

**Understanding the Origins of Product Specificity of Protein
Methyltransferases from QM/MM MD and Free Energy
Simulations**

**A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville**

Yuzhuo Chu

May 2013

Copyright © 2013 by Yuzhuo Chu

All rights reserved.

DEDICATION

To my daughter

Olivia Cao

my husband

Qi Cao

my parents

Junxin Chu and Yumei Zhao

ACKNOWLEDGEMENTS

I wish to express my deep and sincere gratitude to all those who offer great support and help during my Ph.D. study.

My deepest gratitude goes first to my advisor, Dr. Hong Guo, for his constant encouragement, patient guidance and professional suggestions. His comprehensive knowledge, wonderful ideas and enthusiasm in science have led me through the completion of my dissertation.

I sincerely thank my dissertation committee members, Dr. Cynthia Peterson, Dr. Elizabeth Howell, Dr. Jerome Baudry, Dr. Xiaolin Cheng for their professional supports and valuable suggestions during my course of study and the completion of my dissertation.

I want to appreciate all the members of Dr. Hong Guo's lab and my friends at the University of Tennessee for their friendship and help, especially Dr. Qin Xu for his technical help. I also want to appreciate Dr. Guohui Li at Dalian Institute of Chemical Physics, Chinese Academy of Sciences for his valuable instructions and collaborations.

Finally, this work is not possible without the financial support from the department of Biochemistry, Cellular and Molecular Biology and the National Institute for Mathematical and Biological Synthesis at the University of Tennessee, and also a grant from National Science Foundation.

ABSTRACT

Protein lysine methyltransferases (PKMTs) catalyze the methylation of certain lysine residues on histone tails using S-adenosyl-L-methionine (AdoMet) as the methyl donor. Regulation of chromatin structure and gene expression through histone lysine methylation depends on the degree of methylation. Therefore, it is of importance to understand the features of PKMTs that control how many methyl groups would be added to the target lysine (product specificity). In my dissertation, I have applied quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) and free energy simulations to understand the origin of the product specificities of SET8, GLP and their mutants. The relative free energy barriers of the methyl transfers calculated from our simulations are consistent with experimental data on product specificity, suggesting that they may govern the product specificity of PKMTs energetically. Our simulation results on the effects of the Tyr/Phe switch mutations in SET8 and GLP and the Y1124F mutation of GLP suggest that the space around the target lysine may be of importance. Protein arginine methyltransferases (PRMTs) catalyze the methylation of peptidyl arginines using AdoMet as the methyl donor. The product specificity of PRMTs often refers to their ability to form either asymmetric dimethylarginine (ADMA) (type I PRMTs) or symmetric dimethylarginine (SDMA) (type II PRMTs), and different product formation could lead to different biological consequences. The QM/MM MD and free energy simulations have been performed on one member of type I PRMTs, PRMT3. The simulation results

suggest that the interactions of the active site residues make the same nitrogen atom of the guanidine to be the target of both 1st and 2nd methylations, leading to the formation of the asymmetric product. It is also proposed that Glu326 may function as a general base during the methyl transfer. The results of the MD simulations on the reactant complexes demonstrate that the ability of the reactant complexes to form the reactive configurations for the methyl transfer could be an important factor for the product specificity.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION	1
Histone Code and Protein Lysine Methylation.....	1
Product specificity of PKMTs	3
SET8.....	4
GLP	6
Protein Arginine Methylation	8
Quantum Mechanical/Molecular Mechanical (QM/MM) Molecular Dynamics (MD) and Free Energy Simulations	10
QM/MM MD simulations	10
Free energy (potential of mean force) change determinations	13
References	16
Appendix	23
CHAPTER 2 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?	30
Abstract.....	31
Introduction	32
Method	37
Results	43
Discussion.....	47

Conclusion	53
Acknowledgement.....	54
Supporting Information.....	54
References.....	55
Appendix	60
CHAPTER 3 QM/MM MD AND FREE ENERGY SIMULATIONS OF G9A-LIKE PROTEIN (GLP) AND ITS MUTANTS: UNDERSTANDING THE FACTORS THAT DETERMINE THE PRODUCT SPECIFICITY.....	
Abstract.....	73
Introduction	74
Method	78
Results	82
Discussion.....	86
Conclusion	89
References.....	90
Appendix	95
CHAPTER 4 QM/MM MD AND FREE ENERGY SIMULATIONS OF THE METHYLATION REACTIONS CATALYZED BY PROTEIN ARGININE METHYLTRANSFERASE PRMT3.....	
Abstract.....	107
Introduction	108
Method	110

Results	114
1st methylation on $N_{\eta 2}$ vs. $N_{\eta 1}$	114
2nd methylation on $N_{\eta 2}$ vs. $N_{\eta 1}$	116
1st and 2nd methylations on the $N_{\eta 2}$ atom	116
The roles of the “double E” residues	117
2D free energy simulations	119
Discussion	119
Conclusion	121
Acknowledgement	121
Supporting Information	122
References	123
Appendix	128
CHAPTER 5 CONCLUSION	148
References	157
VITA	158

LIST OF FIGURES

Figure 1.1. The post-translational modification sites on histone tails.	23
Figure 1.2. Biological processes associated with histone code.	24
Figure 1.3. PKMTs and methyl-lysine binding proteins.	25
Figure 1.4. Product specificity of PKMTs.	26
Figure 1.5. Active site structures of SET8 bound to AdoHcy and H4K20.....	27
Figure 1.6. Illustration of QM/MM approach in the simulations of enzyme catalysis.	28
Figure 1.7. General scheme for the partition of the system in Stochastic boundary method.....	29
Figure 2.1. The definition of the structural parameters for monitoring the relative orientation of AdoMet and H4K20me1 [H4K20 and H4K20(me)2] in the reactant complex.	60
Figure 2.2. MD results for the wild-type enzyme (SET8).....	61
Figure 2.3. MD results for Y334F.....	63
Figure 2.4. Free energy profiles for methyl transfers	65
Figure 2.5. Free energy simulations of the dissociation of the active-site water molecule from the active site	68
Figure S2.1. MD results for Y334F.....	70
Figure S2.2. Free energy changes along with the statistical errors.....	71

Figure 3.1. The definition of the structural parameters for monitoring the relative orientation of AdoMet and H3K9me1 [H3K9 and H3K9(me)2] in the reactant complex	95
Figure 3.2. MD results for the WT enzyme (GLP)	96
Figure 3.3. MD results for the F1209Y mutant	99
Figure 3.4. Free energy profiles of methyl transfer processes in WT GLP and the F1209Y mutant	101
Figure 3.5. MD results for the Y1124F mutant	102
Figure 3.6. Free energy profiles of methyl transfer processes in the Y1124F mutant.....	104
Figure S3.1. The two-dimensional plots of $r(C_M...N_\zeta)$ and θ distributions of the transition state complexes	105
Figure 4.1.....	128
Figure 4.2. Results from the MD simulations for 1st methylation on $N_{\eta 2}$	130
Figure 4.3. Results from the MD simulations for 1st methylation on $N_{\eta 1}$	131
Figure 4.4. Two-dimensional plots of $r(C_M-N_\eta)$ and θ distributions in the reactant complexes.....	132
Figure 4.5. Free energy profiles of 1st methyl additions to the $N_{\eta 2}$ and $N_{\eta 1}$ atoms, respectively.....	133
Figure 4.6. Results from the MD simulations for 2nd methylation on $N_{\eta 2}$	134
Figure 4.7. Results from the MD simulations for 2nd methylation on $N_{\eta 1}$	135

Figure 4.8. Two-dimensional plots of $r(C_M-N_\eta)$ and θ distributions of the reactant complexes.....	136
Figure 4.9. Free energy profiles of 2nd methyl additions to the N_{η_2} and N_{η_1} atoms	137
Figure 4.10. Free energy profiles of 1st and 2nd methyl additions to N_{η_2} ...	138
Figure 4.11. 2D free energy contours for the methyl transfer from AdoMet to methyl arginine and the proton transfer between methyl arginine and Glu326	139
Figure 4.12. The catalytic mechanism for the methyl additions	140
Figure S4.1. The model system used for comparison.....	141
Figure S4.2. The potential energy curve calculated from SCC-DFTB (green line) and the single-point energy curve calculated by B3LYP/6-31G ** (blue line)	142
Figure S4.3. MD simulations of the product complex of the 1st methylation on N_{η_2} after the deprotonation of Glu326	143
Figure S4.4. The channel opening from solution to Glu326	144
Figure S4.5. MD results for 1st methylation on N_{η_2} with shaking N_{η_2} -HH ₂₂ bond	145
Figure S4.6. MD results for 2nd methylation on N_{η_2} with shaking N_{η_2} -HH ₂₂ bond	146
Figure S4.7. Free energy profiles of 1st and 2nd methyl additions to N_{η_2} atoms with and without shaking the N_{η_2} -HH ₂₂ bond.....	147

LIST OF ABBREVIATIONS

PKMTs:	protein lysine methyltransferases
AdoMet:	S-adenosyl-L-methionine
AdoHcy:	S-adenosyl-L-homocysteine
QM/MM:	quantum mechanical/molecular mechanical
MD:	molecular dynamics
GLP:	G9a-Like-Protein
PRMTs:	protein arginine methyltransferases
ADMA:	asymmetric ω - N^G , N^G -dimethylarginine
SDMA:	symmetric ω - N^G , N^G -dimethylarginine
MMA:	ω - N^G -monomethylarginine
DFT:	density function theory
SCC-DFTB:	self-consistent charge density functional tight binding
PMF:	potential of mean force
WHAM:	Weighted Histogram Analysis Method
SD:	steepest descent
ABNR:	adopted-basis Newton-Raphson
WT:	wild type
TS:	transition state

CHAPTER 1 INTRODUCTION

Histone Code and Protein Lysine Methylation

Nucleosomes are the fundamental packaging units of eukaryotic chromatin. They are composed of 146 bp of DNA wrapping around a histone octamer containing two copies of each core histones, H2A, H2B, H3 and H4 (Luger, Mader et al. 1997; Watson, Baker et al. 2004). These histone proteins have positively charged amino-terminal tails which are exposed on the outside of nucleosomes. These tails are subject to a variety of post-translational covalent modifications, including acetylation, phosphorylation, ubiquitination, sumoylation and methylation (Turner 2005) (**Figure 1.1**) (all figures are in the appendix). These modifications are believed to constitute the “histone code”, which can be read by other protein and leads to distinct downstream events in the regulation of chromatin structure and gene expression (Strahl and Allis 2000; Jenuwein and Allis 2001; Lall 2007) (**Figure 1.2**).

Histone lysine methylation governs a series of important biological events, including heterochromatin formation, X-chromosome inactivation, transcriptional silencing and activation (Martin and Zhang 2005). Methylations have been found on a range of lysine (K) residues in different histones including K4, K9, K27, K36 and K79 in histone H3, K20 in histone H4, *etc.* The methylated lysine residues could further bind to different methyl-lysine binding proteins through their methyl-lysine binding domains (**Figure 1.3**). Several proteins responsible for the methylation of specific residues have been characterized, all but one of these

contain a SET domain and make up the SET-domain protein lysine methyltransferase (PKMT) family. The exception is the proteins in the DOT1 family, which methylate K79 in the globular region of histone H3 and are structurally not related to SET-domain-containing proteins (Feng, Wang et al. 2002).

The SET domain, which is approximately 130 amino acids long, was first identified as a conserved sequence in three *Drosophila melanogaster* proteins: the suppressor of variegation 3-9 (Su(var)3-9) (Tschiersch, Hofmann et al. 1994), the Polycomb-group chromatin regulator Enhancer of zeste (E(z)) (Jones and Gelbart 1993), and the trithorax-group chromatin regulator trithorax (Trx) (Stassen, Bailey et al. 1995). The SET domain PKMTs may be divided into the following families, SUV39, SET1, SET2, EZ, RIZ, SMYD, SUV4-20, and also orphan members like SET7/9 and SET8. Members from each family share a great level of similarity in the SET domain and also have similar motifs surrounding the SET domain (Dillon, Zhang et al. 2005).

The SET-domain PKMTs may transfer methyl group(s) from S-adenosyl-L-methionine (AdoMet) to the amino group of a lysine residue on the histone, leaving a methylated lysine residue and the cofactor byproduct S-adenosyl-L-homocysteine (AdoHcy) (Trievel 2004). The histone substrate and AdoMet bind at opposite faces of the SET domain, the substrate is deposited into a narrow channel in an extended β -strand conformation and the methylene part of the substrate lysine side chain makes van der Waals contacts to the side chains of

several aromatic residues, these interactions might guide the side chain of the target lysine into the channel (Zhang, Yang et al. 2003). The substrate binding channel runs through the core of the SET domain to reach AdoMet to allow the methyl transfer to occur (Dillon, Zhang et al. 2005). The special arrangement of the substrate binding sites has been proposed to permit the processive lysine methylation in some cases without releasing the protein substrate from the SET domain after each round of methylation (Trievel, Beach et al. 2002; Zhang, Yang et al. 2003).

Product specificity of PKMTs

In addition to the property to select different lysine targets for methylation (substrate specificity), PKMTs also differ in the ability to add one, two or three methyl groups to target lysine. This property is called product specificity (**Figure 1.4**). Specific methylation states have been shown to enrich in distinct regions of genes (Li, Carey et al. 2007). Different degrees of lysine methylation could be recognized by certain methyl-lysine binding domains to lead to different events (Taverna, Li et al. 2007). Thus it is important to understand the mechanism underlying the product specificities of these enzymes, which is the key focus of my study. Mutational studies have indicated a specific position within the lysine binding site which seems to control the product specificities of PKMTs. This position is frequently occupied by a tyrosine in the mono-methyltransferases and by a phenylalanine or another hydrophobic amino acid in the di-methyltransferases and tri-methyltransferases (Collins, Tachibana et al. 2005;

Couture, Dirk et al. 2008). The mechanism underlying this Tyr/Phe switch is still unclear despite its biological importance, some suggested this might work through altering the affinity of an active-site water molecule (Couture, Dirk et al. 2008), but evidence is still needed to confirm this suggestion.

SET8

SET8 is a histone H4 Lys-20-specific methyltransferase, which is implicated in cell-cycle-dependent transcriptional silencing (Nishioka, Rice et al. 2002) and mitotic regulation in metazoans (Karachentsev, Sarma et al. 2005). Active-site structure of SET8 bound to the H4K20 peptide (PDB:1ZKK) indicated that the K20 side chain is bound in an extended conformation within a channel composed of several aromatic residues, including Tyr-245, the Tyr/Phe switch residue 334, Tyr-336 in SET8 and also His-18 in H4. The cation - pi interactions between the monopole (cation in lysine or methylated lysine) and the quadrupole (pi systems) might contribute to the binding of the substrate lysine into the active site and also the formation of the methylated lysine product. However, further study is still necessary to determine the role of the cation - pi interactions. The side chain of K20 is oriented towards the methyl group of AdoMet through hydrogen bonding to the Tyr-245 OH group and a tightly bound water molecule. This water molecule is further stabilized in the solvent pocket via hydrogen bonding to the Tyr-334 OH group and the carbonyl oxygen of Gly-294 and Ile-297 (Couture, Collazo et al. 2005; Couture, Dirk et al. 2008) (Figure 1.5). Molecular dynamics studies have suggested that chains of water molecules

might form transiently to be responsible for the deprotonation of the substrate lysine during catalysis (Zhang and Bruice 2007); an invariant active site tyrosine residue has also been suggested to function as the general base in the computational study of SET7/9 (Guo and Guo 2007). pH rate profiles of SET domain PKMTs were also studied previously to investigate the possible general base in the reaction. The pH rate profile of k_{cat} versus pH for SET7/9, an enzyme that is quite similar to SET8, was determined. Analysis of the plot yielded a pKa value of 8.3 and a pH optimum of 9.5. The pKa value could be consistent with that of the conserved Tyr residues within close proximity to the substrate lysine (e.g., Tyr245 in SET8). It was proposed that the positive charges of the AdoMet sulfonium cation and the ϵ -ammonium cation of the Lys side chain close to this conserved Tyr residue might depress the pKa of the Tyr to be around 8 (Trievel, Beach et al. 2002; Dirk, Flynn et al. 2007).

The Tyr/Phe switch mutation (Y334F in SET8) alters SET8 from a mono-methyltransferase to a di-methyltransferase without affecting histone binding or the overall methylation rate (Couture, Collazo et al. 2005). Previous structure observations have suggested that the diminished affinity of the water molecule in the solvent pocket due to the loss of a hydrogen bond as a result of the Y334F mutation might enable a further methyl group addition in the mutant (Couture, Dirk et al. 2007). In **CHAPTER 2**, QM/MM MD and free energy simulations were performed to study the origins of product specificity of human SET8, with comparison to its Y334F mutant. The calculated free energy barriers of the

methyl transfer step are consistent with the product specificity of wild type SET8 and its Y334F mutant, and the relative barriers are proposed to govern the methylation state of SET8's target. Indeed, the free energy barrier of the second methylation step in wild type SET8 was found to be much higher than that of the first methylation step, and the increase of the barrier could stop the second methylation and lead to the mono-methyl lysine product. By contrast, the free energy barrier of the second methylation step in the Y334F mutant is similar to that of the first step, and the mutation alters SET8 to a di-methyltransferase. The affinity of the active-site water molecule was also studied through the free energy simulations of the reactant complexes of wild type SET8 and Y334F; the dissociation of this water molecule was suggested to alter the product specificity of Y334F mutant by leaving more space in the active site. The MD simulations of the reactant complex of methyl transfers in wild type and Y334F mutant also indicate that the ability of the reactant complex to form the reactive configurations could be used to predict the product specificity of SET8.

GLP

G9a-Like-Protein (GLP) is from the SUV39 family. It is the only closely related homolog of G9a, the major mammalian H3-K9 methyltransferase that targets euchromatic regions and essential for murine embryogenesis (Tachibana, Sugimoto et al. 2002). Previous studies show that G9a and GLP form a heteromeric complex and work cooperatively in euchromatin to exert H3-K9 methyltransferase function *in vivo* (Tachibana, Ueda et al. 2005). The F1152Y

mutant at the Tyr/Phe switch site of G9a was found to alter the enzyme from a di-methyltransferase to a mono-methyltransferase; the equivalent residue in GLP is F1209. Mutation of another residue at the active site of G9a (the Y1067F mutant) was found to alter the enzyme to a tri-methyltransferase; the equivalent residue in GLP is Y1124 (Collins, Tachibana et al. 2005).

After the investigation of SET8, we would like to look into whether the product specificity of other PKMTs is also affected by the dissociation of the active-site water molecule during methyl transfer. We are also interested in answering the question as to whether the QM/MM MD and free energy simulations could provide important insights into the changes of the product specificity as a result of the mutations of the residues other than those at the Tyr/Phe switch. The availability of the structure of GLP co-crystallized with the cofactor byproduct AdoHcy and the substrate peptides makes it possible to investigate this enzyme catalysis by computational methods. Our simulations indicate that the free energy barriers of the methyl transfer of wild type GLP and its F1209Y and Y1124F mutants could also be used to predict the product specificity. Similar to the case of SET8, the F1209Y mutation appears to alter the product specificity of GLP through affecting the affinity of the active-site water molecule. For Y1124F, the mutant becomes a tri-methyltransferase, probably due to the loss of hydroxyl group that produces more space around the target nitrogen for additional methyl addition. The work of GLP is listed in **CHAPTER 3**.

Protein Arginine Methylation

Protein arginine methylation occurs in abundant eukaryotic proteins and modulates different cell signaling pathways, including gene transcription, RNA splicing, signal transduction, cell growth and proliferation (Bedford and Clarke 2009). This modification is catalyzed by the family of protein arginine methyltransferases (PRMTs) through the transfer of methyl groups from S-adenosyl-L-methionine (AdoMet) to the guanidine group of arginine residues in protein substrate (Lee, Kim et al. 1977). Currently there are two major types of PRMTs: the both types of enzymes catalyze the formation of ω - N^G -monomethylarginine (MMA), but they differ in the dimethylation products. Type I PRMTs catalyze the formation of asymmetric ω - N^G , N^G -dimethylarginine (ADMA), while type II PRMTs catalyze the formation of symmetric ω - N^G , N'^G -dimethylarginine (SDMA) (Di Lorenzo and Bedford 2011). Different product formation (ADMA or SDMA) of the substrate could lead to different biological consequences (Wysocka, Allis et al. 2006), and therefore it is very important to understand the basic mechanism underlying the methylation processes.

A conserved core region carrying the characteristic motifs of seven-beta strand (Katz, Dlakic et al. 2003) and also additional invariant “double E” and “THW” sequence motifs have been observed in this PRMT family (Zhang, Zhou et al. 2000; Zhang and Cheng 2003). The two invariant glutamates on the “double E” loop have been suggested to be required for the activities of PRMTs, for instance, in the cases of PRMT1 and PRMT5 (Zhang and Cheng 2003; Sun,

Wang et al. 2011). The superimposition of the active site residues of PRMT1 (crystallized with the substrate under pH 4.7) and PRMT3 (crystallized under pH 6.3, at which the enzyme has detectable activity) placed the substrate arginine in between E326 and E335 of PRMT3 (the two invariant glutamates); hydrogen bonding to these two glutamates stabilize the target arginine in the active site (Zhang and Cheng 2003). The charge-charge interactions between E326 and AdoMet also contributes to the good alignment of the methyl group of AdoMet and the target nitrogen atom. The negative charge of E326 has also been suggested to be greatly reduced through the charge-charge interactions with AdoMet and a nearby arginine in the protein, thus the positive charge on the guanidine group has been suggested to be redistributed towards the amino group near E335 and leaving the other amino group to be the possible nucleophile (Zhang and Cheng 2003). The possible general base during the catalysis has also been discussed previously in several PRMTs (Zhang, Zhou et al. 2000; Rust, Zurita-Lopez et al. 2011). A His-Asp relay system has been suggested to function as the general base. However, the location of the His residue ($> 6 \text{ \AA}$ from the substrate guanidine group) makes it less possible to function as the general base; the mutagenesis study also did not support its role as the general base (Rust, Zurita-Lopez et al. 2011). The active-site structure suggested that only the two invariant glutamates on the "double E" loop are within the distance ($< 3 \text{ \AA}$) to remove the proton from guanidine directly and to be possibly function as the general base. pH profiles of PRMT1 and PRMT3

indicated that the pKa of the reaction rates is between 6.0 and 7.0 (Zhang and Cheng 2003), which could correlate to the deprotonation of the substrate guanidine by the general base.

In **CHAPTER 4**, we discuss the results of the methylation reaction catalyzed by PRMT3 from the QM/MM MD and free energy simulations. Our simulation results support previous suggestions that the interactions of the active-site residues might lead to the charge re-distribution on the guanidine group and make $N_{\eta 2}$ the more effective nucleophile. The calculated free energy barriers of the both first and second methyl transfers to $N_{\eta 2}$ are much lower than those to $N_{\eta 1}$. As a result, the both methyl groups are likely to be added to $N_{\eta 2}$, and this might lead to the formation of the asymmetrically methylation product. The both 1D and 2D free energy simulations suggest that the methyl transfer and proton transfer processes might be somehow concerted, and Glu326 on the “double E” loop could serve as the general base of the reaction.

Quantum Mechanical/Molecular Mechanical (QM/MM) Molecular Dynamics (MD) and Free Energy Simulations

QM/MM MD simulations

Proteins are not rigid and static structures. They are actually in constant motions at multiple timescales (Henzler-Wildman and Kern 2007; Zwier and Chong 2010), and these motions may play essential roles for their biological functions. Molecular dynamics simulations are perhaps the most popular

theoretical method to study protein motions and provide complementary results which could not be obtained from experimental studies.

The hybrid QM/MM MD simulations combine the accuracy of QM with the low computational cost of MM. This method is very useful to study enzyme reaction mechanism by treating the reacting part quantum mechanically and the other part molecular mechanically (Warshel and Levitt 1976) (**Figure 1.6**). The MM force fields are based on empirical potentials describing small-amplitude vibrations, torsions, van der Waals interactions and electrostatic interactions. However, they are unable to describe the changes in electronic structures of a system involving, for instance, bond breaking and making, charge transfer, *etc.* (Lin and Truhlar 2007) A number of MM force fields are available to date, for example, AMBER (Du, Wu et al. 2003), CHARMM (Brooks, Bruccoleri et al. 1983), OPLS (Jorgensen, Maxwell et al. 1996), GROMOS (Schuler, Daura et al. 2001), *etc.* QM potentials allow the electron re-distribution in the processes of bond breaking and making and could be used for the simulations of chemical reactions. However, its application is limited to small systems due to the high computational cost of this approach. The available methods for QM simulations include some high level *ab initio* (e.g., MP2) and density functional (e.g., B3LYP) methods (Gao and Truhlar 2002) as well as some semi-empirical QM methods (which are much less time-consuming). In our study, one semi-empirical method, self-consistent charge density functional tight binding (SCC-DFTB) method (Cui, Elstner et al. 2001) was used for the QM simulations. SCC-DFTB is derived from

DFT by a second order expansion of the total DFT energy with respect to charge density variation, relative to a chosen reference density.

The key issue to perform accurate simulations with QM/MM approaches is how to specify the interface between the QM and MM regions, especially when the QM and MM regions are covalently bonded (for example, the C_{α} - C_{β} bond of an amino residue). Several approaches are available to address this issue, including simple fragment methods (Maseras and Morokuma 1995; Svensson, Humbel et al. 1996), link atom methods (Field, Bash et al. 1990), and frozen orbital methods (Gao, Amara et al. 1998). The most widely used approach, link atom method, was adopted in our simulations.

The biological activities of enzymes are often associated with the dynamics of relatively localized region, thus it is inefficient to perform MD simulations of the entire system. To reduce the size of the simulation models and increase the efficiency of calculations, stochastic boundary method (Brooks, Brunger et al. 1985) was used in our simulations. A general scheme of this method is shown in **Figure 1.7**: normally an atom close to the active center is chosen as the reference point; the main system is partitioned into two regions, the reservoir region and the reaction zone. The reservoir region is deleted; the reaction zone is further partitioned into the reaction region and the buffer region. The reaction region contains the part of dynamic interest; the residues in this part are simulated with molecular dynamics, which are separated into QM and MM

part as discussed before. The buffer region simply represents the heat bath and is treated by Langevin dynamics.

Free energy (potential of mean force) change determinations

Potential of mean force (PMF) describes the probability of being in a state along a reaction coordinate (Kirkwood 1935; Kirkwood 1936). The relationship between PMF ($A(\zeta)$) and the probability ($p(\zeta)$) is defined as:

$$A(\zeta) = -k_B T \ln p(\zeta) \quad (1)$$

k_B is Boltzmann's constant, T is the temperature and ζ describes state of molecule (a function of a position on the reaction coordinate). To calculate PMF, MD simulations could be performed to generate samples along a reaction coordinate and get the probability distributions. The problem with a normal simulation is some states (for example, high energy states) are poorly sampled within the available computer time. The umbrella sampling method (Torrie and Valleau 1974) implemented in the CHARMM program could avoid this problem by applying harmonic potential functions along the reaction coordinate to force the system to explore the regions which are hard to be sampled in normal simulations. The harmonic potential function is defined as:

$$U = \frac{1}{E} \times K \times (r - r_0)^E \quad (2)$$

in the CHARMM package, Where E is default to be 2, K is the force constant, r is the reaction coordinate and r_0 is the origin of the harmonic function on the reaction coordinate. The reaction coordinate could be a distance, angle or a linear combination of them changes with the transformation of the system from

one state to another. With a series of r_0 , the configurations along the reaction coordinate could be sampled.

The biased probability distributions of the configurations generated from umbrella sampling could then be statistically translated into relative free energy change (PMF) along the reaction coordinate by Weighted Histogram Analysis Method (WHAM) (Kumar et al. 1992; Roux 1995; Souaille and Roux 2001). It is a useful method to extract all the available information from a series of simulations with different r_0 and force constant K in an optimal way and obtains better estimates by combining the results of all the simulations. The central idea of WHAM consists in constructing an optimal estimate of the unbiased distribution function as a weighted sum over the data from all the simulations and determining the functional form of the weight factors that minimizes the statistical error. The overlapping probability distributions between different simulation windows are critical for PMF calculations in WHAM, thus it is important to select suitable force constants and the origins of the harmonic functions for each simulation window so that the probability distributions can have enough data in the overlapping regions.

The statistical error of the PMF calculation was calculated as the 95% confidence interval in our study, which estimates the difference between the observed average value of many trials and the true value of the probability (Kobark 2003). Based on this approach, the larger the number of trials, the smaller this difference should become. The configurations around the higher

energy state (for example, the transition state) are hard to be sampled and will have lower quantity of snapshots in each bin, which makes the statistical error of this region larger. More simulation windows or smaller intervals could be used around these states to lower the statistical error. An alternative way to calculate the statistical error is to calculate the standard deviation of several PMF calculations (for example, forward and backward PMF calculations based on the same or different reaction coordinate) of the same reaction.

References

- Bedford, M. T. and S. G. Clarke (2009). "Protein arginine methylation in mammals: who, what, and why." Mol Cell **33**(1): 1-13.
- Brooks, B. R., R. E. Bruccoleri, et al. (1983). "CHARMM - a program for macromolecular energy, minimization, and dynamics calculations." J Comput Chem **4**: 187-217.
- Brooks, C. L., A. Brunger, et al. (1985). "Active-site dynamics in protein molecules - a stochastic boundary molecular-dynamics approach." Biopolymers **24**(5): 843-865.
- Collins, R. E., M. Tachibana, et al. (2005). "In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases." J Biol Chem **280**(7): 5563-5570.
- Couture, J. F., E. Collazo, et al. (2005). "Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase." Genes Dev **19**(12): 1455-1465.
- Couture, J. F., L. M. A. Dirk, et al. (2008). "Structural origins for the product specificity of SET domain protein methyltransferases." Proc Natl Acad Sci U S A **105**(52): 20659-20664.
- Cui, Q., M. Elstner, et al. (2001). "A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method." J Phys Chem B **105**(2): 569-585.
- Di Lorenzo, A. and M. T. Bedford (2011). "Histone arginine methylation." Febs Lett **585**(13): 2024-2031.

Dillon, S. C., X. Zhang, et al. (2005). "The SET-domain protein superfamily: protein lysine methyltransferases." Genome Biol **6**(8): 227.

Dirk, L. M., E. M. Flynn, et al. (2007). "Kinetic manifestation of processivity during multiple methylations catalyzed by SET domain protein methyltransferases." Biochemistry **46**: 3905-3915.

Duan, Y., C. Wu, et al. (2003). "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations." J Comput Chem **24**: 1999-2012.

Feng, Q., H. B. Wang, et al. (2002). "Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain." Curr Biol **12**(12): 1052-1058.

Field, M. J., P. A. Bash, et al. (1990). "A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations." J Comput Chem **11**(6): 700-733.

Gao, J. L., P. Amara, et al. (1998). "A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations." J Phys Chem A **102**(24): 4714-4721.

Gao, J. L. and D. G. Truhlar (2002). "Quantum mechanical methods for enzyme kinetics." Annu Rev Phys Chem **53**: 467-505.

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

Henzler-Wildman, K. A. and D. Kern (2010). "Dynamic personalities of proteins." Nature **450**: 964-972.

Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." Science **293**(5532): 1074-1080.

Jones, R. S. and W. M. Gelbart (1993). "The drosophila polycomb-group gene enhancer of zeste contains a region with sequence similarity to trithorax." Mol Cell Biol **13**(10): 6357-6366.

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

Karachentsev, D., K. Sarma, et al. (2005). "PR-Set7-dependent methylation of histone H4 Lys 20 functions in repression of gene expression and is essential for mitosis." Genes Dev **19**(4): 431-435.

