Figure 3.5 One possible deprotonation pathway through E117

Average structures of the reactant complex, near transition state and product state are shown in (A), (B) and (C) respectively. Some average distances are given with standard deviations in angstrom. (D) The free energy for the deprotonation reaction path involving E117. Point A, B and C designate where the average structures in (A), (B) and (C) are taken from the reaction path. The distance difference between Nε···H and O···H is the reaction coordinate used to describe the processing of deprotonation.

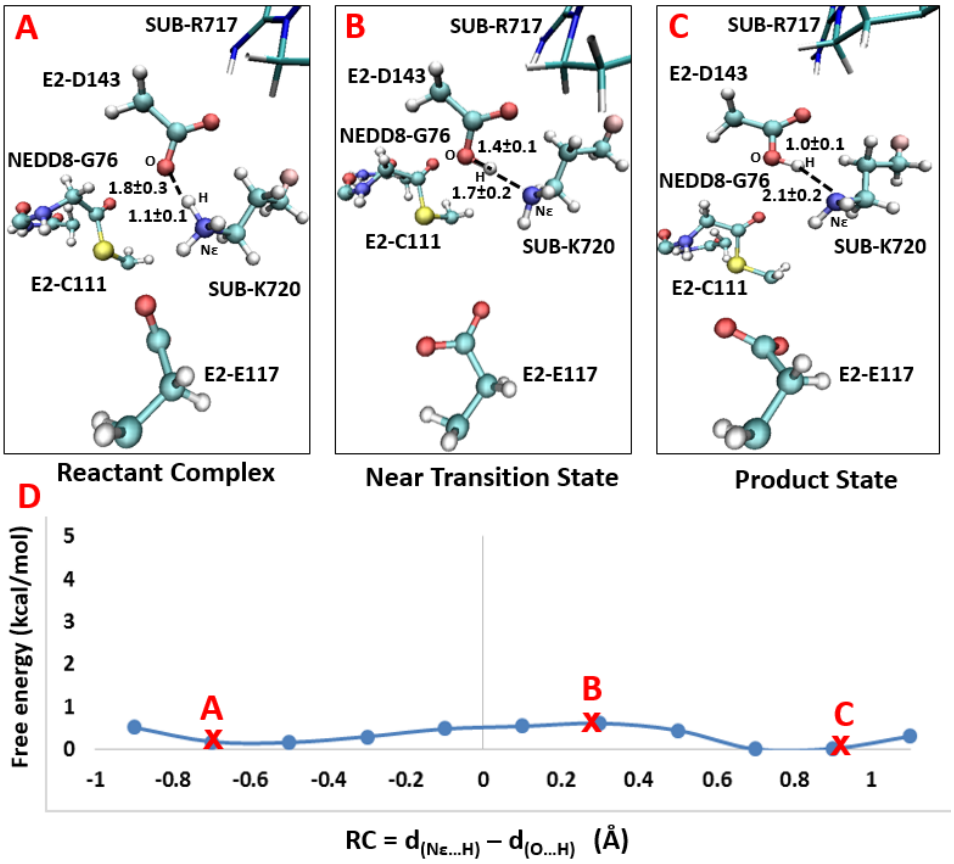


Figure 3.6 One possible deprotonation pathway through D143

Average structures of the reactant complex, near transition state and product state are shown in (A), (B) and (C) respectively. Some average distances are given with standard deviations in angstrom. (D) The free energy of the deprotonation reaction path involving D143. Point A, B and C designate where the average structures in (A), (B) and (C) are taken from the reaction path. The distance difference between Nε···H and O···H is the reaction coordinate used to describe the processing of deprotonation.

