To the graduate Council:

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

Hong Guo, Major Professor

We have read this dissertation
and recommend its acceptance:

Cynthia Peterson

Elizabeth Howll

Robert Hinde

Accepted for the Council:

Carolyn R. Hodges, Vice Provost
and Dean of the Graduate School

(Original signatures are on file with official student records)

Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations on Enzymes

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Qin Xu

August 2008

DEDICATION

To my parents, Lihong Xu and Yueping Bai for your support.

ACKNOWLEDGEMENTS

I would like to express my deep and sincere gratitude to all the people who gave me generous help and support that allow me to complete this dissertation.

I am honored to sincerely thank to my advisor, Prof. Hong Guo, for introducing me to the exciting research fields of computational simulations and for giving me the opportunities to participate in so many interesting projects. Without his excellent guidance, this work would not be possible. His kindly advice and patience are deeply appreciated. I have benefited a lot from his comprehensive knowledges and his enthusiasm in science which have greatly influenced me and inspired me in my graduate education.

I am grateful to my dissertation committe members, Dr. Cynthia Peterson, Dr. Elizabeth Howll, Dr. Robert Hinde, Dr. Nagiza Samatova and Dr. Xiaolin Cheng for the valuable comments, suggestions and support in my progression.

I wish to thank both past and present members of Dr. Guo laboratory for their friendship and help: Dr. Niny Rao, Robert Beahm, Kanan Vyas, Mauricio O. Valenzuela, Glenn Hernandez, and especially Dr. Haobo Guo, who teach me a lot in computational simulations and collaborated with me in many projects.

Finally, I would like to express my deepest acknowledgement to my family. I owe my parents, Lihong Xu and Yueping Bai for their selfless understanding and support to my life. I also thank to my brother, Yong Xu, for his support to the family.

ABSTRACT

The dynamic nature of proteins in solution is often an indispensable factor in biological function such as enzymatic catalysis. Complementary to the conventional structural analysis, computational simulations have the advantage to reflect the dynamic nature of proteins or enzymes. One of the computational simulation methods, the quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) simulations, has been widely applied to the research in structural analysis, ligand-receptor binding and enzymatic catalysis.

In this dissertation, QM/MM MD simulations were applied to the studies on cytidine deaminase (CDA), yeast cytosine deaminase (yCD), and kumamolisin-As, as well as two protein lysine methyltransferases (PKMTs), DIM-5 and SET7/9. In the simulations of the transition state analogue (TSA) binding of zebularine 3, 4-hydrate to CDA and of 4-[R]-hydroxyl-3,4-dihydropyrimidine (DHP) to yCD, proton transfers were observed between the TSA and a catalytic Glu residue in both cases. Such general acid-base mechanism was also observed in the stabilization of the tetrahedral intermediate by a critical Asp residue during the acylation of kumamolisin-As. Moreover, dynamic substrate-assisted catalysis (DSAC) involving the His of the substrate at P1 site was proposed. It was suggested that DSAC may contribute to the transition state stabilizations and substrate specificity of kumamolisin-As. The origin of the product specificities of PKMTs was studied by comparison of QM/MM MD simulations on the first, second and third methyl transfers in the trimethylase DIM5 and the monomethylase SET7/9. The

product specificities of the enzymes can be well explained by population distributions of well-aligned reactive structures and the relative free energy barriers for the methyl transfers. The structural and energetic reasons for the product specificities were discussed and a triplet code based on the relative free energy barriers for the three methyl transfers was proposed in the determination of product specificities of PKMTs.

TABLE OF CONTENTS

Chapter I **1**

General Introduction **1**

Background 2

Dyanmics feature of Enzymes 2

How enzymes work 3

General acid-base catalysis 4

Gene-directed enzyme prodrug therapy (GDEPT) 5

Proposed catalytic mechanism of Sedolisin family 6

Dynamic substrate-assisted catalysis (DSAC) 8

Product specificity of histone methylation 9

Summary of Chapters 11

Reference 16

Appendix 21

Chapter II **34**

Methods **34**

Quantum Mechanical (QM)/Molecular Mechanical (MM) Molecular Dynamics (MD) Simulations 35

Construction of Simulation Models and the conditions in simulations 37

Free Energy Change Determination 40

Reference 43

Appendix 46

Chapter III **47**

Quantum Mechanical/Molecular Mechanical Molecular-Dynamics Simulations of Cytidine Deaminase: from Stabilization of Transition-State Analogs to Catalytic Mechanisms 47

Abstract 48
Abbreviation: 49
Introduction 50
Methods 54
Results 56
ZEB-H_2O in wild-type CDA 56
ZEB-H_2O in E104A 60
Discussions 61
Acknowledgements 65
References 66
Appendix 69

Chapter IV 73

Stabilization of a Transition-State Analogue at the Active Site of Yeast Cytosine Deaminase through a General Acid-Base Mechanism 73

Abstract 74
Abbreviations 74
Introduction 75
Methods 78
Results and Discussions 81
Conclusions 84
Acknowledgements 85
Supporting Information 85
References 87

Appendix 90

Chapter V 100

The Importance of Dynamics in Substrate-Assisted Catalysis and Specificity 100

Abstract 101
Abbreviations: 101
Introduction 102
Results and Discussions 104
Supporting Information: Computational Method 106
Acknowledgment 108
References 110
Appendix 113

Chapter VI 118

The QM/MM molecular dynamics and free energy simulations of the acylation reaction catalyzed by the serine carboxyl peptidase kumamolisin-As 118

Abstract 119
Abbreviations: 120
Introduction 121
Methods 125
Results and Discussion 132
Structural and dynamic properties of the substrate complex, tetrahedral intermediate, and acyl-enzyme 132
Free energy profiles of the acylation 134
Dynamic substrate assisted catalysis (DSAC) 136
Conclusions 139
Acknowledgment 140

References 141
Appendix 144

Chapter VII 149

A Triplet Code for Writing Epigenetic Marks: Application of QM/MM MD and Free Energy Simulations to Understand Product Specificity of Protein Lysine Methyltransferases DIM-5 and SET7/9 149

ABSTRACT 150
Introduction 151
Methods 159
Results 163
Methyl transfers in wild type DIM-5 163
Methyl transfers in F281Y mutant of DIM-5 165
The 3rd methyl transfers in wild type and Y305F mutant of SET7/9 167
The relative free energy barriers for methyl transfers: a triplet code for writing epigenetic marks 168
Discussions 169
Conclusions 172
References 173
Appendix 179

Chapter VIII 188

Conclusions and future directions 188

VITA 193

LIST OF FIGURES

Figure I-1. Thermodynamic parameters of uncatalyzed and CDA-catalyzed cytidine deamination 22
Figure I-2. The X-ray structure of CDA-ZEB-H_2O complex 23
Figure I-3. Gene-directed enzyme prodrug therapy (GDEPT) 24
Figure I-4. Sequence alignment of sedolisins..... 25
Figure I-5. Proposed catalytic mechanism of kumamolisin-As 26
Figure I-6. Substrate-assisted catalysis 27
Figure I-7. Proposed dynamic substrate-assisted catalysis (DSAC) in Kumamolisin-As catalysis..... 28
Figure I-8. Necleosome and histones 29
Figure I-9. Sites of post-translational modifications on histones..... 30
Figure I-10. Association of biological functions and histone modifications 31
Figure I-11. Protein lysine methyltransferases (PKMTs) and methyl-lysine binding proteins..... 32
Figure I-12. Publications on the topic of "histone lysine methylation" 33

Figure II-1. Stochastic boundary system 46

Figure III-1. Reaction catalyzed by CDA and TSA of this reaction 69
Figure III-2. Simulation results of CDA-ZEB-H_2O complex..... 70
Figure III-3. Simulation results of E104A-ZEB-H_2O complex..... 72

Figure VI-1. The reaction catalyzed by yCD and the formation of TSA in yCD..... 90
Figure IV-2. Comparison of yCD-TSA complexes between the average structures from the simulations and the X-ray structure..... 91
Figure IV-3. The distance fluctuations in simulations of yCD-TSA complex 92
Figure IV-4. The nucleophilic attack in the yCD-TSA complex..... 93

Figure IV-5. The proton transfers monitored by the changes of the distances..... 94
Figure IV-6. The changes of the distances during the nucleophilic attack..... 95
Figure IV-S1. The structure of zinc ligation in the bonded model..... 96
Figure IV-S2. Comparison of the performances of the SCC-DFTB and B3LYP/6-31G(d,p) methods..... 97
Figure IV-S3. Comparison of the free energy changes of the nucleophilic attack starting from different initial structures..... 98
Figure IV-S4. The free energy changes for the proton transfer between O_4 and $O_{\varepsilon 1}$ 99

Figure V-1. Average structures of kumamolisin-As complex during the nucleophilic attack..... 113
Figure V-2. The free energy changes of the nucleophilic attacks..... 114
Figure V-S1. The average structure of the substrate complex for RPH*FR 115
Figure V-S2. The average structure of the transition state complex for RPH*FR. 116
Figure V-S3. The average structure of the TI complex for RPH*FR..... 116
Figure V-S4. The average structure of the substrate complex for RPF*FR..... 117
Figure V-S5. The average structure of the TI complex for RPF*FR..... 117

Figure VI-1. The average active-site structures of the kumamolsin-As-substrate (GPH*FF) complex, tetrahedral intermediate and acyl-enzyme obtained from the simulations, and the motions of the three key protons involved in the catalysis. 144
Figure VI-2. The free energy profiles of the acylation and the average structures at the transition states for the nucleophilic attack and the formation of the acyl-enzyme..... 147
Figure VI-3. The free energy changes for the shifts of P_1 His side chain..... 148

Figure VII-1. The methyl transfers in PKMTs 179

Figure VII-2. The average structures and $r(C_M \ldots N_\zeta)/\theta$ distributions of the reactant complexes for methyl transfers in wild type DIM-5 181

Figure VII-3. The average structures and $r(C_M \ldots N_\zeta)/\theta$ distributions of the reactant complexes for methyl transfers in DIM-5 F281Y mutant 183

Figure VII-4 The average structures and $r(C_M \ldots N_\zeta)/\theta$ distributions of the reactant complexes for the 3rd methyl transfers in the wild type and Y305F mutant of SET7/9 185

Figure VII-5. The free energy changes for the methyl transfers in the wild type and the F281Y mutant of DIM-5, as well as in wild type SET7/9 186

LIST OF TABLES

Table I-1. Comparison between the features of sedolisin and subtilisin..... 21
Table I-2. Substrate-assisted catalysis in subtilisin..... 21

Table IV-S1. Parameters for the zinc-bonded model 96

Chapter I

General Introduction

Background

Dyanmics feature of Enzymes

Enzymes are one of the major foci of biochemical studies because of their amazing ability in reaction rate enhancement, as high as 10^{20} times, [1] and high specificity, both for the substrates and the products. It has been realized that the structures of enzymes are critical to their activities since the crystal of urease by James Sumner in 1926. [2] In 1965, the X-ray structure of lysozyme by David Phillips et al. made it possible for direct strcutrural analyses on the function of enzymes. [3] However, enzymes are far beyond rigid and static structures, which were generally identified by the X-ray crystallography method. With the developments in nuclear magnetic resonance (NMR) and other techniques, the dynamic feature of enzymes is becoming better appreciated.

The flexibility of enzyme structures has been pointed out in the famous "induce fit" hypotheis by Koshland 50 years ago, [4] in which substantial conformational changes are induced by substrate binding from the apo-enzyme to the enzyme-substrate (ES) complex (or Michaelis complex, MC)[5]. Recently, growing evidence from NMR expriments confirmed the existence of the dynamic conformational ensembles of ubiquitin and calmodulin, [6-8] which support the "conformational selection" model for protein binding. In this model, the ligand (or substrate) binding to a protein (or enzyme) shifts the dynamic equilibrium of the protein's conformational ensemble in favor of the unbound structure to the ensemble with the bound structure most populated. [9, 10] The recognition of the importance of dynamic features in enzymatic catalysis [11] has been boosted by the developments in new methods, such as computational simulations.

How enzymes work

The catalytic proficicency and specificity of enzymes may involve multiple factors, [11] which may be classified into TS-theories and MC-theories, [5] based on their functions in transition state (TS) stabilization and substrate preorganization, respectively.

As early as 1946, Linus Pauling attributed the efficiency of enzymatic catalysis to the ability of enzymes to stabilize transition state and lower the activation barriers of the reactions. [5, 12] (Figure I-1) Searching for the origin of the transition state stabilization in enzyme catalyzed reactions is one of the major foci in studying the catalytic mechanisms. It is believed that the structure of transition state analogs (TSAs) may partially mimic the altered structure of the substrate at transition state during the enzyme-catalyzed reaction and capture a part of the transition state stabilization. The interactions with the TSA at the active site may be similar to the interactions responsible for the transition state stabilization during the catalysis.[13-20] Therefore, it is possible to utilize the abundant high-resolution crystallographic structures of enzyme-TSA complexes [21, 22] in computational simulations and to obtain important insights into the catalytic power of the enzymes.

As in Figure I-1, given the energy of the transition state complex, the activation energy may also be determined by the binding energy in the formation of Michaelis complex (ES). This energy may have contributions from the reorganization of the substrate structure in the active site of the reactant complex ES. Computer simulations can be useful in understanding of related questions. For instance, in Chapter VII computational methods are used to monitor the geometrical parameters between the

substate and the cofactors and to explain the product specificities of protein lysine methyltransferases (PKMTs) and their mutants.

General acid-base catalysis

General acid-base catalysis involving proton transfers may be one of the most significant mechanisms in enzymatic catalysis.[23-26] Unfortunately, the possible proton transfers in a catalytic process or in an inhibitor binding are often neglected by the studies of static structures, due to the invisibility of protons in high resolution X-ray structures. On the other hand, it is possible to trace the microscopic positions of protons in computational simulations. In simulations, the proton transfers can be observed both in enzymatic catalysis and in the binding of a TSA at the active site.[27-32]

For example, in the X-ray structure of CDA-ZEB-H_2O complex (PDB code: 1CTU), the presumed hydroxyl group has an abnormal longer C_4-O_4 bond of 1.6Å with C_4 and an unexpected low barrier hydrogen bond of 2.5Å with $O_{\varepsilon 1}$ of Glu104, even with a restraint of 1.43 Å for C_4-O_4 bond and a restraint of 2.8Å for the O_4…$O_{\varepsilon 1}$ distance during the refinement. (Figure I-2) Computer simulations allowed us to observe the proton transfer from O_4 to Glu104, leading a more stable structure with the normal C_4-O_4 bond of 1.5Å and the normal hydrogen bond of O_4…H- $O_{\varepsilon 1}$ as 2.7Å. [31] In addition to identification of proton motions in MD simulations, it is possible to quantitatively calculate the free energy change as a result of the stabilizations by the general acid-base mechanism[33-36].

Gene-directed enzyme prodrug therapy (GDEPT)

The gene-directed enzyme/prodrug therapy delivers a DNA construct containg an enzyme encoding gene to target cancer cells. The enzyme expressed from this gene generally has no toxic effect itself, but it can convert a nontoxic compound (termed as prodrug) into a cytoxic drug. It is suggested that this reaction should be different from endogenous pathways so as to avoid cytoxic activation in normal tissues. To ensure a therapeutic effect, the drug should have more than 100 fold higher toxicity than the prodrug. What is more, a proper half-life is required for the drugs to allow the diffusion into surrounding untransfected cells, killing them by the bystander effect. However, the drug in ciculation system must be inactive so as not to kill normal cells. [37] (Figure I-3)

The cytosine deaminase (CD)/5-fluorocytosine (5-FC) system is one of the promising protocol for GDEPT. In this protocol, the enzyme CD is a hydrolytic deaminase converting cytosine to uracil; this enzyme is only found in some bacteria and fungi but not in mammalian cells. It can also convert the nontoxic prodrug 5-FC into an anticancer drug, 5-fluorouracil (5-FU). This toxic nucleotide analogue can be converted by normal cellular enzymes into three possible products, 5-fluorodeoxyuridine-5-monophosphate (5-FdUMP), 5-fluorodeoxyuridine-5-triphosphate (5-FdUTP), and 5-fluorouridine-5-triphosphate (5-FUTP), which may lead to inhibition of thymidylate synthase, formation of 5-FU RNA and formation of 5-FU DNA, and lead to cell death. [37-39] The strong bystander effect of the CD/5-FC system allows lower transfection efficiency of the CD gene. The requirement for prodrug dose has also been lowered by using yeast cytosine deaminase (yCD) from *Saccharomyces cerevisiae*, which has a 22-

fold lower Km for 5-FC than bacteria cytosine deaminase (bCD) from *E. Coli*. [40] Extensive experiments have proved the CD/5-FC system to be useful in killing tumor cells, such as breast cancer, liver cancer, colorectal carcinoma and colon carcinoma. [41-44] A number of protocols using this system are in development, including clinic trials in USA and UK.[37] To understand the role of proton transfers in the yCD-catalyzed deamination and substrate/inhibitor binding by computational simulations is quite interesting and may contribute to the development of therapeutical approaches and drug design.