Katz, J. E., M. Dlakic, et al. (2003). "Automated identification of putative methyltransferases from genomic open reading frames." Mol Cell Proteomics **2**(8): 525-540.

Kirkwood, J. G. (1935). "Statistical mechanics of fluid mixtures". J Chem Phys **3**: 300-313.

Kirkwood, J. G. (1936). "Statistical mechanics of liquid solutions". Chem Rev **19**: 275-307.

Kumar, S., D. Bouzida, et al. (1992). "THE WEIGHTED HISTOGRAM ANALYSIS METHOD FOR FREE-ENERGY CALCULATIONS ON BIOMOLECULES .1. THE METHOD." J Comput Chem **13**(8): 1011-1021.

Lall, S. (2007). "Primers on chromatin." Nat Struct Mol Biol **14**(11): 1110-1115.

Lee, H. W., S. Kim, et al. (1977). "S-adenosylmethionine - protein-arginine methyltransferase - purification and mechanism of Enzyme." Biochemistry **16**(1): 78-85.

Li, B., M. Carey, et al. (2007). "The role of chromatin during transcription." Cell **128**(4): 707-719.

Lin, H. and Truhlar, D. G. (2007). "QM/MM: what have we learned, where are we, and where do we go from here?" Theor Chem Acc **117**: 185-199.

Luger, K., A. W. Mader, et al. (1997). "Crystal structure of the nucleosome core particle at 2.8 Å resolution." Nature **389**(6648): 251-260.

Martin, C. and Y. Zhang (2005). "The diverse functions of histone lysine methylation." Nat Rev Mol Cell Biol **6**(11): 838-849.

Maseras, F., K. Morokuma (1995). "IMOMM- a new integrated ab-initio plus molecular mechanics geometry optimization scheme of equilibrium structures and transition-states." J Comput Chem **16**: 1170-1179.

Nishioka, K., J. C. Rice, et al. (2002). "PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin." Mol Cell **9**(6): 1201-1213.

Roux, B. (1995). "The calculation of the potential of mean force using computer simulations." Comput Phys Commun **91**: 275-282.

Rust, H. L., C. I. Zurita-Lopez et al. (2011). "Mechanistic Studies on Transcriptional Coactivator Protein Arginine Methyltransferase 1." Biochemistry **50**(16): 3332-3345.

Schuler, L. D., X. Daura et al. (2001). "An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase." J Comput Chem **22**: 1205-1218.

Souaille, M. and B. Roux (2001). "Extension to the weighted histogram method: combining umbrella sampling with free energy calculations." Comput Phys Commun **135**: 40-57.

Stassen, M. J., D. Bailey, et al. (1995). "The drosophila-trithorax proteins contain a novel variant of the nuclear receptor-type DNA-binding domain and an ancient conserved motif found in other chromosomal-proteins." Mech Dev **52**(2-3): 209-223.

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature **403**(6765): 41-45.

Sun, L. T., M. Z. Wang et al. (2011). "Structural insights into protein arginine symmetric dimethylation by PRMT5." Proc Natl Acad Sci U S A **108**(51): 20538-20543.

Svensson, M., Humbel, S., Froese, R. D. J., Matsubara, T., Sieber, S., Morokuma, K. (1996). "ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder

reactions and $\text{Pt}(\text{P}(\text{t-Bu})_3)_2 + \text{H}_2$ oxidative addition." J Phys Chem **100**: 19357-19363.

Tachibana, M., K. Sugimoto, et al. (2002). "G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis." Genes Dev **16**(14): 1779-1791.

Tachibana, M., J. Ueda, et al. (2005). "Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9." Genes Dev **19**(7): 815-826.

Taverna, S. D., H. Li, et al. (2007). "How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers." Nat Struct Mol Biol **14**(11): 1025-1040.

Torrie, G. M. and J. P. Valleau (1974). "MONTE-CARLO FREE-ENERGY ESTIMATES USING NON-BOLTZMANN SAMPLING - APPLICATION TO SUBCRITICAL LENNARD-JONES FLUID." Chem Phys Lett **28**(4): 578-581.

Trievel, R. C. (2004). "Structure and function of histone methyltransferases." Crit Rev Eukaryot Gene Expression **14**(3): 147-169.

Trievel, R. C., B. M. Beach, et al. (2002). "Structure and catalytic mechanism of a SET domain protein methyltransferase." Cell **111**(1): 91-103.

Tschiersch, B., A. Hofmann, et al. (1994). "The protein encoded by the drosophila position-effect variegation suppressor gene *su(var)3-9* combines domains of antagonistic regulators of homeotic gene complexes." Embo J **13**(16): 3822-3831.

Turner, B. M. (2005). "Reading signals on the nucleosome with a new nomenclature for modified histones." Nat Struct Mol Biol **12**(2): 110-112.

Warshel, A., M. Levitt (1976). "Theoretical studies of enzymatic reactions - dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme". J Mol Biol **103**: 227-249.

Watson, J. D., T. A. Baker, et al. (2004). Molecular Biology of the Gene, 5th edition. CSHL Press.

Wysocka, J., C. D. Allis, et al. (2006). "Histone arginine methylation and its dynamic regulation." Front Biosci **11**: 344-355.

Zhang, X. and X. D. Cheng (2003). "Structure of the predominant protein arginine methyltransferase PRMT1 and analysis of its binding to substrate peptides." Structure **11**(5): 509-520.

Zhang, X., Z. Yang, et al. (2003). "Structural basis for the product specificity of histone lysine methyltransferases." Mol Cell **12**(1): 177-185.

Zhang, X., L. Zhou, et al. (2000). "Crystal structure of the conserved core of protein arginine methyltransferase PRMT3." Embo J **19**(14): 3509-3519.

Zwier, M. and L. T. Chong (2010). "Reaching biological timescales with all-atom molecular dynamics simulations." Curr Opin Pharmacol **10**: 745-752.

Appendix

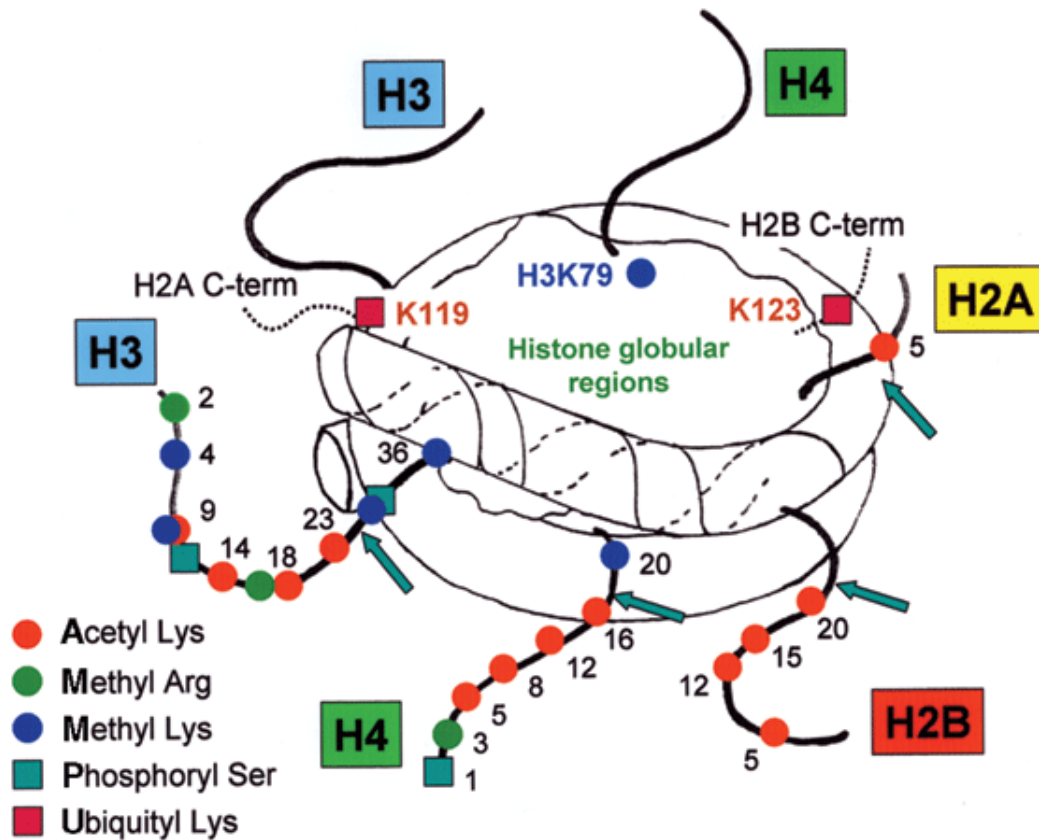


Figure 1.1 The post-translational modification sites on histone tails

Histone tails are subject to a variety of post-translational modifications.

Source: Turner, B. M. (2005). "Reading signals on the nucleosome with a new nomenclature for modified histones." *Nat Struct Mol Biol* **12**(2): 110-112.

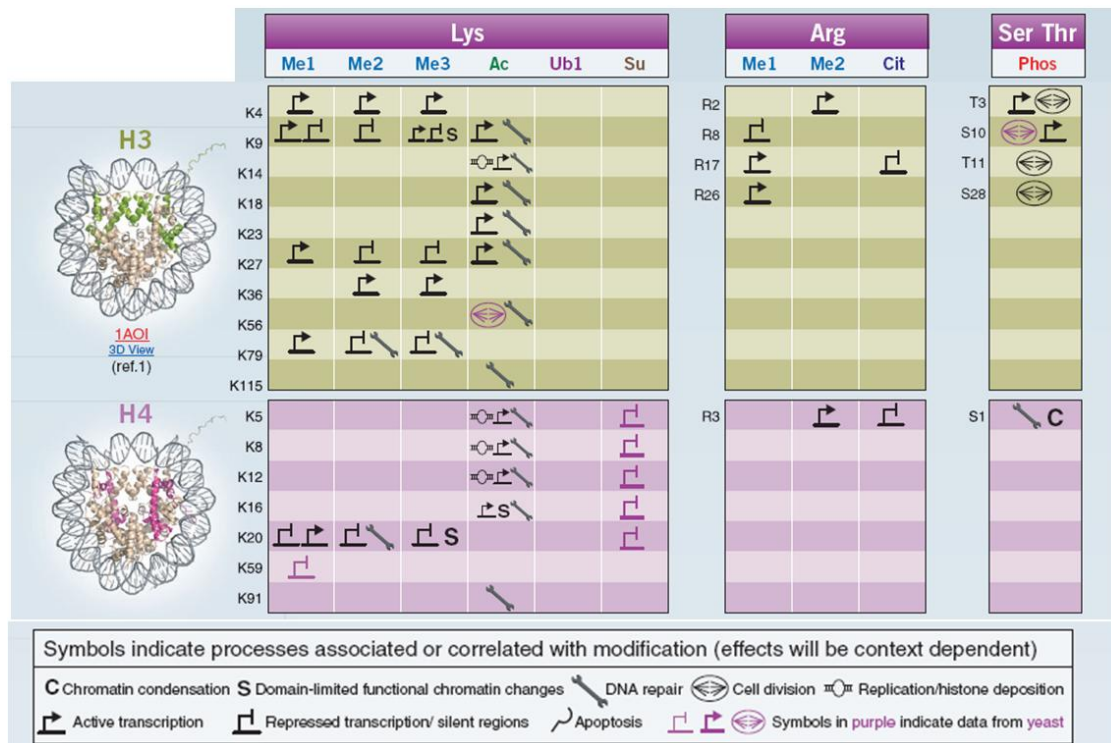


Figure 1.2 Biological processes associated with histone code

Different modifications on histone tails constitute the histone code and are associated with different downstream events.

Source: Lall, S. (2007). "Primers on chromatin." Nat Struct Mol Biol **14**(11): 1110-1115.

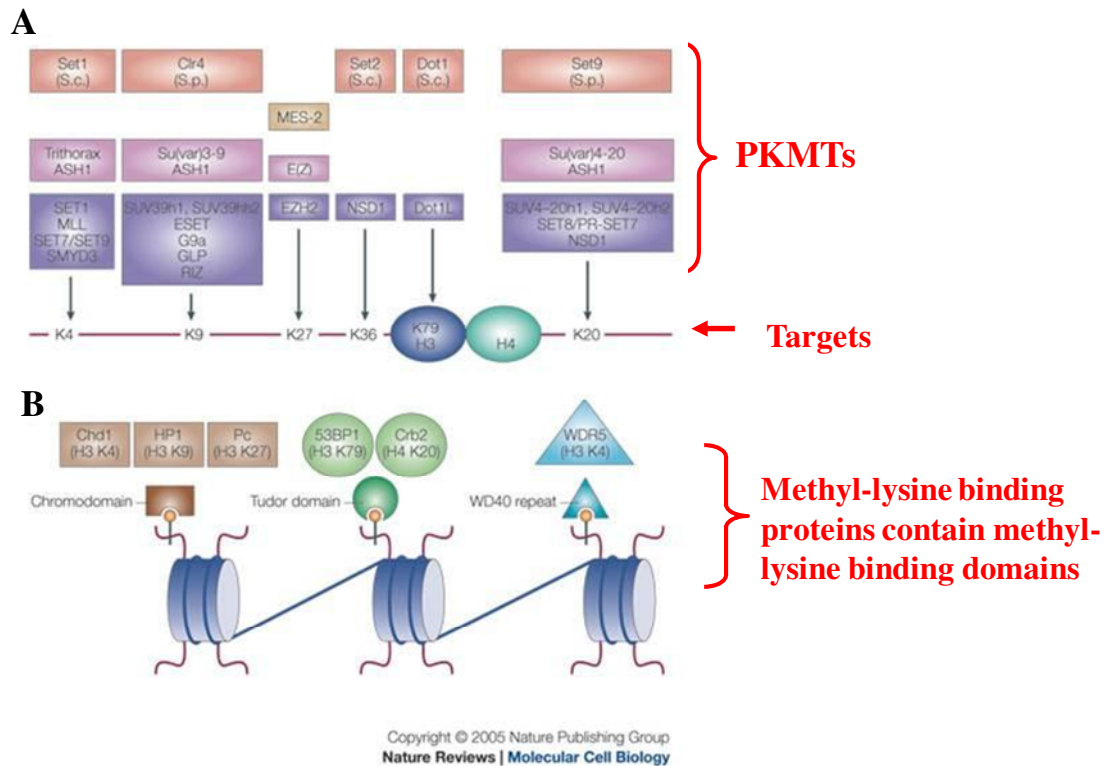


Figure 1.3 PKMTs and methyl-lysine binding proteins

(A) PKMTs target at different lysine residues on histones.

(B) Methyl-lysine binding proteins bind to different methylated lysines through their methyl-lysine binding domains.

Revised from Source: Martin, C. and Y. Zhang (2005). "The diverse functions of histone lysinemethylation." Nat Rev Mol Cell Biol **6**(11): 838-849.

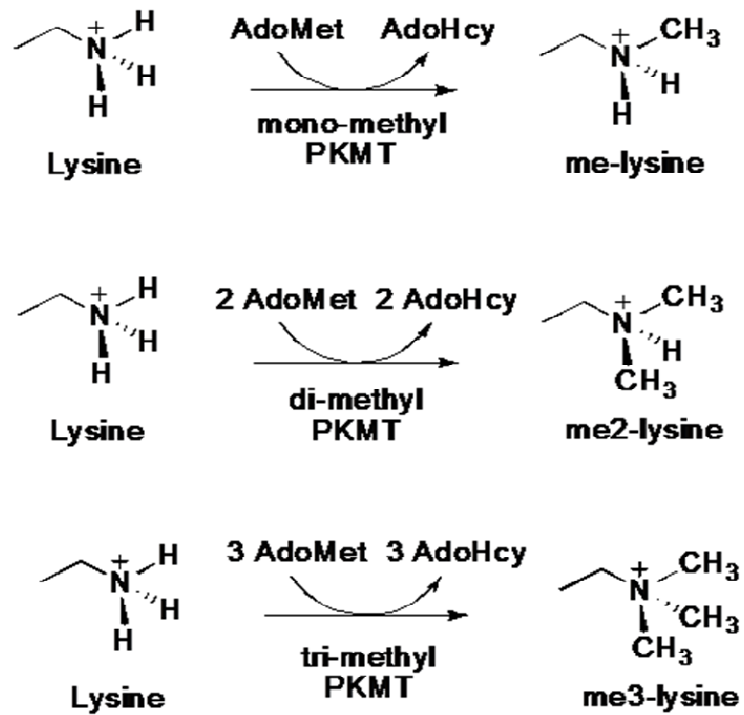


Figure 1.4 Product specificity of PKMTs

PKMTs could add one, two or three methyl groups to target lysine residues.

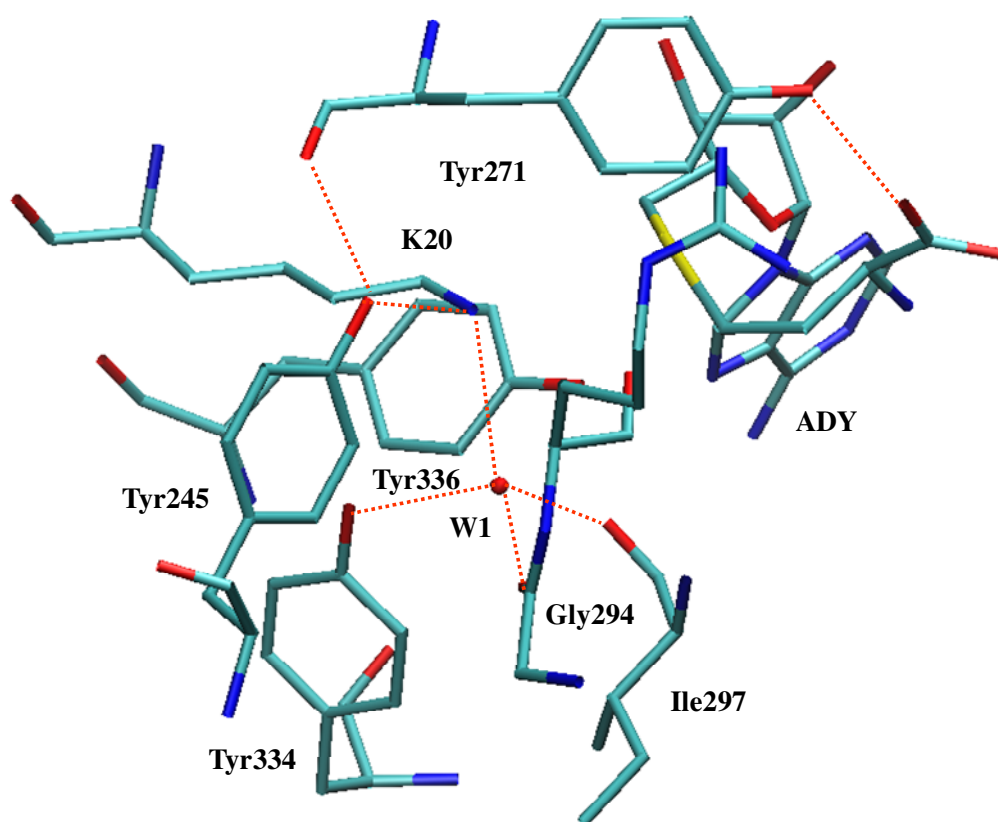


Figure 1.5 Active site structures of SET8 bound to AdoHcy and H4K20 (pdb:1ZKK) (Couture, Dirk et al. 2008)

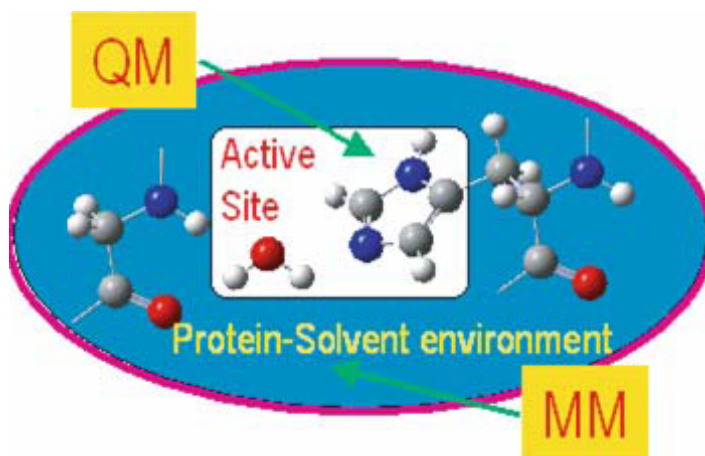


Figure 1.6 Illustration of QM/MM approach in the simulations of enzyme catalysis

The reaction center is treated by QM and the left part is treated by MM.

Source: from Lin, H. and D. G. Truhlar (2007). "QM/MM: what have we learned, where are we, and where do we go from here?" Theor Chem Acc **117**: 185-199.

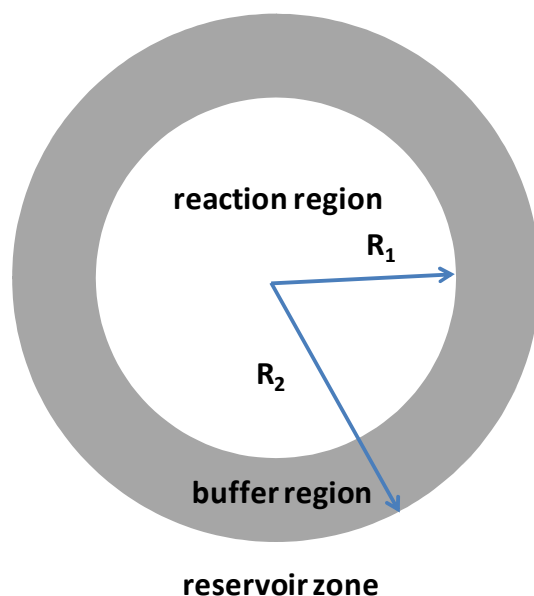


Figure 1.7 General scheme for the partition of the system in stochastic boundary method

The reservoir zone is more than R_2 Å away from the reference point; it has quite less effect on the reaction so it is deleted. The left part is the reaction zone, which is further partitioned into the reaction region (within R_1 Å away from the reference point) and the buffer region (between R_1 and R_2 Å away from the reference point).

CHAPTER 2

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?

A version of this chapter was originally published by Yuzhuo Chu, Qin Xu and Hong Guo:

Yuzhuo Chu, Qin Xu and Hong Guo “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?” *Journal of Chemical Theory and Computation* 6 (2010): 1380-1389.

My contributions to this study include: 1) discussions with the design of the experiments; 2) performing the experiments; 3) analysis of the result and preparing the manuscript. This study was designed by the principle investigator, Dr. Hong Guo, he also wrote the manuscript. Dr. Qin Xu is involved with the discussions of the design of the experiment.

Abstract

Biological consequences of histone lysine methylation depend on the methylation states of the lysine residues on the tails of histone proteins that are methylated by protein lysine methyltransferases (PKMTs). Therefore, the ability of PKMTs to direct specific degrees of methylation (i.e., product specificity) is an important property for regulation of chromatin structure and gene expression. Here the free energy simulations based on quantum mechanical/molecular mechanical (QM/MM) potentials are performed for the first, second and third methyl transfers from S-adenosyl-L-methionine to the ϵ -amino group of the target lysine/methyl lysine in SET8, one of important PKMTs. The key questions addressed in this paper include the energetic origin of the product specificity and the reasons for

the change of the product specificity as a result of the replacement of Tyr334 by Phe. The free energy barriers for the three methyl transfers in SET8 as well as in the mutant obtained from the simulations are found to be well correlated with the experimental observations on the product specificity of SET8 and the change of product specificity as a result of the mutation. The results support the suggestion that the differential free energy barriers for the methyl transfers may determine, at least in part, how the epigenetic marks of lysine methylation are written by the enzymes. Furthermore, the stability of a water molecule to be located at the active site is examined under different conditions using the free energy simulations, and its role in controlling the product specificity is discussed. The QM/MM molecular dynamics (MD) simulations are also performed on the reactant complexes of the first, second and third methyl transfers. The results show that the information on the ability of the reactant complexes to form the reactive configurations for the methyl transfers may be used as useful indicators in the prediction of product specificity for PKMTs.

Introduction

The tails of core histone proteins in the nucleosome are subject to a variety of post-translational covalent modifications, and these modifications can be read by other proteins to lead to distinct downstream events in the regulation of chromatin structure and gene expression (Turner 2005; Lall 2007; Taverna, Li et al. 2007). Protein lysine methyltransferases (PKMTs) catalyze one of such modifications, i.e., histone lysine methylation. Histone lysine methylation can

govern a number of important biological processes, including heterochromatin formation, X-chromosome inactivation, and transcriptional silencing and activation (Martin and Zhang 2005; Jenuwein 2006). Several lysine residues on histone proteins have been identified to be the sites of methylation, including histone H3 Lysine 4 (H3-K4), H3-K9, H3-K27, H3-K36, H3-K79 and H4-K20. In addition to selecting different lysine sites for methylation (i.e., substrate specificity), PKMTs may also differ in their ability to transfer one, two or three methyl groups from S-adenosyl-L-methionine (AdoMet, the methyl donor) to the ϵ -amino group of the target lysine (Xiao, Wilson et al. 2003; Cheng, Collins et al. 2005). This property of the enzymes is called product specificity. Since biological consequences of histone lysine methylation depend on the number of methyl groups added to the lysine residues, understanding the energetic origins of product specificity and developing suitable strategies for manipulation of the signaling properties are of considerable interest.

Biochemical studies and structural analyses of the SET-domain PKMT complexes have identified a tyrosine/phenylalanine switch that can control the product specificity. Comparison of the active site structure of DIM-5 (a trimethylase) (Zhang, Tamaru et al. 2002; Zhang, Yang et al. 2003) with that of SET7/9 (Couture, Collazo et al. 2005; Xiao, Jing et al. 2005) or SET8 (Xiao, Jing et al. 2003) (a mono-methylase) showed that a single amino acid residue occupies a structurally similar position in the enzymes (e.g., F281 in DIM-5 and Y334 in SET8) and is in proximity of the ϵ -amino group of the target lysine

(Zhang, Yang et al. 2003; Couture, Dirk et al. 2008). It has been demonstrated that DIM-5 can be converted from a K-9 tri-methyltransferase to a K-9 mono/di-methyltransferase by the F281→Y mutation (Zhang, Yang et al. 2003) and that the SET7/9 Y305F (Zhang, Yang et al. 2003) or SET8 Y334F mutant (Couture, Collazo et al. 2005; Couture, Dirk et al. 2008) was able to generate dimethylated instead of monomethylated lysine product. In each case, the substrate specificity was not changed, and the mutation had little effect on the overall reaction rate. Understanding the role of the residue at this tyrosine/phenylalanine switch position and the energetic origin for the change of the product specificity as a result of the mutation can provide important insights into the property of the PKMT product specificity. Systematic determinations of the structures for SET8 and its Y334F mutant complexed with unmodified, monomethylated, or dimethylated H4-K20 peptide along with AdoHcy were undertaken (Couture, Dirk et al. 2008) to pinpoint the structural origin for the existence of the tyrosine/phenylalanine switch. It was proposed based on the existence and absence of an active-site water molecule in these structures along with some other information that the Phe/Tyr switch may regulate product specificity through altering the affinity of the observed water molecule and that the dissociation of this water molecule is likely to be essential for the multiple methylation process to proceed (Couture, Dirk et al. 2008). Nevertheless, the energetic interpretations for the product specificity of SET8 and its alternation due to the Y334→F mutation are still lacking, and it is not clear as to how the enzyme's ability to

stabilize/destabilize the water molecule at the active site would change as a result of the mutation and/or methyl addition.

Computer simulations can provide important insights into the energetic origins of the product specificity as well as the effects of mutations. One approach that has been used previously for PKMTs is to perform the free energy (potential of mean force) simulations with hybrid quantum mechanical/molecular mechanical (QM/MM) potentials and to establish the correlations between the free energy profiles of the methyl transfers and the product specificity (Guo and Guo 2007; Xu, Chu et al. 2009) (see below for the discussions of the results based on other computational approaches). In our earlier communication (Xu, Chu et al. 2009), it was demonstrated that for DIM5, SET7/9 and their certain mutants the three free energy barriers for the methyl transfers may be used in each case to explain the product specificity observed experimentally. It was hypothesized (Xu, Chu et al. 2009) that the relative efficiencies of the chemical steps involving the three methyl transfers from AdoMet to the ϵ -amino group of the target lysine in PKMTs may determine, at least in part, the product specificity. The results of the QM/MM molecular dynamics (MD) simulations on the reactant complexes have also been compared for the first, second and third methyl transfers for SET7/9, DIM-5 and their mutants (Xu, Chu et al. 2009). It was shown that a correlation may be established between the formation of the reactive configurations for the three methyl transfers and the product specificities of the enzymes. One problem in the earlier work (Xu, Chu et al. 2009) is that the

experimental structures for the SET7/9 and DIM complexes with *different methylation states* for the target lysine residues do not exist, and some manual modifications had to be introduced in the generation of suitable reactant complexes containing methylated lysine. In order to establish the prediction on the relationship between the efficiency of the methyl transfers and product specificity, additional simulations need to be performed on the PKMT (and the mutant) complexes for which the X-ray structures have been determined at different methylation states for the target lysine residue. SET8 is an excellent system for such investigations because of the recent availability of several experimental structures with unmodified, mono- and di-methylated lysine residue at the active site. Furthermore, the location of the important active-site water molecule has also been clearly identified in the X-ray structures. The simulations based on these structures may not only lead to a better understanding of the energetic origin of the product specificity, but also provide important energetic information concerning the stability of this water molecule at the active site at different stages of methylation in different systems that is believed to be a key property of the enzymes in controlling the product specificity (Couture, Dirk et al. 2008).

Here, we report the results of QM/MM free energy simulations on SET8 and its Y334F mutant. The free energy barriers for the methyl transfers in SET8 and the mutant obtained from the simulations are found to be well correlated with the experimental observations on the product specificities, supporting the

suggestion that the differential free energy barriers for the methyl transfers may determine, at least in part, how the epigenetic marks of lysine methylation are written by the enzymes. Furthermore, the stability of the water molecule at the active site under different conditions (see above) is also examined based on the free energy simulations. The free energy profiles show that the stability of the water molecule at the active site decreases significantly as a result of the Y334→F mutation as well as the methyl addition to the lysine residue. Such changes are likely to make it easier for the water molecule to dissociate from the active site and create the space for further methyl addition. The QM/MM MD simulations are also performed on the reactant complexes of the first, second and third methyl transfers. The results show that the dynamic information on the ability of the reactant complexes to form the reactive configurations for the methyl transfers may be used as useful indicators in the prediction of product specificity for PKMTs, although other factors can be involved as well.

Method

QM/MM free energy (potential of mean force) and MD simulations were applied to determine free energy profiles for the first, second and third methyl transfers from AdoMet to the ϵ -amino group of the target lysine (methyl lysine) and to characterize the active-site dynamics of the reactant complexes of the methyl transfers in SET8 and the Y334F mutant using the CHARMM program (Brooks, Bruccoleri et al. 1983). AdoMet/AdoHcy and lysine/methyl-lysine side chains were treated by QM and the rest of the system by MM. The link-atom

approach (Field, Bash et al. 1990) as implemented in CHARMM was applied to separate the QM and MM regions. Although the QM/MM approach in principle is not required for MD investigations of the reactant complexes, the previous studies on SET7/9 and DIM-5 (Guo and Guo 2007; Xu, Chu et al. 2009) showed that the QM/MM MD approach seems to provide a good description of the active-site dynamic features of the reactant complexes that are consistent with experimental observations concerning the product specificity. A modified TIP3P water model (Jorgensen, Chandrasekhar et al. 1983) was employed for the solvent. The stochastic boundary molecular dynamics method (Brooks, Brunger et al. 1985) was used for the QM/MM MD and free energy simulations. The system was separated into a reaction zone and a reservoir region, and the reaction zone was further divided into a reaction region and a buffer region. The reaction region was a sphere with radius r of 20 Å, and the buffer region extended over $20 \text{ Å} \leq r \leq 22 \text{ Å}$. The reference center for partitioning the system was chosen to be the C δ atom of the target lysine residue/methyl lysine. The resulting systems contained around 5500 atoms, including about 800-900 water molecules.