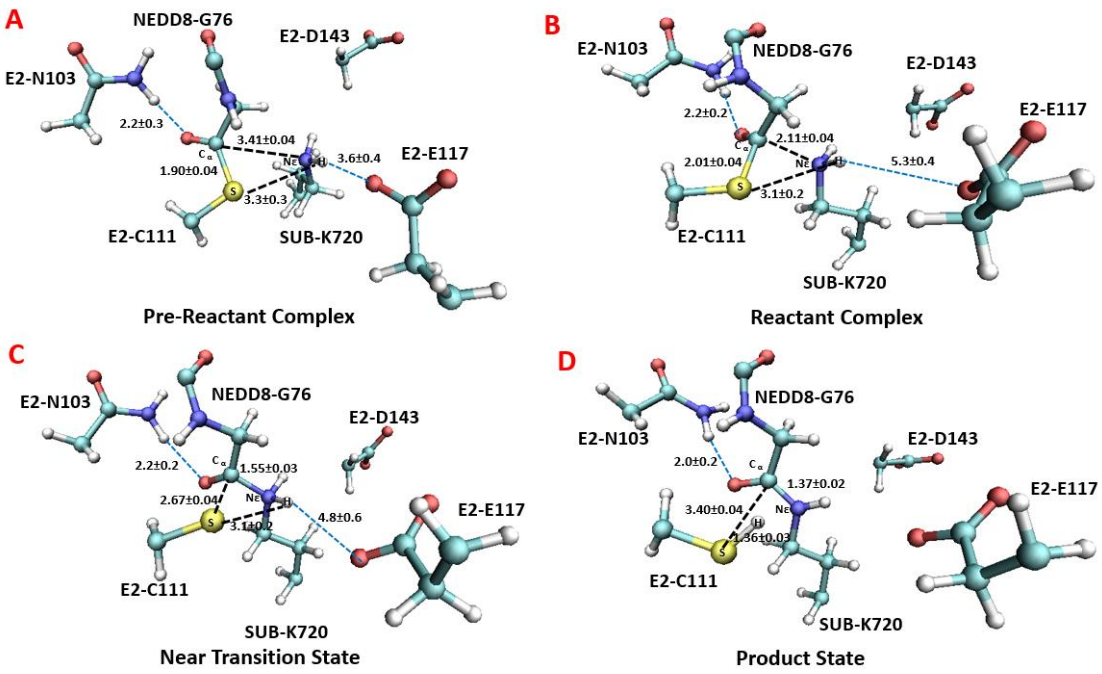


Figure 3.7 Active-site structure during the 2D free-energy simulations of the nucleophilic attack

The average structures around Point A (pre-reactant complex), Point B (reactant complex), Point C (near transition state) and Point D (Product Complex) as designated in Figure 3.8 are shown. Some average distances are given with units of angstrom.

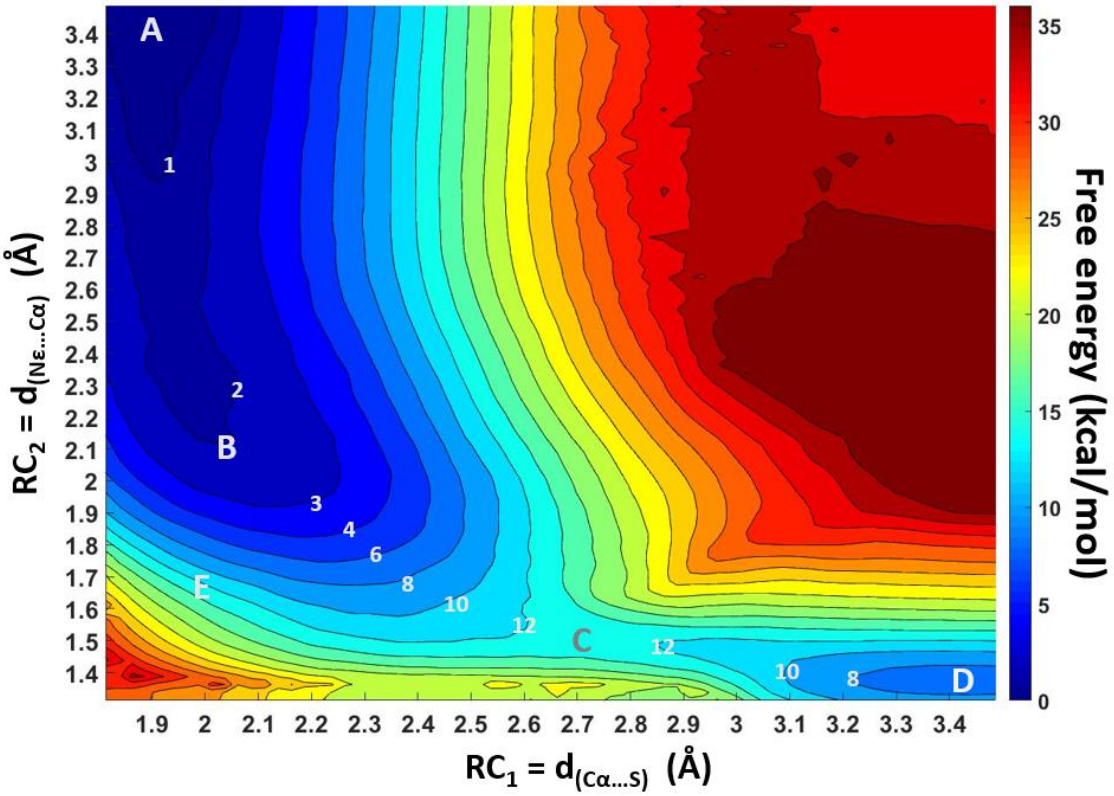


Figure 3.8 2D free-energy contour map for nucleophilic attack

The horizontal axis represents the distance between Cα and Sγ, and the vertical axis represents the distance between Nε and Cα. Points A, B, C and D designate the pre-reactant complex, reactant complex, near transition sate and product complex, respectively. Point E is the hypothetical oxyanion intermediate. Some contour lines are shown with respective energy value. The energy bar is shown on the right.