Proposed catalytic mechanism of Sedolisin family

The serine-carboxyl peptidases (sedolisins) family has a wide variety of sources, such as archaea, bacteria, molds, slime molds (mixomycetes), amoebas, fishes, and mammals. Although much larger in size, sedolisins are preseumed to have a common ancestor with the classical serine peptidases (subtilisins). The sequence alignment of some sedolisin members revealed four totally conserved residues, including the possible catalytic triad Ser-Glu-Asp and an Asp that may contribute to tetrahedral intermediate stabilization.[45] (Figure I-4) The structural fold of these four residues is quite comparable to the catalytic triad Ser-His-Asp and Asn155 in subtilisins. Based on the well-known ping-pong mechanism of classical serine peptidases, the catalytic mechanim of sedolisin family has also been proposed. (Figure I-5) The whole reaction circle can be separated into acylation (left part) and deacylation (right). Each section has a substrate bound at the beginning and a product released at the end. The product release in each section is required to allow the binding of the substrate of the next section. During the acylation,

the serine residue is the catalytic nucleophile to attack the substate peptide bond. The Glu nearby seems to act as the general base to accept the proton from Ser at the time of the nucleophilic attack, and then act as the general acid to protonate the nitrogen forming the leaving N-terminal group. During the deacylation, this Glu still act as the general acid-base to accept a proton from the attacking water molecule and give it back to the deprotonated Ser of the acyl-enzyme. (Figure I-5)

Although the catalytic mechanism of sedolisins and subtilisins are proposed to have many common features, there are two major differences between them. One difference is that in sedolisins, the Glu takes the place of the His in subtilisins to work as the general acid-based to mediate proton transfers from the nucleophile Ser to the N-terminal of the scissile peptide bond and from the attacking water molecule back to the deprotonated Ser in the acyl-enzyme. The other important difference is the mechanism of the tetrahedral intermediate stabilization. In sedolisins, the Asp may act as a general acid to protonate the tetrahedral intermediate, whereas the corresponding Asn creates the oxyanion hole in subtilisins. Therefore, the mechanism of the tetrahedral intermediate stabilization in sedolisins is more similar to that in aspartic peptidase [46] rather than that in subtilisins. Replacement of the His and Asn in subtilisins by the Glu and Asp in sedolisins, respectively, may be related to the acidic pH optimum of the sedolisin family. (Table I-1) A possible explanation for the evolutions of sedolisins might be the demand for some organisms of the serine peptidase activity under acidic environments. To study the similarities and differences in the catalytic mechanisms of sedolisins, subtilisins and

aspartic peptidases, details of the catalytic process by sedolisins are indispensable. Thus computational simulations monitoring the proton transfers might be of great help.

Dynamic substrate-assisted catalysis (DSAC)

Substrate-assisted catalysis (SAC) is the process in which one or more functional groups from the substrate, in addition to those from the enzyme, contribute to the rate acceleration for the enzyme-catalyzed reaction. SAC has been observed in many cases of both natural and engineered enzymes. [47-53] In the case of SAC in natural enzymes, removal of a functional group of the substrate may inactivate the enzymatic catalysis. In the case of SAC in engineered enzymes, addition on the subtrate with a functional group that similar to mutated catalytic group of the enzyme may at least partially restore the activity of the enzyme. [48] (Figure I-6 a, b) For example, in GTPase catalysis, it was found that any mutation on the catalytic Gln resiude would inactivate the GTP hydrolysis. However, a GTP analogue with an aromatic amine correctly at the mutated active site can help to restore the catalytic activity of GTPase.[48] (Figure I-6c) Another example is the SAC in subtilisin catalysis, in which the P_2-His or P_1'-His can partially restore the activity of H64A mutant. Thus the broadly specificity of wild type subtilisin is converted into His specificity in the H64A mutant. [49] (Table I-2) Similarly, the product-assisted catalysis has also been reported in the catalysis by human 8-oxoguanine DNA glycosylase/β-lyase (hOGG1). The product of the glycosylase activity, 8-oxoguanine (oxoG) was suggest to work as a general acid-base in the subsequent β-lyase catalytic cascade of hOGG1. [54]

Many of previous studies on SAC only focused on the cases with well-positioned substrate groups that can participate in SAC directly without significant conformational changes. Here we introduce dynamic substrate-assisted catalysis (DSAC), in which substrate groups require conformational changes triggered by bond breaking and making events so as to contribute to the rate acceleration. As proposed in Figure I-7, the bond breaking and making events, especially the protonation and deprotonation of Asp164 may trigger the back and forth shifts of the P_1 His side chain between the hydrogen bond with the carbonyl of P_2 Pro and the salt bridge with the carboxylate of Asp164, which may help to lower the free energy barriers at different stages of the catalysis. It is reasonable to assume that DSAC might have been overlooked in many enzyme catalyses, since its dynamic feature may not be well reflected in static X-ray structures. Fortunately, this dynamic feature can be well reflected in computational simulations, as demonstrated by our QM/MM MD simulations on Kumamolisin-As.

Product specificity of histone methylation

It is believed that most eukaryotic DNA are wrapped around histone proteins in nucleosome.[55, 56] (Figure I-8) The N-terminal tails of the core histones are suggested to be not only important in the stabilizations of the the DNA wrapping around the histone octamer, but also indispensable in the formation of the higher order structure, 30-nm fiber.[55] The tails of the histone proteins are subject to a variety of post-translational covalent modifications.[57] (Figure I-9) These modifications form the so-called epigenetic histone code[58, 59] read by other proteins, which may lead to distinct downstream events in

the regulation of chromatin structure and gene expression. [60] (Figure I-10) One such modification is histone lysine methylation catalyzed by protein lysine methyltransferases (PKMTs).[61, 62] (Figure I-11) The publications related to "histone lysine methylation" in recent years are summarized in Figure I-12. There have been several lysine residues on histone proteins identified as the methylation sites, including H3-K4 (Lys4 of histone H3), H3-K9, H3-K27, H3-K36, H3-K79 and H4-K20 (Lys20 of histone H4). The preferences of PKMTs for the sites of methylation are called substrate specificity. The methylated lysine residues can then serve as the docking sites for methyl-lysine binding domains and effector proteins to mediate the downstream events in the regulation of chromatin structure and gene expression.[62] As far as we know, most of the PKMTs contain the SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain. The only exception is Dot1p, whose structural scaffolding is totally different from the SET domain PKMTs. A number of crystal structures of the complexes of the SET domain PKMTs (including SET7/9, SET8, Clr4, DIM-5, vSET and pLSMT) have been determined,[60] in which the substrate lysine and the cofactor are on the opposite sides of the enzymes, with the target lysine side chain accessing the active site and the cofactor through a narrow channel. The crystal structures of the non-SET domain PKMTs (yeast Dot1p and human Dot1L) have also been determined.[63, 64]

In addition to the substrate specificity, PKMTs may also differ in the ability to transfer one, two or three methyl groups from S-adenosyl-L-methionine (AdoMet) molecule(s) to the ε-amino group of the target lysine. This is because each of the three protons on the ε-amino group of the target lysine may be replaced by a methyl group.

This difference in degree of methylation is called product specificity of PKMTs, and PKMTs can therefore be classified into mono-methylases, di-methylases and tri-methylases. (Figure VII-1A) For example, the mono-methylase human SET7/9 generates exclusively monomethyl H3 Lys4.[65] Another SET domain PKMT, the tri-methylase DIM-5 of *N. crassa*, generates primarily trimethyl H3 Lys9, which marks chromatin regions for DNA methylation.[66, 67] For the SET domain di-methylases and tri-methylases, the methylation processes are believed to proceed processively without the release of the intermediates from the active sites.[67, 68] (Figure VII-1B)

Like the substrate specificity, the products that differ in the degree of lysine methylation (mono-, di- or tri-methylation) may lead to distinct biological consequences.[60] Therefore, it is quite interesting to determine the origin of the product specificity and to develop suitable computational models or techniques for the prediction of the product specificity. Nevertheless, our understanding of product specificity is still lacking.

Summary of Chapters

I performed the simulations reported in chapter III, IV, V, VI and VII, with some extensive discussions with my advisor, Prof. Hong Guo and one of my lab members, Dr. Haobo Guo. Three additional papers for which I am one of the co-authors are not included in this thesis. A short summary of the chapters are listed as below.

In chapter I, the first section is a general introduction for the concepts related to our research, and the second section has a short summary of the contents for each chapter.

In chapter II there is a general description of the computational and mathematical methods used in our studies and analyses. The first section is an introduction of the background about some quantum mechanical and the molecular mechanical approaches developed for computer simulations, especially the SCC-DFTB/CHARMM22 method that we used in our QM/MM MD simulations. The second section contains some descriptions of the techniques we used to set up our simulations models, including the link atom method and the stochastic boundary system as well as some basic parameters (such as the integration steps and the simulations temperatures). The third section provides a summary of the umbrella sampling method and the WHAM program, the approaches for determination of the free energy change along a selected reaction coordinate. In addition, this section describes the techniques we used for estimation of statistical errors and energy correction based on comparison with higher level of QM calculations on small models. For more detailed discussions of the methods used, please check the Method section in the later chapters.

In chapter III, we discuss the QM/MM MD simulations on the CDA-ZEB-H_2O (cytidine deaminase-zebularine-3, 4-hydrate) complex. Understanding the origin of the tight binding of the transition state analogue (TSA) at the active site may provide important insights into the mechanism by which the enzyme stabilizes the transition state of the reaction. In these simulations, the features of the X-ray structure were successfully

reproduced. The simulation of the E104A mutant confirmed the critical function of Glu104 in the stabilization of the tetrahedral intermediate.

In chapter IV, a complex similar to CDA-ZEB-H_2O was investigated. The similar result from yCD-DHP (yeast cytosine deaminase-hydroxyl-3, 4-dihydropyrimidine) complex confirmed our consideration of the importance of the proton transfers between TSA and the catalytic residue. Similar to our earlier result on CDA, [19] the transition state analogue DHP is unstable at the active site of yCD and tends to exist in an alkoxide-like structure instead of a hydroxyl form, with the proton of the hydroxyl transferred to the carboxylate of the catalytic residue Glu64. When this proton transfer is prevented, the tetrahedral intermediate with a hydroxyl was estimated to be 8-9 kcal/mol less stable.

In chapter V, the nucleophilic attack process in another enzyme, Kumamolisin-As, was simulated. One of the aims is to study the possible role of proton transfer from Asp164 to the substrate in the stabilization of the tetrahedral intermediate (TI). As expected, the proton on Asp164 was transferred to the tetrahedral intermediate during the late stage of the nucleophilic attack. Prevention of this proton transfer resulted in an unstable tetrahedral intermediate complex. Correspondingly, the free energy simulations indicate a much higher energy with no energetic minimum for the TI complex if the proton is fixed on Asp164. There was a surprise finding that the side chain of His at P_1 site of the substrate undergoes a significant conformational change. The original hydrogen bond of this His with the carbonyl of Pro at P_2 site is broken, and the side chain undergoes a rotation to interact with the unprotonated carboxylate of Asp164. Therefore, the side chain of P_1 His seems to contribute to the stabilization of the tetrahedral

intermediate. This substrate-assisted catalysis (SAC) may at least partially contribute to the substrate specificity of Kumamolisin-As. Actually, replacement of this His to Phe resulted in 5kcal/mol higher energetic barrier for the nucleophilic attack according to the free energy simulations.

In chapter VI, the simulations on the catalytic mechanism of Kumamolisin-As were extended to the whole process of acylation. The difference between the distances of C-N and C…O_γ(Ser278) was used as the reaction coordinate, and both the nucleophilic attack by Ser278 and the breaking of the scissile peptide bond could be described in the same umbrella sampling simulations. The results of these simulations confirmed that Asp164 acts as a general acid-base to protonate/deprotonate the tetrahedral intermediate, which resulted in an energetic minimum during the acylation. With the proton fixed on Asp164, a 10kcal/mol higher energetic barrier was estimated by free energy simulations, and no energetic minimum was obtained near the area of the tetrahedral intermediate. The results confirmed the earlier result on the important contribution of the Asp164 in stabilization of the tetrahedral intermediate through a general acid-base mechanism. In addition, comparison of the activation barriers for the transition states of nucleophilic attack and the peptide bond breaking seems to suggest that the nucleophilic attack on the substrate may not be the rate-limiting step of acylation. Accompanying the back and forth proton transfers between the tetrahedral intermediate and Asp164, the positively charged side chain of P_1 His is shifted back and forth between the negatively charged carboxylate of Asp164 and the carbonyl of P_2 Pro. This dynamic feature of P_1 His side chain allows it to participate in the stabilization of complex at different stages of the acylation. To

distinguish it from the conventional SAC, we define the effect of P_1 His side chain as dynamic substrate-assisted catalysis (DSAC).

In chapter VII, QM/MM MD simulations were applied to the 1st, 2nd and 3rd methyl transfers from AdoMet to the target lysine/methyl lysine in the *N. crassa* DIM-5 (a trimethylase) and human SET7/9 as well as their mutants. The results suggest that there are some correlations between the product specificity and the structures of the reactant complexes, the relative difference in free energy barriers for methyl transfers, the distributions for distance $r(C_M \ldots N_\zeta)$ and angle θ (the distance for N_ζ to attack the methyl group and angle between the orientation of the electron lone pair on N_ζ relative to the C_M-S_δ bond). The relative difference in free energy barriers for methyl transfers can be summarized into a triplet code, which can pretty well explain the product specificity switch by F281Y mutation in DIM-5 and Y305F mutant in SET7/9.

In chapter VIII, a general conclusion of our QM/MM MD simulations on enzymes is summarized, and some future directions of the simulations in next step are suggested.

Reference

(1) Lad, C.; Williams, N. H.; Wolfenden, R. V. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5607-5610.

(2) Sumner, J. B. *J. Boil. Chem.* **1926**, *69*, 435-441.

(3) Blake, C. C.; Johnson, L. N.; Mair, G. A.; North, A. C. T.; Phillips, D. C.; Sarma, V. R. *Proc. R. Soc. London B* **1967**, *167*, 378-388.

(4) Koshland, D. E., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98-104.

(5) Marti, S.; Roca, M.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J. *Chem. Soc. Rev.* **2004**, *33*, 98-107.

(6) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128-132.

(7) Lange, O. F.; Lakomek, N.; Fares, C.; Schroder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmuller, H.; Griesinger, C.; Groot, B. L., de. *Science* **2008**, *320*, 1471-1475.

(8) Gsponer, J.; Christodoulou, J.; Cavalli, A.; Bui, J. M.; Richter, B.; Dobson, C. M.; Vendruscolo, M. *Structure* **2008**, *16*, 736-746.

(9) Ma, B.; Kumar, S.; Tsai, C. J.; Nussinov, R. *Protein Eng. Des. Sel.* **1999**, *12*, 713-720.

(10) Boehr, D. D.; Wright, P. E. *Science* **2008**, *320*, 1429-1430.

(11) Ringe, D.; Petsko, G. A. *Science* **2008**, *320*, 1428-1429.