The SCC-DFTB method (Elstner, Porezag et al. 1998; Cui, Elstner et al. 2001) implemented in CHARMM was used for the QM atoms and the all-hydrogen CHARMM potential function (PARAM27) (MacKerell, Bashford et al. 1998) was used for the MM atoms. High level ab initio methods (e.g., B3LYP and MP2) are too time-consuming to be used for MD and free energy simulations.

The results of the SCC-DFTB and B3LYP/6-31G** methods for the description of the methyl transfer in a small model system have been compared in the earlier studies (Guo and Guo 2007; Xu, Chu et al. 2009) using an energy minimization-based approach. This comparison allowed us to understand the performance of the semi-empirical method in the description of the bond breaking and making for the system under investigation. It was found that although the SCC-DFTB optimized geometries along the reaction pathway seemed to be rather close to those from B3LYP/6-31G**, there are some systematic deviations of the SCC-DFTB method in the description of the energetics of the methyl transfer. To correct the errors due to the deficiency of SCC-DFTB method, the empirical correction introduced in the earlier studies (Guo and Guo 2007; Xu, Chu et al. 2009) was applied to the free energy curves obtained from the potential of mean force simulations in the present work (see below). In the earlier study of DIM-5 (Xu, Chu et al. 2009), it was shown that the energy curves from the corrected SCC-DFTB and B3LYP/6-31G** were very close, supporting the use of the approach with the empirical correction. It should be pointed out that the reason that the simple empirical correction can be used in this and previous studies is because the bond breaking and making events in this and the previous papers all involve simple and similar S_N2 methyl transfer processes so that most of the errors is expected to be cancelled out. Moreover, the relative free energy barriers, as opposed to the absolute barriers, are expected to be more important in the determination of the product specificity. The relative free energy barriers

are expected to be less sensitive to the choice of the QM method due to the cancellation of the errors. Indeed, the results of our earlier simulations (Guo and Guo 2007) were confirmed by the use of a quite different QM/MM approach (Hu and Zhang 2006; Hu, Wang et al. 2008) with a difference of only about 1 kcal/mol in the relative free energy barriers of the first and second methyl transfers in SET7/9.

The initial coordinates for the reactant complexes of the first, second and third methyl transfers were based on the crystallographic complexes (PDB codes: 1ZKK, 3F9W, 3F9X, and 3F9Y) of SET8 and its Y334F mutant containing AdoHcy and short H4K20, H4K20me1 and H4K20me2 peptides (Couture, Collazo et al. 2005; Couture, Dirk et al. 2008). In all the cases, a methyl group was manually added to AdoHcy to form AdoMet. In addition, for the reactant complex of the second methyl transfer in wild-type, a methyl group was manually added to the target lysine in the X-ray structure (1ZKK) to form the methylated lysine. 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 used for integration of the equation of motion, and the coordinates were saved every 50 fs for analyses. 1.5 ns QM/MM MD simulations were carried out for each of the reactant complexes of the first, second and third methyl transfers, and the data from the final 0.5 ns were used to generate the distribution maps of $r(\text{C}_M\text{-N}_\xi)$ and θ in each case (see below). As

discussed in the previous studies and mentioned earlier in this paper, the S_N2 methyl transfer from AdoMet to H4-K20, H4-K20me1 or H4-K20me2 is presumably more efficient if the S-CH₃ group of AdoMet is well aligned with the lone pair of electrons on N_ξ in the reactant complex; i.e., with a small θ angle and relatively short C_M-N_ξ distance (Guo and Guo 2007; Xu, Chu et al. 2009). Here θ is defined as the angle between the direction of the C_M-S_δ bond (r_2) and the direction of the electron lone pair (r_1) (see **Figure 2.1**). Therefore, we determined the distributions of $r(C_M-N_\xi)$ and θ from the QM/MM MD trajectories to obtain the information about the relationship between these distributions and product specificity. Moreover, the histogram method was used to calculate the probability density distributions of $r(C_M-N_\xi)$ and θ and the relative free energies as functions of $r(C_M-N_\xi)$ and θ . For $r(C_M-N_\xi)$, histograms with bin width of 0.1 Å were used, and the probability density in the i th histogram is as follows: $p_i = N_i/N$ (N is the total number of the configurations from the MD simulations, and N_i is the occurrence number in the i th histogram). To calculate the probability density distribution of θ , the histograms with width of 10° were used, and N_i was weighted by $1/A_i$, where A_i is the area of the i th histogram of θ . Thus, the probability density in the i th histogram of θ is as follows: $N_i/(N \times A_i)$. The relative free energy of the i th histogram of $r(C_M-N_\xi)$ and θ were calculated through $W_i = -k_B T \times \ln p_i$, where k_B is the Boltzmann's constant, and T is the temperature.

The umbrella sampling method (Torrie and Valleau 1974) implemented in the CHARMM program along with the Weighted Histogram Analysis Method

(WHAM) (Kumar, Bouzida et al. 1992) was applied to determine the change of the free energy (potential of mean force) as a function of the reaction coordinate for the methyl transfer from AdoMet to H4-K20, H4-K20me1 or H4-K9me2 in the wild-type and mutated enzyme. The reaction coordinate was defined as a linear combination of $r(C_M-N_\xi)$ and $r(C_M-S_\delta)$ [$R = r(C_M-S_\delta) - r(C_M-N_\xi)$] (see **Figure 2.1**). For each methyl transfer process, twenty windows were used, and for each window 50 ps production runs were performed after 20 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 statistical errors for the free energy profiles were also estimated and were found to be quite small (see the Supporting Information). In addition, the umbrella sampling method was applied to determine the free energy profiles for the movement of the active-site water molecule (W1) in wild-type and Y334F at different methylation states. The reaction coordinate was defined as the distance between the water oxygen and sulfur atom of AdoMet [i.e., $r(O_{w1} \cdots S_\delta)$]. Twenty windows were used for the change of the water position, and for each window 20 ps production runs were performed after 20 ps equilibration. The force constants of the harmonic biasing potentials used in the PMF simulations were 20 to 50 kcal mol⁻¹ Å⁻². The distance for which the simulations were performed is in the range of 3-15 Å. The simulations were not performed for longer distances, because the use of the stochastic boundary method in the present study may limit the flexibility of the water molecule as it moves closer to the boundaries of the models (at ~22-25 Å from sulfur atom).

Results

The average active-site structure of the reactant complex for the first methyl transfer in SET8 is given in **Figure 2.2(A)**. **Figure 2.2(A)** shows that the active-site structure has the lone pair of electrons on N_ζ of the target lysine well aligned with the methyl group of AdoMet. This is further demonstrated by the large population of the structures with relatively short $r(C_M...N_\zeta)$ distances and small values of the θ angle as well as the free energy plots generated from the results of the simulations (**Figure 2.2(B)**). The average distance between N_ζ and the methyl group (C_MH_3) is approximately 3.0 Å and the angle is mainly in the range of 0-30°. **Figure 2.2(A)** also shows that Tyr245 forms a hydrogen bond with the ϵ -amino group of the target lysine, and this hydrogen bond may help to orientate the direction of the electron lone pair toward the methyl group of AdoMet. A water molecule (W1) forms stable hydrogen bonds with the both ϵ -amino group of H4K20 and Tyr334 (the important tyrosine/phenylalanine switch residue, see above). **Figure 2.2(D)** and **2.2(E)** show that for the reactant complex of the second methyl transfer the average distance between N_ξ and the methyl group (~ 4.5 Å) and the values of θ (mainly in the range of 45-120°) become significantly larger compared to those for the first methyl transfer (**Figure 2.2(A)** and **2.2(B)**). Thus, the S-CH₃ group of AdoMet cannot be well aligned with the lone pair of electrons on N_ξ for the second methyl transfer, suggesting that the efficiency of the corresponding methyl transfer may be significantly compromised. Indeed, **Figure 2.2(E)** shows that the free energy cost for

producing the structure like the one in **Figure 2.2(A)** is approximately 4-5 kcal/mol [i.e., ~3 kcal/mol for changing $r(\text{C}_M \dots \text{N}_\zeta)$ from 4.5 Å to 3 Å and 1.5 kcal/mol for changing θ to less than 30°].

The average structures for the first, second and third methyl transfers in Y334F are given in **Figure 2.3**. As is evident from **Figure 2.3(C)**, the lone pair of electrons on N_ζ of the methyl lysine is well aligned with the methyl group of AdoMet for the second methyl transfer in Y334F. This is in contrast to the case for the second methyl transfer in wild-type for which the two cannot be well aligned (see above). The results suggest that the efficiency of the second methyl transfer is likely to be significantly enhanced due to the improvement of the reactant structure for the methyl transfer, although other factors may be involved as well (see below). For the third methyl transfer in Y334F (**Figure 2.3(E)**), the $\text{S}-\text{CH}_3$ group of AdoMet cannot be well aligned with the lone pair of electrons on N_ξ in the reactant complex as indicated by the long $r(\text{C}_M \dots \text{N}_\xi)$ distance (~4.6 Å) and large values of the θ angle (45-150°). Thus, the corresponding methyl transfer is unlikely to be efficient. One of the key structural changes at the active site is the relocation of the active-site water molecule (W1). Indeed, W1 occupies a position in **Figure 2.3(C)** that is completely different from the one in **Figure 2.2(D)** (i.e., the complex for the second methyl transfer in wild-type) and is not in close contact with the target lysine/methyl lysine and the Y334/F334 residue anymore. For the third methyl transfer in Y334F, W1 has been pushed away from the active site during the MD simulations and is not visible in **Figure 2.3(F)**.

The free-energy profiles for the first and second methyl transfers in SET8 are plotted in **Figure 2.4(A)** as a function of the reaction coordinate; the free energy barrier for the first methyl transfer was calculated to be 15.8 kcal/mol. This free energy barrier is within the error limit of the average barrier from earlier single-point MP2/6-31G+G(d,p)/MM calculations (13.9 ± 2.3 kcal/mol) on SET8 (Zhang and Bruice 2008). It should be pointed, however, that cautions must be exercised when the energetic data obtained based on very different computational approaches are compared. As is evident from **Figure 2.4(A)**, the free energy barrier for the second methyl transfer is much higher than that of the first methyl transfer (by as much as 6.5 kcal/mol). Thus, the second methyl transfer is much less efficient compared to the first methyl transfer process, and this is consistent with the experimental findings that SET8 is a mono-methylase (Couture, Collazo et al. 2005; Xiao, Jing et al. 2005; Couture, Dirk et al. 2008). The earlier single-point MP2/6-31G+G(d,p)/MM calculations led to an average barrier that is as much as 20 kcal/mol higher for the second methyl transfer than that for the first methyl transfer (Zhang and Bruice 2008). The ab initio QM [HF(6-31G*/3-21G*)]/MM free energy simulations for the first and second methyl transfers in two different PKMTs, SET7/9 and Rubisco LSMT, have also been performed previously (Hu, Wang et al. 2008). For SET7/9, the free energy barriers for the first and second methyl transfers were calculated to be 22.5 ± 0.5 kcal/mol and 26.2 ± 0.5 kcal/mol, respectively (Hu, Wang et al. 2008). Comparing with our earlier results on SET7/9 (Guo and Guo 2007) and taking into account

the suggestion that the HF(6-31G*/3-21G*) method may overestimate the barriers by about 3 kcal/mol (Hu, Wang et al. 2008), these data indicate that the corrected SCC-DFTB method might underestimate the barriers, although other factors may affect the results as well. However, as mentioned earlier, the relative free energy barriers, as opposed to the absolute barriers, are expected to be more important in the determination of the product specificity. For SET7/9, the difference for the two methods in the description of the differential free energy barriers is only about 1 kcal/mol.

Figure 2.4(B) plots the free energy profiles for the methyl transfers in the Y334F mutant. Unlike the wild-type enzyme, the free energy profiles for the both first and second methyl transfers are rather similar with relatively low barriers. Thus, if the first methyl transfer from AdoMet to the target lysine can be catalyzed by Y334F, the second methyl transfer to mono-methyl lysine would also be possible. By contrast, the free energy barrier for the third methyl transfer in Y334F is considerably higher. The results are consistent with the experimental observations that Y334F is a di-methylase (Couture, Collazo et al. 2005; Xiao, Jing et al. 2005; Couture, Dirk et al. 2008) and support the suggestion (Guo and Guo 2007; Xu, Chu et al. 2009) that the relative free energy barriers for the methyl transfers are likely to be important energetic factors controlling the product specificity.

Figures 2.5(A) and **2.5(B)** plot the free energy profiles as a function of the distance between the oxygen atom of W1 and the sulfur atom of AdoMet in the

reactant complexes for the first and second methyl transfers, respectively, in the wild-type enzyme. As is evident from **Figures 2.5(A)** and **2.5(B)**, the most stable location for W1 is at the active site in these complexes with a distance of about 5 Å to the sulfur atom, consistent with the average structures observed from the MD simulations (**Figures 2.2(A)** and **2.2(D)**). The free energy profiles further show that W1 seems to be held tightly by the active site interactions, as the energetic cost for W1 to move away from the active site in each case is quite high. **Figure 2.5(D)** shows that for the reactant complex of the second methyl transfer in Y334F, the most stable position for W1 has moved away from the active site (to a location with a distance of about 10-11 Å from the sulfur atom). With the removal of W1, the active site of Y334F becomes less crowded and is able to accommodate the second methyl group on the target lysine. The second methyl transfer can therefore proceed, and the mutant becomes a di-methylase (see below).

Discussion

The key question on the product specificity of PKMTs concerns the factor that controls *the methylation state of the product*. This is in contrast with many other investigations on enzyme-catalyzed reactions which concentrate on the effects of enzymes in the reduction of the activation barriers in going from solution to the enzyme active sites. Thus, a convenient reference reaction for understanding the product specificity would be the process involving the first methyl transfer in the wild-type enzyme (see below). The existence of a relatively

high barrier for one of the methyl transfer processes may lead to the termination of further methyl addition and therefore determine the product specificity of the enzyme. If none of the three free-energy barriers for the methyl transfers is significantly high, the enzyme might be able to catalyze all the three methyl transfers and could be a tri-methylase (e.g., as in the case of DIM-5). This proposal is consistent with some previous computational results that suggested that the product specificity is probably to be mainly controlled by the methyl transfer reaction step (Hu and Zhang 2006; Guo and Guo 2007; Hu, Wang et al. 2008; Xu, Chu et al. 2009). An alternative explanation of the product specificity is based on the formation of a water channel observed during the MD simulations of PKMTs (Zhang and Bruice 2007; Zhang and Bruice 2007; Zhang and Bruice 2007; Zhang and Bruice 2008; Zhang and Bruice 2008; Zhang and Bruice 2008). However, the dramatic increase of the energy barrier from the first to the second methyl transfer in SET8 obtained by the same authors (Zhang and Bruice 2008) (see above) raises the question concerning the true event that prevents the further methylation in PKMTs (e.g., dimethylation by SET8).

It was proposed in our earlier communication that two different free energy triplets, $(0, \Delta_{2-1W}, \Delta_{3-1W})$ and $(\Delta_{M-W}, \Delta_{2-1M}, \Delta_{3-1M})$ for wild-type and mutated enzyme, respectively, may be used in the description of the product specificity of PKMTs and their mutants (Xu, Chu et al. 2009). Here the free energy barrier for the first methyl transfer in the wild-type enzyme (i.e., the reference reaction) is taken as the zero (For the detailed explanation of the parameters, see **Figure**

2.4). For SET8 and its Y334F mutant studied in this work, the corresponding energy triplets can be written as (0, 6.5, x) and (-1.7, -1.1, 9), respectively, which reflect the fact that SET8 is a mono-methylase and the mutant is a di-methylase (see below). The kinetic study on Y334F (Couture, Dirk et al. 2008) suggested that the activation barrier to produce the H4K20me2 product from the H4K20me1 substrate is about 2 kcal/mol higher than the barrier to produce the mono-methyl product from the unmodified H4K20 substrate. **Figure 2.4(B)** shows that the second methyl transfer has a free energy barrier that is slightly lower than that for the first methyl transfer (by ~ 1 kcal/mol). The simulation data are therefore consistent with the suggestion (Couture, Dirk et al. 2008) that methyl lysine reorientation and deprotonation between turnovers may constitute a rate-limiting step in catalysis. However, it should be pointed out that, as far as the product specificity is concerned, the key question is what causes the stop of further methyl addition during histone lysine methylation (which is different from the question concerning the rate-limiting step of the enzyme-catalyzed process). The results of the simulations from the present work on SET8 and our previous studies on DIM-5 and SET7/9 (Xu, Chu et al. 2009) (and certain mutants) suggest that one of the key energetic factors for the specific product specificities of PKMTs is presumably due to a significant increase of the free energy barrier for one of the methyl transfers in the enzymes. Thus, the reason that SET8 is mono-methylase is probably due to the fact that the energy barrier for the second methyl transfer is too high that stops further methylation. The Y334→F mutation

on SET8 effectively reduces the barrier for the second methyl transfer so that the second methyl transfer can proceed. Since the free energy barrier for the third methyl transfer is very high, and the addition of the third methyl group cannot proceed, leading to a di-methylase. The similar arguments may be made for SET7/9 and DIM-5.

It is of interest to note that the free energy data on the ability of SET8 (Y334F) to catalyze the first (first and second) methyl transfer and the inability of SET8 (Y334F) to catalyze the second (third) methyl transfer are already reflected from the MD simulations in **Figure 2.2(B)** and **2.2(E)** (**Figure 2.3(A)**, **2.3(C)** and **2.3(E)**). Similar observations have also been made previously (Hu and Zhang 2006; Guo and Guo 2007; Hu, Wang et al. 2008; Zhang and Bruice 2008; Xu, Chu et al. 2009). Thus, the dynamic information on the ability of the reactant complexes to form the reactive configurations for the methyl transfers may be used as useful indicators in the prediction of product specificity for PKMTs, although further tests are still necessary to establish the correlations. This result is of importance, because performing the MD simulations is much easier than undertaking the QM/MM free energy simulations. Examination of the structures at the transition states (TSs) in **Figure 2.2** and **2.3** shows that all these structures are rather similar; e.g., $r(\text{C}_M \dots \text{N}_Z)$ and $r(\text{S}_\delta \dots \text{C}_M)$ are 2.1 and 2.3 Å, respectively, in each of the structures. It is of interest to note these structures are rather close to the TS structures generated earlier from the ab initio QM [HF(6-31G*/3-21G*)]/MM free energy simulations for two different PKMTs, SET7/9 and LSMT

(Hu, Wang et al. 2008). As discussed earlier, the structures for the corresponding reactant complexes, on the other hand, can be significantly different. Indeed, the structures for the reactant complexes of the first methyl transfer in wild type (**Figure 2.2(A)**) and the first and second methyl transfers in Y334F (**Figure 2.3(A)** and **2.3(C)**, respectively) are rather similar to the corresponding TS structures (**Figure 2.2(C)**, **2.3(B)** and **2.3(D)**, respectively) with a $r(\text{C}_M \dots \text{S}_\delta)$ distance about 4.8 Å. For these cases, a part of the TS stabilization is probably already reflected on the reactant state through the generation of such TS-like conformation. By contrast, the structures of the reactant complexes for the second methyl transfer in wild-type (**Figure 2.2(D)**) and the third methyl transfer in Y334F (**Figure 2.3(E)**) are significantly distorted from the corresponding TS structure (**Figure 2.2(F)** and **2.3(F)**, respectively), and the $r(\text{C}_M \dots \text{S}_\delta)$ distances are around 6.3 Å. Therefore, additional energetic cost would be required to generate the TS-like structures from these structures, and this could lead to relatively high activation barriers for the corresponding methyl transfers. It should be pointed out that the free energy costs for generating the TS-like structures alone seem not sufficient to explain the increases of the barriers for the methyl transfers. Indeed, **Figure 2.2(B)** shows that the free energy cost for producing the structure similar to the one in **Figure 2.2(A)** is approximately 4-5 kcal/mol, while the free energy barrier increases by 6.5 kcal/mol. Therefore, other factors may be involved as well. It is of interest to note from **Figure 2.2(A)** and **2.2(C)** that the hydrogen bond distances involving the ϵ -amino group (with Tyr245 and

W1) decrease significantly as the system reaches the transition state. This indicates that the corresponding interactions are strengthening and may play an important role in the transition state stabilization for the methyl transfer as well.

It was proposed that the Phe/Tyr switch may regulate product specificity through altering the affinity of W1 and the dissociation of this water molecule is essential for the multiple methylation process to proceed (Couture, Dirk et al. 2008). Although the structural information on the presence and absence of W1 under different conditions is of considerable interest, question remains concerning how the affinity of W1 would change as a result of the mutation (or methyl addition) and what would be its stability to be located in the active site. The energetic information concerning the stability of W1 at the active site under different conditions is of fundamental importance for the determination of the role of this water molecule in preventing further methylation. **Figure 2.5(B)** shows that W1 is rather stable in the reactant complex of the second methyl transfer in the wild-type enzyme and the free energy cost for its removal is quite high. The high stability is presumably achieved through the interaction involving Tyr334 (**Figure 2.2(D)**). The large energetic cost can make the addition of the second methyl group much more difficult, presumably because the active site becomes too crowded without the removal of W1. This could contribute to the significant increase of the free energy barrier from the first to the second methyl transfer and stop the second methyl addition, although additional simulations are still necessary. **Figure 2.5(D)** shows that for the reactant complex of the second

methyl transfer in Y334F, the most stable location of W1 has changed and W1 moved away from the active site (see also **Figure 2.3(D)**). Thus, this water molecule could not interfere with the methyl transfer process anymore. The mutant is now able to catalyze the second methyl transfer and becomes a di-methylase.

Conclusion

The QM/MM free energy simulations have been performed for the first and second methyl transfers from AdoMet to the target lysine/methyl lysine in SET8 and for the first, second and third methyl transfers in its Y334F mutant (involving the replacement of the tyrosine/phenylalanine switch residue). The two free energy barriers for the methyl transfers in SET8 and the three free energy barriers in the mutant obtained from the simulations have been found to be well correlated with the experimental observations on their product specificities. The results indicated that the significant increase of the free energy barrier for the second methyl transfer in SET8 (for the third methyl transfer in Y334F) might stop further methyl addition, and this could be the reason that SET8 (Y334F) is a mono-methylase (di-methylase). The results support an earlier suggestion (Xu, Chu et al. 2009) that the differential free energy barriers for the methyl transfers may determine, at least in part, how the epigenetic marks of lysine methylation are written by the enzymes. The QM/MM molecular dynamics (MD) simulations are also performed on the reactant complexes of the first and second methyl transfers in SET8 and the first, second and third methyl transfers in Y334F. The

results showed that the dynamic information on the ability of the reactant complexes to form the reactive configurations for the methyl transfers might be used as useful indicators in the prediction of product specificity for PKMTs. The stability of the water molecule at the active site has also been examined based on the free energy simulations. The free energy profiles suggested that the stability of the water molecule at the active site decreases significantly as a result of the Y334→F mutation as well as the methyl addition to the lysine residue. The decrease of the stability of W1 to be located at the active site as a result of the Y334→F mutation is likely to make it easier for the water molecule to dissociate from the active site and create space for further methyl addition.

Acknowledgement

We thank Dr. Hao-Bo Guo, Mr. Mauricio Valenzuela and Professors Xiaodong Cheng and Jeremy C. Smith for useful discussions and Martin Karplus for a gift of the CHARMM program. This work was supported by the National Science Foundation (Grant number: 0817940 to HG). We are also grateful for the computer time from Indiana University TeraGrid Resource.

Supporting Information

The estimates of the free energies for formation of the reactive conformations for Y334F and the statistical errors in the potential of mean force simulations are available.

References

- Brooks, B. R., R. E. Bruccoleri, et al. (1983). "CHARMM - A PROGRAM FOR MACROMOLECULAR ENERGY, MINIMIZATION, AND DYNAMICS CALCULATIONS." J Comput Chem **4**(2): 187-217.
- Brooks, C. L., A. Brunger, et al. (1985). "ACTIVE-SITE DYNAMICS IN PROTEIN MOLECULES - A STOCHASTIC BOUNDARY MOLECULAR-DYNAMICS APPROACH." Biopolymers **24**(5): 843-865.
- Cheng, X., R. E. Collins, et al. (2005). "Structural and sequence motifs of protein (histone) methylation enzymes." Annu Rev Biophys Biomol Struct **34**: 267-294.
- Couture, J. F., E. Collazo, et al. (2005). "Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase." Genes Dev **19**(12): 1455-1465.
- Couture, J. F., L. M. A. Dirk, et al. (2008). "Structural origins for the product specificity of SET domain protein methyltransferases." Proc Natl Acad Sci U S A **105**(52): 20659-20664.
- Cui, Q., M. Elstner, et al. (2001). "A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method." J Phys Chem B **105**(2): 569-585.
- Elstner, M., D. Porezag, et al. (1998). "Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties." Phys Rev B **58**(11): 7260-7268.

Field, M. J., P. A. Bash, et al. (1990). "A COMBINED QUANTUM-MECHANICAL AND MOLECULAR MECHANICAL POTENTIAL FOR MOLECULAR-DYNAMICS SIMULATIONS." J Comput Chem **11**(6): 700-733.

Guo, H. B. and H. Guo (2007). "Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity." Proc Natl Acad Sci U S A **104**(21): 8797-8802.

Hu, P., S. Wang, et al. (2008). "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 **130**(12): 3806-3813.

Hu, P. and Y. K. Zhang (2006). "Catalytic mechanism and product specificity of the histone lysine methyltransferase SET7/9: An ab initio QM/MM-FE study with multiple initial structures." J Am Chem Soc **128**(4): 1272-1278.

Jenuwein, T. (2006). "The epigenetic magic of histone lysine methylation." Febs J **273**(14): 3121-3135.

Jorgensen, W. L., J. Chandrasekhar, et al. (1983). "COMPARISON OF SIMPLE POTENTIAL FUNCTIONS FOR SIMULATING LIQUID WATER." J Chem Phys **79**(2): 926-935.

Kumar, S., D. Bouzida, et al. (1992). "THE WEIGHTED HISTOGRAM ANALYSIS METHOD FOR FREE-ENERGY CALCULATIONS ON BIOMOLECULES .1. THE METHOD." J Comput Chem **13**(8): 1011-1021.

Lall, S. (2007). "Primers on chromatin." Nat Struct Mol Biol **14**(11): 1110-1115.

MacKerell, A. D., D. Bashford, et al. (1998). "All-atom empirical potential for molecular modeling and dynamics studies of proteins." J Phys Chem B **102**(18): 3586-3616.

Martin, C. and Y. Zhang (2005). "The diverse functions of histone lysine methylation." Nat Rev Mol Cell Biol **6**(11): 838-849.

Taverna, S. D., H. Li, et al. (2007). "How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers." Nat Struct Mol Biol **14**(11): 1025-1040.

Torrie, G. M. and J. P. Valleau (1974). "MONTE-CARLO FREE-ENERGY ESTIMATES USING NON-BOLTZMANN SAMPLING - APPLICATION TO SUBCRITICAL LENNARD-JONES FLUID." Chem Phys Lett **28**(4): 578-581.

Turner, B. M. (2005). "Reading signals on the nucleosome with a new nomenclature for modified histones." Nat Struct Mol Biol **12**(2): 110-112.

Xiao, B., C. Jing, et al. (2005). "Specificity and mechanism of the histone methyltransferase Pr-Set7." Genes Dev **19**(12): 1444-1454.

Xiao, B., C. Jing, et al. (2003). "Structure and catalytic mechanism of the human histone methyltransferase SET7/9." Nature **421**(6923): 652-656.

Xiao, B., J. R. Wilson, et al. (2003). "SET domains and histone methylation." Curr Opi Struct Biol **13**(6): 699-705.

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

Zhang, X. and T. C. Bruice (2007). "Catalytic mechanism and product specificity of rubisco large subunit methyltransferase: QM/MM and MD investigations." Biochemistry **46**(18): 5505-5514.

Zhang, X. and T. C. Bruice (2007). "Histone lysine methyltransferase SET7/9: Formation of a water channel precedes each methyl transfer." Biochemistry **46**(51): 14838-14844.

Zhang, X. and T. C. Bruice (2008). "Enzymatic mechanism and product specificity of SET-domain protein lysine methyltransferases." Proc Natl Acad Sci U S A **105**(15): 5728-5732.

Zhang, X., H. Tamaru, et al. (2002). "Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase." Cell **111**(1): 117-127.

Zhang, X., Z. Yang, et al. (2003). "Structural basis for the product specificity of histone lysine methyltransferases." Mol Cell **12**(1): 177-185.

Zhang, X. D. and T. C. Bruice (2007). "A quantum mechanics/molecular mechanics study of the catalytic mechanism and product specificity of viral histone lysine methyltransferase." Biochemistry **46**(34): 9743-9751.

Zhang, X. D. and T. C. Bruice (2008). "Mechanism of product specificity of AdoMet methylation catalyzed by lysine methyltransferases: Transcriptional factor p53 methylation by histone lysine methyltransferase SET7/9." Biochemistry **47**(9): 2743-2748.

Zhang, X. D. and T. C. Bruce (2008). "Product specificity and mechanism of protein lysine methyltransferases: Insights from the histone lysine methyltransferase SET8." Biochemistry **47**(25): 6671-6677.

Appendix

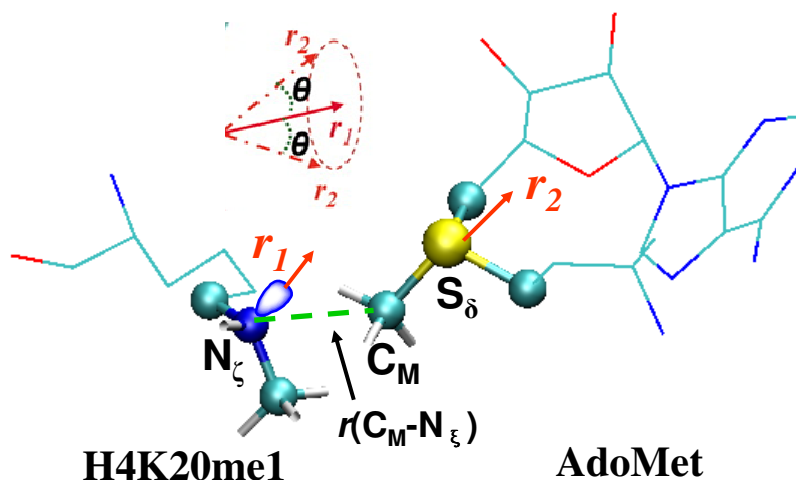


Figure 2.1 The definition of the structural parameters for monitoring the relative orientation of AdoMet and H4K20me1 [H4K20 and H4K20(me)₂] in the reactant complex

The efficiency of the methyl transfer may be related to the distributions of $r(\text{C}_\text{M} \dots \text{N}_\zeta)$ and θ in the reactant complexes. θ is defined as the angle between the two vectors r_1 and r_2 . Here r_1 is the direction of the lone pair of electrons on N_ζ and r_2 is the vector pointing from C_M to S_δ . The reaction coordinate for calculating the free energy profiles for the methyl transfers is $R = r(\text{C}_\text{M} \dots \text{S}_\delta) - r(\text{C}_\text{M} \dots \text{N}_\zeta)$.