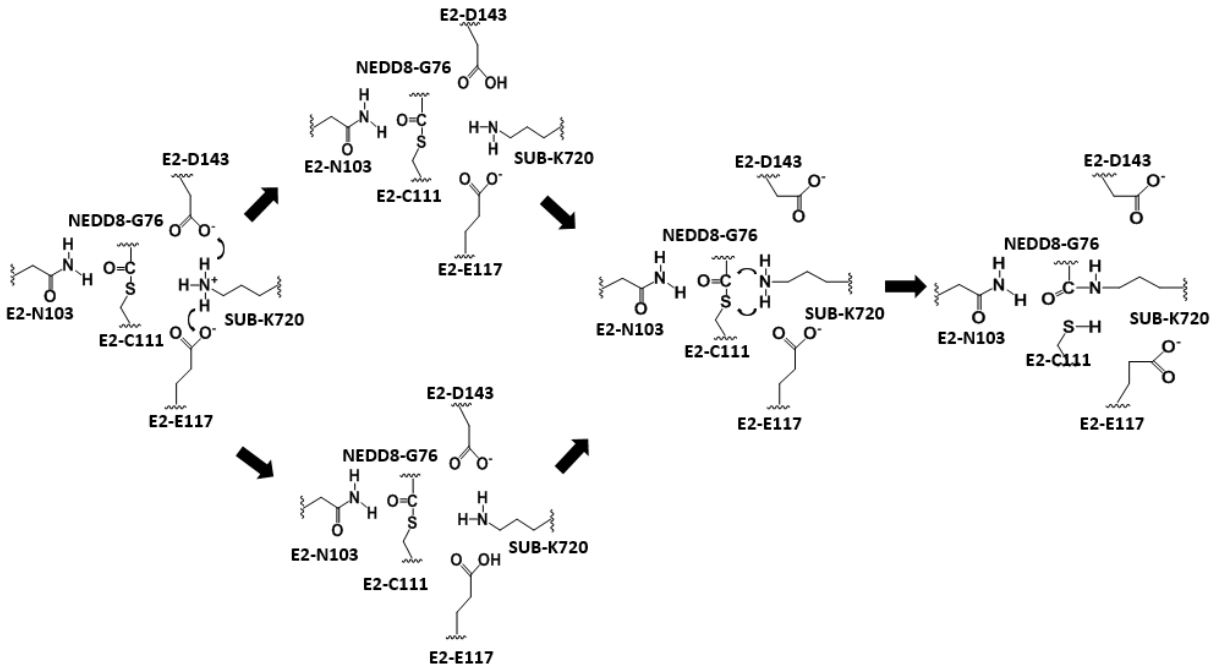


Figure 3.9 Proposed mechanism of NEDD8 transfer catalyzed in the E3-E2~NEDD8-Substrate-CoE3 complex

CONCLUSIONS

Summary

Understanding the catalytic mechanism of enzymes is of fundamental and practical importance. The dynamics during catalysis such as proton transfer and side chain movement, as well as synchronicity of concerted reactions, cannot be directly observed through structural analysis by X-ray crystallography combined with mutagenesis. QM/MM free energy simulations are capable of monitoring those motions and examining the making or breaking of covalent bonds. Moreover, the transition state structures as well as free energy barriers in the enzymatic reactions can be obtained.

In this dissertation, the QM/MM MD simulations, as well as free energy simulations through calculations of potential of mean force using umbrella sampling and the weighted histogram analysis method, were employed to study the catalytic mechanisms of two SAM dependent N-methyltransferase, DXMT and PRMT5, and the ubiquitination-like neddylation process in the RBX1(E3)-UBC12(E2)~NEDD8(UBL)-CUL1(Substrate)-DCN1(CoE3) complex.

Simulations of DXMT has monitored the proton transfer process from the hydroxyl group of both 7-methylxanthine (7mX) and theobromine (Tb) to the general base catalyst His160. It showed that the proton transfer was completed before the end of the methylation process. The timely proton transfer was shown capable of

lowering the free energy barrier for the 7mX methylation and for the Tb methylation by 3.7 and 12.7 kcal/mol, respectively. The two-dimensional(2D) free energy map describing both the proton transfer and methylation process further indicated that the proton transfer was completed before the transition state of methylation was reached, and the completion of proton transfer could thus promote the final attachment of the positive charged methyl group to the substrate. Furthermore, the product promiscuity seems to depend on the relative position and orientation of the side chain of the catalytic His160 with respect to the acceptor nitrogen of substrate 7mX/Tb as observed from the MD simulations.