(12) Garcia-Viloca, M.; Gao, J.; Karplus, M. Truhlar, D. G. *Science* **2004**, *303*, 186-194.

(13) Wolfenden, R. *Nature* **1969**, *223*, 704-705.

(14) Wolfenden, R. *Acc. Chem. Res.* **1972**, *5*, 10-18.

(15) Wolfenden, R.; Kati, W. M. *Acc. Chem. Res.* **1991**, *24*, 209-215.

(16) Wolfenden, R. *Annu. Rev. Biophys. Bioeng.* **1976**, *5*, 271-306.

(17) Radzicka, A; Wolfenden, R. *Methods Enzymol.* **1995**, *249*, 284-312.

(18) Schramm, V. L. *Annu. Rev. Biochem.* **1998**, *67*, 693-720.

(19) Morrison, J. F.; Walsh, C. T. *Adv. Enzymol. Relat. Area Mol. Biol.* **1988,** *61*, 201-301.

(20) Mader, M. M.; Bartlett, P. A. *Chem. Rev.* **1997**, *97*, 1281-1301.

(21) Lolis, E.; Petsko, G. A. *Annu. Rev. Biochem.* **1990**, *59*, 597-630.

(22) Lipscomb, WN; Strater, N. *Chem. Rev.* **1996**, *96*, 2375-2433.

(23) Fersht, A. Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding, 2nd ed., W. H. Freeman and Company, New York, 1999.

(24) Jencks, W. P. *Chem. Rev.* **1972**, *72*, 705-718.

(25) Kirby, A. J. *Acc. Chem. Res.* **1997**, *30*, 290-296.

(26) Schowen, K. B.; Limbach, H.-H.; Denisov, G. S.; Schowen, R. L. *Biochim. Biophys. Acta.* **2000,** *1458*, 43-62.

(27) Guo, H. B.; Wlodawer, A.; Guo, H. *J. Am. Chem. Soc.* **2005**, *127*, 15662-15663.

(28) Xu, Q.; Guo, H.; Wlodawer, A.; Guo, H. *J. Am. Chem. Soc.* **2006**, *128*, 5994-5995.

(29) Guo, H. B.; Wlodawer, A.; Nakayama, T.; Xu, Q.; Guo, H. *Biochem.* **2006**, *45*, 9129-9137.

(30) Xu, Q.; Guo, H. B.; Wlodawer, A.; Nakayama, T.; Guo, H. *Biochem.* **2007**, *46*, 3784-3792.

(31) Guo, H. B.; Rao, N.; Xu, Q.; Guo, H. *J. Am. Chem. Soc.* **2005**, *127*, 3191-3197.

(32) Xu, Q.; Guo, H. B.; Gorin, A.; Guo, H. *J. Phys. Chem. B* **2007**, *111*, 6501-6502.

(33) Rajamani, R; Naidoo, K. J.; Gao, J. *J. Comput. Chem.* **2003**, *24*, 1775-1781.

(34) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comp. Chem.* **1992**, *13*, 1011-1021.

(35) Roux B. *Comput. Phys. Comm*. **1995**, *91*, 275-282.

(36) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys*. **1977**, *23*, 327-341.

(37) Greco, O.; Dachs, G. U. *J. Cell. Physiol.* **2001**, *187*, 22-36.

(38) Aghi, M.; Hochberg, F.; Breakefield, X. O. *J. Gene Med.* **2000**, *2*, 148-164.

(39) Morris, S. M. *Mutat. Res*. **1993**, *297*, 39–51.

(40) Kievit, E.; Bershad, E.; Ng, E.; Sethna, P.; Dev, I.; Lawrence, T. S.; Rehemtulla, A. *Cancer Res*. **1999**, *59*, 1417–1421.

(41) Liu, Y.; Ye, T.; Maynard, J.; Akbulut, H.; Deisseroth, A. *Cancer Gene Ther.* **2006**, *13*, 346-356.

(42) Hwang, L. H. *J. Biomed. Science* **2006**, *13*, 453-468.

(43) Huber, B. E.; Austin, E. A.; Good, S. S.; Knick, V. C.; Tibbels, S.; Richards, C. A. *Cancer Res.* **1993**, *53*, 4619–4626.

(44) Hirschowitz, E. A.; Ohwada, A.; Pascal, W. R.; Russi, T. J.; Crystal, R. G. *Hum. Gene Ther*. **1995**, *6*, 1055–1063.

(45) Wlodawer, A.; Li, M.; Gustchina, A.; Oyama, H.; Dunn B. M.; Oda, K. *Acta Biochim. Pol*. **2003**, *50*, 81-102.

(46) Dunn, B. M. *Chem. Rev.* **2002**, *102*, 4431-4458.

(47) Dall'Acqua, W.; Carter, P. *Protein Sci.* **2000**, *9*, 1-9.

(48) Kosloff, M.; Selinger, Z. *Trends Biochem. Sci.* **2001**, *26*, 161-166.

(49) Carter, P.; Wells, J. A. *Science* **1987**, *237*, 394-399.

(50) Weinger, J. S.; Parnell, K. M.; Doner, S.; Green, R.; Strobel, S. A. *Nat. Struct. Mol. Biol*. **2004**, *11*, 1101-1106.

(51) Jogl, G.; Tong, L. *Cell*, **2003**, *112*, 113-122.

(52) Ryan, M.; Liu, T.; Dahlquist, F. W.; Griffith, O. H. *Biochem*., **2001**, *40*, 9743-9750.

(53) Breldenbach, M. A.; Brunger, A. T. *Nature*, **2004**, *432*, 925-929.

(54) Fromme, J. C.; Bruner, S. D.; Yang, W.; Karplus, M.; Verdine, G. L. *Nat. Struct. Mol. Biol*. **2003**, *10*, 204-210.

(55) Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997**, *389*, 251-260.

(56) Watson, J. D.; Baker, T. A.; Bell, S. P.; Gann, A.; Levine, M.; Losick, R. *Molecular Biology of the Gene*, 5th edition. CSHL Press, **2004**.

(57) Turner, B. M. *Nat. Struct. Mol. Biol.* **2005**, *12*, 110-112.

(58) Strahl, B. D.; Allis, C. D. *Nature*, **2000**, *403*, 41-45.

(59) Turner, B. M. *Cell* **2002**, *111*, 285-291.

(60) Lall, S. *Nature Struct. Mol. Biol.* **2007**, *14*, 1110-1115.

(61) Rea, S.; Eisenhaber, F.; O'Carroll, N.; Strahl, B. D.; Sun, Z. W.; Schmid, M.; Opravil, S.; Mechtler, K.; Ponting, C. P.; Allis, C. D.; Jenuwein, T. *Nature* **2000**, *406*, 593-599.

(62) Martin, C.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 838-849.

(63) Min, J.; Feng, Q.; Li, Z.; Zhang, Y.; Xu, R.M. *Cell* **2003**, *112*, 711–723.

(64) Sawada, K.; Yang, Z.; Horton, J. R.; Collins, R. E.; Zhang, X.; Cheng, X. *J. Biol. Chem.* **2004**, *279*, 43296–43306.

(65) Xiao, B.; Jing, C.; Wilson, J. R.; Walker, P. A.; Vasisht, N.; Kelly, G.; Howell, S.; Taylor, I. A.; Blackburn, G. M.; Gamblin, S. J. *Nature* **2003**, *421*, 652-656.

(66) Zhang, X.; Tamaru, H.; Khan, S. I.; Horton, J. R.; Keefe, L. J.; Selker, E. U.; Cheng, X. *Cell* **2002**, *111*, 117-127.

(67) Zhang, X.; Yang, Z.; Khan, S. I.; Horton, J. R.; Tamaru, H.; Selker, E. U.; Cheng, X. *Mol. Cell* **2003**, *12*, 177-185.

(68) Dirk, L. M.; Flynn, E. M.; Diwtzel, K.; Couture, J.; Trievel, R.; Houtz, R. L. *Biochem.* **2007**, *46*, 3905–3915.

Appendix

Table I-1. Comparison between the features of sedolisin and subtilisin

The members of sedolisin family and of subtilisin family share some similar features. Both of them have a catalytic triad, which is Ser-Glu-Asp in sedolisin and Ser-His-Asp in subtilisin. In sedolisin, an Asp may play an important role in the stabilization of the tetrahedral intermediate (TI), while the corresponding residue is an Asn in subtilisin. In addition, these residues have a similar structural fold in sedolisin and in subtilisin, suggesting a similar ping-pong mechanism in catalysis. On the other hand, the difference in the conserved residues between sedolisin and subtilisin may lead to their different optimum pH conditions. Sedolisin favors acidic condition, while subtilisin favors neutral or alkaline conditions.

	Sedolisin	Subtilisin
Catalytic triad	Ser-Glu-Asp	Ser-His-Asp
TI stabilization	Asp	Asn
pH optimum	Acidic	Neutral or alkaline

Table I-2. Substrate-assisted catalysis in subtilisin

P_2-His or P_1'-His can partially restore the activity of H64A mutant. Thus the broadly specificity of wild type subtilisin is converted into His specificity in the H64A mutant. (From Ref. 49)

Sub-strate P2 residue	Ser24Cys			Ser24Cys:His64Ala		
	k_{cat} (sec^{-1})	K_m (μM)	k_{cat}/K_m ($sec^{-1} M^{-1}$)	k_{cat} (sec^{-1})	K_m (μM)	k_{cat}/K_m ($sec^{-1} M^{-1}$)
Ala	8.1	10	8.0×10^5	8.1×10^{-6}	32	0.25
Gln	7.0	39	1.8×10^5	3.0×10^{-5}	150	0.20
His	4.6	23	2.0×10^5	1.6×10^{-2}	380	42

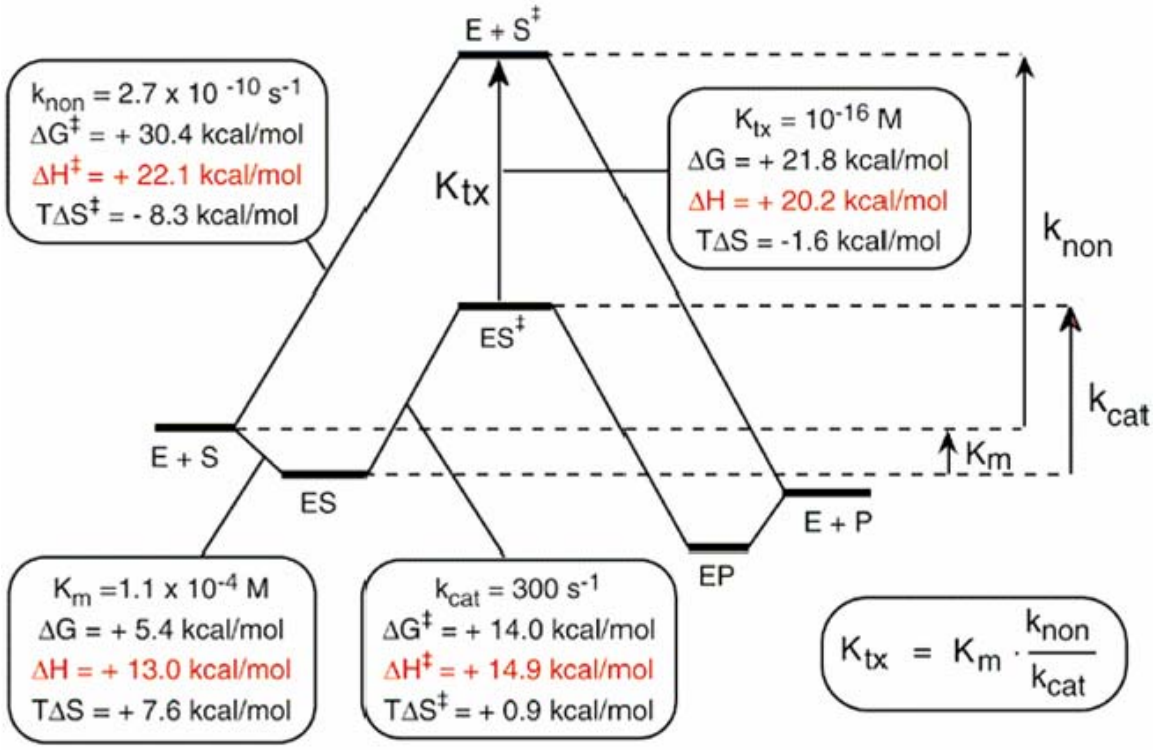


Figure I-1. Thermodynamic parameters of uncatalyzed and CDA-catalyzed cytidine deamination

It is generally believed that the efficiency of enzymes mainly lies in their abilities to stabilize the transition state and lower the activation barriers of the reactions. An example of CDA catalysis is shown here. The energetic barrier of catalyzed reaction was determined to be 16.4kcal/mol lower than that of uncatalyzed reaction at 25ºC. (From Ref. 32)

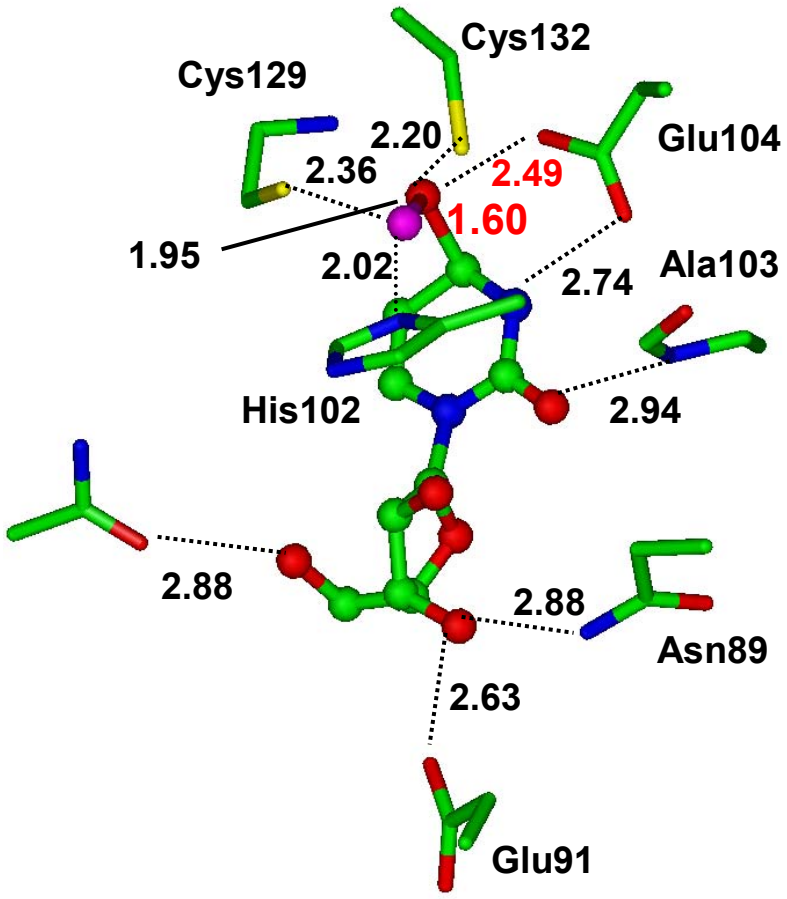


Figure I-2. The X-ray structure of CDA-ZEB-H_2O complex

The X-ray structure of CDA-ZEB-H_2O complex has an abnormally longer C_4-O_4 bond as 1.6Å and an unexpected low barrier hydrogen bond of 2.5Å between O_4 and $O_{\varepsilon 1}$ of Glu104, (shown in red) even with a restraint of 1.43 Å for C_4-O_4 bond and a restraint of 2.8Å for the $O_4 \ldots O_{\varepsilon 1}$ distance during the refinement.