Figure 2.2 MD results for the wild-type enzyme (SET8)

(A) The average active-site structure of the reactant complex for the first methyl transfer. SET8 is shown in balls and sticks, and AdoMet and the H4K20 sidechain are in sticks. Hydrogen atoms are not shown for clarity except for those on N_ζ and transferable methyl group. Hydrogen bonds are indicated by red dotted lines, and the distances related to the reactant coordinate are also shown. **(B)** **Left:** the two-dimensional plot of $r(C_M...N_\zeta)$ and θ distributions based on the 1.5-ns simulations of the reactant complex for the first methyl transfer; **Middle:** the free-energy change as a function of $r(C_M...N_\xi)$ obtained from the distributions; **Right:** the free-energy change as a function of θ obtained from the distributions. **(C)** The average structure near the transition state for the first methyl transfer obtained from the free energy (potential of mean force) simulations. **(D)** The average structure of the reactant complex for the second methyl transfer. **(E)** **Left:** the two-dimensional plot of $r(C_M...N_\zeta)$ and θ distributions of the reactant complex for the second methyl transfer; **Middle:** the free-energy change as a function of $r(C_M...N_\xi)$ obtained from the distributions; **Right:** the free-energy change as a function of θ obtained from the distributions. **(F)** The average structure near the transition state for the second methyl transfer.

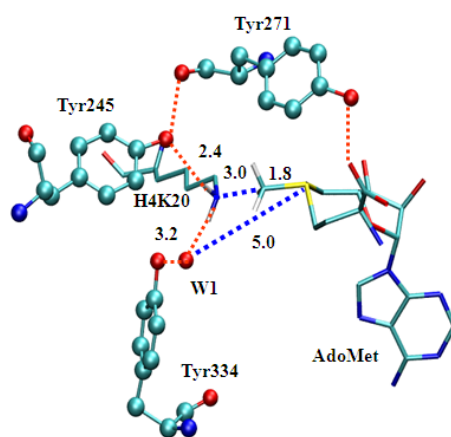
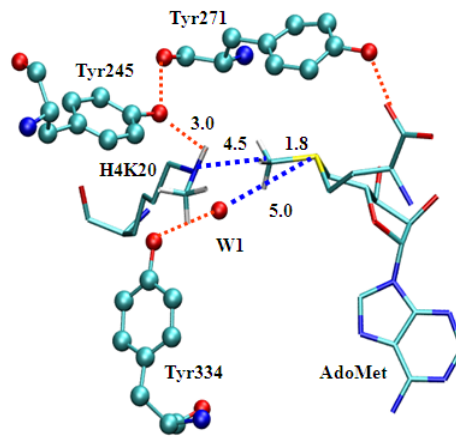
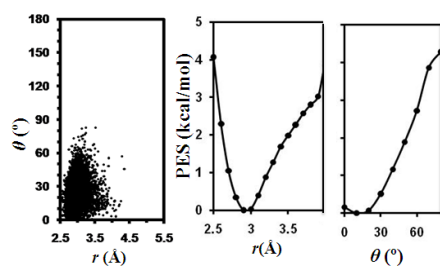
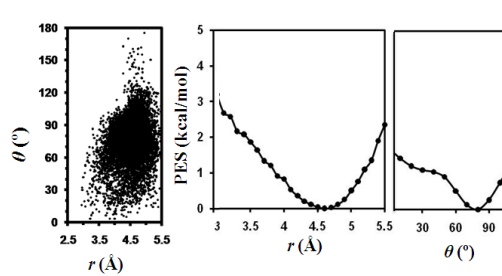
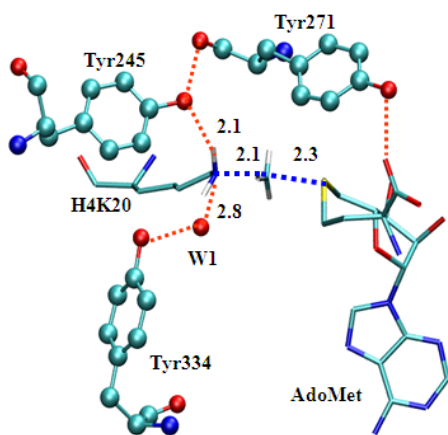
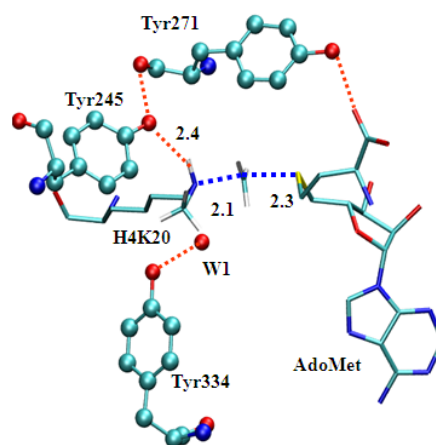
A**D****B****E****C****F**

Figure 2.3 MD results for Y334F

(A) Left: the average active-site structure of the reactant complex for the first methyl transfer; **Right:** the two-dimensional plot of $r(C_M...N_Z)$ and θ distributions of the reactant complex for the first methyl transfer. **(B)** The average structure near the transition state for the first methyl transfer. **(C) Left:** the average structure of the reactant complex for the second methyl transfer; **Right:** the two-dimensional plot of $r(C_M...N_Z)$ and θ distributions of the reactant complex for the second methyl transfer. **(D)** The average structure near the transition state for the second methyl transfer. **(E) Left:** the average structure of the reactant complex for the third methyl transfer; **Right:** the two-dimensional plot of $r(C_M...N_Z)$ and θ distributions of the reactant complex for the third methyl transfer. **(F)** The average structure near the transition state for the third methyl transfer.

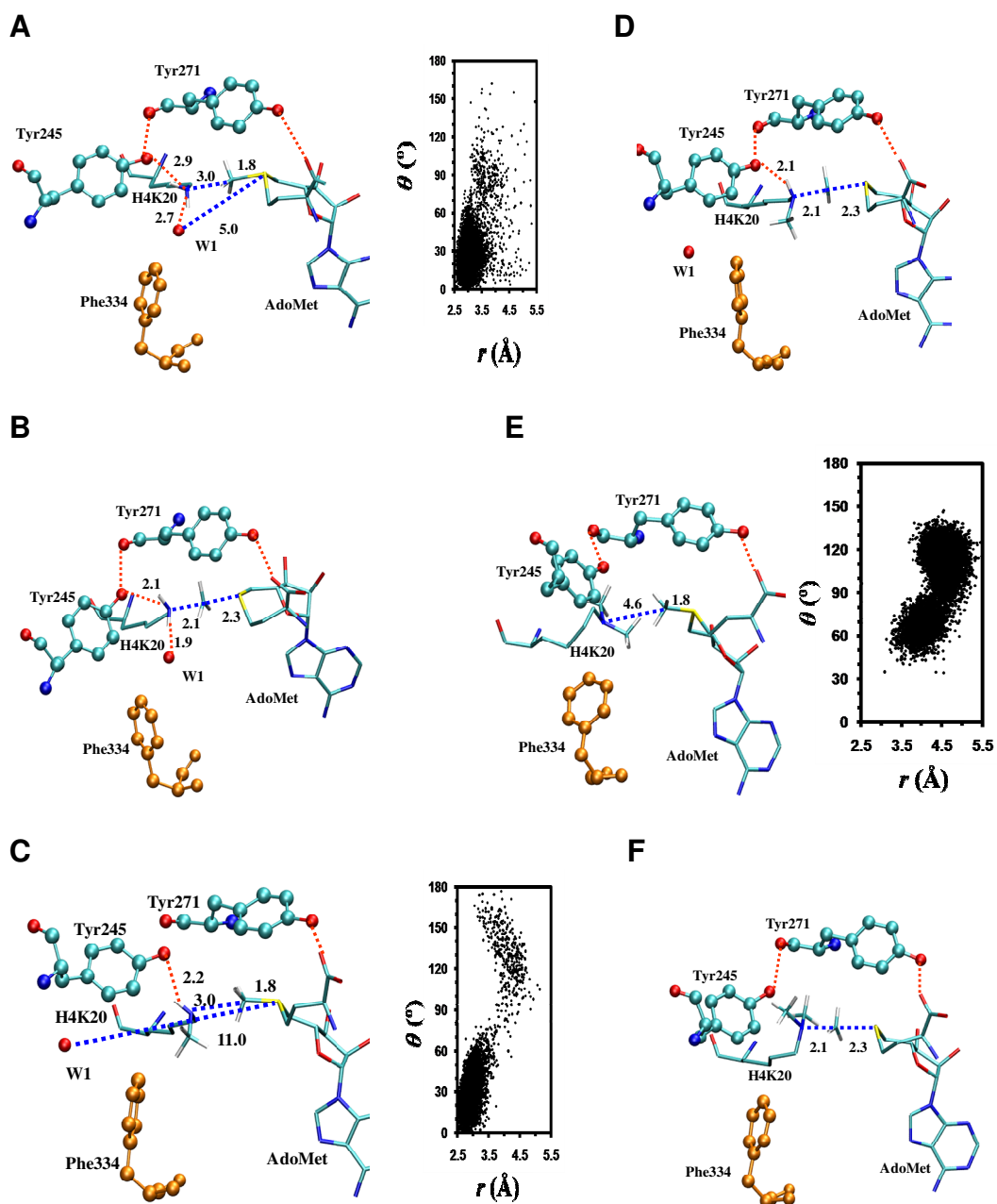
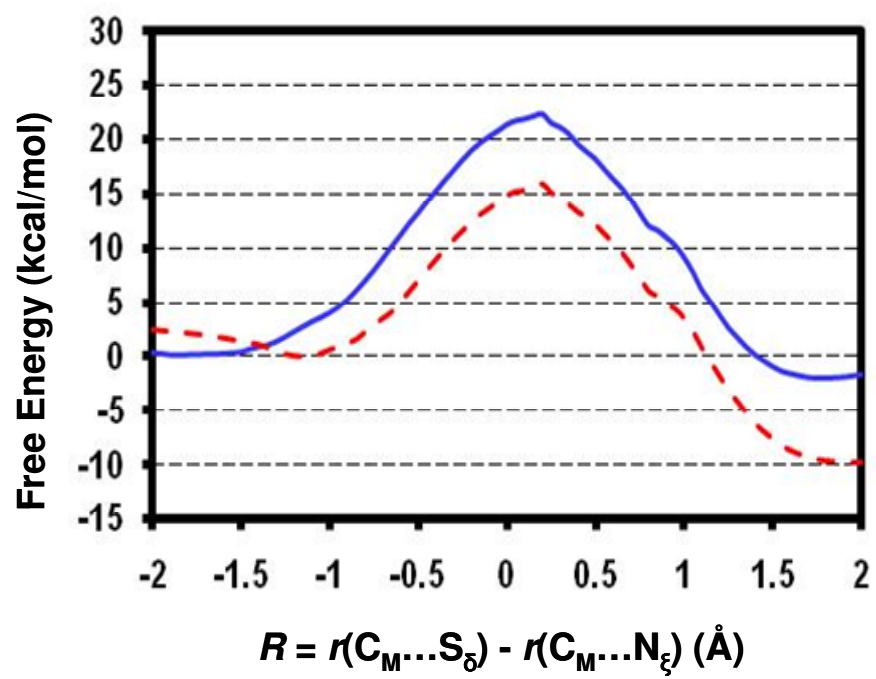


Figure 2.4 Free energy profiles for methyl transfers

(A) Free energy (potential of mean force) changes for the first and second transfers from AdoMet to H4-K20 and H4-K20me1, respectively, as a function of the reaction coordinate [$R = r(C_M \dots S_\delta) - r(C_M \dots N_\zeta)$] in the wild-type SET8. The first methyl transfer: red and dotted line with a free energy barrier of 15.8 kcal/mol. The second methyl transfer: blue and solid line with a free energy barrier of 22.3 kcal/mol (or about 6.5 kcal/mol higher than that of the first methyl transfer). Differences in the free energy barriers may be represented by two energy triplets, $(0, \Delta_{2-1W}, \Delta_{3-1W})$ and $(\Delta_{M-W}, \Delta_{2-1M}, \Delta_{3-1M})$, for the wild-type and mutated enzyme, respectively. For wild-type enzyme the second (Δ_{2-1W}) and third (Δ_{3-1W}) parameters are the differences in the free energy barriers between the second and first and between the third and first methyl transfers, respectively. For the mutated enzyme, the first parameter (Δ_{M-W}) is the difference in the free energy barriers for the first methyl transfer in the wild-type and mutant. The second (Δ_{2-1M}) and third (Δ_{3-1M}) parameters are the differences in the free energy barriers between the second and first and between the third and first methyl transfers, respectively, in the mutant. For SET8, $(0, \Delta_{2-1W}, \Delta_{3-1W}) = (0, 6.5, x)$ (x indicates the undetermined relative barrier in the energy triplet). **(B)** The free energy changes for the first, second and third methyl transfers as a function of the reaction coordinate in the Y334F mutant. The first methyl transfer: red and dotted line with a free energy barrier of 14.1 kcal/mol. The second methyl transfer: blue and solid line with a free energy barrier of 13 kcal/mol. The third

methyl transfer: green and dot-dashed line with a free energy barrier of 23.1 kcal/mol. (Δ_{M-W} , Δ_{2-1M} , Δ_{3-1M}) = (-1.7, -1.1, 9).

A



B

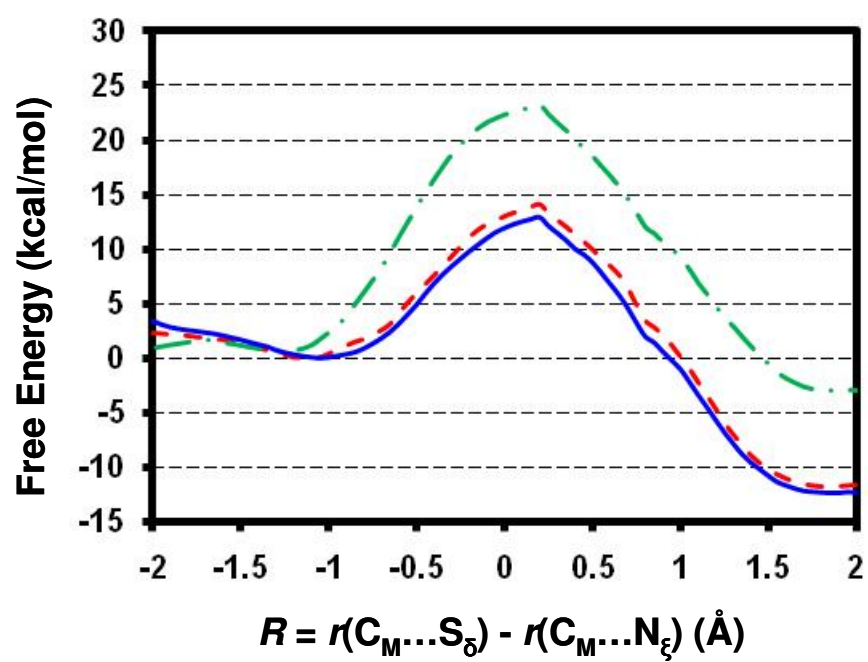
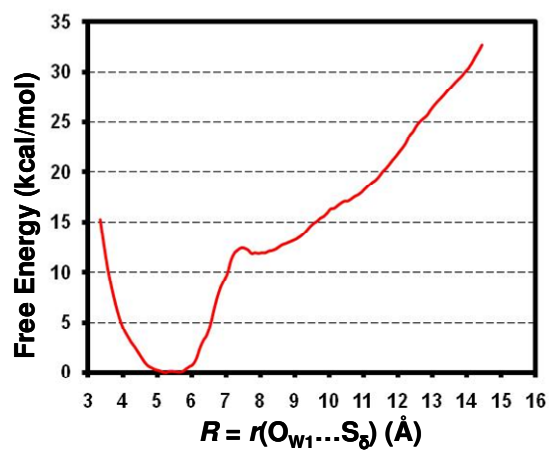
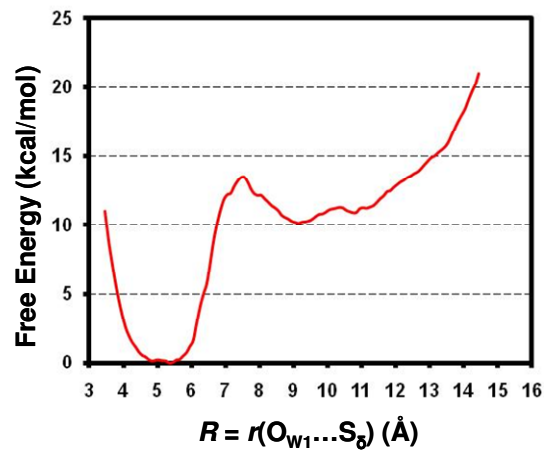
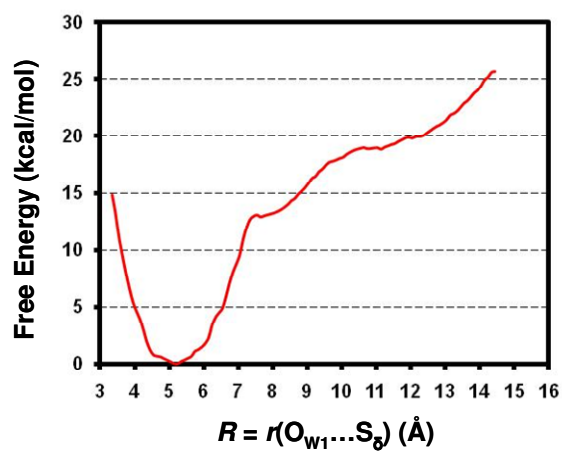
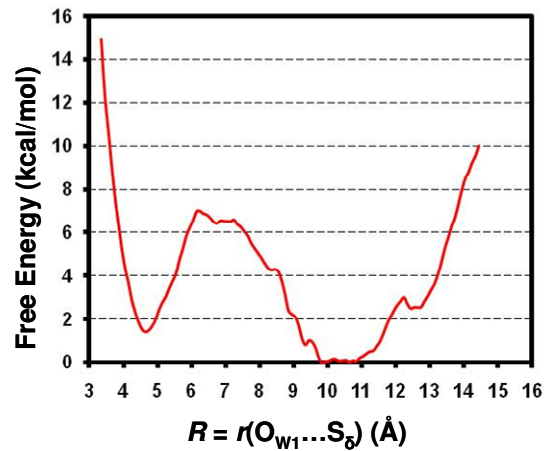


Figure 2.5 Free energy simulations of the dissociation of the active-site water molecule from the active site

(A) Free energy (potential of mean force) change as a function of the distance between the oxygen atom of the active-site water molecule (W1) and sulfur atom of AdoMet [i.e., $r(\text{O}_{\text{W1}} \cdots \text{S}_\delta)$] in the reactant complex for the first methyl transfer in the wild-type enzyme. W1 is stable in the active site, as the free energy minimum is at around 5 Å (i.e., at the similar position observed in the MD simulations in Figure 2.2(A)). High energy at the longer $r(\text{O}_{\text{W1}} \cdots \text{S}_\delta)$ distance suggests that it would be difficult for W1 to move away from the active site. **(B)** Free energy change in the reactant complex of the second methyl transfer in the wild-type enzyme. The simulation data shows that W1 is also stable in the active site. **(C)** Free energy change in the reactant complex of the first methyl transfer in Y334F. **(D)** Free energy change in the reactant complex of the second methyl transfer in Y334F. The most stable position for W1 has changed (now 10-11 Å from the sulfur atom), suggesting that W1 is not in the active site.

A**C****B****D**

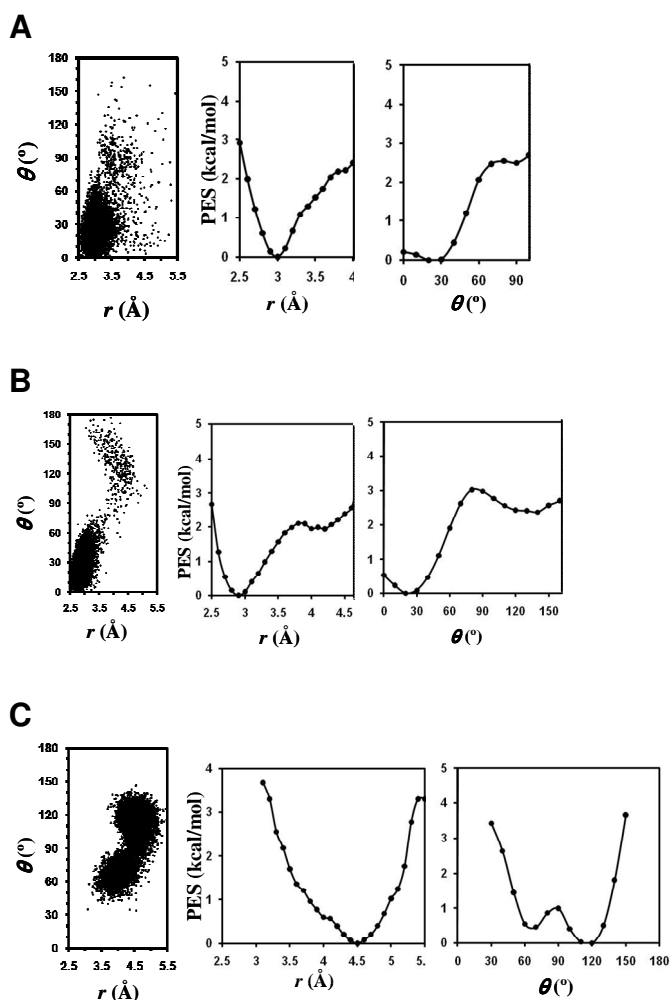


Figure S2.1 MD results for Y334F

(A) Left: the two-dimensional plot of $r(C_M \dots N_Z)$ and θ distributions for the reactant complex of the first methyl transfer; **Middle:** the free-energy change as a function of $r(C_M \dots N_Z)$; **Right:** the free-energy change as a function of θ . **(B)** The two-dimensional plot and the free-energy changes as functions of $r(C_M \dots N_Z)$ and θ for the reactant complex of the second methyl transfer. **(C)** The two-dimensional plot and the free-energy changes as functions of $r(C_M \dots N_Z)$ and θ for the reactant complex of the third methyl transfer.

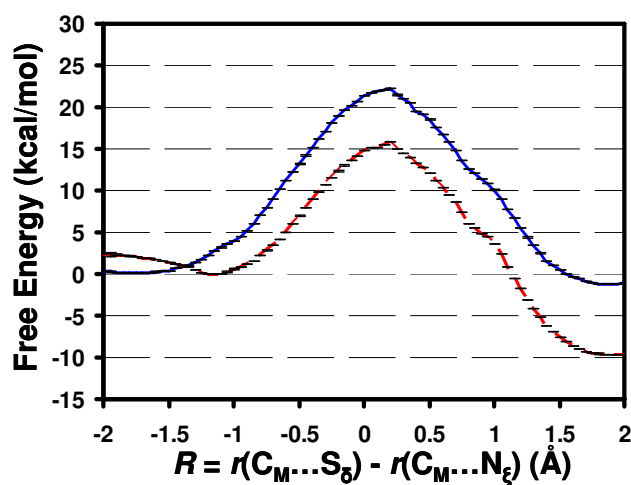
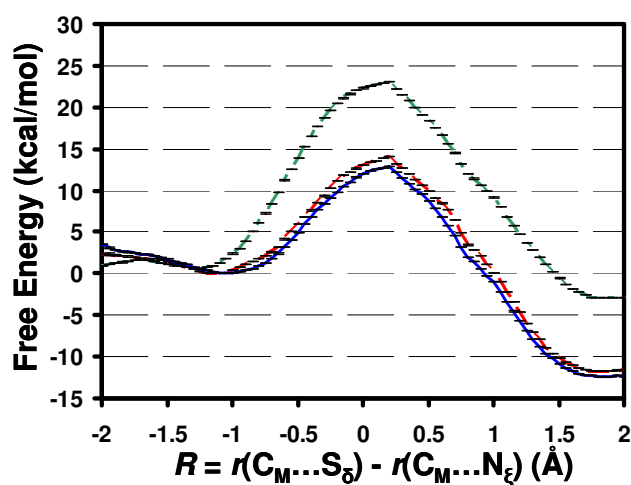
A**B**

Figure S2.2 Free energy changes along with the statistical errors

(A) Free energy (potential of mean force) changes for the first and second transfers as a function of the reaction coordinate [$R = r(C_M \dots S_\delta) - r(C_M \dots N_\xi)$] in the wild-type SET8 along with the statistical errors in the potential of mean force simulations. **(B)** Free energy changes for the first, second and third transfers as a function of the reaction coordinate in Y334F along with the statistical errors.

CHAPTER 3

QM/MM MD AND FREE ENERGY SIMULATIONS OF G9A-LIKE PROTEIN (GLP) AND ITS MUTANTS: UNDERSTANDING THE FACTORS THAT DETERMINE THE PRODUCT SPECIFICITY

A version of this chapter was originally published by Yuzhuo Chu, Jianzhuang Yao and Hong Guo:

Yuzhuo Chu, Jianzhuang Yao and Hong Guo “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* 7(5) (2012): e37674.

Yuzhuo Chu and Dr. Hong Guo designed the experiments; Yuzhuo Chu performed the experiments; Yuzhuo Chu and Dr. Hong Guo analyzed the results and prepared the manuscript. Jianzhuang Yao is involved in the discussions of the design of the experiments.

Abstract

Certain lysine residues on histone tails could be methylated by protein lysine methyltransferases (PKMTs) using S-adenosyl-L-methionine (AdoMet) as the methyl donor. Since the methylation states of the target lysines play a fundamental role in the regulation of chromatin structure and gene expression, it is important to study the property of PKMTs that allows a specific number of methyl groups (one, two or three) to be added (termed as product specificity). It has been shown that the product specificity of PKMTs may be controlled in part by the existence of specific residues at the active site. One of the best examples is a Phe/Tyr switch found in many PKMTs. Here quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) and free energy simulations are performed on wild type G9a-like protein (GLP) and its F1209Y and Y1124F

mutants for understanding the energetic origin of the product specificity and the reasons for the change of product specificity as a result of single-residue mutations at the Phe/Tyr switch as well as other positions. The free energy barriers of the methyl transfer processes calculated from our simulations are consistent with experimental data, supporting the suggestion that the relative free energy barriers may determine, at least in part, the product specificity of PKMTs. The changes of the free energy barriers as a result of the mutations are also discussed based on the structural information obtained from the simulations. The results suggest that the space and active-site interactions around the ϵ -amino group of the target lysine available for methyl addition appear to among the key structural factors in controlling the product specificity and activity of PKMTs.

Introduction

The tails of histone proteins are subject to a variety of post-translational modifications, and these modifications are believed to be a part of the histone code for chromatin regulation (Strahl and Allis 2000). One of such modifications is the methylation of a range of lysine (K) residues on histones, including K4, K9, K27, K36, K79 on histone H3 and K20 on histone H4 (Marmorstein 2003). Catalyzed by protein lysine methyltransferases (PKMTs), histone lysine methylations have been found to affect a variety of important biological processes, including heterochromatin formation, X-chromosome inactivation, transcriptional silencing and activation (Martin and Zhang 2005; Jenuwein 2006). PKMTs may be classified based on their ability to transfer one, two or three

methyl groups from S-adenosyl-L-methionine (AdoMet, the methyl donor) to the ϵ -amino group of target lysine, a special property of the enzymes that is termed as product specificity (Strahl and Allis 2000; Martin and Zhang 2005). Since different methylation states may lead to different downstream events (Turner 2005; Lall 2007; Taverna, Li et al. 2007), it is of fundamental importance to understand the determinant of the product specificity, including the energetic and structural factors that control the product specificity.

A number of experimental (Zhang, Tamaru et al. 2002; Xiao, Jing et al. 2003; Zhang, Yang et al. 2003; Collins, Tachibana et al. 2005; Couture, Collazo et al. 2005; Xiao, Jing et al. 2005; Couture, Dirk et al. 2008; Del Rizzo, Couture et al. 2010; Wu, Min et al. 2010; Xu, Wu et al. 2011) and computational studies (Hu and Zhang 2006; Guo and Guo 2007; Wang, Hu et al. 2007; Hu, Wang et al. 2008; Zhang and Bruice 2008; Zhang and Bruice 2008; Xu, Chu et al. 2009; Chu, Xu et al. 2010; Bai, Shen et al. 2011) have been performed to understand the origin of the product specificity for PKMTs. Structural and mutational studies have identified a Phe/Tyr switch located at the active site of many SET domain PKMTs (Xiao, Jing et al. 2003; Zhang, Yang et al. 2003; Collins, Tachibana et al. 2005; Couture, Collazo et al. 2005). It has been shown previously that the product specificity of the enzyme may depend on whether this position is occupied by a Phe or Tyr residue. Indeed, it has been observed that this position tends to be occupied by a tyrosine residue for mono-methyltransferases (e.g. Y305 in SET7/9 (Del Rizzo, Couture et al. 2010) and Y334 in SET8 (Couture,

Collazo et al. 2005; Couture, Dirk et al. 2008) and by a phenylalanine for di- or tri-methyltransferases (e.g. F281 in DIM-5 (Zhang, Yang et al. 2003)). Moreover, the substitution at the Phe/Tyr switch position could lead to the change of the product specificity that is consistent with the observations on the wild-type enzymes (see above). For example, the F281→Y mutation for DIM-5 altered the enzyme from a tri-methyltransferase to a mono-/di-methyltransferase (Zhang, Yang et al. 2003), while the Y305→F mutation for SET7/9 (Del Rizzo, Couture et al. 2010) or Y334→F mutation for SET8 (Couture, Collazo et al. 2005; Couture, Dirk et al. 2008) was found to switch the enzyme from a mono-methyltransferase to a di-methyltransferase. For H3K9 G9a methyltransferase, the F→Y mutant at the Phe/Tyr switch site (F1152 in G9a) was found to alter the enzyme from a di-methyltransferase to a mono-methyltransferase (Collins, Tachibana et al. 2005); the equivalent residue in its closely relative G9a-like-protein (GLP) is F1209. Previous studies on SET7/9, SET8 and DIM-5 (Couture, Dirk et al. 2008; Xu, Chu et al. 2009; Chu, Xu et al. 2010; Del Rizzo, Couture et al. 2010) have suggested that this Phe/Tyr switch may affect product specificity through altering the affinity of an active site water molecule, as the dissociation of this water molecule is likely to make it easier for further methyl addition (Couture, Dirk et al. 2008; Xu, Chu et al. 2009; Chu, Xu et al. 2010; Del Rizzo, Couture et al. 2010). It would be of interest to examine whether this is the case for GLP or G9a as well.

The alternations of the product specificity of PKMTs by single mutations have also been observed for residues located at some other positions of the

active site (Zhang, Yang et al. 2003; Del Rizzo, Couture et al. 2010; Wu, Min et al. 2010), although detailed studies for the origin of the change of the product specificity due to these mutations are still lacking. For GLP there is a Tyr residue (Y1124) hydrogen-bonding to the methyl accepting nitrogen in the X-ray structures of GLP complexed with the substrate, H3K9me and H3K9me2 peptides (Wu, Min et al. 2010); the equivalent residue in G9a is Y1067. It has been found that the Y1067→F mutant for G9a is able to increase the enzyme's ability to add a third methyl group to the target lysine (even though it is not at the Phe/Tyr switch position) and change the enzyme to a tri-methyltransferase (Wu, Min et al. 2010). Further studies are still necessary to determine the origin of the product specificity change for this case.