The general base catalysis was also observed during the two-step symmetric methylation on substrate arginine catalyzed by PRMT5. A glutamate in the active site, Glu435, utilize its deprotonated carboxylic side chain to accept a proton from the substrate arginine during the 1^{st} and the 2^{nd} methylation. Furthermore, simulations have identified some key residues that contribute to the transition state stabilization. The transition states for the two-step methylations were shown to be both stabilized through strong double salt bridge interactions between the substrate arginine and the two neighboring Glu435 and Glu444 residues, while those interactions were less dominant in the reactant states. The simulations also illustrated the origin of product specificity for the symmetric di-methylation (SDMA) as compared to asymmetric di-methylation (ADMA). Unlike ADMA PRMTs which energetically favors a single ω-guanidino nitrogen ($N_{\eta 2}$) of arginine as the target

for both the 1st and 2nd methylations, SDMA PRMT5 was found to have an energetic preference of targeting $N_{\eta 1}$ for the first methylation and subsequently targeting a different ω-guanidino nitrogen ($N_{\eta 2}$) for the second methylation. The first and second methyl transfers were estimated to have the free energy barriers of 19–20 and 18–19 kcal/mol, respectively, which are consistent with the existing k_{cat} values.

In the simulations of neddylation in the E3-E2~NEDD8-Substrate-CoE3 complex, the synchronicity of two chemical processes, the breaking of the thioester bond and formation of the isopeptide bond, was observed from the 2D free energy map. The proceedings of the thioester bond breaking and isopeptide bond making were shown to be highly concerted, making the formation of an oxyanion intermediate unnecessary. Two residues in the active site, Glu117 and Asp143 from E2, were found to work as the potential general base catalyst to deprotonate the substrate lysine with almost no free energy barrier. Before reaching the product state, the substrate lysine seems to work as a general acid catalyst to transfer a proton to the leaving Cys111 from E2, releasing the newly formed Substrate~NEDD8 from E2. The relative free energy barrier is estimated to be 14-15 kcal/mol, which is consistent with the high reactive activity observed in experiments.

To sum up, the QM/MM MD and free energy simulations applied in this dissertation well support the importance of dynamics during enzymatic catalysis. In all the three

cases studied, the general acid/based catalyzed proton transfer appears to be capable of lowering the free energy barrier. The synchronicity of the proton transfer and main nucleophilic attack reactions were also investigated. The proton transfer was completed before the transition state was reached in the cases of N_3 and N_1 methylation by DXMT, and $N_{\eta 1}$ methylation by PRMT5. And its completion might also fall behind the transition state, such as during $N_{\eta 2}$ methylation by PRMT5 and during neddylation in the E3-E2~NEDD8-Substrate-CoE3 complex. Structure based analyses have led to the proposal that hydrogen bond interactions may be the key reason for the transition state stabilization. But the results from the simulations performed in this dissertation seem to suggest the previous proposal might not work for some cases.

Future directions

DXMT has a close homolog xanthosine methyltransferase (XMT) which converts xanthosine (XR) to 7mX by removing the ribose from XR and adding a methyl group to N_7 of XR. On the pathway of caffeine biosynthesis, XMT functions on the upstream of DXMT, producing 7mX as substrate for DXMT. XMT has an active-site similar to DXMT except the catalytic His160 in DXMT is substituted by a glutamine, Gln161 in XMT. The Q161H mutant could change the substrate specificity of XMT and make it able to catalyze both N_7 and N_3 methylation, while N_3 methylation is also catalyzed by DXMT. It would be interesting to simulate enzymatic catalysis by XMT and Q161H mutant of XMT to gain further understanding of product specificity of enzymes involved in caffeine biosynthesis.

It could help to answer question like whether theobromine biosynthesis occurs evolutionarily as precursor to biosynthesis of caffeine.

The QM/MM simulations has been proved working well for enzymatic catalysis in local regions, especially for describing active-site conformational change and chemical reactions. However, remote regions might also have its influence on enzymatic activity through allosteric regulation. In the E3-E2~NEDD8-Substrate-CoE3 complex, the E3 ligase and assistant CoE3 are not close to the active site of NEDD8 transfer. However, they play essential roles of recruiting and connecting the substrate cullin with the E2~NEDD8 intermediate, building a quaternary complex favored by the enzymatic reaction. With or without the existence of E3 and CoE3, a microsecond (ms) level MM MD simulations followed by QM/MM free energy simulations might help to obtain insights into how E3 and CoE3 allosterically maintain an ideal active site configuration for neddylation.

VITA

Yufei Yue was born in Chengdu, China in 1985. He attended Sichuan University in 2003 and graduated with a Bachelor of Science degree in Biotechnology in 2007. He continued his graduate study in Plant Genetics Lab at Sichuan University. In 2010, he joined the Department of Biochemistry, Cellular and Molecular Biology at the University of Tennessee Knoxville for his Ph.D. study. He then started pursing a Master of Science degree in Statistics in 2012. He is graduating in August 2016.