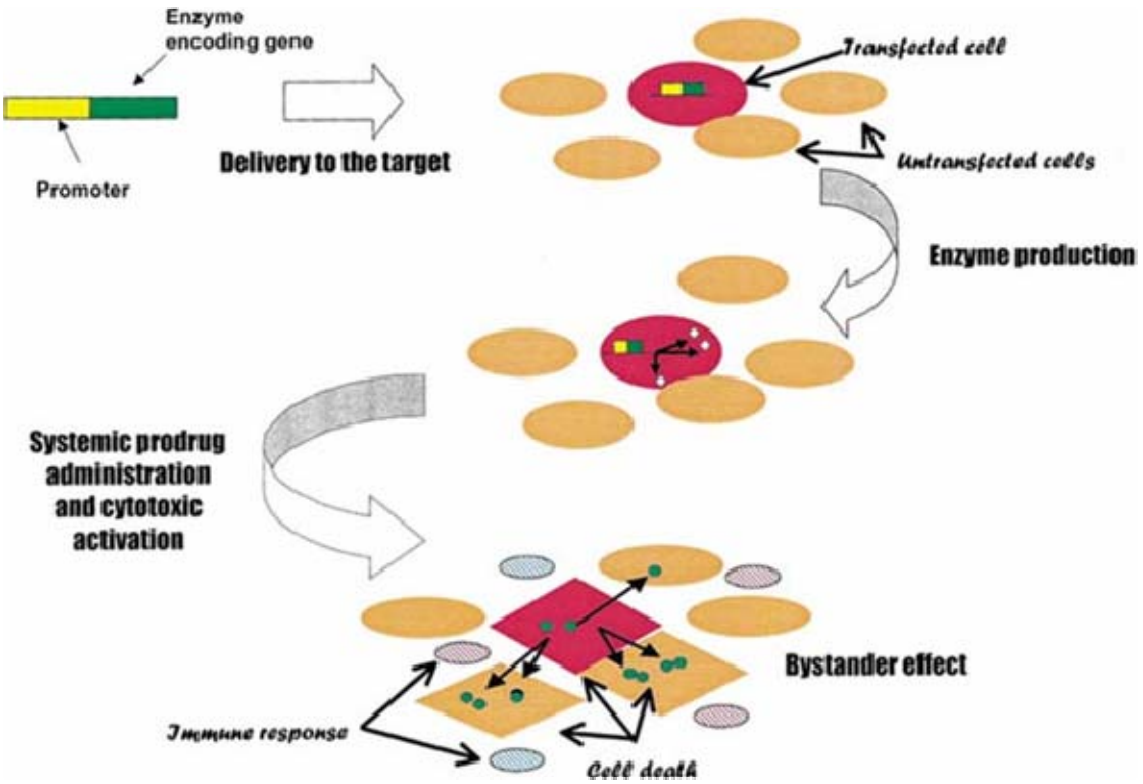


Figure I-3. Gene-directed enzyme prodrug therapy (GDEPT)

A DNA containing an enzyme-encoding gene is delivered to the tumour using viral or non-viral vectors. Only a fraction of the target cells will express the foreign gene and synthesize the enzyme. After systemic injection of the prodrug, cytotoxic activation takes place only at the cells with the enzymes. The bystander effect allows the activated drugs to diffuse into neighbouring untransfected cells and killed them. (), transfected cell; (), untransfected cell; (), enzyme; (), activated drug; (), dead cells; (), immune response. (From Ref. 37)

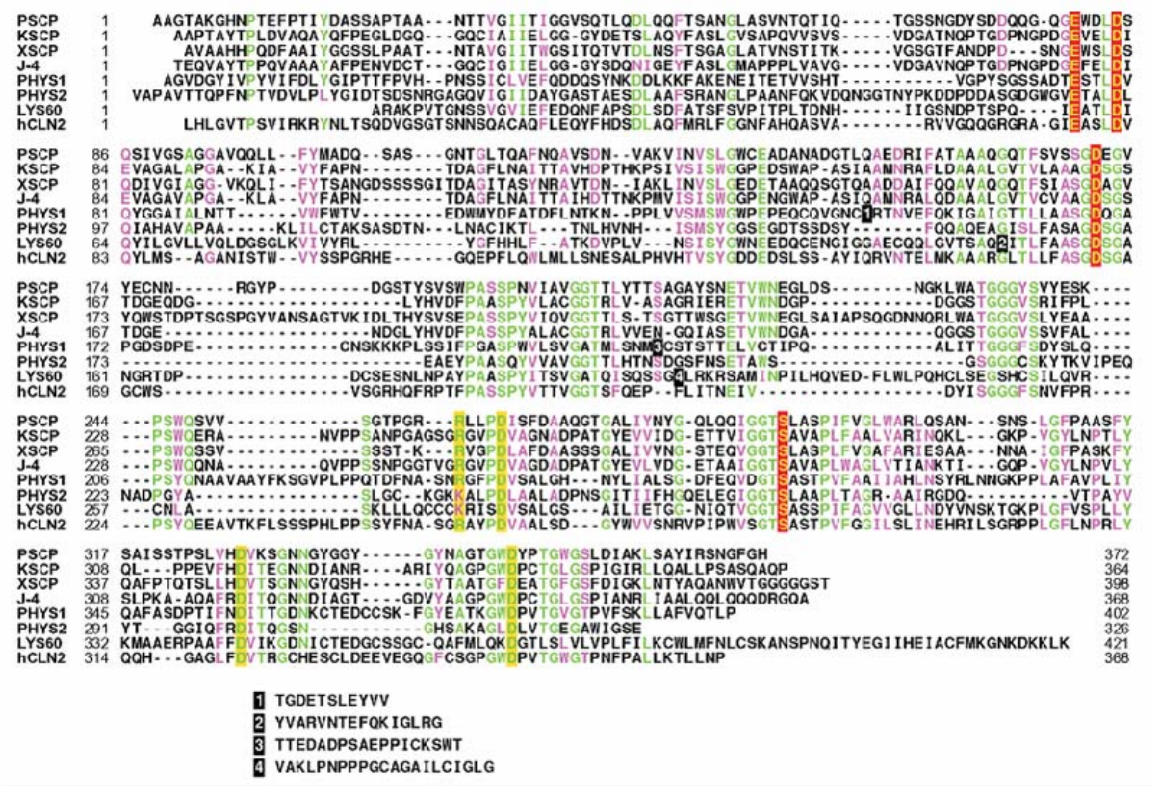


Figure I-4. Sequence alignment of sedolisins

The sequence alignment of some sedolisin members revealed four totally conserved residues, including the possible catalytic triad S-E-D (in the fourth and the first section) and the D (in the second section) that may contribute to the stabilization of tetrahedral intermediate, as marked in red. (From Ref. 45)

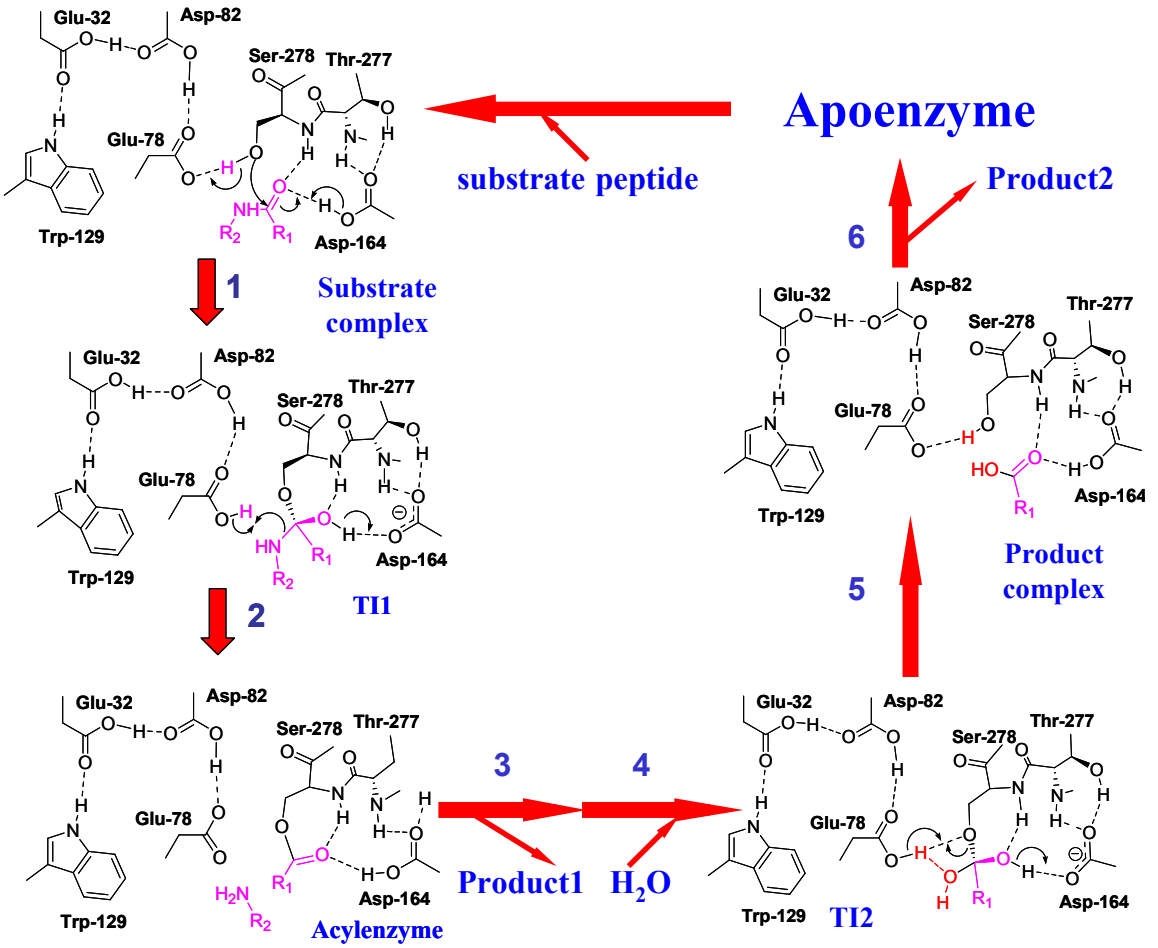


Figure I-5. Proposed catalytic mechanism of kumamolisin-As

Similar to the ping-pong mechanism in classical serine peptidase, the whole reaction circle of can be separated into two sections, acylation (left part) and deacylation (right). Each section has a different substrate bound at the beginning and a different product released at the end. The product release in each section is required to allow the binding of the substrate of the next section. During the acylation, the serine residue is the catalytic nucleophile to attack the substate peptide bond, while the nearby Glu acts as the general base to accept the proton from Ser at the time of the nucleophilic attack, and then act as the general acid to protonate the nitrogen forming the leaving N-terminal group. During the deacylation, this Glu still act as the general acid-base to accept a proton from the attacking water molecule and give it back to the deprotonated Ser of the acyl-enzyme. At the same time, the Asp may act as a general acid to protonate the oxygen of the tetrahedral intermediate (TI) and contribute to the TI statbilization.

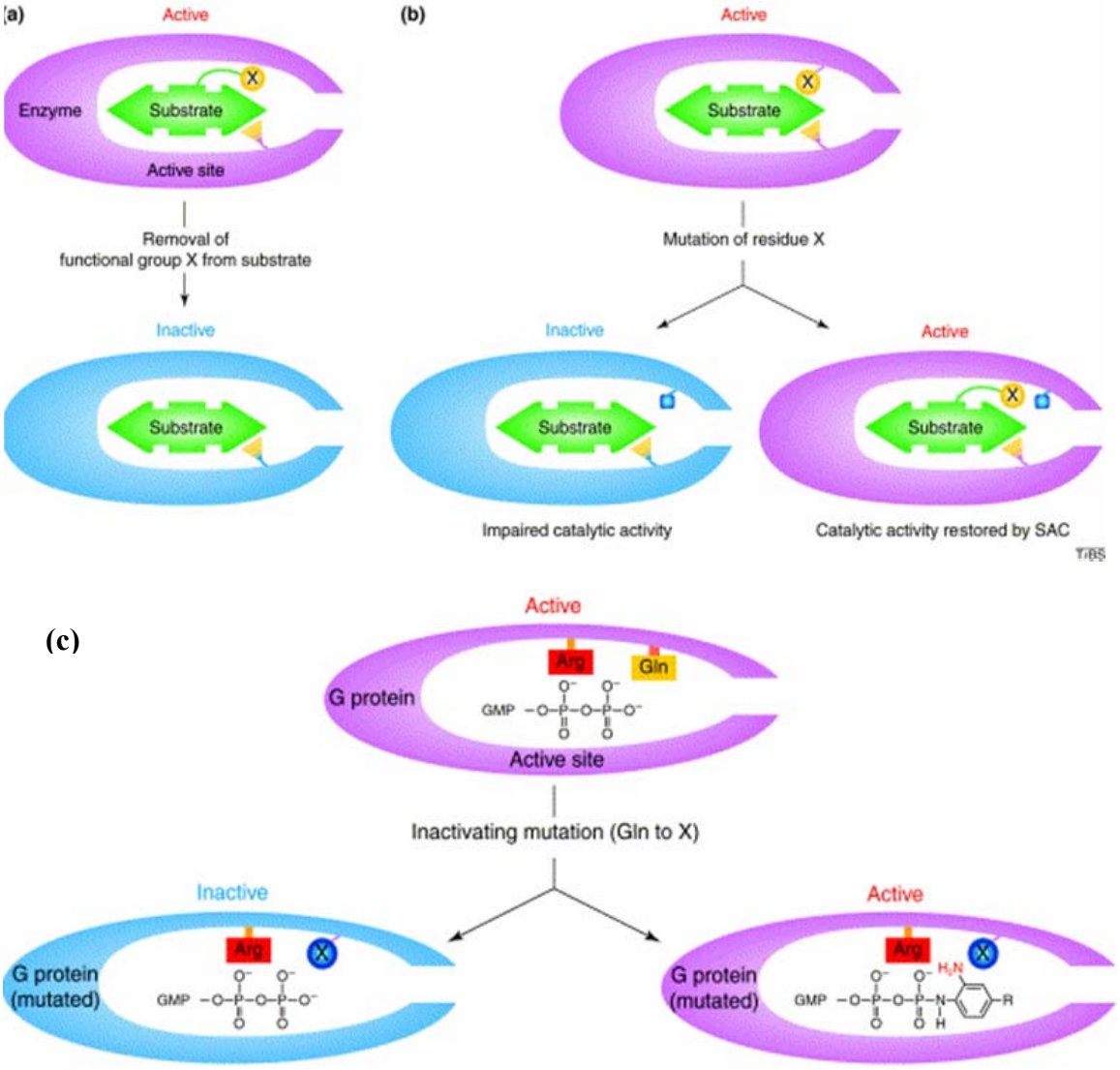


Figure I-6. Substrate-assisted catalysis

(a) SAC in natural enzyme: Removal of a functional group of the substrate may inactivate the enzymatic catalysis. (b) SAC in engineered enzymes: Addition on the subtrate with a functional group that similar to mutated catalytic group of the enzyme may at least partially restore the activity of the enzyme. (c) SAC in GTPase catalysis: Any mutation on the catalytic Gln resiude would inactivate the GTP hydrolysis by GTPase. However, a GTP analogue with an aromatic amine correctly positioned at the mutated active site can help to restore the catalytic activity of GTPase. (From Ref. 48)

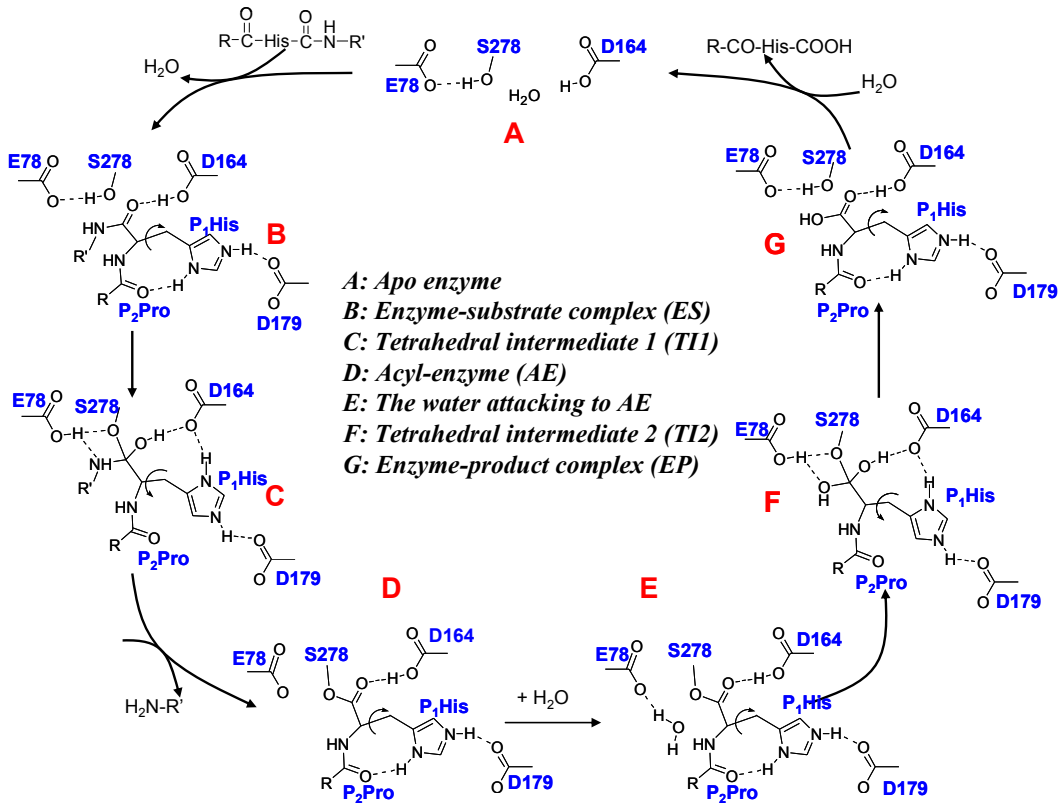


Figure I-7. Proposed dynamic substrate-assisted catalysis (DSAC) in Kumamolisin-As catalysis

The bond breaking and making events, especially the protonation and deprotonation of Asp164 may trigger the back and forth shifts of the P_1 His side chain between the hydrogen bond with the carbonyl of P_2 Pro (B, enzyme-substrate complex; D, E, acyl-enzyme; G, enzyme-product complex) and the salt bridge with the carboxylate of Asp164 (C, tetrahedral intermediate 1; F, tetrahedral intermediate 2), which may help to lower the free energy barrier for different step at different stage with different interactions.