Hybrid quantum mechanical/molecular mechanical (QM/MM) free energy simulations have been performed previously on various SET domain PKMTs, including DIM-5, SET7/9, SET8 and their mutants involving the residue at the Phe/Tyr switch site (Guo and Guo 2007; Xu, Chu et al. 2009; Chu, Xu et al. 2010). One of the main purposes was to study the energetic origin of the product specificity and identify the structural factors which might control the methylation states of target lysine residues (Guo and Guo 2007; Xu, Chu et al. 2009; Chu, Xu et al. 2010). For instance, it has been proposed that the existence of a relatively high free energy barrier for one or more methyl transfer processes could stop the continuation of the methyl addition and may therefore determine how the epigenetic marks of lysine methylation are written (Guo and Guo 2007; Xu, Chu

et al. 2009; Chu, Xu et al. 2010). However, questions remain whether the previous conclusions could be extended to mutations involving the residues other than those at the Phe/Tyr switch site. In this paper, the results of the QM/MM molecular dynamics (MD) and free energy simulations are reported for different methyl transfer processes of wild type (WT) GLP as well as its F1209Y and Y1124F mutants. The free energy barriers of the methyl transfer processes obtained from simulations are well correlated with experimental observations of product specificities (Collins, Tachibana et al. 2005; Wu, Min et al. 2010), supporting previous conclusions that the relative free energy barrier for the methyl transfer process may determine, at least in part, the number of methyl groups added by the enzyme (Guo and Guo 2007; Xu, Chu et al. 2009; Chu, Xu et al. 2010). The dynamic information of reactant complexes gathered from the QM/MM MD simulations were also examined, and the results suggest that the ability of the reactant complex to form the reactive configuration may be used as an important indicator for the product specificity of PKMTs.

Method

The QM/MM MD and free energy (potential of mean force) simulations were applied to determine the free energy profiles of different methyl transfer processes catalyzed by WT GLP, F1209Y and Y1124F mutants. MD simulations were also performed on the reactant complexes of the methyl transfers for each system to study their dynamic properties related to the methyl transfers. AdoMet and the target lysine side chains were treated by QM and the rest of the system

were treated by MM. The link-atom approach (Field, Bash et al. 1990) was implemented with CHARMM (Brooks, Bruccoleri et al. 1983) to separate the QM and MM regions. A modified TIP3 water model (Jorgensen, Chandrasekhar et al. 1983; Neria, Fischer et al. 1996) was used for the solvent. The stochastic boundary molecular dynamics method (Brooks, Brunger et al. 1985) was used for the QM/MM MD and free energy simulations. The system was separated into a reaction zone and a reservoir zone, and the reaction zone was further divided into a reaction region and a buffer region. The reaction region was a sphere with radius r of 20 Å, and the buffer region had r equal to $20 \text{ Å} \leq r \leq 22 \text{ Å}$. The reference center for partitioning the system was chosen to be the C₆ of the target K-9. The resulting systems contained around 5500 atoms, including about 800-900 water molecules. There is a crystal water molecule in the active site of the X-ray structure, and no additional water at the active site was introduced through the solvation process.

The SCC-DFTB method (Cui, Elstner et al. 2001) implemented in CHARMM (Brooks, Bruccoleri et al. 1983) was used for the QM atoms and the all-hydrogen CHARMM potential function (PARAM27) (MacKerell, Bashford et al. 1998) was used for the MM atoms. In our earlier studies, the results of the SCC-DFTB and B3LYP/6-31G** methods for the description of the methyl transfer in a small model system were compared using energy minimization-based approach (Guo and Guo 2007; Xu, Chu et al. 2009; Chu, Xu et al. 2010). This comparison allowed us to understand the performance of the semi-empirical method in the

description of the bond breaking and making and derive an empirical correction for the free energy curves obtained from the potential of mean force simulations. The results indicated that the energy curves from the corrected SCC-DFTB and B3LYP/6-31G** were very close, supporting the use of this approach to make the first correction to the bond breaking and making events involving the simple and similar S_N2 methyl transfer processes. The similar approach of this “first-order” correction to the free energy curves was also adopted in this work. For a more detailed discussion, see previous work by our group (Guo and Guo 2007). It should be pointed out that cares must be exercised when applying this procedure to study other methyl transfer processes involving different methyl donors and acceptors (such as the methyl transfer in protein arginine methyltransferases). In such cases, a model system containing the corresponding methyl donor and acceptor should be used instead for determining the deficiency of the SCC-DFTB method in the description of methyl transfer. The results of SCC-DFTB calculations on the methyl transfer for such model systems can be compared with the results from high-level ab initio (e.g., B3LYP/6-31G**) calculations to obtain reasonable corrections for such methyl transfer process.

The initial coordinates for the reactant complexes of the first, second and third methyl transfers were based on the crystallographic complexes (PDB codes: 2RFI, 3HNA) of GLP containing AdoHcy and short H3K9me1 or H3K9me2 peptide (Wu, Min et al. 2010). A methyl group was added to AdoHcy to form AdoMet. The initial coordinates for F1209Y and Y1124F mutants were

generated by changing the corresponding amino acid manually. For the first methyl transfer step, methyl group in H3K9me1 was removed to generate H3K9. The initial structures for the entire stochastic boundary system were optimized using the steepest descent (SD) and adopted-basis Newton-Raphson (ABNR) methods. The systems were gradually heated from 50.0 to 283.15 K in 50 ps and equilibrated at 283.15 K for 500 ps. A 1-fs time step was used for integration of the equation of motion, and the coordinates were saved every 50 fs for analyses. After 500-ps of equilibration were performed, 1-ns QM/MM MD simulations were carried out for each of the reactant complexes of the first, second and third methyl transfers. As discussed in the previous studies (Guo and Guo 2007; Xu, Chu et al. 2009; Chu, Xu et al. 2010), the S_N2 methyl transfer from AdoMet to K9, K9me1 or K9me2 is presumably more efficient if the S-CH₃ group of AdoMet is well aligned with the lone pair of the electrons of N_ξ in the reactant complex; i.e., with a small θ angle and relatively short C_M-N_ξ distance. Therefore, we determined the distributions of $r(\text{C}_M\text{-N}_\xi)$ and θ from the MD trajectories; θ was defined as the angle between the direction of the C_M-S_δ bond (r_2) and the direction of the electron lone pair on N_ξ (r_1) (see **Figure 3.1**).

The umbrella sampling method (Torrie and Valleau 1974) implemented in the CHARMM program along with the Weighted Histogram Analysis Method (WHAM) (Kumar, Bouzida et al. 1992) was applied to determine the change of the free energy (potential of mean force) as a function of the reaction coordinate for the methyl transfer from AdoMet to H3-K9, H3-K9me1 or H3-K9me2 in the

WT and mutated enzyme. The reaction coordinate was defined as a linear combination of $r(\text{C}_\text{M}-\text{N}_\xi)$ and $r(\text{C}_\text{M}-\text{S}_\delta)$ [$R = r(\text{C}_\text{M}-\text{S}_\delta) - r(\text{C}_\text{M}-\text{N}_\xi)$] (see **Figure 3.1**). For each methyl transfer process, twenty windows were used, and for each window 50-ps production runs were performed after 20-ps equilibration. The force constants of the harmonic biasing potentials used in the PMF simulations were from 50 to 400 kcal mol⁻¹ Å⁻². More sophisticated simulation methods (Glowacki, Paci et al. 2009; Ricchiuto, Brukhno et al. 2011) may be used in the future to generate the free energy surfaces of methyl transfers that also depend on angular coordinates.

Results

The active-site structure along with the average values of some structural parameters of the reactant complex for the first methyl transfer in WT GLP is shown in **Figure 3.2(A)**. As is clear from this figure, the lone pair of electrons on N_ξ of the target lysine is well aligned with the methyl group of AdoMet in the structure, a condition that was found to make the methyl transfer relatively easier to occur. The distribution plot from the MD simulations for the reactant complex (**Figure 3.2(B)**) also shows that there are large populations of the structures with relatively short $r(\text{C}_\text{M}-\text{N}_\xi)$ distance and small values of the θ angle; the average $r(\text{C}_\text{M}-\text{N}_\xi)$ distance is around 3.0 Å and the average θ angle is about 30°. **Figure 3.2(A)** and **3.2(G)** (the structure near the transition state obtained from the free energy simulations along with the average values of some structural parameters) show that Tyr1124 and an active-site water molecule (W1) form stable hydrogen

bonds with the ϵ -amino group of target lysine, and these hydrogen bonds may help to orientate the electron lone pair and achieve a good alignment with the methyl donor group of AdoMet. The active-site structure along with the average values of some structural parameters of the reactant complex for the second methyl transfer in WT GLP is given in **Figure 3.2(C)**. Similar to the case for the first methyl transfer, a good alignment between the lone pair of electrons and the methyl group from AdoMet can also be achieved in this reactant complex (also see **Figure 3.2(D)**). Examination of the structure in **Figure 3.2(C)** shows that while Tyr1124 still forms a hydrogen bond with the ϵ -amino group of target lysine as in the case for the first methyl transfer, the hydrogen bond with W1 disappears as a result of losing the hydrogen-bond donating capacity for the methylated lysine. It is of interest to note from the average structure near the transition state (**Figure 3.2(H)**) that W1 has dissociated from the active site during the methyl transfer. The dissociation of this water presumably produces additional space in the active site that makes the second methyl transfer easier. The active-site structure along with the average values of some structural parameters of the reactant complex for the third methyl transfer is shown in **Figure 3.2(E)**. As is demonstrated in this figure, the lone pair of electrons cannot be well aligned with the methyl group; the average $r(C_M-N_\epsilon)$ distance is about 4.5 Å and the values of θ angle become significantly larger with a broad distribution (**Figure 3.2(F)**).

The active-site structure along with the average values of some structural parameters of the reactant complex for the first methyl transfer in the F1209Y

mutant is shown in **Figure 3.3(A)**. As is the case for the wild type enzyme, the lone pair of electrons on N_ξ of target lysine is well aligned with the methyl group of AdoMet in the reactant complex for the first methyl transfer. **Figure 3.3(C)** and **3.3(D)** show that for the second methyl transfer, the chances for the θ angle to reach 0-15° (i.e., near the line arrangement) become smaller in the mutant than in wild-type. Moreover, the free energy cost for generating the linear configuration is as much as 2-3 kcal/mol (only about 1 kcal/mol in wild-type, see **Figure 3.2(D)**). This energy cost may be related to the relatively higher free energy barrier for the second methyl transfer in F1209Y (see below). The structure near the transition state (**Figure 3.3(F)**) shows that unlike in the case of wild-type, W1 remains in the active site during the second methyl transfer. The failure for the dissociation of W1 from the active site is presumably due to the presence of the hydrogen bond between W1 and Y1209 (that stabilizes the water molecule), and the existence of W1 is likely to make it more difficult for further methyl addition.

The free energy profiles for methyl transfer processes in WT GLP and the F1209Y mutant are plotted in **Figure 3.4(A)** and **3.4(B)**, respectively, as a function of the reaction coordinate [$R = r(C_M \dots S_\delta) - r(C_M \dots N_\zeta)$]. **Figure 3.4(A)** shows that the increase of the free energy barrier from the first or second methyl transfer to that of the third methyl transfer in WT GLP is considerably large (6-9 kcal/mol), consistent with the fact that WT GLP is a di-methyltransferase (Collins, Tachibana et al. 2005). For the F1209Y mutant, there is a significant

increase of the free energy barrier from the first to second methyl transfer (3.6 kcal/mol) and the chance for observing dimethyl lysine from F1209Y is expected to be much smaller compared to that for wild-type. The results are consistent with the experimental observation that F1209Y mutation altered the enzyme from a dimethyltransferase to a mono-methyltransferase (Collins, Tachibana et al. 2005). These data agree with the previous conclusions that the existence of a relatively high free energy barrier could stop the further addition of methyl group and may therefore control the product specificity of PKMTs (Guo and Guo 2007; Xu, Chu et al. 2009; Chu, Xu et al. 2010), at least in many cases.

The active site structures along with the average values of some structural parameters of the reactant complexes for the first, second and third methyl transfer of the Y1124F mutant are shown in **Figure 3.5(A), 3.5(C), 3.5(E)**, respectively. All the structures have the lone pair of electrons on N_{ξ} of target lysine well aligned with the methyl donor group of AdoMet. The distribution plots for these structures also show large population of structures with relatively short $r(C_M-N_{\xi})$ distances and small values of the θ angle; i.e. the average $r(C_M-N_{\xi})$ distances are around 3.0 Å and the average angles are in the range of 0-30° (**Figure 3.5(B), 3.5(D), 3.5(F)**). The distributions of the $r(C_M-N_{\xi})$ distance and θ angle for the second methyl transfer is broader compared with that observed in wild type enzyme. This could be caused by the loss of the stabilization from the hydrogen bond involving Tyr1124. Nevertheless, the free energy costs for generating the linear configuration is rather small (**Figure 3.5(D)**). The

substitution of Tyr1124 by Phe may be able to lead to additional space around the ϵ -amino group of the target lysine and make it easier to add the third methyl group to target lysine. Indeed, the free energy profiles of all three methyl transfer steps in the Y1124F mutant in **Figure 3.6** show that all the three methyl transfers have similar and relatively low free energy barriers, consistent with the experimental observation that the Y1124F mutant is a tri-methyltransferase (Wu, Min et al. 2010).

Discussion

The results of the simulations show that the structures of the reactant complexes for the first and second methyl transfers in WT (**Figures 3.2(A), 3.2(C)**), the first methyl transfer in F1209Y (**Figure 3.3(A)**), and the first, second and third methyl transfers in Y1124F (**Figures 3.5(A), 3.5(C), 3.5(E)**) are rather similar to the corresponding structures near transition state (TS). The distribution plots (**Figure 3.2(B), 3.2(D), 3.3(B), 3.5(B), 3.5(D), 3.5(F)**) of these structures show large population of structures with relatively short $r(\text{C}_M\text{-N}_\epsilon)$ distances and small values of the θ angle, which are close to the distribution plots (**Figure S3.1**) of the TS structures. This is in contrast to the case of the third methyl transfer in wild-type and the second methyl transfer in F1209Y mutant where the reactant structures (**Figure 3.2(E), 3.3(C)**) are significantly distorted from the corresponding TS structures (**Figure 3.2(I), 3.3(F)**). The distribution plots (**Figure 3.2(F), 3.3(D)**) of these reaction complexes show large population of structures with either relatively long $r(\text{C}_M\text{-N}_\epsilon)$ distances or large values of the θ angle, which

are considerably different from the corresponding distribution plot (**Figure S3.1**) of the TS structures. Thus, the results are consistent with the previous suggestion (Guo and Guo 2007; Xu, Chu et al. 2009; Chu, Xu et al. 2010) that one of the reasons for the existence of the relatively low free energy barrier is likely owed to the fact that a part of TS stabilization is already reflected on the reactant state through the generation of the TS-like conformation. Thus, a good alignment of the electron lone pair on N ϵ of target lysine with the methyl group of AdoMet is likely to lead to a relatively efficient methyl transfer, while a poor alignment could make the methyl transfer more difficult.

Two free energy triplets, (0, Δ_{2-1W} , Δ_{3-1W}) and (Δ_{M-W} , Δ_{2-1M} , Δ_{3-1M}) have been proposed (Guo and Guo 2007; Xu, Chu et al. 2009; Chu, Xu et al. 2010) to describe the energetics of the methyl transfers for PKMTs and their mutants, respectively, and to determine their product specificity. Here the free energy barrier of the first methyl transfer step in WT enzyme is taken as the zero. For WT enzyme, the second (Δ_{2-1W}) and third (Δ_{3-1W}) parameters are the differences in the free energy barriers between the second and first and between the third and first methyl transfers, respectively. These two parameters therefore provide information about the relative efficiencies of the second and third methyl transfers compared to the first methyl transfer. For instance, if Δ_{2-1W} is significant (e.g., in the case of SET8 (Chu, Xu et al. 2010)), the enzyme may stop the methylation process leading to a mono-methylase. For the mutated enzyme, the first parameter (Δ_{M-W}) is the difference in the free energy barriers for the first methyl

transfer in the WT and mutant. The second (Δ_{2-1M}) and third (Δ_{3-1M}) parameters are the differences in the free-energy barriers between the second and first and between the third and first methyl transfer, respectively (e.g., defined similar to the cases for WT). Based on such definitions, the energy triplets for WT, F1209Y and Y1124F mutants of GLP are (0, 2.4, 8.7), (0.4, 4.0, x) and (0.5, 2.5, -0.1), respectively (**Figure 3.4 and 3.6**). The significant increase of the free energy barrier from the first or second methyl transfer to that of the third methyl transfer in WT GLP is consistent with the observation that the enzyme is a di-methylase based on the simulations. For the F1209Y mutant, the free energy barrier of the second methyl transfer is around 3.6 kcal/mol higher than the first methyl transfer, and the mutant is more likely to be mono-methylase. It is of interest to note that the free energy barriers of all three methyl transfers in the Y1124F mutant are similar, suggesting that the replacement of Y1124 by F changes the enzyme from a di-methylase into a tri-methylase. All these results are consistent with previous experimental observations (Collins, Tachibana et al. 2005; Wu, Min et al. 2010).

The structural information generated from the simulations shows that the dissociation of W1 might occur during the second methyl transfer in both WT GLP and the Y1124F mutant (**Figure 3.2(H), 3.5(H)**). For the second methyl transfer in the F1209Y mutant, however, W1 cannot dissociate from the active site, presumably, due to its formation of the hydrogen bond with Tyr1209. Consequently, the free energy barrier increases significantly for the addition of the second methyl group, leading to a mono-methylase. These results are

consistent with the earlier suggestion that the dissociation of the active site water molecule is important for further addition of methyl group to target lysine (Couture, Dirk et al. 2008; Xu, Chu et al. 2009; Chu, Xu et al. 2010; Del Rizzo, Couture et al. 2010). For the Y1124F mutant, additional space in the active site may be generated by the removal of the hydroxyl group on Y1124, making it easier for the addition of the third methyl group.

Conclusion

In this study, the QM/MM MD and free energy simulations have been performed on different methyl transfer steps from AdoMet to target lysine/methyl-lysine in WT GLP as well as its F1209Y and Y1124F mutants. The reactant complexes for different methyl transfers generated from QM/MM MD simulations support the suggestion that the ability of the reactant complex to form the TS-like reactive configuration (in which the electron lone pair on N_ξ of target lysine is well aligned with the methyl group of AdoMet) is important for the methyl addition. The free energy barriers of methyl transfers were also calculated. The results show that the relative free energy barriers for the three methyl transfers may determine the product specificity for WT GLP and its mutants. The role of the active site water molecule in controlling product specificity was also discussed.

References

- Bai, Q. F., Y. L. Shen, et al. (2011). "Modeling a New Water Channel That Allows SET9 to Dimethylate p53." Plos One **6**(5): e19856.
- Brooks, B. R., R. E. Bruccoleri, et al. (1983). "CHARMM - a program for macromolecular energy, minimization, and dynamics calculations." J Comput Chem **4**: 187-217.
- Brooks, C. L., A. Brunger, et al. (1985). "Active-site dynamics in protein molecules - a stochastic boundary molecular-dynamics approach." Biopolymers **24**(5): 843-865.
- Chu, Y. Z., Q. Xu, et al. (2010). "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 **6**(4): 1380-1389.
- Collins, R. E., M. Tachibana, et al. (2005). "In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases." J Biol Chem **280**(7): 5563-5570.
- Couture, J. F., E. Collazo, et al. (2005). "Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase." Genes Dev **19**(12): 1455-1465.
- Couture, J. F., L. M. A. Dirk, et al. (2008). "Structural origins for the product specificity of SET domain protein methyltransferases." Proc Natl Acad Sci U S A **105**(52): 20659-20664.
- Cui, Q., M. Elstner, et al. (2001). "A QM/MM implementation of the self-consistent

charge density functional tight binding (SCC-DFTB) method." J Phys Chem B **105**(2): 569-585.

Del Rizzo, P. A., J. F. Couture, et al. (2010). "SET7/9 Catalytic Mutants Reveal the Role of Active Site Water Molecules in Lysine Multiple Methylation." J Bio Chem **285**(41): 31849-31858.

Field, M. J., P. A. Bash, et al. (1990). "A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations." J Comput Chem **11**(6): 700-733.

Glowacki, D. R., E. Paci, et al. (2009). "Boxed Molecular Dynamics: A Simple and General Technique for Accelerating Rare Event Kinetics and Mapping Free Energy in Large Molecular Systems." J Phys Chem B **113**(52): 16603-16611.

Guo, H. B. and H. Guo (2007). "Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity." Proc Natl Acad Sci U S A **104**(21): 8797-8802.

Hu, P., S. Wang, et al. (2008). "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 **130**(12): 3806-3813.

Hu, P. and Y. K. Zhang (2006). "Catalytic mechanism and product specificity of the histone lysine methyltransferase SET7/9: An ab initio QM/MM-FE study with multiple initial structures." J Am Chem Soc **128**(4): 1272-1278.

Humphrey W, D. A. and K. Schulten (1996). "VMD: visual molecular dynamics." J Mol Graph **14**(33-38): 27–38.

Jenuwein, T. (2006). "The epigenetic magic of histone lysine methylation." Febs J **273**(14): 3121-3135.

Jorgensen, W. L., J. Chandrasekhar, et al. (1983). "Comparision of simple potential functions for simulating liquid water." J Chem Phys **79**(2): 926-935.

Kumar, S., D. Bouzida, et al. (1992). "The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method." J Comput Chem **13**(8): 1011-1021.

Lall, S. (2007). "Primers on chromatin." Nat Struc Mol Biol **14**(11): 1110-1115.

MacKerell, A. D., D. Bashford, et al. (1998). "All-atom empirical potential for molecular modeling and dynamics studies of proteins." J Phys Chem B **102**(18): 3586-3616.

Marmorstein, R. (2003). "Structure of SET domain proteins: a new twist on histone methylation." Trends Biochem Sci **28**(2): 59-62.

Martin, C. and Y. Zhang (2005). "The diverse functions of histone lysine methylation." Nat Rev Mol Cell Biol **6**(11): 838-849.

Neria, E., S. Fischer, et al. (1996). "Simulation of activation free energies in molecular systems." J Chem Phys **105**(5): 1902-1921.

Ricchiuto, P., A. V. Brukhno, et al. (2011). "Communication: Conformation state diagram of polypeptides: A chain length induced alpha-beta transition." J Chem Phys **135**(6).

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature **403**(6765): 41-45.

Taverna, S. D., H. Li, et al. (2007). "How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers." Nat Struc Mol Biol **14**(11): 1025-1040.

Torrie, G. M. and J. P. Valleau (1974). "Monte-carlo free-energy estimates using non-boltzmann sampling - application to subcritical lennard -jones fluid." Chem Phys Lett **28**(4): 578-581.

Turner, B. M. (2005). "Reading signals on the nucleosome with a new nomenclature for modified histones." Nat Struc Mol Biol **12**(2): 110-112.

Wang, S. L., P. Hu, et al. (2007). "Ab initio quantum mechanical/molecular mechanical molecular dynamics simulation of enzyme catalysis: The case of histone lysine methyltransferase SET7/9." J Phys Chem B **111**(14): 3758-3764.

Wu, H., J. R. Min, et al. (2010). "Structural Biology of Human H3K9 Methyltransferases." Plos One **5**(1): e8570.

Xiao, B., C. Jing, et al. (2005). "Specificity and mechanism of the histone methyltransferase Pr-Set7." Genes Dev **19**(12): 1444-1454.

Xiao, B., C. Jing, et al. (2003). "Structure and catalytic mechanism of the human histone methyltransferase SET7/9." Nature **421**(6923): 652-656.

Xu, Q., Y. Z. Chu, et al. (2009). "Energy Triplets for Writing Epigenetic Marks: Insights from QM/MM Free-Energy Simulations of Protein Lysine Methyltransferases." Chem Eur J **15**: 12596 – 12599.

Xu, S. T., J. Wu, et al. (2011). "Structural and biochemical studies of human lysine methyltransferase Smyd3 reveal the important functional roles of its post-SET and TPR domains and the regulation of its activity by DNA binding." Nucleic Acids Res **39**(10): 4438-4449.

Zhang, X. and T. C. Bruice (2008). "Enzymatic mechanism and product specificity of SET-domain protein lysine methyltransferases." Proc Natl Acad Sci U S A **105**(15): 5728-5732.

Zhang, X., H. Tamaru, et al. (2002). "Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase." Cell **111**(1): 117-127.

Zhang, X., Z. Yang, et al. (2003). "Structural basis for the product specificity of histone lysine methyltransferases." Mol Cell **12**(1): 177-185.

Zhang, X. D. and T. C. Bruice (2008). "Product specificity and mechanism of protein lysine methyltransferases: Insights from the histone lysine methyltransferase SET8." Biochemistry **47**(25): 6671-6677.

Appendix

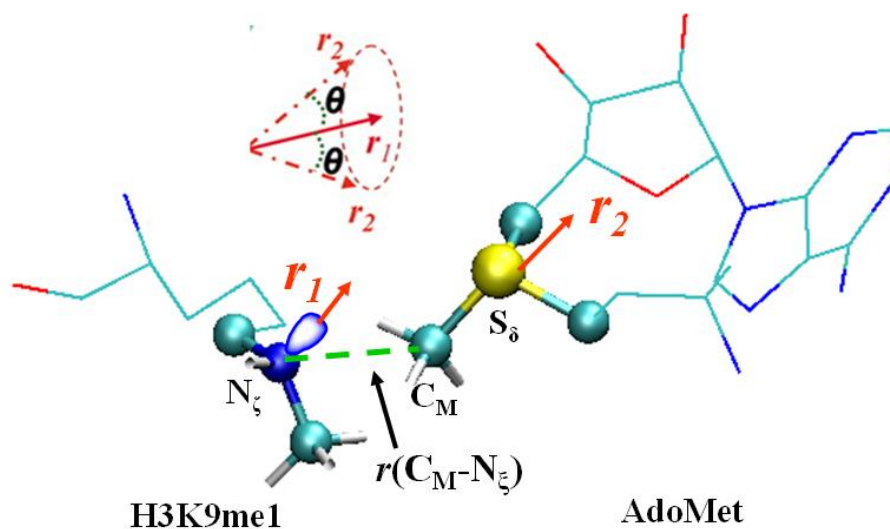


Figure 3.1 The definition of the structural parameters for monitoring the relative orientation of AdoMet and H3K9me1 [H3K9 and H3K9(me)₂] in the reactant complex

The efficiency of the methyl transfer may be related to the distributions of $r(C_M \dots N_\xi)$ and θ in the reactant complexes. θ is defined as the angle between the two vectors r_1 and r_2 . Here r_1 is the direction of the lone pair of electrons on N_ξ and r_2 is the vector pointing from C_M to S_δ . The reaction coordinate for calculating the free energy profiles for the methyl transfers is $R = r(C_M \dots S_\delta) - r(C_M \dots N_\xi)$.

Figure 3.2 MD results for the WT enzyme (GLP)

(A) A representative active-site structure along with the average values of some structural parameters of the reactant complex for the first methyl transfer. GLP is shown in balls and sticks, and AdoMet and the H3K9 sidechain are in sticks. Hydrogen atoms are not shown for clarity except for those on N_ζ and transferable methyl group. Hydrogen bonds are indicated by red dotted lines, and the distances related to the reactant coordinate are also shown. **(B) Left:** the two-dimensional plot of $r(C_M...N_\zeta)$ and θ distributions based on the 1.5-ns simulations of the reactant complex for the first methyl transfer; **Middle:** the free-energy change as a function of $r(C_M...N_\zeta)$ obtained from the distributions; **Right:** the free-energy change as a function of θ obtained from the distributions. **(C)** The active-site structure along with the average values of some structural parameters of the reactant complex for the second methyl transfer. **(D) Left:** the two-dimensional plot of $r(C_M...N_\zeta)$ and θ distributions of the reactant complex for the second methyl transfer; **Middle:** the free-energy change as a function of $r(C_M...N_\zeta)$ obtained from the distributions; **Right:** the free-energy change as a function of θ obtained from the distributions. **(E)** The structure of the reactant complex for the third methyl transfer. **(F) Left:** the two-dimensional plot of $r(C_M...N_\zeta)$ and θ distributions of the reactant complex for the third methyl transfer; **Middle:** the free-energy change as a function of $r(C_M...N_\zeta)$ obtained from the distributions; **Right:** the free-energy change as a function of θ obtained from the distributions. **(G)** A representative active-site structure along with the average values of some

structural parameters near the transition state for the first methyl transfer obtained from the free energy (potential of mean force) simulations. **(H)** The structure along with the average values of some structural parameters near the transition state for the second methyl transfer. **(I)** The structure near the transition state for the third methyl transfer. All images were made by VMD (Humphrey and Schulten 1996). The distances shown on the structures are the calculated average distances from the trajectories over the 50-ps production runs in the corresponding window.