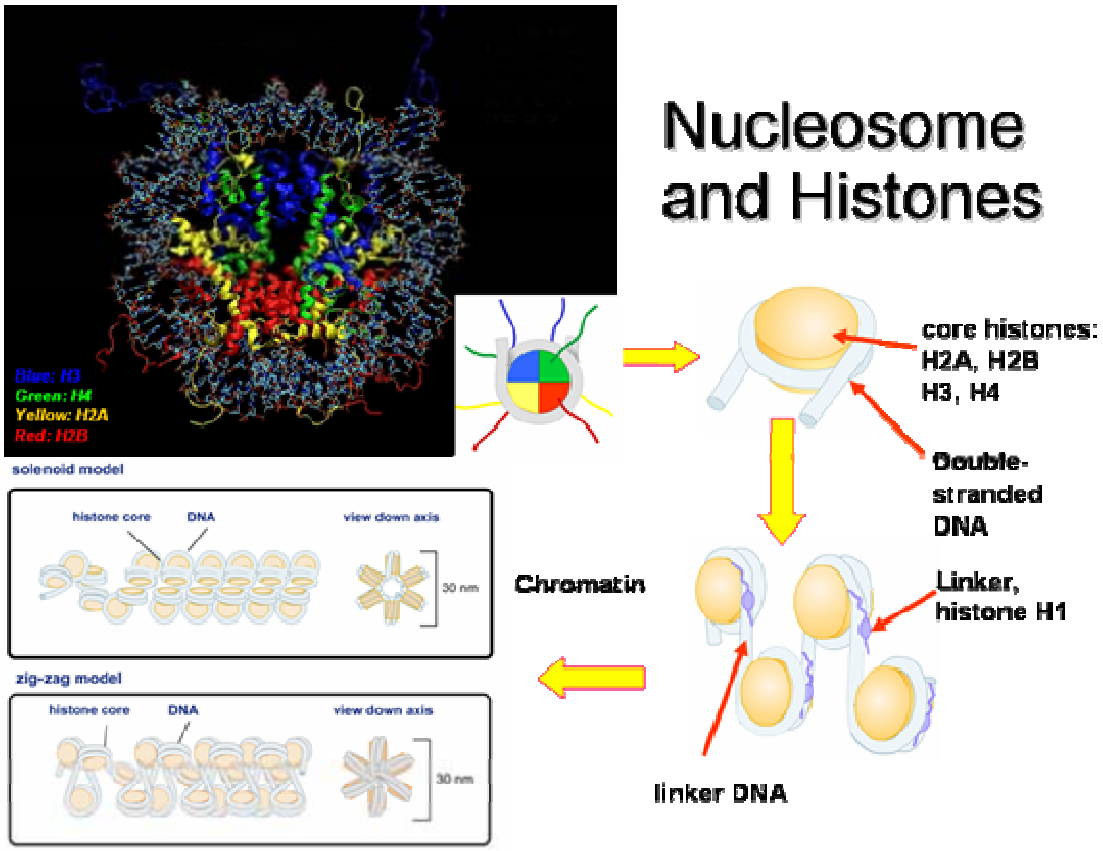


Figure I-8. Necleosome and histones

The nucleosome is the fundamental building block of eukaryotic chromatin, and histone proteins may play an important role in packaging DNA into nucleosome and futher into higher order chromatin structures. (Revised from Ref. 55, 56)

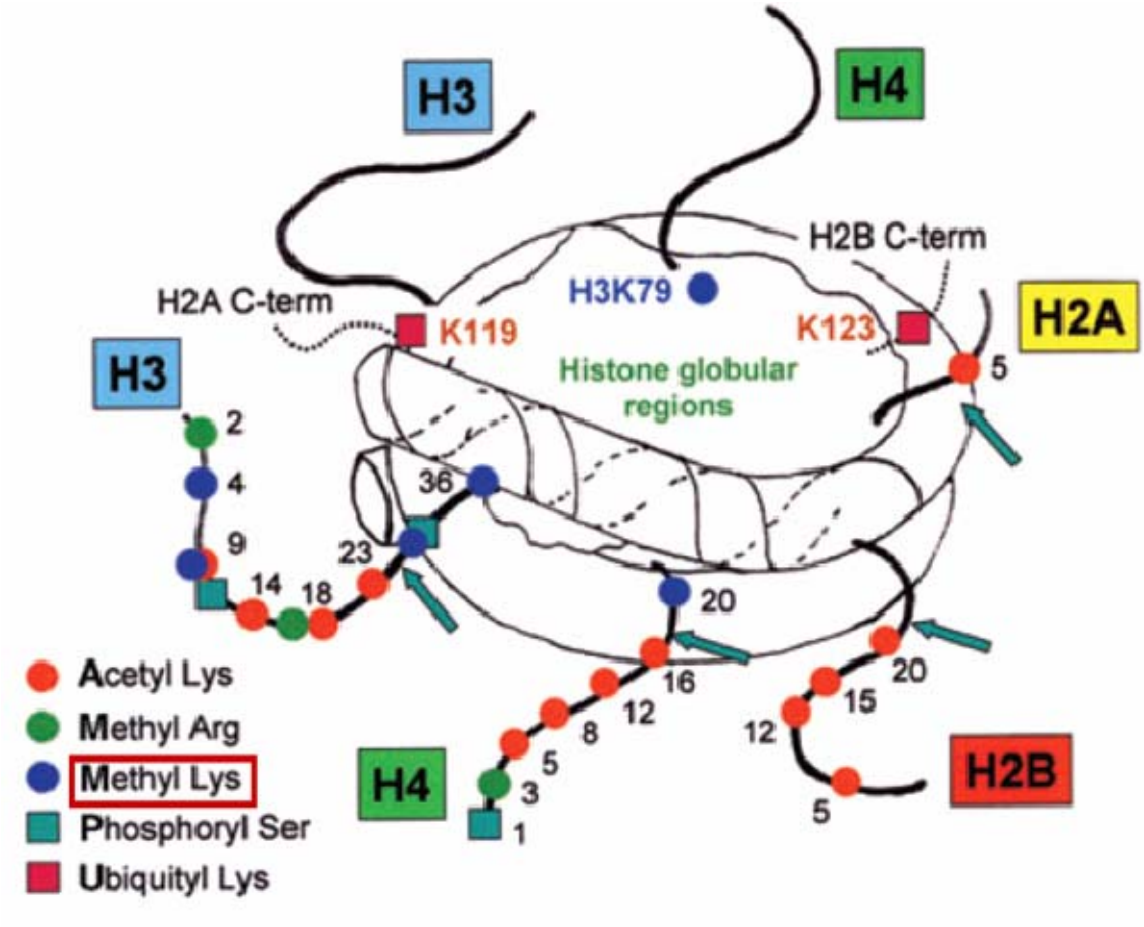


Figure I-9. Sites of post-translational modifications on histones

The tails of the histone proteins are subject to a variety of post-translational covalent modifications. The methylations on lysines are indicated as blue dots. (From Ref. 57)

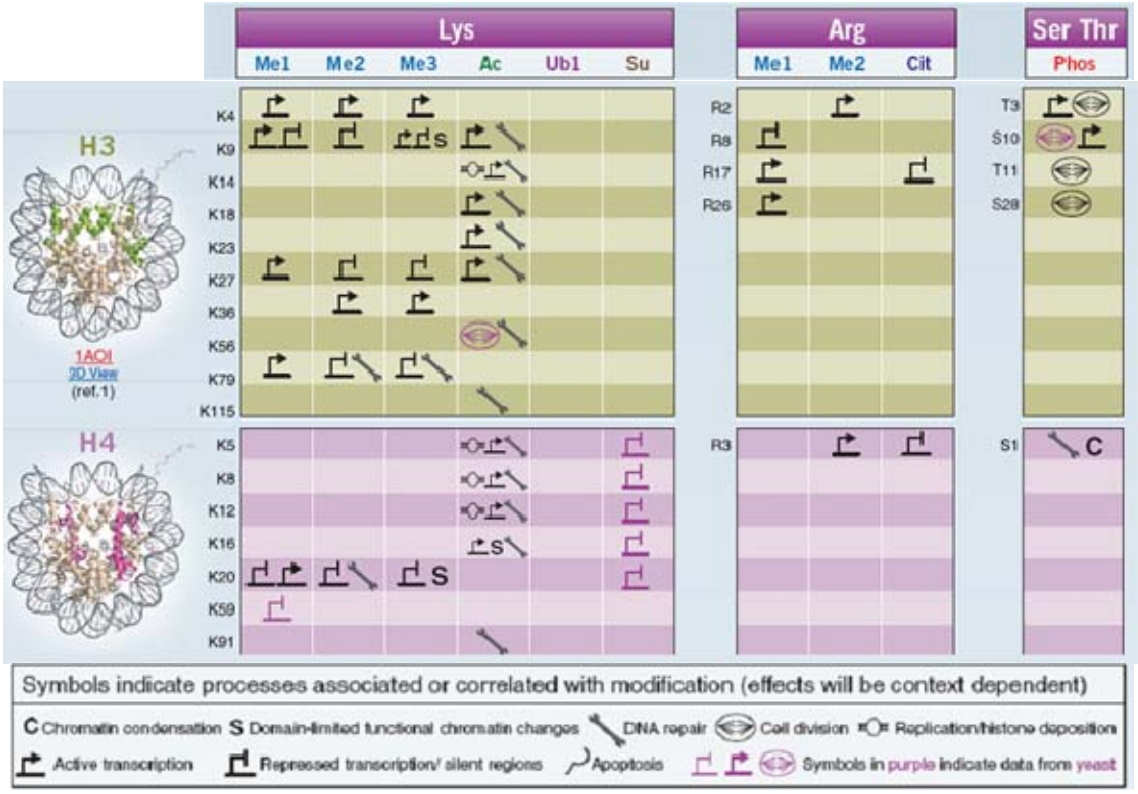


Figure I-10. Association of biological functions and histone modifications

The covalent post-translational modifications on histone tails may be associated with distinct biological downstream events in the regulation of chromatin structure and gene expression. Only the modifications on histone 3 and histone 4 are shown here. Under the list, the symbols of the biological functions are explained. The abbreviation of the associated or correlated covalent modifications on histones are: **Me**, methylation; **Ac**, acetylation; **Ub1**, monoubiquitination; **Su**, sumoylation; **Cit**, citrullination; **Phos**, phosphorylation. (Revised from Ref. 60)

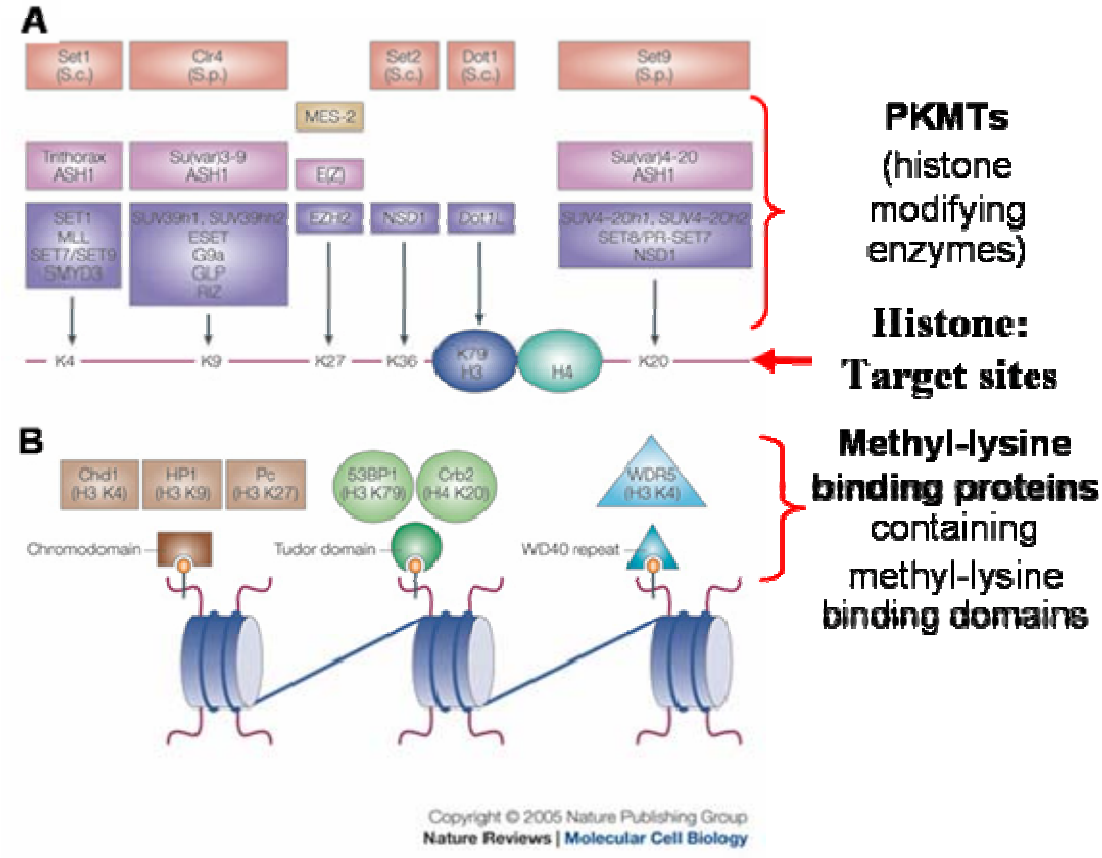


Figure I-11. Protein lysine methyltransferases (PKMTs) and methyl-lysine binding proteins

(A) The protein lysine methyltransferases are grouped based on their target sites (H3-K4, H3-K9, H3-K27, H3-K36, H3-K79 and H4-K20) and their source (yeast: red; worm: yellow; fly: pink; mammalian: purple). (B) The methyl-lysine binding proteins bind the methylated lysines through their methyl-lysine binding domains, such as the chromodomain (orange), the tudor domain (green) and the WD40-repeat (blue). (Revised from Ref. 62; more PKMTs, methyl-lysine binding domains and methyl-lysine binding proteins are summarized in Ref. 60)

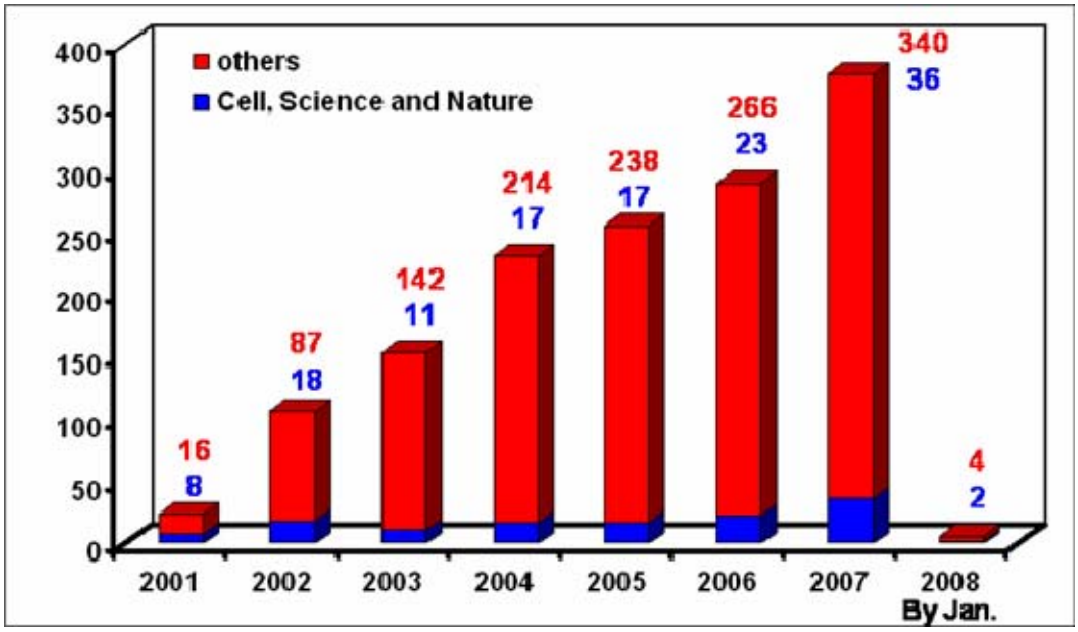


Figure I-12. Publications on the topic of "histone lysine methylation"

The numbers of publications on the topics related to the lysine methylation by PKMTs are quickly increasing recent years. The numbers in blue are those published in *Cell, Science and Nature*; the numbers in red are those published in other journals.

Chapter II

Methods

Quantum Mechanical (QM)/Molecular Mechanical (MM) Molecular Dynamics (MD) Simulations

The QM/MM hybrid potential combines the speediness of Molecular Mechanical (MM) potentials and the generality of Quantum Mechanical (QM) potentials and can be used to model enzymatic reactions involving bond breaking and making.

Currently, there are a number of force fields widely used for the Molecular Mechanical (MM) descriptions, such as AMBER (Assisted Model Building and Energy Refinement), CHARMM (Chemistry at HARvard Macromolecular Mechanics), GROMOS (GROningen MOlecular Simulation), and OPLS (Optimized Potential for Liquid Simulations).[2-6] The common advantage of these empirical approaches is their simplicity, which makes them applicable to the macroscopic systems like proteins, DNAs and other biological molecules. In our simulations of the catalytic mechanisms, the all-atom CHARMM22 force field[7] was used for the enzyme and the modified TIP3P model[8] implemented into the CHARMM program was used for the solvent water molecules. The parameters of this CHARMM force field treat all atoms explicitly.[7] In contrast, the older versions, such as the extended-atom force field CHARMM19, only treats polar hydrogens explicitly, while the nonpolar hydrogens were treated as part of extended atoms of the alkyl groups. The parameters of CHARMM22 were gradually refined to a force field that is well-balanced between the bonding and nonbonding interactions and among the solvent-solvent, solvent-solute, and solute-solute interactions. In order to simplify the calculation on the solvent molecules, the explicit water model TIP3P uses three point charges to describe water molecules and ignores the internal interactions.[8]

Solvent models can also include implicit solvent models, such as GBMV (Generalized Born using Molecular Volume) module[9, 10] of CHARMM, which looks as the solvents as a continuous medium instead of individual solvent molecules.

Quantum Mechanical (QM) potentials are critical for the simulations with permission of electron redistribution caused by the bond-breaking, bond-making, or other processes. There have been many *ab initio* approaches developed, such as Hartree-Fock (HF) theory, MP2 theory, and density functional theory (DFT) methods.[2] A widely used hybrid-functional version of DFT, B3LYP (Becke, three-parameter, Lee-Yang-Parr)[11] that implemented in Gaussian program was used in our lab for calculations on small model systems. However, the *ab initio* QM methods are generally very time-consuming. Consequently, many semi-empirical methods have been applied to the quantum mechanical calculations, such as MNDO (Modified Neglect of Diatomic Overlap), AM1 (Austin Model 1), PM3 (Parameterized Model 3), and SCC-DFTB (Self-Consistent Charge Density Functional Tight Binding) [12, 13]. In these semi-empirical approaches, the SCC-DFTB method is derived from density functional theory with a second-order expansion of the DFT total energy in the variations of charge density variation relative to a chosen reference.[13] This semi-empirical density-functional method is as fast as AM1 and PM3, but gives more reliable results of energies and vibration frequencies in a number of biological models, such as those of the triosephosphate isomerase (TIM) [14]. The energies determined by SCC-DFTB are comparable to the high-level DFT methods, such as B3LYP, for some systems with or without MM potentials. The application of SCC-DFTB/CHARMM method has also been applied to the study of enzyme chorismate

mutase and proved to be successful.[15, 16] Therefore, in our simulations we had this approach used in the quantum mechanical calculations, which generally include the substrate/ligand and the side chains of catalytic residues.

Construction of Simulation Models and the conditions in simulations

The X-ray structures of the enzymes studied were obtained from the Protein Data Bank.[33] The hydrogen coordinates were also built onto the x-ray structures by CHARMM program. In the construction of the model for kumamolisin-As-substrate complex, the structure of pro-kumamolisin (PDB ID: 1T1E) was superimposed onto the structure of kumamolisin-As-inhibitor complex (PDB ID: 1SIO) using the convenient program MOE[1]. With a 5 amino acid peptide of the linker fragment isolated from 1T1E as the substrate, and with the inhibitor deleted from 1SIO, the resulting system was revised and minimized in gas phase by MOE, to be the substrate complex model.

In order to apply QM/MM methods in our simulations, the simulation models are divided into QM part and MM part, for which SCC-DFTB and CHARMM force field are used in the calculations, respectively. For the hybrid QM/MM method one key question is how to deal with the interface between the QM part and the MM part, especially with covalent bonds between the two parts. Actually, in many cases it is unavoidable to have covalent bonds on QM/MM interface, such as the C_α-C_β bonds of the catalytic residues. To deal with such covalent bonds, several methods have been developed in earlier works, such as link-atom methods, capping-atom methods and hybrid-orbital methods. [2, 17-19] Adapted Link-atom method is the most widely used method for semi-empirical QM

functions. Although often criticized for addition of extra, unphysical atoms, this approach has the advantages of simplicity and convenience. In this method, generally imaginary hydrogen atoms are placed at about 1Å from the QM atom along the covalent bonds on the boundary. These link atoms are included in the QM part, acting as the MM atom on the other end of the covalent bond, so as to satisfy the valency of the QM atom in the QM calculations. On the other hand, these dummy atoms have no interactions with the MM atoms calculated.[34]

For example, in the CDA-ZEB-H_2O complex, it was believed that the interactions between TSA and Glu104 are only electrostatic interactions or low barrier hydrogen bond.[20] Thus in our first model, only the transition state analogue was included in the QM region, which only permits nonbonding interactions between the QM part and the MM part. The QM/MM MD simulations reproduced the structure consistent to the X-ray structure, including the abnormal longer C_4-O_4 bond as 1.6Å and the relatively shorter low barrier hydrogen bond between O_4…$O_{\varepsilon 1}$(Glu104) as 2.5Å. However, in this model the proton transfers between the TSA and the catalytic residue Glu104 are prohibited. Therefore, the QM part was expanded to the side chains of the catalytic residue using the link atom approach. The computational simulations on this new model resulted in much more stable structures before and after the nucleophilic attack. In the structure of the tetrahedral intermediate after the nucleophilic attack, a normal length of C_4-O_4 bond less than 1.5Å was observed; the proton of zinc-hydroxide was totally transferred to the carboxylate of Glu104, forming a normal hydrogen bond with O_4…$O_{\varepsilon 1}$(Glu104) distance

as 2.7Å. This more reasonable structure was estimated to be 8kcal/mol more stable than the model without permission for the proton transfers.[21]

Stochastic Boundary (SB) system[22] was used in our simulations of the catalytic mechanisms of enzymes, in order to further reduce the size of our simulations models. A typical SB system is shown in Figure II-1. Normally an atom close to the active center is selected as the reference point, which may also be the center of mass of the whole model or selected groups. The deleted part is called as the reservoir region, while the part left is called as the reaction zone. The reaction zone is further divided into the reaction region and the buffer region around it. The atoms in the buffer region were propagated by Langevin dynamics, in which the friction constants for the protein atoms can be calculated (e.g., 250ps^{-1} and for the water molecules as 62ps^{-1}). The atoms in the reaction region were propagated by molecular dynamics, and generally separated into QM part and MM part as described above.

Energy minimizations and dynamic equilibrations were then performed on the constructed simulation models. With QM part fixed, the constructed models are generally first minimized by 500 steps of steepest descent (SD) method[23], which can improve a very poor structure very fast but is relatively inefficient in convergency. The better structure is further minimized by 5000 to 10000 steps of conjugate gradient technique (CONJ) [24] or adapted basis Newton Raphson (ABNR) [4] method to search for a stable minimum. Following minimizations, the structure is simulated by dynamic equilibrations. In these simulations, the temperature of the system is gradually increased from 50K to 300K (or the most reactive temperature of the enzyme) and kept equilibrated at that

temperature for several hundred picoseconds to several nanoseconds. During the simulations, the SHAKE algorithm[25] is applied as constraint for the covalent bonds involving hydrogen atoms, so as to allow 1fs time step to be used for integration of the equations of motion.

Free Energy Change Determination

In order to determine the free energy change along a certain reaction coordinate (Potential of Mean Force), we used Umbrella Sampling methods[26, 27] and the weighted histogram analysis method (WHAM) program[28-30], which have been implemented into CHARMM.

The configurations far from the equilibrium, such as those around the transition states, are hard to be sampled in a normal simulation. The umbrella sampling methods can force the system to explore these regions of coordinate space by using a harmonic potential function. This harmonic function has been defined as

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

in the CHARMM package, where E is default to be 2, K is the force constant, r is the reaction coordinate selected for PMF, and r_0 is the origin of the harmonic function on the selected reaction coordinate. The selected reaction coordinate may be a critical structural parameter that changes in the chemical transformation (like distance, bond angle, or dihedral angle), as well as the combination of different variables, such as the difference between two distances $r = r(\text{C-N}) - r(\text{C}\ldots\text{O}_\gamma)$ in our simulations of the acylation of kumamolisin-AS. This harmonic function can force the system to explore specific regions

of the reaction coordinate, r_0, especially the transition state of the energetic barrier that is hard to be sampled without a bias in potential. With a series of r_0, generally in about 20 simulation windows in our researches, the system can be forced to explore the configurations along the selected reaction coordinate.

Based on the data of umbrella sampling, the WHAM program statistically translates the probability distributions of the configurations within a series of bins along the selected reaction coordinate into the relative free energy change as the function of this reaction coordinate.[28-31] Because the overlapping probability distributions between different simulation windows are critical to WHAM in the evaluation of the relative free energy change between different regions of the reaction coordinate, it is important to select probable 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 95% confidence interval in the free energy determined by WHAM was calculated based on the difference between the upper and lower limits.[32] According to this approach, the statistical error of the PMF is mainly related to the distribution probability in each bin. That is, the bins with higher distribution density have lower statistical errors. Generally the configurations around the transition state of the free energy barrier are hard to be sampled and results in lower quantity of snapshot frames in each bins, which makes the confidence intervals of the statistical error at this region tend to be larger. Therefore, more simulation windows or smaller intervals in r_0 around the transition state were adopted so as to lower the statistical error.

In addition to the statistical error in PMF determination, a systematic error might be possible from the semi-empirical SCC-DFTB method. It was suggested to compare the result of SCC-DFTB method with higher level *ab initio* or DFT approach on a model system in new applications so as to correct the energy.[14] The energy correction for our simulations were mainly based on comparison between the single point energies determined by SCC-DFTB and by B3LYP/6-31G(d,p) on simplified model systems, which were generally a part of or same as the QM region in the SCC-DFTB/CHARMM simulations. The single point configurations along the reaction coordinate were generated through a series of constrained iterative minimizations on the model system in gas phase using SCC-DFTB. When r_0 was changed stepwise from the reactant state to the product state, or reverse, a series of snapshot configurations could be sampled along the reaction coordinate. Then the single point potential energies of the generated structures were determined by SCC-DFTB and B3LYP/6-31G(d,p) separately, and the potential energy surfaces (PESs) along the reaction coordinate were compared. If possible, the SCC-DFTB energies were fit to the PES by B3LYP/6-31G(d,p) with certain corrections, which were then applied to the relative free energies obtained from the QM(SCC-DFTB)/MM umbrella sampling simulations.

Reference

(1) MOE, Chemical Computing Group Inc., 1010 Sherbrooke Street West, Suite 910, Montreal, Quebec, Canada H3A 2R7.

(2) Gao, J.; Truhlar, D. G. *Annu. ReV. Phys. Chem*. **2002**, *53*, 467-505.

(3) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999-2012.

(4) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.

(5) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657-1666.

(6) Schuler, L. D.; Daura, X.; Gunsteren, W. F. van. *J. Comput. Chem* **2001**, *22*, 1205-1218.

(7) Mackerell, A. D., Jr.; Bashford, D.; Bellott, M.; Bunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E. III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; W.-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586-3616.

(8) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. P. *J. Chem. Phys.* **1983**, *79*, 926-935.

(9) Lee, M. S.; Salsbury, F. R, Jr.; Brooks, C. L., III. *J. Chem. Phys.* **2002**, *116*, 10606-10614.

(10) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III. *J. Comput. Chem.* **2003**, *24*, 1348-56.

(11) Kim, K.; Jordan, K. D. *J. Phys. Chem.* **1994**, *98*, 10089-10094.

(12) Stewart, J. J. P. *J. Comput. Aided Mol. Des.* **1990**, *4*, 1-105.

(13) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260-7268.

(14) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M, *J. Phys. Chem. B* **2001**, *105*, 569-585.

(15) Guo, H.; Cui, Q.; Lipscomb, W.L.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9032-9037.

(16) Guo, H.; Cui, Q.; Lipscomb, W.L.; Karplus, M. *Angew. Chem., Int. Ed.* **2003**, *42*, 1508-1511.

(17) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720-1735.

(18) Field, M. J.; Albe, M.; Bret, C.; Martin, F. P.; Thomas, A. *J. Comput. Chem.* **2000**, *21*, 1088-1100.

(19) Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43-52.

(20) Xiang, S.; Short, S. A.; Wolfenden, R.; Carter, C. W., Jr. *Biochemistry* **1995**, *34*, 4516-4523.

(21) Guo, H. B.; Rao, N.; Xu, Q.; Guo, H. *J. Am. Chem. Soc.* **2005**, *127*, 3191-3197.

(22) Brooks C. L., III; Brunger, A.; Karplus, M. *Biopolymers* **1985**, *24*, 843-865.

(23) Arfken, G. The Method of Steepest Descents. §7.4 in *Mathematical Methods for Physicists*, 3rd ed. Orlando, FL: Academic Press, **1985**, 428-436.