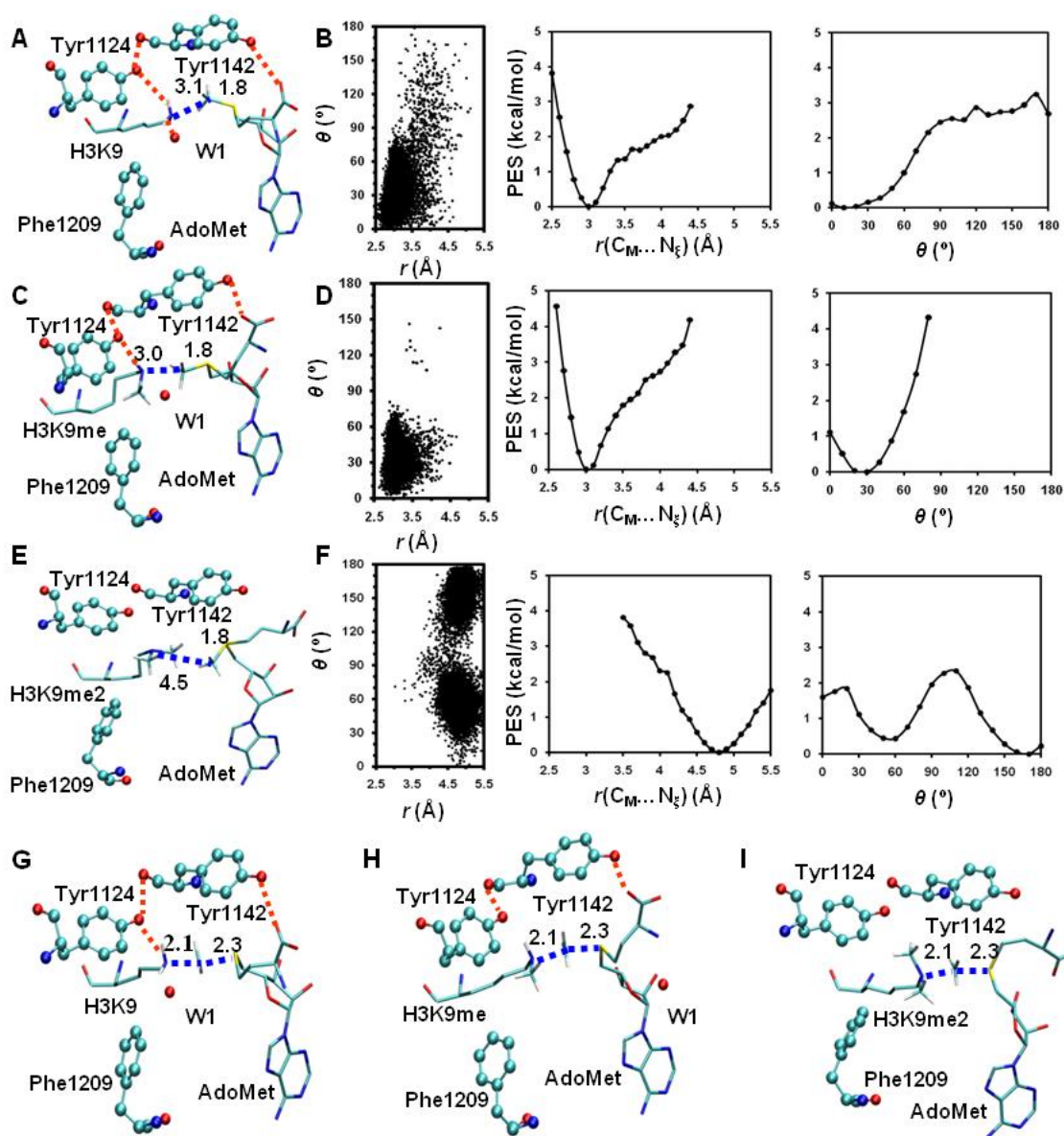
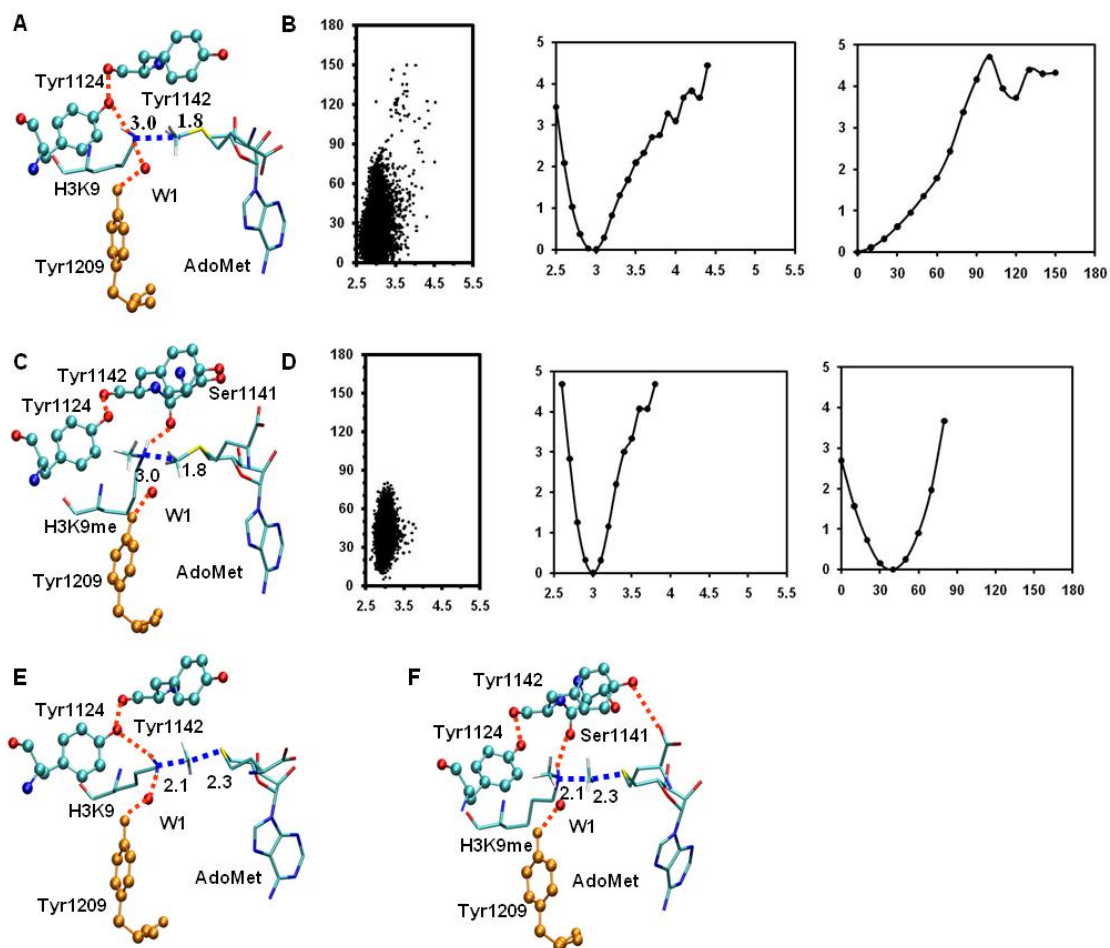


Figure 3.3 MD results for the F1209Y mutant

(A) A representative active-site structure along with the average values of some structural parameters of the reactant complex for the first methyl transfer. (B) **Left:** the two-dimensional plot of $r(C_M...N_\zeta)$ and θ distributions of the reactant complex for the first methyl transfer; **Middle:** the free-energy change as a function of $r(C_M...N_\xi)$ obtained from the distributions; **Right:** the free-energy change as a function of θ obtained from the distributions. (C) The structure of the reactant complex for the second methyl transfer. (D) **Left:** the two-dimensional plot of $r(C_M...N_\zeta)$ and θ distributions of the reactant complex for the second methyl transfer; **Middle:** the free-energy change as a function of $r(C_M...N_\xi)$ obtained from the distributions; **Right:** the free-energy change as a function of θ obtained from the distributions. (E) The structure near the transition state for the first methyl transfer. (F) The structure near the transition state for the second methyl transfer.



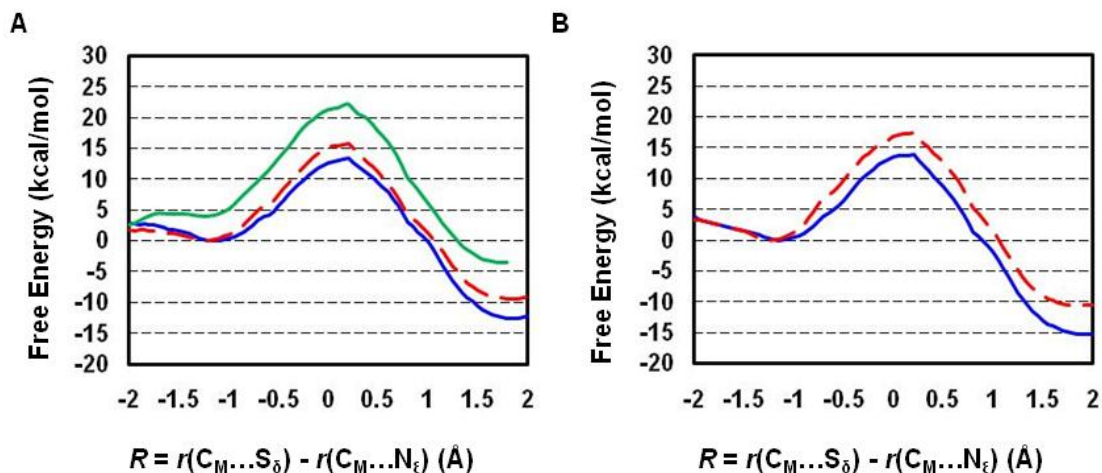
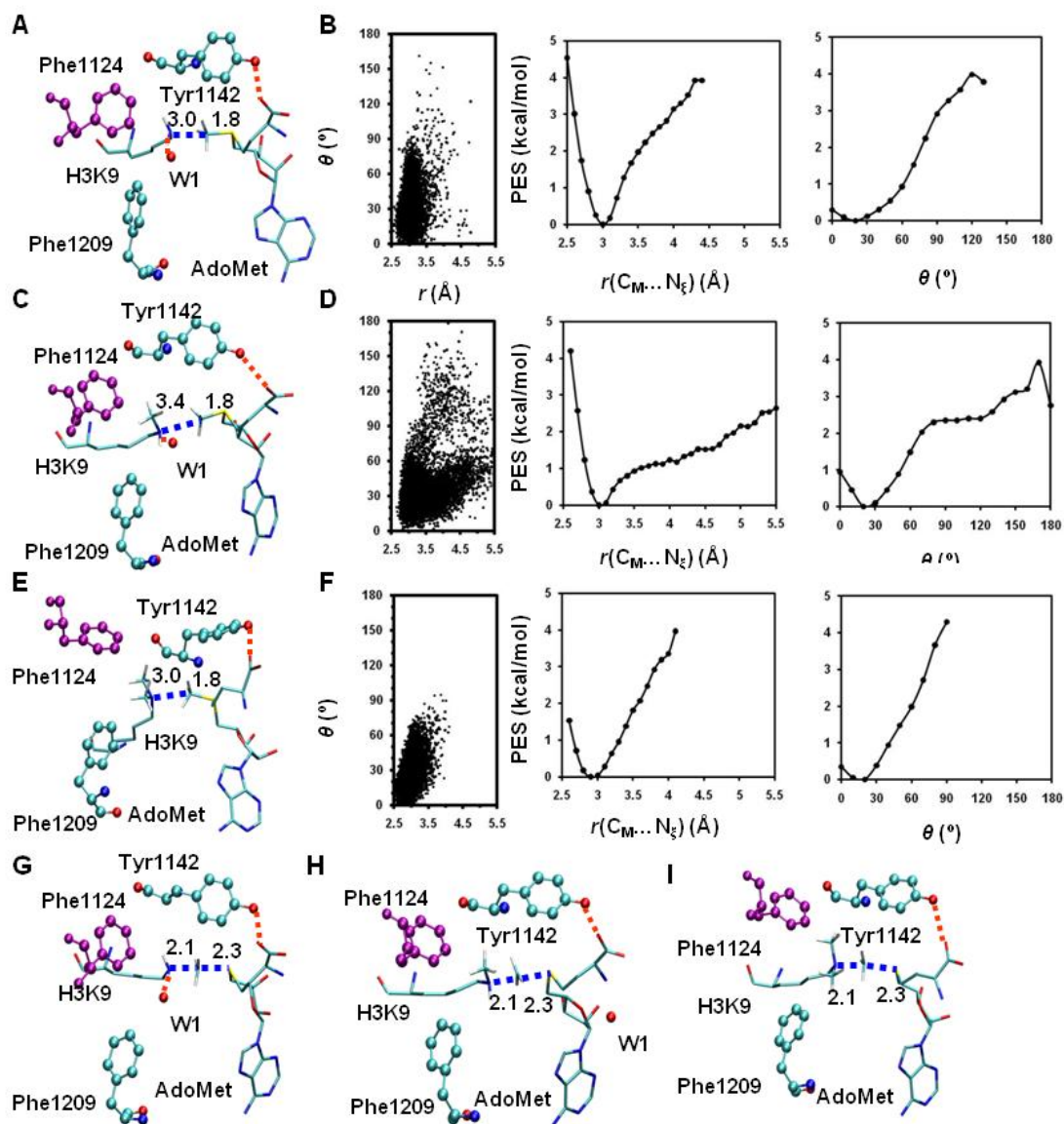


Figure 3.4 Free energy profiles of methyl transfer processes in WT GLP and the F1209Y mutant

(A) The free energy (potential of mean force) changes for the first, second and third methyl transfers from AdoMet to H3-K9, H3-K9me1 and H3-K9me2, respectively, as a function of the reaction coordinate [$R = r(C_M...S_\delta) - r(C_M...N_\xi)$] in the wild-type GLP. The first methyl transfer: blue and solid line with a free energy barrier of 13.4 kcal/mol; the second methyl transfer: red and dashed line with a free energy barrier of 15.8 kcal/mol (or about 2.4 kcal/mol higher than that of the first methyl transfer); the third methyl transfer: green and solid line with a free energy barrier of 22.1 kcal/mol (or about 8.7 kcal/mol higher than that of the first methyl transfer). **(B)** The free energy changes for the first and second methyl transfers as a function of the reaction coordinate in the F1209Y mutant. The first methyl transfer: blue and solid line with a free energy barrier of 13.8 kcal/mol; the second methyl transfer: red and dashed line with a free energy barrier of 17.4 kcal/mol (or about 3.6 kcal/mol higher than that of the first methyl transfer).

Figure 3.5 MD results for the Y1124F mutant

(A) A representative active-site structure along with the average values of some structural parameters of the reactant complex for the first methyl transfer. (B) **Left:** the two-dimensional plot of $r(C_M...N_\zeta)$ and θ distributions of the reactant complex for the first methyl transfer; **Middle:** the free-energy change as a function of $r(C_M...N_\xi)$ obtained from the distributions; **Right:** the free-energy change as a function of θ obtained from the distributions. (C) The structure of the reactant complex for the second methyl transfer. (D) **Left:** the two-dimensional plot of $r(C_M...N_\zeta)$ and θ distributions of the reactant complex for the second methyl transfer; **Middle:** the free-energy change as a function of $r(C_M...N_\xi)$ obtained from the distributions; **Right:** the free-energy change as a function of θ obtained from the distributions. (E) The structure of the reactant complex for the third methyl transfer. (F) **Left:** the two-dimensional plot of $r(C_M...N_\zeta)$ and θ distributions of the reactant complex for the third methyl transfer; **Middle:** the free-energy change as a function of $r(C_M...N_\xi)$ obtained from the distributions; **Right:** the free-energy change as a function of θ obtained from the distributions. (G) A representative active-site structure along with the average values of some structural parameters near the transition state for the first methyl transfer. (H) The structure near the transition state for the second methyl transfer. (I) The structure near the transition state for the third methyl transfer.



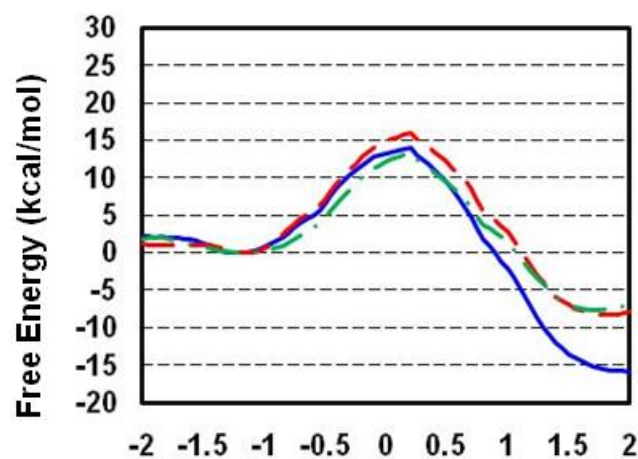


Figure 3.6 Free energy profiles of methyl transfer processes in the Y1124F mutant

The first methyl transfer: blue and solid line with a free energy barrier of 13.9 kcal/mol; the second methyl transfer: red and dashed line with a free energy barrier of 15.9 kcal/mol; the third methyl transfer: green and dashed line with a free energy barrier of 13.3 kcal/mol.

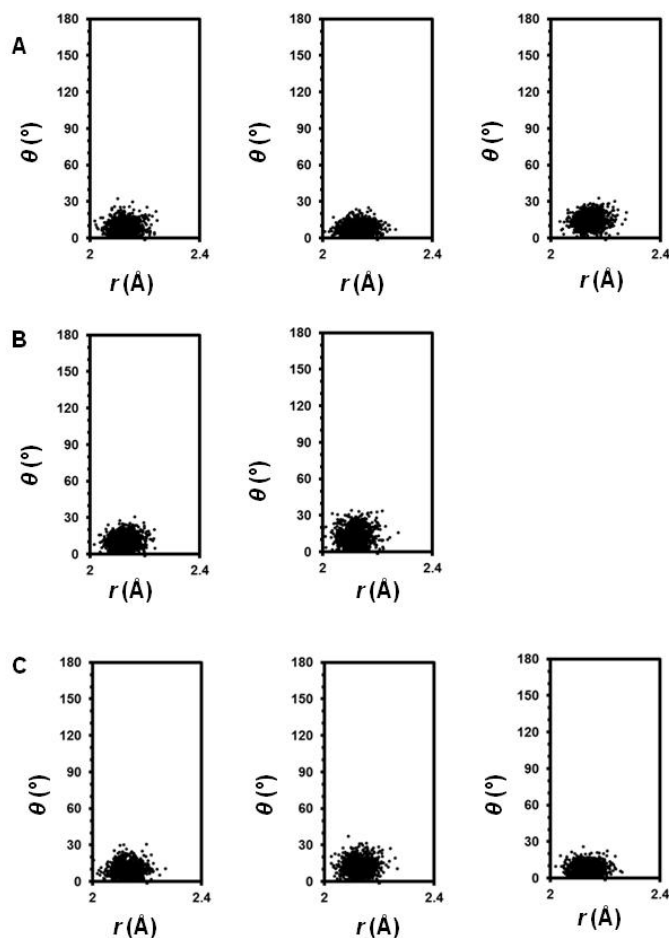


Figure S3.1 The two-dimensional plots of $r(C_M...N_Z)$ and θ distributions of the transition state complexes

(A) Plots for WT GLP, **left:** first methyl transfer; **middle:** second methyl transfer; **right:** third methyl transfer. **(B)** Plots for F1209Y mutant, **left:** first methyl transfer; **right:** second methyl transfer. **(C)** Plots for Y1124F mutant, **left:** first methyl transfer; **middle:** second methyl transfer; **right:** third methyl transfer. All distribution plots are based on the 50-ps production runs of the transition state complexes.

CHAPTER 4

QM/MM MD AND FREE ENERGY SIMULATIONS OF THE METHYLATION REACTIONS CATALYZED BY PROTEIN ARGININE METHYLTRANSFERASE PRMT3

A version of this chapter was accepted for publication by Yuzhuo Chu, Guohui Li and Hong Guo:

Yuzhuo Chu, Guohui Li and Hong Guo “QM/MM MD and Free Energy Simulations of the Methylation Reactions Catalyzed by Protein Arginine Methyltransferase PRMT3.” *Canadian Journal of Chemistry*

Yuzhuo Chu and Dr. Hong Guo designed the experiments; Yuzhuo Chu performed the experiments; the results were analyzed by Yuzhuo Chu, Dr. Guohui Li and Dr. Hong Guo. The manuscript was prepared by Yuzhuo Chu and Dr. Hong Guo.

Abstract

Protein arginine N-methyltransferases (PRMTs) catalyze the transfer of methyl group(s) from S-adenosyl-L-methionine (AdoMet) to the guanidine group of arginine residue in abundant eukaryotic proteins. Two major types of PRMTs have been identified in mammalian cells. Type I PRMTs catalyze the formation of asymmetric ω - N^G , N^G -dimethylarginine (ADMA), while Type II PRMTs catalyze the formation of symmetric ω - N^G , N^G -dimethylarginine (SDMA). The two different methylation products (ADMA or SDMA) of the substrate could lead to different biological consequences. Although PRMTs have been the subject of extensive experimental investigations, the origin of the product specificity remains unclear. In this study, quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) and free energy simulations are performed to study the reaction mechanism for one of Type I PRMTs, PRMT3, and to gain insights into the

energetic origin of its product specificity (ADMA). Our simulations have identified some important interactions and proton transfers involving the active site residues. These interactions and proton transfers seem to be responsible, at least in part, in making the $N_{\eta 2}$ atom of the substrate arginine the target of the both 1st and 2nd methylations, leading to the asymmetric dimethylation product. The simulations also suggest that the methyl transfer and proton transfer appear to be somehow concerted processes and that Glu326 is likely to function as the general base during the catalysis.

Keywords: protein arginine methyltransferases, product specificity, QM/MM MD and free energy simulations, reaction mechanism.

Introduction

One specific post-translational modification, protein arginine methylation, occurs in abundant proteins in eukaryotes. This modification has been suggested to be involved in modulating different cell signaling pathways, including gene transcription, RNA splicing, signal transduction, cell growth and proliferation (Bedford and Clarke 2009). The family of protein arginine N-methyltransferases (PRMTs) catalyze the transfer of methyl groups from S-adenosyl-L-methionine (AdoMet) to the guanidine group of arginine residues in protein substrates (Lee, Kim et al. 1977). Two major types of PRMTs have been identified in mammalian cells. The both types of enzymes could catalyze the formation of ω - N^G -monomethylarginine (MMA), but they differ in the dimethylation products: i.e., Type I PRMTs catalyze the formation of asymmetric ω - N^G , N^G -dimethylarginine

(ADMA), while Type II PRMTs catalyze the formation of symmetric ω - N^G , N^G -dimethylarginine (SDMA) (**Figure 4.1(A)**) (Di Lorenzo, Bedford et al. 20110).

PRMTs belong to a highly conserved family of proteins in eukaryotes with a conserved core region. They contain characteristic motifs of seven-beta strand (Katz, Dlakic et al. 2003) as well as invariant “double E” and “THW” sequence motifs (Zhang and Cheng 2003; Zhang, Zhou et al. 2000). Two invariant glutamates on the “double E” loop have been suggested to be essential for PRMTs activities, for example, in the cases of PRMT1 and PRMT5 (Zhang and Cheng 2003; Sun, Wang et al. 2011). It was believed that the interactions of these two Glu residues with the substrate might play a role in determining the charge distribution of the guanidine group and making $N_{\eta 2}$ a more effective nucleophile (Zhang and Cheng 2003; Zhang, Zhou et al. 2000). However, the question remains as to whether they may play some other role as well during the methyl transfers. Although PRMTs have been the subject of extensive experimental investigations and are of fundamental importance in many biological processes, our understanding of their catalytic mechanism is still lacking. For instance, although different arginine methylations (e.g., ADMA or SDMA) have been implicated for different biological function (Wysocka, Allis et al. 2006), the factors that control such product specificity have not been clearly identified. Thus, it is of considerable interest to understand the catalytic mechanism and origin of the product specificity of these enzymes using computational approaches.

The crystal structures of several PKMTs have been solved, including those for rat PRMT1 (Zhang and Cheng 2003) and rat PRMT3 (Zhang, Zhou et al. 2000) of Type I PRMTs and *C. elegans* PRMT5 (Sun, Wang et al. 2011) of Type II PRMTs. The existence of these structures makes it possible to apply the methods of computer simulations to investigate the catalytic mechanisms of PRMTs and to address the questions concerning the product specificity of different types of PRMTs. In this study, we investigate the methylation reaction catalyzed by one of Type I PRMTs, PRMT3, using quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) and free energy simulations. The results of the simulations may be applied to other Type I PRMTs as well. Our simulations have identified some important active-site interactions that seem to make the N_{η2} atom the site of both 1st and 2nd methylations, leading to the formation of the asymmetric di-methylation product. It has also been demonstrated that the methyl transfer and proton transfer appear to be somehow concerted processes with Glu 326 on “double E” loop functioning as the general base in the enzyme-catalyzed reaction.

Method

The QM/MM MD and free energy (potential of mean force) simulations were applied for studying the methyl transfer reactions. The CHARMM program (Brooks, Bruccoleri et al. 1983) was used for the simulations. The self-consistent charge density functional tight-binding method (SCC-DFTB) (Cui, Elstner et al. 2001) and CHARMM force field (MacKerell, Bashford et al. 1998) were used for

the QM and MM atoms, respectively. Although high-level *ab initio* calculations (e.g., B3LYP/6-31G** or MP2/6-31G**) have been widely used for addressing the questions for chemical and biological systems, including the systems with cooperative hydrogen bonding (Guo and Salahub 1998), proton sponges (Guo and Salahub 2001), polyacetylene (Guo and Paldus 1997) and peptides with carbon hydrogen bonds (Guo, Beahm et al. 2004), these approaches are still too time consuming for generating two-dimensional free energy maps of enzyme-catalyzed reactions, as we did in this study. The SCC-DFTB approach has been used previously on a number of systems by us (Guo, Wlodawer et al. 2005; Guo, Rao et al. 2005; Xu, Guo et al. 2007; Guo and Guo 2007; Xu, Chu et al. 2009; Chu, Xu et al. 2010), and the results seem to be reasonable. We have compared the energetic results obtained from the SCC-DFTB and B3LYP/6-31G** calculations for the methyl transfer in a simple model system that resembles the methyl transfer in PRMTs (see Supplementary Materials). It was found that the potential energy function obtained from SCC-DFTB is quite similar to that obtained from the B3LYP/6-31G** calculations. Therefore, no empirical correction seems to be necessary for this study. This is in contrast with our earlier studies of PKMTs (Guo and Guo 2007; Xu, Chu et al. 2009; Chu, Xu et al. 2010) where a systematic error in the SCC-DFTB potential energy function for the methyl transfer in a model system was observed (when compared to the B3LYP/6-31G** calculations) and an empirical correction was therefore introduced.

The $-\text{CH}_2-\text{CH}_2-\text{S}^+(\text{Me})-\text{CH}_2-$ part of AdoMet, the side chains of the substrate arginine /methyl arginine and the two conserved glutamate residues (Glu326 and Glu335) on the “double E” loop were treated by QM and the rest of the system were treated by MM. The link-atom approach (Field, Bash et al. 1990) implemented in the CHARMM program was used to separate the QM and MM regions. The all-hydrogen potential function (PARAM27) (MacKerell, Bashford et al. 1998) was used for the MM atoms. A modified TIP3 water model (Jorgensen, Chandrasekhar et al. 1983) was used for the solvent. The stochastic boundary molecular dynamics method (Brooks, Brunger et al. 1985) was used for the QM/MM MD and free energy simulations. The system was separated into a reaction zone and a reservoir zone, and the reaction zone was further divided into a reaction region and a buffer region. The reaction region was a sphere with radius r of 22 Å, and the buffer region had r equal to $20 \text{ Å} \leq r \leq 22 \text{ Å}$. The reference center for partitioning the system was chosen to be N_σ of the substrate arginine. The resulting systems contained around 5500 atoms, including about 800-900 water molecules.

The initial coordinates for the enzyme were obtained from the X-ray structure of PRMT3 complexed with AdoHcy (PDB ID: 1F3L, resolution 2.03 Å) (Zhang, Zhou et al. 2000). The substrate peptidyl arginine was put into the active site of PRMT3 by the superposition of 1F3L with the X-ray structure of PRMT1 complexed with AdoHcy and substrate peptidyl-Arg (PDB ID: 1ORI, resolution 2.5 Å) (Zhang and Cheng 2003). The methyl group was added manually to

AdoHcy to generate AdoMet. The reactant complex for the 2nd methyl transfer was built by adding one methyl group to the N_{η2} atom. The initial structures for the entire stochastic boundary system 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 and equilibrated for 50 ps. A 1-fs time step was used for integration of the equation of motion, and the coordinates were saved every 50 fs for analyses. After 50-ps of equilibration were performed, 1.5-ns QM/MM MD simulations were carried out for each of the reactant complexes. As discussed in previous studies of protein lysine methyltransferases (PKMTs) (Guo and Guo 2007; Xu, Chu et al. 2009; Chu, Xu et al. 2010), the S_N2 methyl transfer from AdoMet to the target is presumably more efficient if the S-CH₃ group of AdoMet is well aligned with the target; i.e., with a small θ angle and relatively short C_M...N_η distance in PRMTs. Therefore, we determined the distributions of $r(\text{C}_\text{M} \dots \text{N}_\eta)$ and θ from the MD trajectories; θ was defined as the angle of C_M-S_δ-N_η (**Figure 4.1(B)**).

The umbrella sampling method (Torrie and Valleau 1974) implemented in the CHARMM program along with the Weighted Histogram Analysis Method (WHAM) (Kumar, Bouzida et al. 1992) was applied to determine the changes of the free energy (potential of mean force) for the methyl and proton transfers. The 1D free energy curves were obtained for the process of methyl transfers from AdoMet to the substrate arginine and methyl arginine. The linear combination of $r(\text{C}_\text{M} \dots \text{N}_\eta)$ and $r(\text{C}_\text{M} \dots \text{S}_\delta)$ [$R_x = r(\text{C}_\text{M} \dots \text{S}_\delta) - r(\text{C}_\text{M} \dots \text{N}_\eta)$] was used as the reaction

coordinate for the 1D free energy simulations (**Figure 4.1(B)**). For each window of the methyl transfer process, 50 ps production runs were performed after 50 ps equilibration. The force constants of the harmonic biasing potentials used in the 1D free energy simulations were in the range of 100 to 1400 kcal mol⁻¹ Å⁻². In addition to the 1D free energy curves, the 2D free energy map was also obtained for the di-methylation. The second reaction coordinate in the 2D free energy simulations is the proton transfer coordinate between the HH₂₂ atom on N_{η2} and the OE₁ atom of Glu326 (atoms labeled as shown in **Figure 4.2**) [the first coordinate is still $R_x = r(\text{C}_M\text{-S}_\delta) - r(\text{C}_M\text{-N}_\eta)$]. The linear combination of $r(\text{HH}_{22}\text{-N}_{\eta2})$ and $r(\text{HH}_{22}\text{-OE}_1)$ [$R_y = r(\text{HH}_{22}\text{-N}_{\eta2}) - r(\text{HH}_{22}\text{-OE}_1)$] was used as the reaction coordinate for the proton transfer process. For the 2D free energy map, 50 ps production runs were performed after 100 ps equilibration for each window. The force constants were in the range of 100 to 1000 kcal mol⁻¹ Å⁻² for R_x and was 200 kcal mol⁻¹ Å⁻² for R_y .

Results

1st methylation on N_{η2} vs. N_{η1}

In this study, the methyl additions on the N_{η2} vs. N_{η1} atoms were evaluated by the QM/MM MD and free energy simulations. The active-site structures of the reactant complex, transition state and product complex for the 1st methyl addition on N_{η2} and N_{η1} are shown in **Figure 4.2 and 4.3**, respectively. The active-site structure of the reactant complex shows that the N_{η2} atom is well aligned with the

methyl group of the AdoMet, and this good alignment could make the methyl transfer easier. The arginine residue on the substrate seems to be stabilized in the active site through the interactions with several active-site residues, including the hydrogen bond interactions with E326 and E335. E326 also interacts with AdoMet and helps to direct the methyl group of AdoMet towards the $N_{\eta 2}$ atom. These active-site interactions may lead to the good alignment of the $N_{\eta 2}$ atom and the methyl group of AdoMet. The result is also reflected by the two dimensional distribution plots of $r(C_M-N_\eta)$ and θ (**Figure 4.4, Left**), which show that there are a large number of structures with relatively short $r(C_M-N_\eta)$ distances and small θ angles.

The free energy barriers for the 1st methyl additions to the $N_{\eta 2}$ and $N_{\eta 1}$ atoms, respectively, are compared in **Figure 4.5**. Figure 5 shows that the free energy barrier of the methyl addition to the $N_{\eta 1}$ atom is about 8 kcal/mol higher than that to the $N_{\eta 2}$ atom, suggesting that the 1st methyl group is likely to be added to the $N_{\eta 2}$ atom. Therefore, the product complex in **Figure 4.2(C)** is the most likely product from the 1st methylation reaction. Further MD simulations on the product complex after the deprotonation of E326 indicated that E326 could return to the active position after deprotonation and play a role as the general base for the 2nd methylation step as well (see **Supplementary Materials, Figure S4.3**). A channel that opens to the solvent near E326 was observed (see **Supplementary Materials, Figure S4.4**); the solvent water could possibly enter through this channel to reach E326 for the deprotonation of this residue.

Additional simulations are required in the future to address the question concerning the deprotonation process.

2nd methylation on $N_{\eta 2}$ vs. $N_{\eta 1}$

Next, we examine the 2nd methylation reaction and investigate the mechanism of the formation of the asymmetric product from the PRMT3-catalyzed reaction. The methyl addition on the $N_{\eta 2}$ vs. $N_{\eta 1}$ atoms for the 2nd methylation reaction was evaluated. The active-site structures of the reactant complexes (**Figure 4.6(A)**) show that the $N_{\eta 2}$ atom is well aligned with the methyl group of AdoMet, presumably, with the help of the similar active-site interactions observed in the 1st methylation step. Relative poor alignment was observed with $N_{\eta 1}$ (**Figure 4.7(A)**). The two dimensional distribution plots of $r(C_M-N_\eta)$ and θ (**Figure 4.8**) support the suggestion that the methyl group of AdoMet is better aligned with the $N_{\eta 2}$ atom than with $N_{\eta 1}$. The free energy barriers of the methyl additions to $N_{\eta 2}$ and $N_{\eta 1}$ atoms are compared in **Figure 4.9**. As is evident from Figure 9, the free energy barrier for the methyl addition to the $N_{\eta 1}$ atom is about 7 kcal/mol higher than that to the $N_{\eta 2}$ atom. Thus the 2nd methyl group is also likely to be added to $N_{\eta 2}$, leading to the asymmetric product (**Figure 4.6(C)**).