(24) Flectcher, R.; Reeves, C. M. *Comput. J.* **1964**, *7*, 149-154.

(25) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys*. **1977**, *23*, 327-341.

(26) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578-581.

(27) Rajamani, R.; Naidoo, K. J.; Gao, J. *J. Comput. Chem.* **2003**, *24*, 1775-1781.

(28) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comp. Chem.* **1992**, *13*, 1011-1021.

(29) Roux, B. *Comput. Phys. Comm*. **1995**, *91*, 275-282.

(30) Grossfield, A. An implementation of WHAM: the Weighted Histogram Analysis Method. http://dasher.wustl.edu/alan/

(31) Boczko, E. M.; Brooks, C. L. *J. Phys. Chem*. **1993**, *97*, 4509-4513.

(32) Kobrak, M. N. *J. Comput. Chem.* **2003**, *24*, 1437-1446.

(33) PDB, RCSB Protein Data Bank, http://www.rcsb.org/pdb.

(34) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700-733.

Appendix

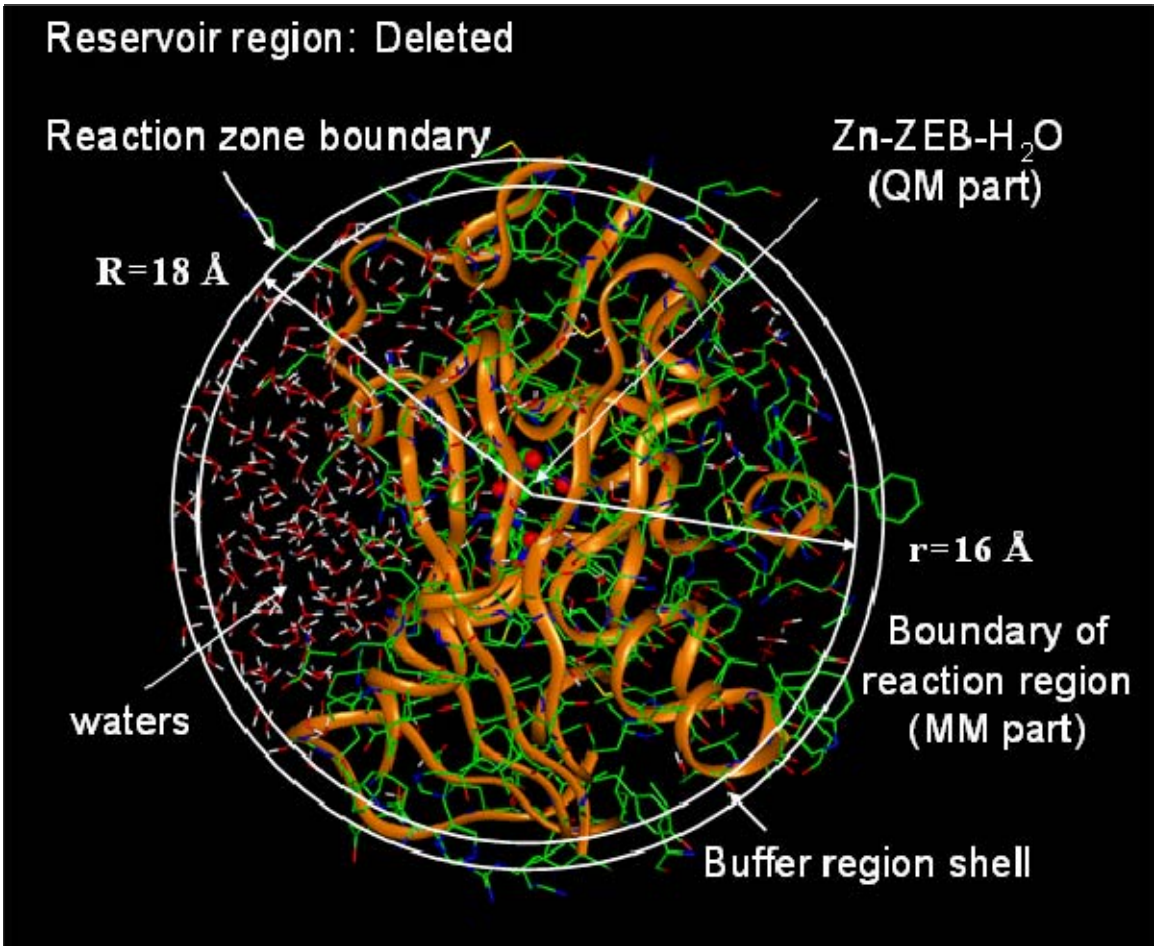


Figure II-1. Stochastic boundary system

A sample of stochastic boundary system for CDA-ZEB-H^2O model is shown. Using the N_1 atom of ZEB-H_2O as the reference center, the out most reservoir region is more than 18 Å away from the reactive site. It has quite less effect on the reaction so it is deleted. The region left is called reaction zone, which is further separated into the reaction region within 16Å and the buffer region between 16Å and 18Å. The atoms in the buffer region were restrained by Langevin dynamics. The atoms in the reaction region are divided into quantum mechanical part and molecular mechanical part, and treated by SCC-DFTB and CHARMM force field, respectively.

Chapter III

Quantum Mechanical/Molecular Mechanical Molecular-Dynamics Simulations of Cytidine Deaminase: from Stabilization of Transition-State Analogs to Catalytic Mechanisms

Qin Xu and Hong Guo

Revised from *J. Phys. Chem. B* **2004**, *108*, 2477-2483

Abstract

Cytidine deaminase catalyzes the hydrolytic deamination of cytidine to uridine and accelerates the rate of the reaction by 10^{11}-fold. The enzyme is strongly inhibited by the transition-state analog inhibitor zebularine 3, 4-hydrate, and a fraction of transition state stabilization is captured by the interaction of this inhibitor at the active site. QM/MM molecular dynamics and free energy simulations are performed for *E. coli* cytidine deaminase (CDA) complexed with zebularine 3, 4-hydrate to understand the origin of the structural stability of zebularine 3, 4-hydrate in the active site and elucidate the mechanism of the CDA-catalyzed reaction. It is shown that the existence of Glu-104 is essential for the structural stability and integrity of zebularine 3, 4-hydrate. The simulations also reveal that the covalent bond between C_4 and the 4-OH group in ZEB-H_2O undergoes transient bond-breaking and making on a picosecond scale in the active site, resembling the process of the nucleophilic attack by the zinc hydroxide group on C_4 during the catalysis. The role of the active-site interactions in stabilizing zebularine 3, 4-hydrate and transition state is discussed. The results demonstrate that understanding the stability and integrity of TSAs in the active sites may provide important insights into the origin of transition state stabilization.

Abbreviation:

CDA, cytidine deaminase; ZEB, zebularine, (Pyrimidin-2-one ribonucleotide); ZEB-H_2O, zebularine 3, 4-hydrate; TSA, transition state analog; TS, transition state; QM/MM, quantum mechanical/molecular mechanical; MD, molecular dynamics; SCC-DFTB, self-consistent charge density functional tight-binding.

Introduction

Enzymes are the most proficient catalysts. Their catalytic power lies in the enzyme's ability to stabilize transition state (TS) and to lower the energy barriers for enzyme-catalyzed reactions, although other factors may be involved as well. Understanding the forces responsible for transition state stabilization is of fundamental importance for a microscopic description of the catalytic mechanism, but also for the design of effective enzyme inhibitors or improvement of existing enzymes (or abzymes) for catalyzing chemical reactions. It has been recognized[1-3] that a molecule that mimics the altered structure of substrate in the transition state should bind to the enzyme tightly, and such a transition state analog (TSA) may capture a significant fraction of transition state stabilization. The enzyme-TSA complexes, where the inhibitors are designed based on possible catalytic mechanisms, can therefore be used as models for understanding the properties and structural features of transition states (which have a very short lifetime and are difficult to study). A number of the TSA inhibitors have been identified.[4-8] It has been shown that some of them can indeed bind to enzymes significantly stronger than the corresponding substrates, as originally hypothesized.[1] The precise arrangements of many TSA inhibitors in the enzyme active sites have also been determined by X-ray crystallography[9-10] which led to significant insights into the structural origins of enzyme catalysis as well as the possible roles of active-site interactions in transition state stabilization.

The ability of well-designed TSAs to mimic the transition state or intermediate structures and to capture a significant fraction of transition state affinity implies that the

structural stability and integrity of TSAs in the active sites may depend, at least in some of the cases, on the same forces that are responsible for transition state stabilization during the enzyme-catalyzed reactions. Removal of certain key residues crucial for the stability and integrity of the TSA inhibitors could also affect the transition state stabilization and lead to a reduced catalytic efficiency. Thus, determination of the conditions for existence of stable TSAs in the enzyme active sites should be of considerable interest, as the problem of understanding the origin of transition state stabilization might be simplified to a stability problem concerning TSAs in the enzymes' active sites. The large number of high resolution structures of enzyme-TSA complexes[9-10] can be used directly as the initial coordinates in computational studies; they may also serve as an important testing ground for theoretical predictions. Although all the questions concerning transition state stabilization and catalytic mechanisms are unlikely to be solved based on our understanding of the stability of TSAs in the active sites alone, important insights may be obtained from such studies.

Here Quantum mechanical/molecular mechanical (QM/MM) molecular dynamics simulations are performed for *E. coli* cytidine deaminase (CDA)[11] complexed with the transition state analog zebularine 3,4-hydrate (ZEB-H_2O) (Figure III-1B) to elucidate the origin of the stabilization of zebularine 3,4-hydrate by the active-site interactions. One advantage of the QM/MM treatments of the enzyme complex is that they allow the QM part of the system (e.g., ZEB-H_2O and zinc in this study) to undergo chemical transformations during molecular dynamics (MD) simulations in the search for stable chemical structure(s) under a given environment. The free energy approach can be used

to determine the free energy changes associated with chemical and conformational transformations. However, it is extremely time consuming to perform free energy calculations with *ab intio* methods, and semi-empirical approaches need to be used. There are different semi-empirical methods available. In this paper, we use a semi-empirical density functional method, SCC-DFTB,[25a] to describe the QM part, because this approach has been extensively tested against high level quantum mechanical methods for a range of active site models containing zinc.[25b] It was found that the SCC-DFTB approach reproduced structural and energetic properties rather reliably for the active sites of these enzymes.[25b] The SCC-DFTB/MM approach has been tested previously for chorismate and used to study the enzyme chorismate mutase.[28-29]

CDA catalyzes the hydrolytic deamination of cytidine to uridine (Figure III-1A) and has been a subject of extensive investigations.[11-24] It is one of the efficient enzymes known with a rate acceleration of 10^{11}-fold.[22] It has been shown that the deamination proceeds stepwise through a tetrahedral intermediate (formed by the nucleophilic attack by the zinc hydroxide group on C_4) with ammonia elimination as the major rate-determining step.[24] The catalysis depends critically on the presence of Glu104, and the mutation of Glu104 to Ala reduces k_{cat} by 10^8-fold.[14] The side chain of Glu104 seems to play multiple roles and is thought to provide all the necessary proton transfer functions during the catalysis (i.e., generating the zinc hydroxide nucleophile, protonating the pyramidine ring nitrogen and shuttling the proton from the 4-OH group to leaving amino group).[12] Glu104 is believed to be important for transition state stabilization, probably through the formation of a strong hydrogen bond with the 4-OH group of the tetrahedral

intermediate.[11] The observation that the mutation of Glu104 to Ala increases the enzyme's affinities toward substrate cytidine and product uridine by factors of 30 and 120, respectively, has led to the suggestion that Glu104 may be involved in the ground state destabilization as well.[13]

Zebularine (ZEB. See Figure III-1B) and 5-fluorozebularine (FZEB) bind CDA in hydrated forms (ZEB-H_2O and FZEB-H_2O, respectively),[11-12] and the hydration processes seem to be catalyzed by CDA by a mechanism similar to the formation of the tetrahedral intermediate during the CDA-catalyzed reaction.[10] Interestingly, the ^{19}F-NMR resonance[13] of 5-[^{19}F]FZEB showed that whereas the inhibitor is bound by wild-type enzyme as the 3, 4-hydrated species, it is bound by the E104A mutant without modification in a form that resembles the substrate in the ground state. The detailed structural origin for the different behavior is still not clear. CDA is strongly inhibited by ZEB-H_2O and FZEB-H_2O; e.g., the enzyme's affinity (K_i = 1.2 x 10^{-12}M) for ZEB- H_2O exceeds the affinity for the product uridine and the inhibitor 3, 4-dihydrozebularine (DHZ) (for which the 4-OH group of ZEB-H_2O is replaced by a hydrogen atom) by a factor of approximately 10^8.[23] Thus, the enzyme appears to have a very strong ability to stabilize the rare 3, 4-hydrated species in the active site that resemble the tetrahedral intermediate formed by the nucleophilic attack by the zinc hydroxide group on C_4.[12] In this chapter, the stability of ZEB-H_2O is studied, and it is shown that the presence of Glu104 is essential for the existence of zebularine 3, 4-hydrate in the CDA active site. The QM/MM molecular dynamics simulations demonstrate that whereas the covalent bond between C_4 and the 4-OH group in ZEB-H_2O exists in wild-type CDA, it is broken

in the E104A mutant followed by a proton transfer from the 3-NH group to the hydroxide. The simulations also show that, unlike conventional covalent bonds, the bond between C_4 and the 4-OH group in the wild-type CDA-ZEB-H_2O complex undergoes large fluctuations and breaks transiently during the simulations. The QM/MM MD simulations of the CDA-ZEB-H_2O complex seem to be able, not only to capture the structural features of the CDA-ZEB-H_2O complex, but also to provide some important insights into the dynamic events associated with the nucleophilic attack by the zinc hydroxide group on C_4 during the catalysis.

Methods

A fast semi-empirical density-functional approach (SCC-DFTB),[25] recently implemented in the CHARMM program,[27] was used for QM/MM molecular-dynamics simulations.[26] The SCC-DFTB approach has been extensively tested against high level quantum mechanical methods for a range of active site models containing zinc.[25b] It was found that the SCC-DFTB approach reproduced structural and energetic properties rather reliably.[25b] The QM(SCC-DFTB)/MM molecular-dynamics simulations have been used recently to study the conformational transitions of substrate in chorismate mutase.[28-29] The initial coordinates for the simulations were obtained from the crystal structures of the CDA-TSA complex (1CTU),[11] which have the transition state analog ZEB-H_2O at the active site. It has been suggested[19] that there might be a trapped water molecule in the CDA-ZEB-H_2O complex occupying the site of the leaving group (-NH_2 group) of the altered substrate in the transition state. Therefore, a water molecule was manually docked

into this site. The MD simulations were performed on the CDA-ZEB-H_2O complexes with and without this additional water, and the results are rather similar. ZEB-H_2O and zinc were treated by QM and the rest of the system by MM. To study the effects of the interactions involving Glu104 on the stability and integrity of ZEB-H_2O, Glu104 was simply replaced by Ala based on the X-ray structure of the CDA-ZEB-H_2O complex; the structure for the E104A mutant is not available. The all-hydrogen potential function (PARAM22)[30] was used for MM atoms. A modified TIP3P water model[31-32] was employed for the solvent. The stochastic boundary molecular-dynamics method[33] was used for the QM/MM MD simulations. The system was separated into a reaction zone and a reservoir region, which was deleted; the reaction zone was further divided into the reaction region and the buffer region. The reference point for partitioning the system was chosen as the N_1 atom of ZEB-H_2O (Figure III-1B). The reaction region was a sphere with radius 16 Å, and the buffer region has R equal to 16 Å $\leq$ R $\leq$ 18 Å. Inside the reaction region, the atoms were propagated by molecular dynamics, whereas atoms in the buffer region were propagated by Langevin dynamics. The friction constants for the Langevin dynamics were 250ps^{-1} for the protein atoms and 62ps^{-1} for the water molecules. A 1fs time step was used for integration of the equations of motion, and the results of the simulations were saved on every 50fs. In initiating the runs, 500 steps of minimization using the steepest descent method were performed for the protein and solvent atoms. Then 4,000 steps of minimization were performed for the entire stochastic boundary system with the Adapted Basis Newton Raphson methods. The temperature of the system was gradually increased from 50K to 300K (30ps) and equilibrated at 300K

(40ps). The simulations were performed on the resulting systems involving ZEB-H_2O for several hundred picoseconds to 1ns. The umbrella sampling method[34] implemented in the CHARMM program along with the Weighted Histogram Analysis Method (WHAM) [35] was applied to determine the change of the free energy (potential of mean force) as a function of the distance (from 1.45Å to 3.1Å) between C_4 and O_4 in ZEB-H_2O; harmonic umbrella potentials with a force constant of 100kcal/mol-Å^2 were used. Twelve windows were used to go from 1.45Å to 3.10Å. The trajectories obtained from the umbrella sampling were used to study how the active site interactions change as functions of the C_4-O_4 distance. That is, for each window in the umbrella sampling, the average distances for the $O_{\varepsilon 2}$...H_3, $O_{\varepsilon 1}$...H (O_4) and Zn...O_4 interactions as well as for the C_4-O_4 bond were obtained from the corresponding trajectory. The average distances from different trajectories were then plotted as functions of the C_4-O_4 distance. The umbrella sampling method was also used to determine the relative stability of the structures before and after the dehydration of ZEB-H_2O in E104A; the difference between r_1 (O_4...H_3) and r_2 (N_3-H_3) was used as the reaction coordinate with a force constant of 100kcal•mol^{-1} •radian^{-2}.