1st and 2nd methylations on the $N_{\eta 2}$ atom

The free-energy barriers for the 1st and 2nd methylations on the $N_{\eta 2}$ atom are compared in **Figure 4.10**. Figure 4.10 shows that the two methyl transfers have similar free energy barriers; the difference is only 2 kcal/mol. Our simulation

results indicate that the 2nd methylation step is more efficient compared with the 1st step, and this is consistent with experimental results that showed that the major product of the reaction was ADMA. Previous studies also suggested that ω -MMA formation was not obligatory (Obianyo, Osborne et al. 2008). ω -MMA might be further methylated to form ADMA before released from the active site. These results along with the previous suggestion that both methyl groups were added to the N_{n2} atom explain as to why the asymmetric product is formed from the PRMT3-catalyzed reaction.

The roles of the “double E” residues

The possible involvement of a general base during the catalysis has been discussed previously for some PRMT enzymes (Zhang, Zhou et al. 2000; Rust, Zurita-Lopez et al. 2011). A His-Asp relay system was originally thought to play this role (Zhang, Zhou et al. 2000). However, the His residue was found to be located at > 6 Å from the substrate guanidine group, making it less likely to act as the general base. Moreover, the mutagenesis study did not support the role of the His residue as the general base (Rust, Zurita-Lopez et al. 2011). The examination of the active site indicated that only the two conserved glutamates on the “double E” loop seem to be within a distance (< 3 Å) that allows them to remove the proton from the guanidine directly. To investigate the roles of these two residues, the side chains of these two residues were included in our QM system to allow the proton transfer from guanidine to the carboxyl group of these residues. The results of our simulations are consistent with an earlier suggestion

that the conserved E326 residue might function as the general base.²⁸ At the transition state, the N-H bond of the attacked nitrogen is broken, and the proton is transferred to the carboxyl group of E326 (**Figure 4.2, 4.6**). The methyl and proton-transfer processes seem to follow a somewhat concerted mechanism to reach the TS state, as indicated from **Figure 4.11**. As can be seen from our simulation results, E326 also functions as the general base even in the simulations of methyl transfers to N_{η1} atoms (**Figure 4.3, 4.7**). In these cases, E326 was further away from the target nitrogen in the reactant complexes, and it moved towards the N_{η1} atom during the reaction to function as the general base. E335 is much closer to the target nitrogen in these cases. However, its negative charge is likely to be reduced significantly through the interactions with the positive charges on the N_E and N_{η1} atoms (Zhang and Cheng 2003; Zhang, Zhou et al. 2000). This is probably the reason that it does not function as the general base. To further confirm that proton is transferred during the methyl transfer reaction, the methylation reactions were also simulated with the N-H bond involving the N_{η2} atom fixed using the Shake algorithm to prevent the proton transfer (see **Figures S4.5, S4.6**). The free-energy barriers of the reactions became much higher (6-8 kcal/mol higher) (see **Figure S4.7**). These results support the suggestion that the proton transfer is required for the methyl transfer and E326 acts as the general base.

2D free energy simulations

2D free energy simulations for the 2nd methyl transfer to the N_{η2} atom were performed to explore the relationship between the methyl and proton-transfer processes. As shown in **Figure 4.11**, the results of 2D free energy simulations seem to suggest that the methyl and proton transfers are somehow concerted. The proton on the N_{η2} atom is transferred to the carboxyl group of E326 near the transition state.

Discussion

One important question for PRMTs concerns the factors that determine their product specificity (ADMA or SDMA). Several experimental studies (Obianyo, Osborne et al. 2008; Rust, Zurita-Lopez et al. 2011; Gui, Wooderchak et al. 2011; Branscombe, Frankel et al. 2001) have been performed to address this question. For Type I enzyme, it was proposed that the active site methionine residues might prevent the binding of ω -MMA in a conformation which allows the formation of symmetric product. For Type II enzymes, this position is occupied by the residues with smaller side chains, such as Ser as observed in PRMT5 (Branscombe, Frankel et al. 2001). Although the kinetic studies of the mutants of these methionines in PRMT1 (one Type I PRMT) led to a change of distribution of mono- and di- methylated products, the formation of SDMA was not observed. This seems to suggest that these methionines may not be responsible for the failure of the formation of SDMA in the Type I enzymes (Obianyo, Osborne et al. 2008; Rust, Zurita-Lopez et al. 2011).

In this paper, QM/MM MD and free energy simulations were performed to investigate the energetic origin of the product specificity of PRMT3, a Type I PRMT that generate ADMA instead of SDMA. It has been discussed in our previous studies of PKMTs (Guo and Guo 2007; Xu, Chu et al. 2009; Chu, Xu et al. 2010) that the product specificity of PKMTs could be reflected from MD and free energy simulations. It would be of considerable interest to see if it would be true for PRMTs as well. As demonstrated in this work, the product specificity could indeed be reflected from the MD and free energy simulations.

As can be seen in the structures of the reactant complexes of the 1st and 2nd methylations (**Figure 4.2(A), 4.6(A)**), the N_{η2} atom can be well aligned with AdoMet. Thus, the structures of the reactant complexes of the methyl additions to the N_{η2} atom are rather similar to the corresponding transition state (TS) structures (**Figure 4.2(B), 4.6(B)**). As discussed in our earlier papers (Guo, Wlodawer et al. 2005; Guo, Rao et al. 2005; Xu, Guo et al. 2007; Guo and Guo 2007; Xu, Chu et al. 2009; Chu, Xu et al. 2010), the corresponding methylation reactions are expected to be more efficient, as a part of the TS stabilization would be already reflected in the reactant complexes. In contrast, for the methylation reactions involving N_{η1} the reactant complexes (**Figure 4.3(A), 4.7(A)**) are significantly distorted from the corresponding TS structures (**Figure 4.3(B), 4.7(B)**). Additional energies could be required to generate the TS-like structures, leading to higher free energy barriers and less efficient methyl transfer reactions.

The free energy barriers of the methyl transfer reactions involving the $N_{\eta 2}$ and $N_{\eta 1}$ atoms also indicated that the methyl additions to $N_{\eta 2}$ would be more efficient than to $N_{\eta 1}$. Thus for the both 1st and the 2nd methylations the methyl groups would be added to the $N_{\eta 2}$ atom.

The roles of the two invariant glutamates on “double E” were also studied. Consistent with previous suggestions (Zhang and Cheng 2003; Zhang, Zhou et al. 2000), our simulation results seem to suggest that the both “double E” residues may contribute to the alignment of the $N_{\eta 2}$ atom with AdoMet. We also proposed that E326 may function as the general base for the reactions. The results of 1D and 2D free energy simulations are therefore consistent with the mechanisms of the methyl transfer processes that are given in **Figure 4.12**.

Conclusion

In this study, QM/MM MD and free energy simulations were performed to study the reaction mechanism of PRMT3. Our results suggest that the both 1st and 2nd methyl groups would be added to the $N_{\eta 2}$ atom and form the asymmetric di-methylation product. The free-energy barriers of these reactions are consistent with the fact that PRMT3 is a dimethylase. We also proposed that E326 may act as the general base for the methylation reactions.

Acknowledgement

This work is supported in part by the National Science Foundation Award (grant no. 0817940 to H.G.) and in part by National Sciences Foundation of China

(grant no. 31070641 to G.L.) and National 973 Program of China (grant no. 2012CB721000 to G.L.). Y.Z.C is supported by a fellowship from the National Institute for Mathematical and Biological Synthesis (NIMBioS), an Institute sponsored by the National Science Foundation, the U.S. Department of Homeland Security, and the U.S. Department of Agriculture through NSF Award EF-0832858, with additional support from The University of Tennessee, Knoxville.

Supporting Information

Comparison of SCC-DFTB and B3LYP/6-31G**:

A simplified model system was constructed for comparison of the performances of SCC-DFTB and B3LYP/6-31G**. This model contains a trimethyl-sulfonium as the methyl donor and a methylguanidine as the acceptor (**Figure S4.1**). A series of constrained minimizations were performed on the model system in the gas phase to map the potential energy from the reactant to product along the reaction coordinate [$r = r(\text{S}_{\delta} \dots \text{C}_M) - r(\text{N}_{\eta 2} \dots \text{C}_M)$] using SCC-DFTB. The harmonic potential with force constant $k = 1000 \text{ kcal/mol/\AA}^2$ was applied. **Figure S4.2** shows the SCC-DFTB potential energy curve along the methyl transfer path and the single point energies calculated by B3LYP/6-31G**. The B3LYP/6-31G** single-point energy calculations on the structures from the SCC-DFTB energy minimizations agree very well with to the potential energy curve obtained from the SCC-DFTB calculations.

References

- Bedford, M. T. and S. G. Clarke (2009). "Protein Arginine Methylation in Mammals: Who, What, and Why." Mol Cell **33**(1): 1-13.
- Branscombe, T. L., A. Frankel et al. (2001). "PRMT5 (Janus kinase-binding protein 1) catalyzes the formation of symmetric dimethylarginine residues in proteins." J Biol Chem **276**(35): 32971-32976.
- Brooks, B. R., R. E. Bruccoleri, et al. (1983). "CHARMM - a program for macromolecular energy, minimization, and dynamics calculations." J Comput Chem **4**: 187-217.
- Brooks, C. L., A. Brunger, et al. (1985). "Active-site dynamics in protein molecules - a stochastic boundary molecular-dynamics approach." Biopolymers **24**(5): 843-865.
- Chu, Y. Z., Q. Xu, et al. (2010). "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 **6**(4): 1380-1389.
- Cui, Q., M. Elstner, et al. (2001). "A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method." J Phys Chem B **105**(2): 569-585.
- Di Lorenzo, A. and M. T. Bedford (2011). "Histone arginine methylation." Febs Lett **585**(13): 2024-2031.

Field, M. J., P. A. Bash, et al. (1990). "A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations." J Comput Chem **11**(6): 700-733.

Gui, S. Y., W. L. Woolderchak et al. (2011). "Investigation of the Molecular Origins of Protein-arginine Methyltransferase I (PRMT1) Product Specificity Reveals a Role for Two Conserved Methionine Residues." J Biol Chem **286**(33): 29118-29126.

Guo, H. and D. R. Salahub (1998). "Cooperative hydrogen bonding and enzyme catalysis." Angew Chem Int Ed **37**(21): 2985-2990.

Guo, H. and D. R. Salahub (2001). "Origin of the high basicity of 2,7-dimethoxy 1,8-bis-(dimethylamino)naphthalene: Implications for enzyme catalysis." Theochem **547**: 113-118.

Guo, H. and J. Paldus (1997). "Estimates of the structure and dimerization energy of polyacetylene from ab initio calculations on finite polyenes." Int J Quantum Chem **63**(2): 345-360.

Guo, H. B. and H. Guo (2007). "Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity." Proc Natl Acad Sci U S A **104**(21): 8797-8802.

Guo, H. B., A. Wlodawer et al. (2005). "A general acid-base mechanism for the stabilization of a tetrahedral adduct in a serine-carboxyl peptidase: A computational study." J Am Chem Soc **127**(45): 15662-15663.

Guo, H. B., N. Rao et al. (2005). "Origin of tight binding of a near-perfect

transition-state analogue by cytidine deaminase: Implications for enzyme catalysis." J Am Chem Soc **127**(9): 3191-3197.

Guo, H. B., R. F. Beahm et al. (2004). "Stabilization and destabilization of the C-delta-h center dot center dot center dot O=C hydrogen bonds involving proline residues in helices." J Phys Chem B **108**(46): 18065-18072.

Humphrey W. D. A. and K. Schulten (1996). "VMD: visual molecular dynamics." J Mol Graph **14**: 33-38.

Jorgensen, W. L., J. Chandrasekhar, et al. (1983). "Comparision of simple potential functions for simulating liquid water." J Chem Phys **79**(2): 926-935.

Katz, J. E., M. Dlakic et al. (2003). "Automated identification of putative methyltransferases from genomic open reading frames." Mol Cell Proteomics **2**(8): 525-540.

Kobrak, M. N. (2003). "Systematic and statistical error in histogram-based free energy calculations." J Comput Chem **24**(12): 1437-1446.

Kumar, S., D. Bouzida, et al. (1992). "The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method." J Comput Chem **13**(8): 1011-1021.

Lee, H. W., S. Kim et al. (1977). "S-ADENOSYLMETHIONINE – PROTEIN ARGININE METHYLTRANSFERASE - PURIFICATION AND MECHANISM OF ENZYME." Biochemistry **16**(1): 78-85.

MacKerell, A. D. and D. Bashford (1998). "All-atom empirical potential for molecular modeling and dynamics studies of proteins." J Phys Chem B

102(18): 3586-3616.

Obiany, O., T. C. Osborne et al. (2008). "Kinetic mechanism of protein arginine methyltransferase 1." Biochemistry **47**(39): 10420-10427.

Rust, H. L., C. I. Zurita-Lopez et al. (2011). "Mechanistic Studies on Transcriptional Coactivator Protein Arginine Methyltransferase 1." Biochemistry **50**(16): 3332-3345.

Sun, L. T., M. Z. Wang et al. (2011). "Structural insights into protein arginine symmetric dimethylation by PRMT5." Proc Natl Acad Sci U S A **108**(51): 20538-20543.

Torrie, G. M. and J. P. Valleau (1974). "Monte-carlo free-energy estimates using non-boltzmann sampling - application to subcritical lennard -jones fluid." Chem Phys Lett **28**(4): 578-581.

Wysocka, J., C. D. Allis et al. (2006). "Histone arginine methylation and its dynamic regulation." Front Biosci **11**: 344-355.

Xu, Q., H. B. Guo et al. (2007). "The QM/MM molecular dynamics and free energy simulations of the acylation reaction catalyzed by the serine-carboxyl peptidase kumamolisin-As." Biochemistry **46**(12): 3784-3792.

Xu, Q., Y. Z. Chu, et al. (2009). "Energy Triplets for Writing Epigenetic Marks: Insights from QM/MM Free-Energy Simulations of Protein Lysine Methyltransferases." Chem Eur J **15**: 12596 – 12599.

Zhang, X. and X. D. Cheng (2003). "Structure of the predominant protein arginine methyltransferase PRMT1 and analysis of its binding to substrate peptides."

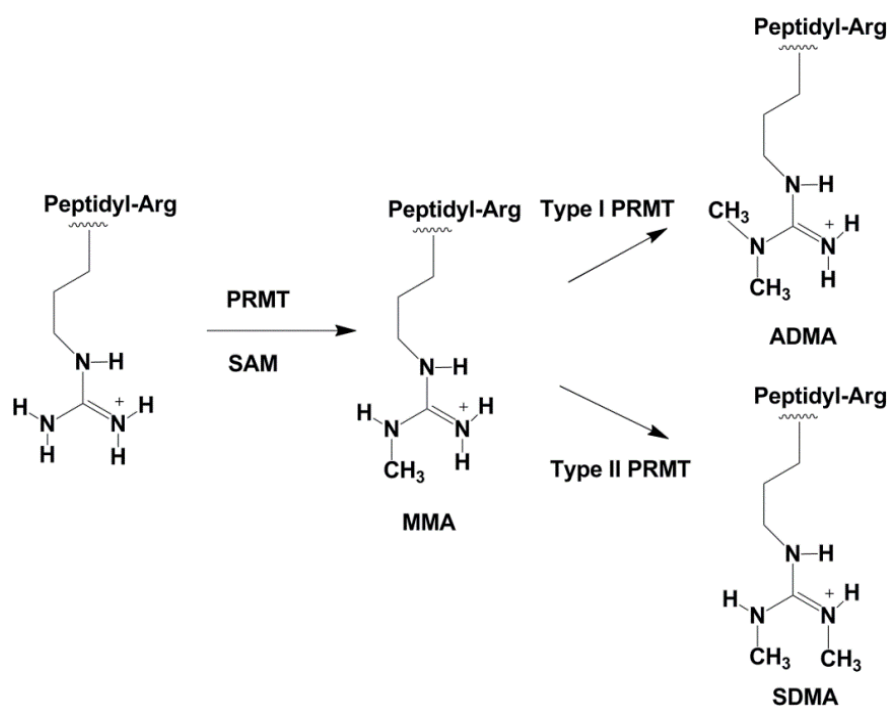
Structure **11**(5): 509-520.

Zhang, X., L. Zhou et al. (2000). "Crystal structure of the conserved core of protein arginine methyltransferase PRMT3." Embo J. **19**(14): 3509-3519.

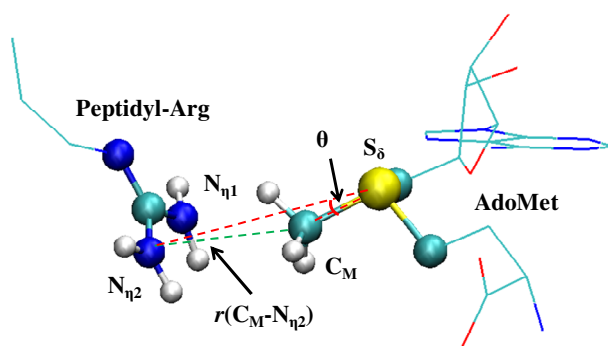
Appendix

Figure 4.1 (A) The reactions catalyzed by Type I and Type II PRMTs. **(B)** Definitions of the structural parameters for monitoring the relative orientation of AdoMet and target peptidyl-Arg in the reactant complexes. The efficiency of the methyl transfer may be related to the distributions of $r(C_M-N_\eta)$ and θ in the reactant complexes. θ is defined as the angle of $C_M-S_\delta-N_\eta$. The reaction coordinate for calculating the free energy profiles for the methyl transfers is $R = r(C_M-S_\delta) - r(C_M-N_\eta)$. (All images were generated by VMD (Humphrey and Schulten 1996)).

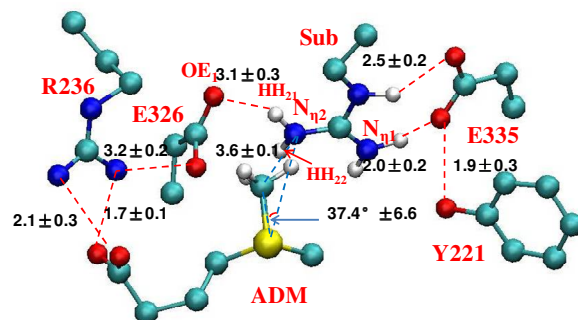
A



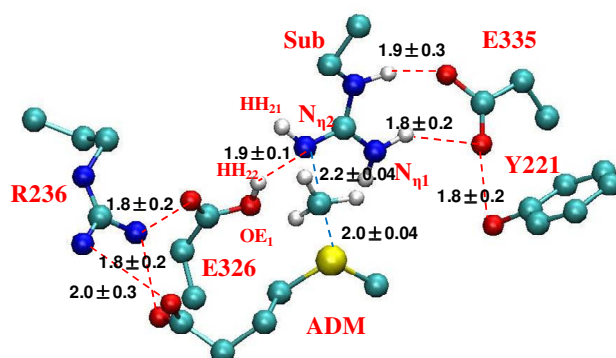
B



A



B



C

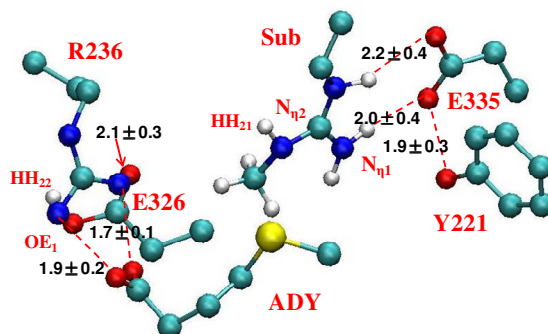
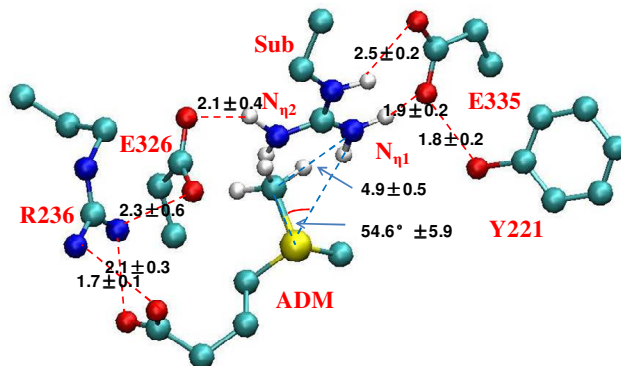


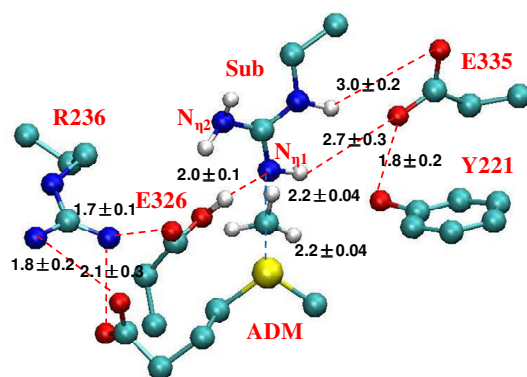
Figure 4.2 Results from the MD simulations for 1st methylation on N_{η2}

(A) The active-site structure of the reactant complex. (B) The active-site structure near the transition state. (C) The active-site structure of the product complex.

A



B



C

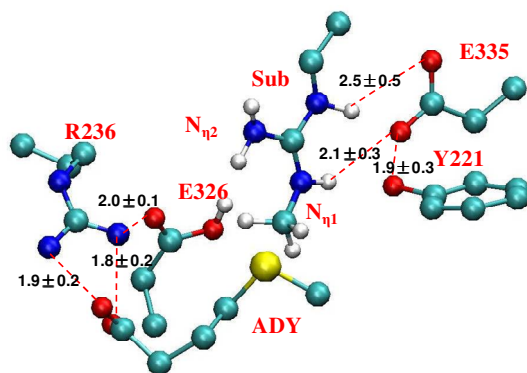


Figure 4.3 Results from the MD simulations for 1st methylation on $N_{\eta 1}$

(A) The active-site structure of the reactant complex. **(B)** The active-site structure near the transition state. **(C)** The active-site structure of the product complex.

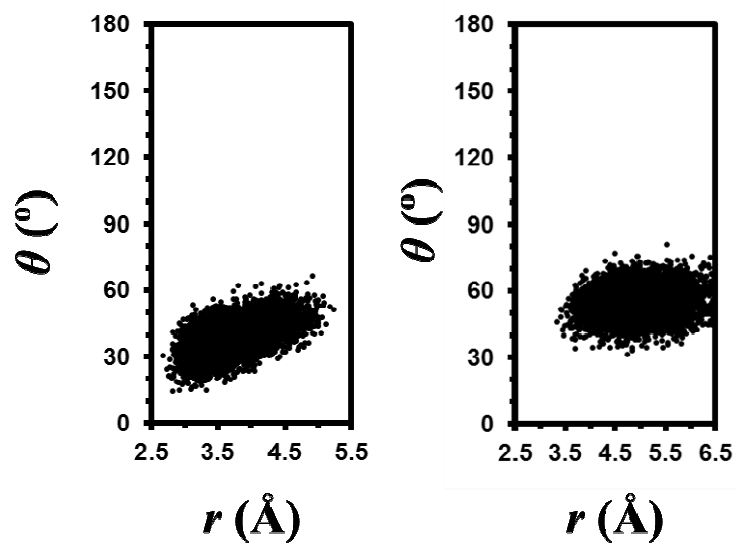


Figure 4.4 Two-dimensional plots of $r(\text{C}_M\text{-N}_{\eta})$ and θ distributions in the reactant complexes

Left: 1st methylation on $\text{N}_{\eta 2}$; **Right:** 1st methylation on $\text{N}_{\eta 1}$.

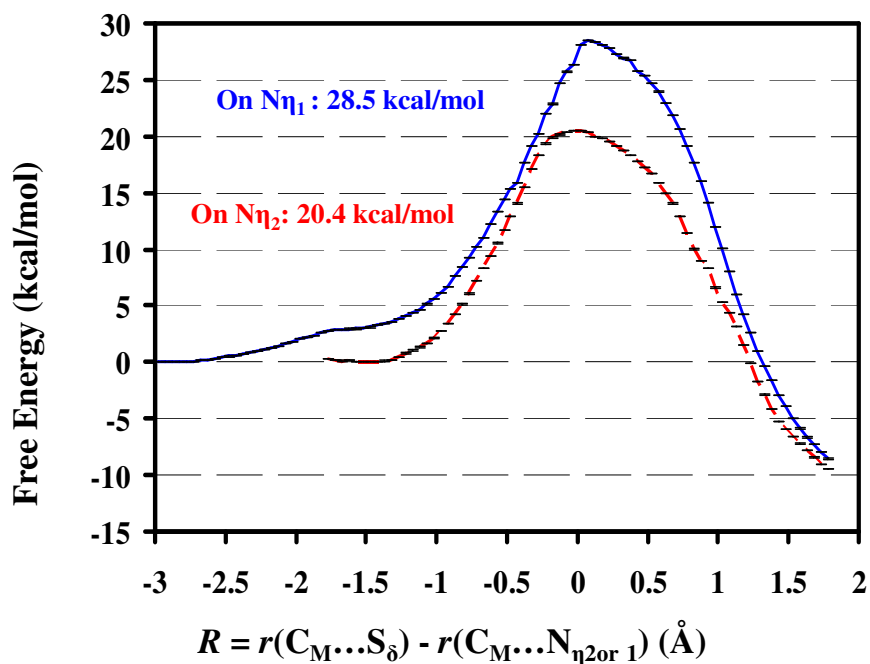
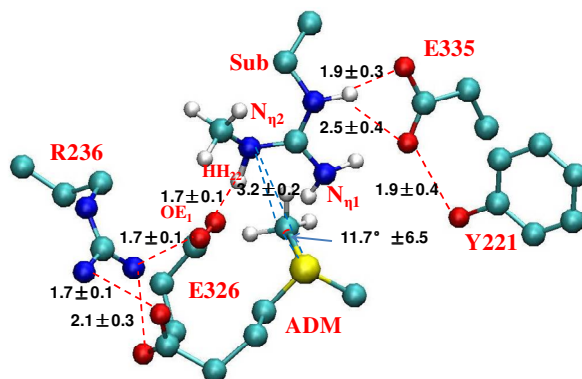


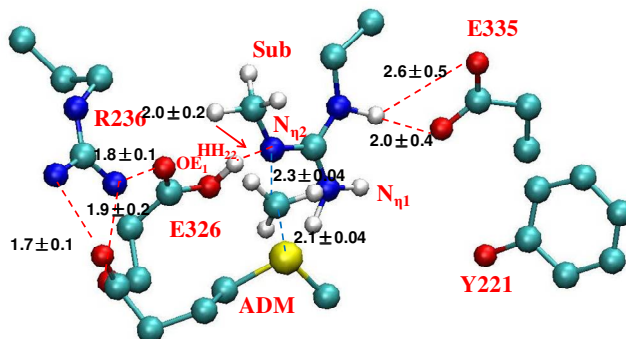
Figure 4.5 Free energy profiles of 1st methyl additions to the $N_{\eta 2}$ and $N_{\eta 1}$ atoms, respectively.

To $N_{\eta 2}$: red line with a free energy barrier of 20.4 kcal/mol; To $N_{\eta 1}$: blue line with a free energy barrier of 28.5 kcal/mol. The statistical errors of the free energies were calculated (Kobrak 2003) and given, and the errors were found to be less than 0.03 kcal/mol.

A



B



C

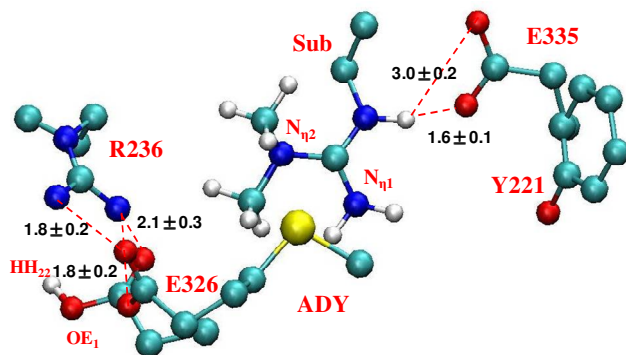
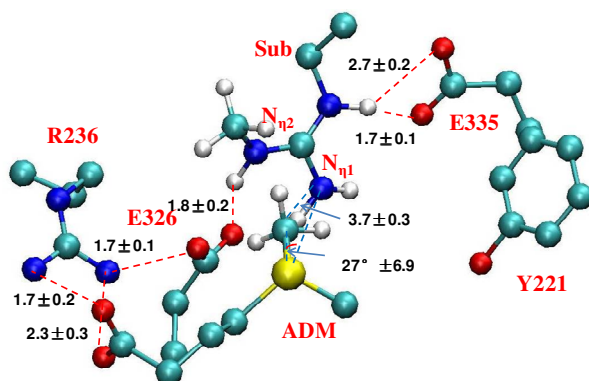


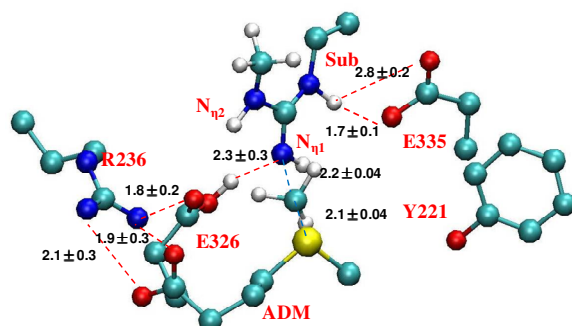
Figure 4.6 Results from the MD simulations for 2nd methylation on N_{η2}

(A) The active-site structure of the reactant complex. **(B)** The active-site structure near the transition state. **(C)** The active-site structure of the product complex.

A



B



C

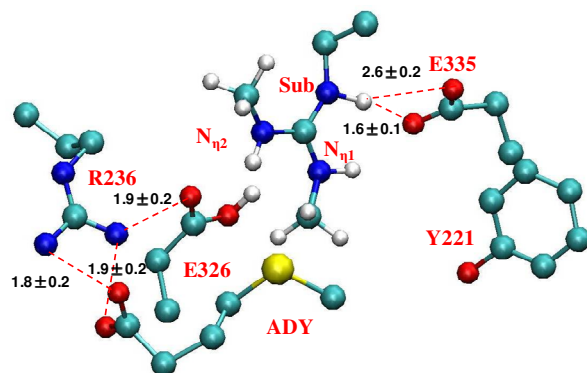


Figure 4.7 Results from the MD simulations for 2nd methylation on N_{η1}

(A) The active-site structure of the reactant complex. (B) The active-site structure near transition state. (C) The active-site structure of the product complex.

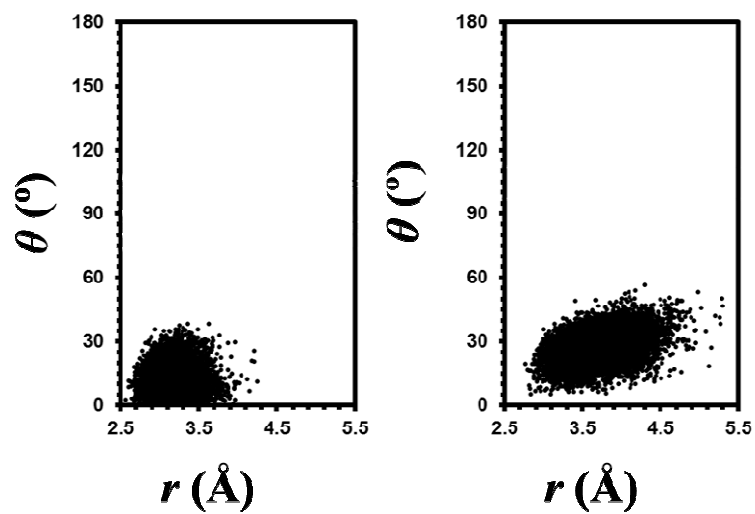


Figure 4.8 Two-dimensional plots of $r(\text{C}_M\text{-N}_\eta)$ and θ distributions of the reactant complexes

(A) 2nd methylation on $\text{N}_{\eta 2}$. **(B)** 2nd methylation on $\text{N}_{\eta 1}$.

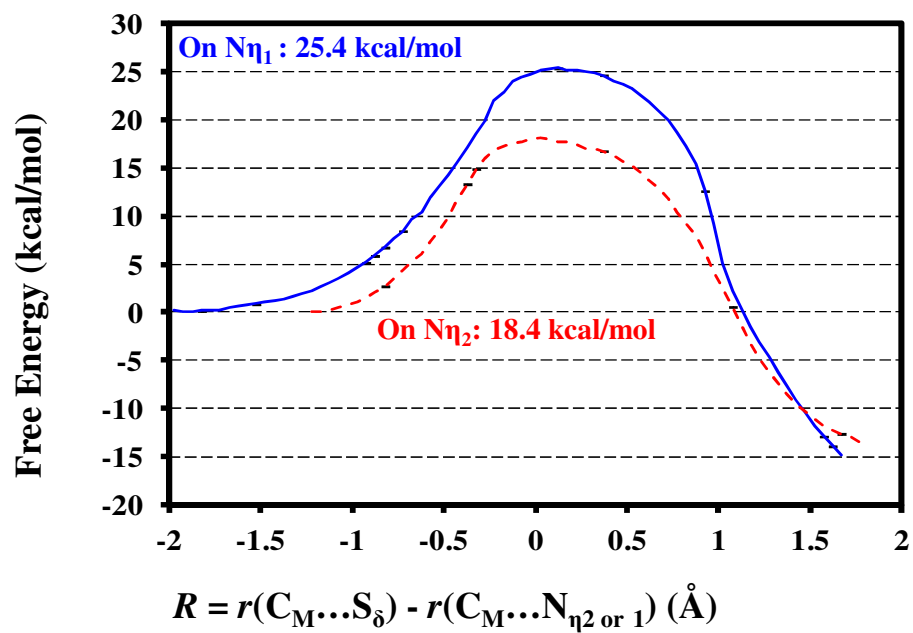


Figure 4.9 Free energy profiles of 2nd methyl additions to the N_{η2} and N_{η1} atoms, respectively

To N_{η2}: red line with a free energy barrier of 18.4 kcal/mol; To N_{η1}: blue line with a free energy barrier of 25.4 kcal/mol.

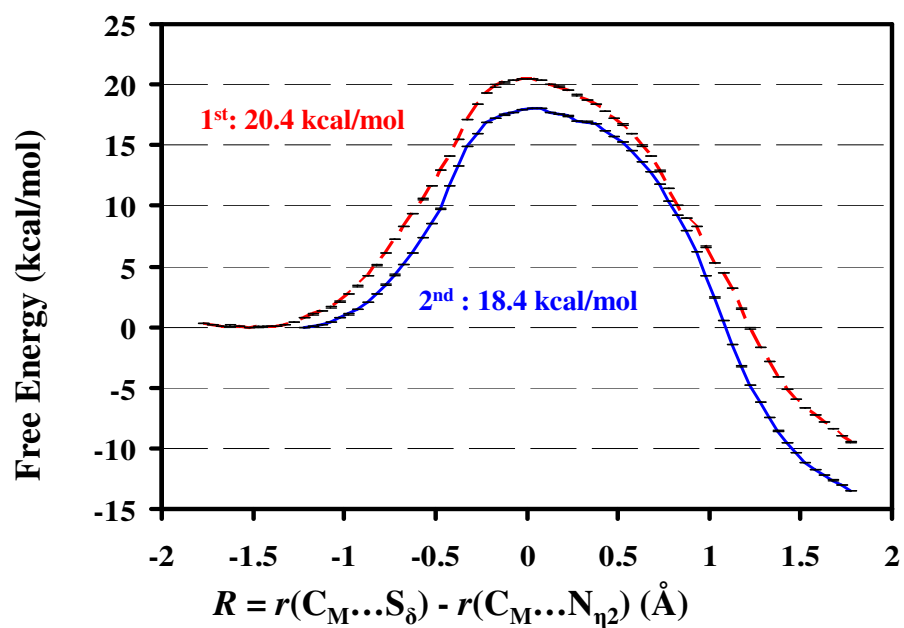


Figure 4.10 Free energy profiles of 1st and 2nd methyl additions to $\text{N}_{\eta 2}$

1st methylation: red line with a free energy barrier of 20.4 kcal/mol; 2nd methylation: blue line with a free energy barrier of 18.4 kcal/mol.

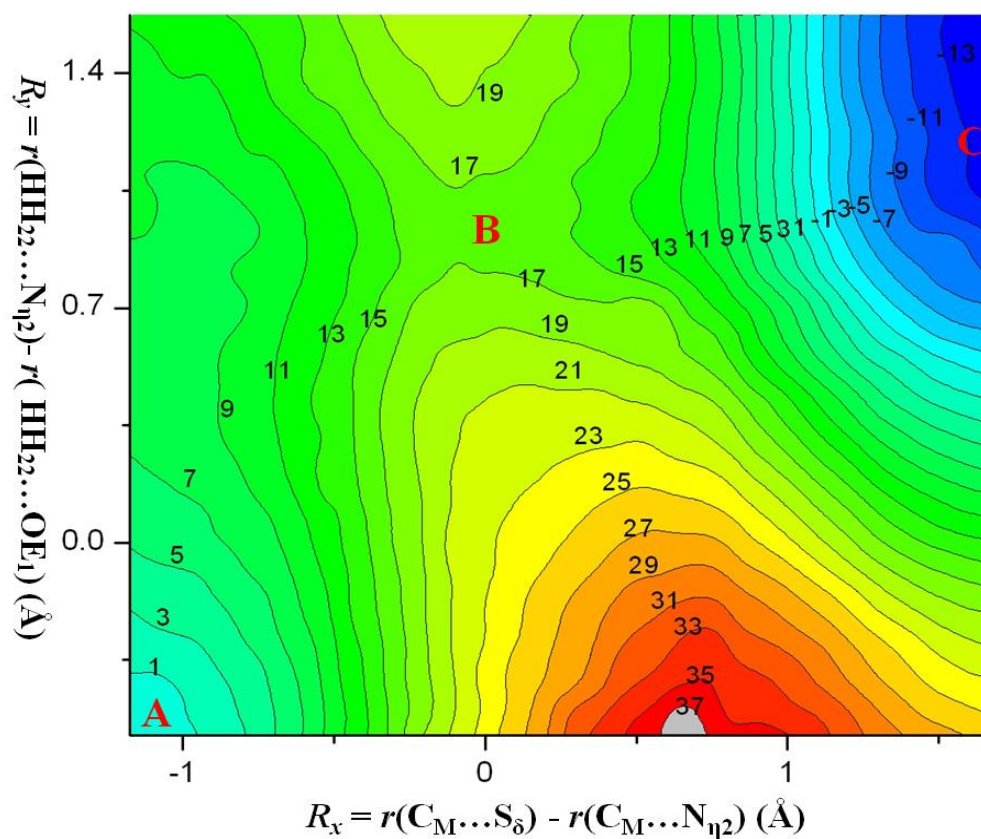
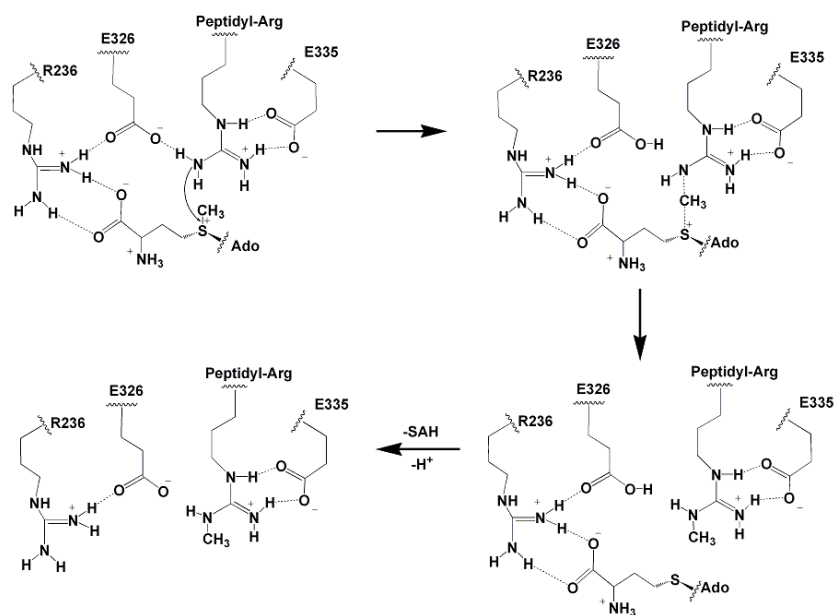


Figure 4.11 2D free energy contours for the methyl transfer from AdoMet to methyl arginine and the proton transfer between methyl arginine and Glu326

Point A, B, C designate the reactant, transition state and product complexes, respectively.

A



B

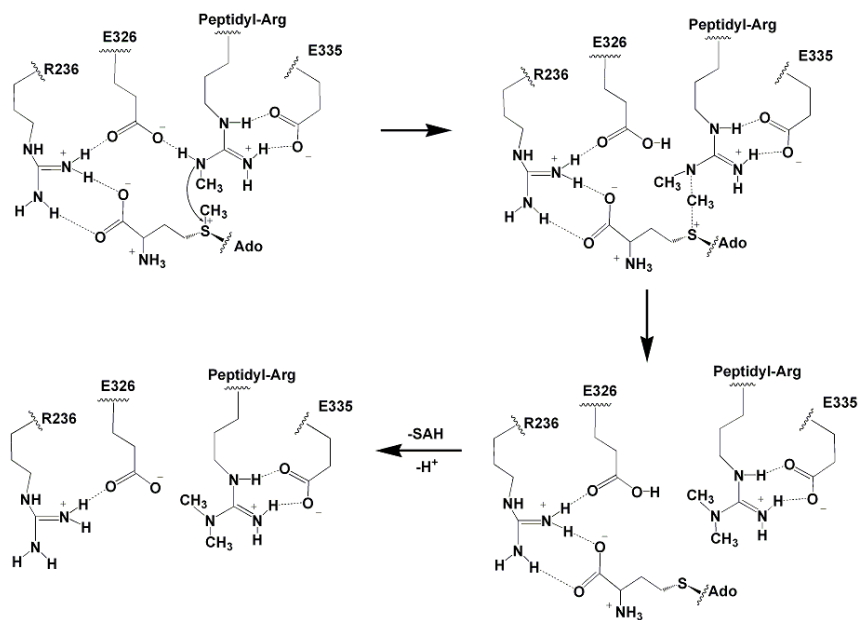


Figure 4.12 The catalytic mechanism for the methyl additions

(A) 1st methylation. **(B)** 2nd methylation.

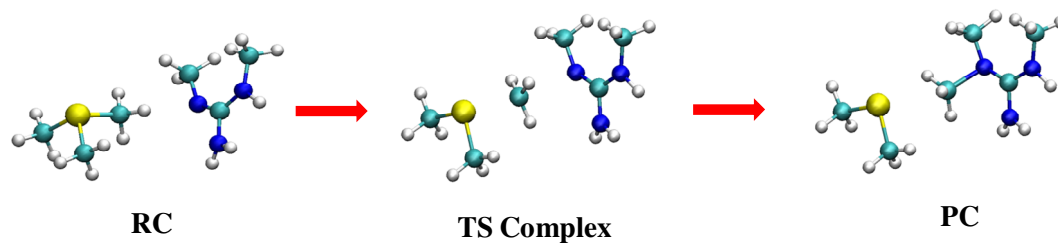


Figure S4.1 The model system used for comparison

The methyl group was transferred from the sulfonium S_6 -atom to the guanidine $N_{\eta 2}$ -atom. All structures were obtained by constrained minimizations using SCC-DFTB.

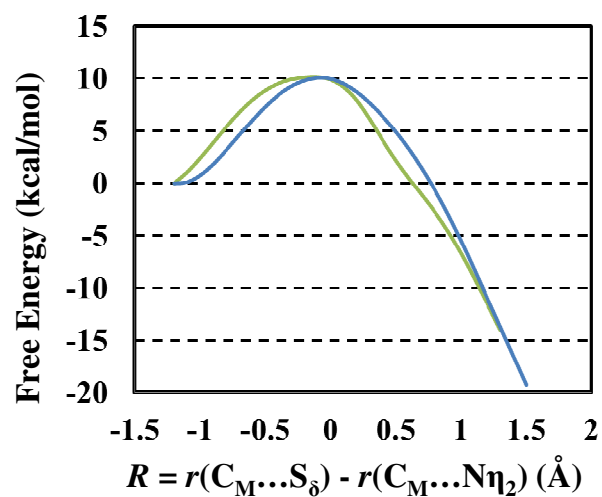


Figure S4.2 The potential energy curve calculated from SCC-DFTB (green line) and the single-point energy curve calculated by B3LYP/6-31G ** (blue line)

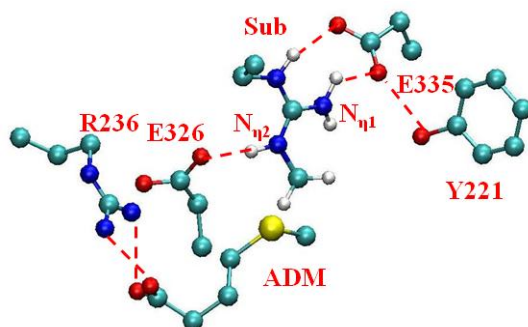


Figure S4.3 MD simulations of the product complex of the 1st methylation on N η_2 after the deprotonation of Glu326

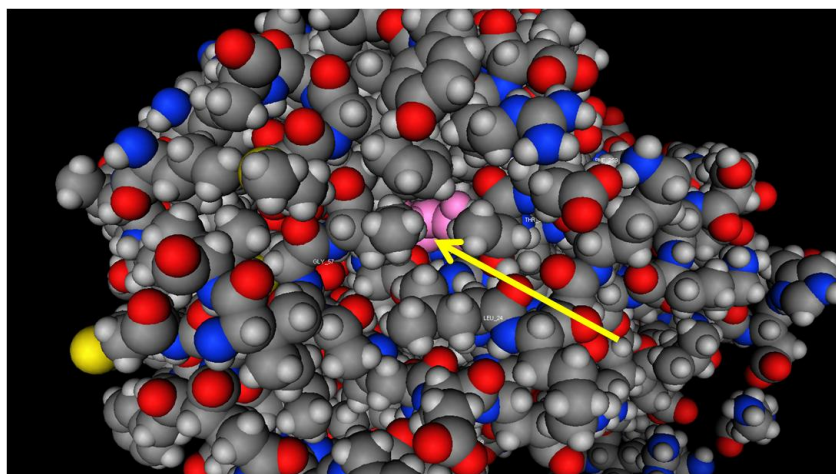


Figure S4.4 The channel opening from solution to Glu326

Glu326 is colored in pink.

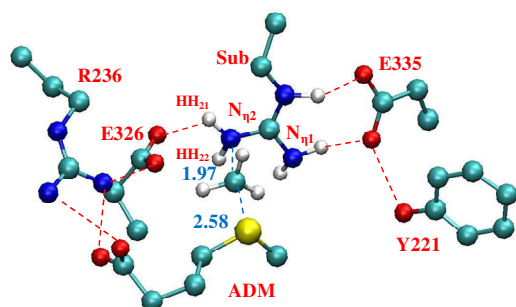
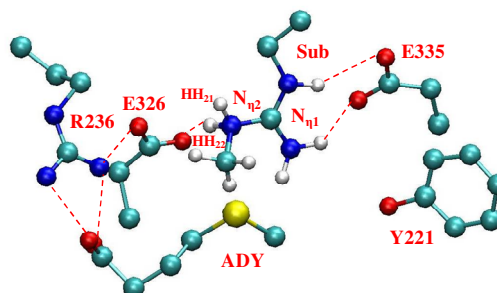
A**B**

Figure S4.5 MD results for 1st methylation on N_{η2} with shaking N_{η2}-HH₂₂ bond

(A) the active-site structure of the transition state complex. **(B)** the active-site structure of the product complex (The structure of the reactant complex is not shown since it is the same as that of without shaking N_{η2}-HH₂₂ bond, as shown in **Figure 4.2(A)**).

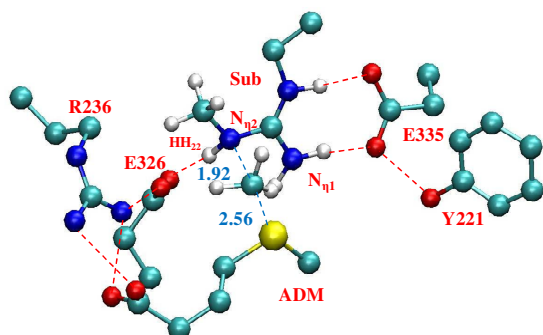
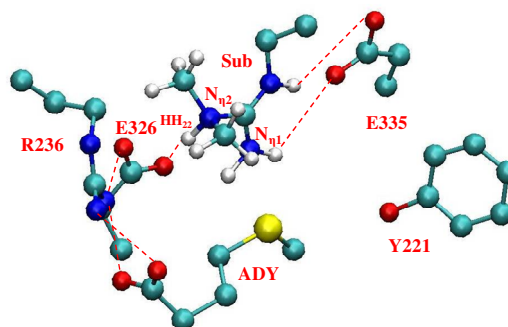
A**B**

Figure S4.6 MD results for 2nd methylation on $N_{\eta 2}$ with shaking $N_{\eta 2}$ - HH_{22} bond

(A) the active-site structure of the transition state complex. **(B)** the active-site structure of the product complex (The structure of the reactant complex is not shown since it is the same as that of without shaking $N_{\eta 2}$ - HH_{22} bond, as shown in **Figure 4.6(A)**).

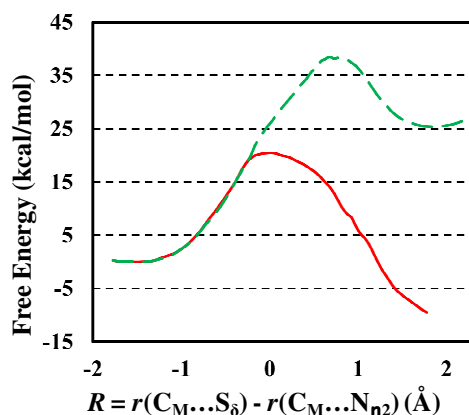
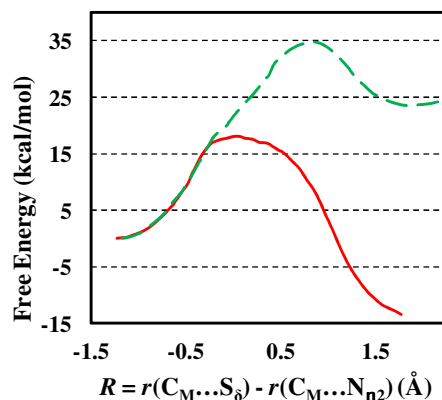
A**B**

Figure S4.7. Free energy profiles of 1st and 2nd methyl additions to N_{η2} atoms with and without shaking the N_{η2}-HH₂₂ bond

(A) 1st methyl additions to N_{η2} atoms. Without shaking: red and solid line with a free energy barrier of 20.4 kcal/mol; with shaking: green and dashed line with a free energy barrier of 38.5 kcal/mol. **(B)** 2nd methyl additions to N_{η2} atoms. Without shaking: red and solid line with a free energy barrier of 18.4 kcal/mol; with shaking: green and dashed line with a free energy barrier of 34.8 kcal/mol.

CHAPTER 5 CONCLUSION

My dissertation focuses on the investigation of the origin of product specificity and properties of certain protein methyltransferases, including histone lysine methyltransferases (SET8 and GLP), and also protein arginine methyltransferases (PRMT3), by QM/MM MD and free energy simulations.

Different degrees of histone-lysine methylation could be recognized by different methyl-lysine binding proteins to lead to different biological outcomes. Therefore, it is very important to understand the mechanism underlying the methylation processes. During my Ph.D. study, I first worked on the enzyme human SET8, which is a histone H4-Lys20 mono-methyltransferase. The availability of its structure co-crystallized with substrates in different methylation states made it possible to perform computational studies to investigate the origin of product specificity of wild type SET8 and also its Y334F mutant (Tyr/Phe switch). Our QM/MM MD and free energy simulation results have indicated that the free energy barriers of the methyl transfer step could govern the methylation state of SET8's target. Y334F mutant could switch SET8 to a dimethyltransferase since the free energy barrier of the second methylation step has been lowered to be similar to the first step, allowing the addition of the second methyl group. By contrast, for wild type SET8 the free energy barrier of the second methylation step is much higher than that of the first methylation step, stopping the further methylation after one methyl group is added to the target lysine. Our simulation results have also indicated that the Y334F mutation alters the product specificity

of SET8 through changing the affinity of a crystallized active-site water molecule. In wild type SET8, this water molecule is stabilized in the active site through hydrogen-bond interactions with the hydroxyl group of Y334, and also the carbonyl oxygen of Gly-294 and Ile-297. The mutation of Y334 to F leads to the loss of the stabilizing interaction for this water molecule from the hydroxyl group of Y334, makes it easier for this water molecule to dissociate from the active site and produces more space around the target lysine for the second methyl addition. The affinity of this active-site water molecule is further examined by free energy simulations with different reactant complexes (e.g., for the first and second methylations of wild type SET8 and its Y334F mutant). The results have also suggested that this water molecule would dissociate from the active site during the second methylation step in Y334F mutant, generating more space in the active site around the target lysine and allowing the second methyl transfer. The MD simulations of the reactant complex of each methyl transfer in wild type and Y334F mutant have also shown that the ability of the reactant complex to form the reactive configurations could be used to predict the product specificity of SET8, with the good alignment of the methyl group and the target nitrogen favoring the methyl transfer. In the first methylation of wild type SET8, the first and second methylations of Y334F mutant, the structure of the reactant complex is quite similar to the corresponding transition state structure. By contrast, for the second methylation of wild type SET8 the structure of the reactant complex is distorted from the corresponding transition state structure, and additional energy

would be required to generate the transition-state like structure, leading to higher free energy barrier of this methylation step.

Next, I started to investigate whether the similar computational approaches (QM/MM MD and free energy simulations) could be applied to other PKMTs to understand the origin of product specificity of wild type enzyme, the mutants for the Tyr/Phe switch as well as some other residues located at the active site. I started to work on mammalian GLP, which forms a heteromeric complex with its close homolog G9a and works cooperatively to exert H3-K9 methyltransferase function *in vivo*. The results of our MD simulations have indicated that the mutation at the Tyr/Phe switch (F1209Y) also alters the product specificity of GLP through altering the affinity of the crystallized active-site water molecule. In wild type enzyme, this water molecule could dissociate from the active site during the second methylation step, leaving more space around the target lysine for further methyl addition. For the F1209Y mutant, this active-site water molecule could be further stabilized at the active site through hydrogen-bond interactions with the hydroxyl group of Y1209, thus the space around the target lysine is too crowded for accepting the second methyl group. As to the Y1124F mutant, more space is generated around the target lysine due to the loss of the hydroxyl group of Y1124, and this could allow the target lysine to accept three methyl groups and alter GLP to a tri-methyltransferase. Together, the simulation results of F1209Y and Y1124F mutants suggest that the space around

the target lysine is one of the key structural factors which control the product specificity of GLP.

The free energy barriers of each methyl transfer step in wild type GLP and its mutants are also consistent with their product specificities. In wild type GLP, the free energy barriers of the first and second methylation steps are similar; as to the F1209Y mutant, the free energy barrier of the second methylation step is much higher compared to that of the first step, and this might stop the second methyl transfer. The free energy barrier of the third methylation step in Y1124F has also been lowered to be similar to those of the first and second steps, consistent with the fact that it is a tri-methyltransferase. These results support the suggestions from the SET8 study that the free energy barrier of the methyl transfer step could govern the product specificity of PKMTs energetically.

Next, I worked on one member of type I PRMTs, PRMT3, which would catalyze the formation of asymmetrically dimethylated peptidyl arginine instead of symmetrically dimethylated peptidyl arginine. As suggested from previous studies (Wysocka, Allis et al. 2006), different product formation (ADMA or SDMA) of the substrate could lead to different biological consequences, thus it is important to study the origin of the product specificity and properties of these enzymes. The free energy barriers of the first and second methyl additions to $N_{\eta 2}$ and $N_{\eta 1}$ atoms of the target guanidine group were compared. The results showed that the barriers of methyl additions to the $N_{\eta 1}$ atom are much higher than those to the $N_{\eta 2}$ atom. As a result, the both methyl groups would be added to the $N_{\eta 2}$ atom,

leading to the formation of the asymmetrically methylated product. The structural information obtained from our simulations also indicates that the active site interactions, including the interactions between the active site residues (mostly the two conserved glutamate residues, Glu326 and Glu335 on the "double E" loop), AdoMet and the substrate arginine, help align the $N_{\eta 2}$ atom well with the methyl group of AdoMet. These results are also consistent with previous suggestions that the re-distribution of charge makes the $N_{\eta 2}$ atom to be the mostly likely target of methyl addition (Zhang and Cheng 2003). 1D and 2D free energy simulations also suggest that the deprotonation of the substrate arginine is somehow concerted with the methyl transfer process and that Glu326 might function as the general base of the reaction.

In conclusion, during my Ph.D. study, QM/MM MD and free energy simulations approaches were applied to several protein methyltransferases to investigate the origin of product specificity and properties of these enzymes. However, there are still many outstanding questions to be answered in these areas. For PKMTs, the key question on the product specificity concerns the identification of the factors that control the methylation state of the product. My computational studies suggested that the product specificity may be determined in part by the existence of one or more energetically unfavorable methyl transfer steps that terminate the full methyl addition process. One of the important tasks for determining the product specificity and developing strategies for manipulating this property would be to identify the energetically unfavorable steps that the

enzyme may not catalyze and study how it shifts from one step to another under different circumstances (e.g., through mutations). I have demonstrated in this dissertation that computational approaches based on the QM/MM MD and free energy simulations are powerful tools for determining the energetics of reaction processes and can be used to produce the combined energetic and structural information for different reactant and product states as well as unstable/transient states (e.g., the transition state) that may not be directly measurable with current experimental techniques. I have used the QM/MM methods successfully to understand the product specificity for several PKMTs and the effects of the Tyr/Phe switch mutations. The further examination of this type of questions is of importance, because PKMTs play central roles in the regulation of gene expression and are considered as compelling targets for drug discovery efforts. Manipulating the methylation events of the enzyme catalysis might be used as an attractive strategy. In this regard, it is interesting to note that the strategy of using the lysine mimic in combination with an established inhibitor (BIX-01294) of G9a and GLP via a slow methylation process was able to enhance the potency of the inhibitor in vitro as well as reduce cell toxicity in vivo (Chang et al. 2010).

In addition to determine the free energy barriers for the methyl transfers, I have also showed that there seems to be a good correlation between the product specificity and ability of the reactant complexes to sample the reactive in-line conformations for the methyl transfers in the active site. The potential value of this approach is that it might provide a faster way of predicting the product

specificity. However, there may be some weaknesses and uncertainties associated with the use of the MD simulation data based on the reactant complexes alone and the results from such MD simulations may not be able to replace the data from the QM/MM free energy simulations. The usefulness and limitations of this approach will still be determined. Previous structural studies have suggested that steric constraints within the active sites as well as the presence/absence of the active-site water molecule(s) are important structural factors in the determination of the number of methyl groups to be added by the enzyme. A better understanding of the role of these structural factors on the product specificity is still necessary. Moreover, other hypotheses concerning the origin of product specificity of PKMTs (e.g., the product specificity may be controlled by the formation of a water channel) will also need to be examined further.

For PRMTs, biological consequences of arginine methylation may differ depending on whether the arginine residue is asymmetrically or symmetrically dimethylated. Nevertheless, the factors that lead to the different di-methylation products have not been clearly identified despite of extensive experimental investigations on the enzymes. It has been hypothesized that, similar to the existence of the Tyr/Phe switch in PKMTs, certain active-site residue(s) might control the product specificity for PRMTs (e.g., asymmetric vs. symmetric dimethylation). Thus, the mutation of these residues might allow PRMTs to change between the Type I and II enzymes. However, experimental research along this

direction has not been very successful. For instance, two of the conserved active-site residues in Type I PRMTs that have been hypothesized to play such a role are Met155 and Met48 (the residue numbers are based on the rat PRMT1 sequence); the corresponding residues in Type II PRMTs are much smaller (e.g., alanine or serine). It was originally thought that the methionine might provide steric bulk to block the binding of mono-methylated arginine in a conformation suitable for symmetric di-methylation. However, recent experimental study showed that M155A and M48A mutants produced only mono-methyl-arginine and asymmetric d-imethyl-arginine instead of symmetric di-methyl-arginine as hypothesized, although the distribution of mono- and di-methylated products was changed as a result of the mutations (Sun et al. 2011). This suggests that removing the steric bulk afforded by either Met155 or Met48 in the active-site of PRMT1 is not sufficient to transform PRMT1 into a Type II PRMT. Interestingly, the replacement of Phe327 in the active site of PRMT5 (a Type II PRMT) by a methionine was able to convert PRM5 to an enzyme that catalyzes both symmetric and asymmetric di-methylation. The exact reason for this change of the product specificity is still not clear. Future QM/MM MD and free energy simulations along with high-level QM/MM calculations can be performed to determine the free energy/energy barriers for mono-methylation as well as asymmetric and symmetric di-methylation of arginine catalyzed by PRMTs (e.g., PRMT1, PRMT3 and PRMT5) and their mutants. Studies can also be performed to find certain mutants that may be able to alter the product specificity of wild-

type enzymes. The prediction from the simulations can then be confirmed experimentally. Additional simulations may be performed on the product complexes of mono-methylation to determine the energetics of the conformational changes necessary for generating the reactive configurations for the next methyl addition as well as the effects of certain mutations on the energetics of such conformational changes. As demonstrated already from the study of PKMTs and PRMT3, computational approaches are powerful tools for understanding and predicting product specificity of methyltransferases and may be able to provide important insights into the product specificity of other PRMTs.

References

- Chang, Y. Q., T. Ganesh, et al. (2010). "Adding a Lysine Mimic in the Design of Potent Inhibitors of Histone Lysine Methyltransferases." J Mol Biol **400** (1): 1-7.
- Sun, L. T., M. Z. Wang, et al. (2011). "Structural insights into protein arginine symmetric dimethylation by PRMT5." Proc Natl Acad Sci U S A **108**(51): 20538-20543.
- Wysocka, J., C. D. Allis, et al. (2006). "Histone arginine methylation and its dynamic regulation." Front Biosci **11**: 344-355.
- Zhang, X. and X. D. Cheng (2003). "Structure of the predominant protein arginine methyltransferase PRMT1 and analysis of its binding to substrate peptides." Structure **11**(5): 509-520.

VITA

Yuzhuo Chu was born in Meihekou, Jilin, China at 1983. She attended Jilin University, China in 2001 and got her bachelor's degree majoring in Biotechnology in 2005. She then entered Washington State University and got her Master's degree majoring in Chemistry in December 2007. She then joined the department of Biochemistry, Cellular and Molecular Biology at the University of Tennessee for her Ph.D. study. She is graduating in May 2013.