Results

ZEB-H_2O in wild-type CDA

An important test for the methods of the QM/MM MD simulations is to see if the structures generated from the simulations agree with the structure determined by X-ray crystallography. In Figure III-2A, an average active-site structure based on 2,000 frames in the trajectory is compared with that of the X-ray structure.[11] As can be seen from

Figure III-2A, the active site structure of the CDA-ZEB-H_2O complex from the QM/MM simulations is rather close to the X-ray structure, [11] suggesting the computational approaches used in this study are meaningful. In particular, the 4-OH group of ZEB-H_2O is coordinated to the zinc ion and hydrogen bonded to $O_{\varepsilon 1}$ of Glu104. The C_4-O_4 bond has an average distance of about 1.6Å. $O_{\varepsilon 2}$ of Glu104 receives a hydrogen bond from the 3-NH group. The additional ligands to zinc include Cys129, Cys132 and His102. The distances for some of the interactions between the QM (e.g., zinc) and MM atoms (e.g., the atoms from Cys129, Cys132 and His102) seem to be slightly longer than those in the X-ray structure, but the differences are rather small.

The fluctuation of the C_4-O_4 distance of ZEB-H_2O as a function of time is plotted in Figure III-2B; the fluctuation of the C_3'-O_3' distance is also given for comparison. Figure III-2B shows that the C_4-O_4 bond is essentially intact with an average distance of ~1.57 Å. This average distance is somewhat longer than the distance for a normal C-O bond (e.g., for the C_3'-O_3' bond of ZEB-H_2O, the average distance is about 1.45 Å). A relatively long C_4-O_4 distance (1.6 Å) was also obtained in the X-ray structure refinement of the CDA-ZEB-H_2O complex, [11] even though a restraint of 1.43Å was imposed for this bond. The result of the QM/MM MD simulations is therefore consistent with the experimental assignment. It is interesting to note that the distance fluctuations for the C_4-O_4 bond are rather large. A striking observation from Figure III-2B is that the C_4-O_4 bond undergoes transient bond breaking and making on a picoseconds scale during the simulations; the bond distance approaches or exceeds 2.0Å about 2-3% of the times. Figure III-2C shows the free energy (potential of mean force) as a function of the C_4-O_4

distance. Consistent with the earlier observation, there is a second minimum near r(C_4-O_4) = 2.1Å, corresponding to the transient structures with the relatively long C_4-O_4 distances. The existence of two minima indicates that there is equilibrium between two different types of the structures for ZEB-H_2O in the active site.

The changes of the $O_{\varepsilon 2}$ (Glu104)...H_3 and $O_{\varepsilon 1}$ (Glu104)...H(O_4) hydrogen bond distances as functions of time are shown in Figure III-2D. The role of these two hydrogen bonds has been a subject of previous discussions. It was proposed[11-14] that the interactions involving Glu104 play an important role in the transition state stabilization, probably, through the formation of a strong hydrogen bond between the carboxylate and 4-OH groups. As can be seen from Figure III-2D, the two hydrogen bonds involving Glu104 are rather stable during the simulations. The average distance for the $O_{\varepsilon 1}$...H(O_4) hydrogen bond (1.6Å) is shorter than that for the $O_{\varepsilon 2}$...H_3(N_3) hydrogen bond (1.75Å), consistent with the X-ray structure.[11] It would be of interest to study how the interactions between ZEB-H_2O and CDA change as the C_4-O_4 distance decreases, in a process that mimics the nucleophilic attack by the zinc hydroxide group on C_4 during the CDA-catalyzed reaction. In Figure III-2E, the average distances for the $O_{\varepsilon 2}$...H_3, $O_{\varepsilon 1}$...H(O_4) and Zn...O_4 interactions obtained from the trajectories in the umbrella sampling (See Methods and Figure III-2C above) are given as functions of the C_4-O_4 distance. As is evident from Figure III-2E, the $O_{\varepsilon 1}$...H(O_4) interaction is very sensitive to the C_4-O_4 distance. When the C_4-O_4 bond is broken (e.g., with a distance greater than 2.1Å), the $O_{\varepsilon 1}$...H(O_4) interaction is rather weak with an average distance > 2.2 Å; the 4-OH group is on the top of the carboxylate plane and does not have the angles for favorable hydrogen bonding

interactions with Glu104 (not shown). When the C_4-O_4 bond is formed (i.e., with a distance of 1.5-1.7Å), $O_{\varepsilon 1}$ of Glu104 and 4-OH form a strong hydrogen bond with an average hydrogen bond distance of l.5-1.7Å. The $O_{\varepsilon 2}...H_3$ hydrogen bond also strengthens with the formation of the C_4-O_4 bond, but the change is much smaller. It is interesting to note from Figure III-2E that the distance between O_4 and Zn increases (from 1.81Å to 1.95Å) with decreasing the C_4-O_4 distance (from 2.3Å to 1.5Å). The average distance of 1.95Å between O_4 and Zn is the same as that of the X-ray structure[11] for the CDA-ZEB-H_2O complex, whereas the distance of 1.81Å is similar to the distance (1.80Å) in the CDA-DHZ complex.[11] The distances for the interactions of Zn with Cys129, Cys132 and His102 were also examined. It was found that these distances increase by about 0.05Å with the decrease of the C_4-O_4 distance. These changes are smaller than that for the Zn-O_4 distance (0.15Å), and quantum mechanical description of the enzyme side chains may be necessary to obtain more reliable results.[11]

It is believed that the pyramidine ring nitrogen at position 3 accepts a proton from Glu104 during the hydration of ZEB and the CDA-catalyzed reaction.[12] In order to see the effect of this protonation on the nucleophilic attack by the zinc hydroxide, the simulations were performed on the complex where the proton was placed on $O_{\varepsilon 2}$ rather than 3-N. It was found that the C_4-O_4 bond was broken during the energy minimization, leading to ZEB and the zinc hydroxide group. The zinc hydroxide group was unable to approach C_4 with a distance of less than 2Å during the MD simulations (data not shown); the average C_4-O_4 distance is 2.2Å. Therefore, the results of the simulations suggest that the protonation of 3-N would make the nucleophilic attack much easier.

ZEB-H_2O in E104A

To obtain additional information concerning the role of Glu104 in stabilizing ZEB-H_2O, Glu104 was replaced by Ala. Figure III-3A plots the active site structures representing three different stages observed during the MD simulations; only the interactions involving the pyramidine ring are shown. The structure on the left represents the early stage of the simulations. As can be seen from this structure, the C_4-O_4 bond is broken and the 4-OH group exists as a zinc hydroxide ion. The figure in the middle shows the active structure observed in the next stage of the simulations. The zinc activated hydroxide oxygen has moved significantly and is now in the position to interact with the 3-NH group. The figure on the right shows the active site structure in the final stage; this structure was found to be stable during the rest of the simulations. As can be seen from this figure, the hydroxide ion has already picked up the proton from N_3 and moved back to the initial position (i.e., on the top of C_4), leading to the dehydration of ZEB-H_2O into zebularine and a water molecule. The umbrella sampling method was used to determine the free energy change associated with the process given in Figure III-3A. It was found that the free energy decreases by about 20kcal/mol as the proton transfers from 3-N to the hydroxide ion. Thus, the results of the simulations suggest that the hydroxide ion is unstable without the presence of Glu104. In Figure III-3B, the O_4-H_3, N_3-H_3 and C_4-O_4 distances are used to monitor the process shown in Figure III-3A. As can be seen from Figure III-3B, the O_4-H_3 distance (red line) first changes from a distance of 3Å to 1.5Å so that O_4 can interact with 3-NH. It then changes to 1Å, indicating that the

covalent bond is formed between O_4 and H_3. At the same time the N_3-H_3 bond (blue line) is broken.

Fig. 3C shows the free energy (potential of mean force) as a function of the C_4-O_4 distance in the E104A mutant (i.e., before the proton transfer discussed earlier); the free energy curve was obtained in the same way as the one given in Figure III-2C except that Glu104 was replaced by Ala. Consistent with the earlier observation, the C_4-O_4 bond becomes unstable and cannot be formed in the mutant. It is interesting to note from Figure III-3C that the mutation of Glu104 to Ala shifted the minimum on the free energy curve to the other location at $r(C_4\text{-}O_4) \sim 2.1$Å identified earlier in the wild type enzyme.

Discussions

One of the key events in the CDA-catalyzed reaction is the nucleophilic attack by the zinc hydroxide group on C_4, leading to the formation of the hypothetical tetrahedral intermediate (Figure III-1A). Thus, an important question is how CDA would be able to stabilize this intermediate state and to lower the energy barrier for the reaction. The ability of zebularine 3, 4-hydrate (ZEB-H_2O) to mimic the intermediate structure and to capture a fraction of the affinity for the transition state implies that the forces responsible for the structural stability and integrity of ZEB-H_2O in the active site may also be responsible for the transition state stabilization. Thus, understanding how the enzyme is able to stabilize this rare 3, 4-hydrated species in the active site may provide important insights into the mechanism of the CDA-catalyzed reaction. The results of the QM/MM MD simulations reported here demonstrate that ZEB- H_2O is generally stable in the

active site of the wild-type enzyme with an average structure close to the one observed experimentally (Figure III-2A). However, its C_4-O_4 bond undergoes transient bond-breaking and making (Figure III-2B); the bond-making process resembles the nucleophilic attack by the hydroxide group on C_4 of ZEB in the hydration process. Thus, the dynamic behavior of the C_4-O_4 bond seems to be quite different from conventional covalent bonds (compare, for example, the fluctuations of the C_4-O_4 and C_3'O_3' bonds in Figure III-2B). It would be of interest to confirm this prediction experimentally and examine if this interesting dynamic behavior exists in other enzyme-TSA complexes. The observation from the X-ray crystallography[11] that this bond is unusually long in the TSA complex is consistent with the simulation results (See above). However, it does not provide a direct support for the dynamic behavior. The umbrella sampling simulations showed that there is equilibrium between two different types of the structures in the active site with different C_4-O_4 distances (Figure III-2C). These two types of the structures correspond to the stages before and after the nucleophilic attack, respectively. Because the hydration of ZEB seems to be catalyzed by CDA by a mechanism similar to the formation of the tetrahedral intermediate, the process of the nucleophilic attack by the hydroxide group on C_4 during the catalysis may have the similar dynamic behavior.

The question is how the active site interactions are able to shift the equilibrium and lead to the formation of zebularine 3, 4-hydrate and, presumably, the tetrahedral intermediate as well during the catalysis. Examination of the interactions as functions of the C_4-O_4 distance revealed that the interaction that undergoes the most significant change during the formation of the C_4-O_4 bond is the hydrogen bond between

$O_{\varepsilon 1}$(Glu104) and H(O_4) (See Figure III-2E). In particular, $O_{\varepsilon 1}$ and the 4-OH group can form a strong hydrogen bond only when the C_4-O_4 distance is short (l.5-1.7Å). The corresponding differential hydrogen bonding energy should be able to make an important contribution to the stability of ZEB-H_2O in the active site and, presumably, to transition state stabilization as well. The role of the $O_{\varepsilon 1}$(Glu104)...H(O_4) hydrogen bond in transition state stabilization has been a subject of previous discussions. It was shown that the hydrogen bond distance for this interaction in the CDA-ZEB complex is 0.2Å shorter than the corresponding distance in the CDA-DHZ complex (i.e., between $O_{\varepsilon 1}$ and a trapped water). This observation has led to the suggestion that the strength of this hydrogen bond may increase with the development of the C_4-O_4 covalent character and the energetic effect would be used for the transition state stabilization. The results of this study support this previous suggestion and provide direct computational evidences that the nucleophilic attack by the zinc hydroxide group on C_4 leads to the strengthening of the $O_{\varepsilon 1}$(Glu104)...H(O_4) hydrogen bond.

It was suggested that one of the reasons for the increase of the strength of the $O_{\varepsilon 1}$(Glu104)...H(O_4) hydrogen bond is that the formation of the C_4-O_4 bond converts the zinc-bound oxygen into a species more closely resembling an alkoxide group than a hydroxyl group. As a result, the pK_a of the 4-OH proton would be depressed toward to that of the carboxylate, leading to a stronger interaction between these two groups. Other factors may be involved as well. For instance, the formation of the C_4-O_4 bond leads to smaller fluctuations for the OH group, making its interaction with $O_{\varepsilon 1}$ more stable. Moreover, the 3-NH and 4-OH groups of ZEB-H_2O (i.e., with an unbroken C_4-O_4 bond)

seem to be in ideal relative positions to match OE_1 and OE_2 of the carboxylate group from Glu104 (See Figure III-2A). The simultaneous formations of these two strong hydrogen bonds may help to hold the two fragments of ZEB-H_2O together and prevent the C_4-O_4 bond from being broken.

The results discussed above indicate that the interactions involving Glu104 are essential for stabilizing ZEB-H_2O in the active site of the wild-type enzyme. The QM/MM MD simulations on the E104A-ZEB-H_2O complex provide additional insights concerning the role of Glu104. It was shown that when Glu104 was replaced by Ala, the covalent bond between C_4 and the 4-OH group in ZEB-H_2O was broken followed by a proton transfer from the 3-NH group to the hydroxide, leading to the dehydration of ZEB-H_2O into ZEB and a water molecule (Figure III-3A, B). Thus, ZEB-H_2O is likely to be rather unstable in the E104A mutant. It has been observed from ^{19}F-NMR resonance[13] that whereas the wild-type enzyme binds FZEB (whose structure is very close to that of ZEB) as the 3, 4-hydrated species (FZEB-H_2O), the E104A mutant binds the inhibitor without modification in a form that resembles the substrate in the ground state. Our results are therefore consistent with the experimental observations. Moreover, the simulations suggest that the active site water may not be in split form in E104A. This is in contrast with the case for the wild type enzyme where the water is believed to be in split form with the hydroxide ion on zinc and proton on Glu104.[19-20] The different binding modes might contribute to the loss of the activity as a result of the mutation; [14] they could also be related to the different affinities of the wild-type enzyme and E104A mutant toward substrate cytidine.[13]

The results of the simulations demonstrated that understanding the stability and integrity of well-designed tight binding TSAs in the enzymes active sites may provide important insights into the origin of transition state stabilization during the catalysis.

Acknowledgements

This work was supported by a start-up fund from the Center of Excellence in Structural Biology, University of Tennessee. We are grateful for the computational resources from the Center for Computational Sciences, Oak Ridge National Laboratory, the National Center for Supercomputing Applications (NCSA), and the Scientific Computing Facilities of Boston University. We thank Prof. Martin Karplus for a gift of the CHARMM program and Prof. Qiang Cui for the help with the WHAM analysis.

References

(1) Wolfenden, R. *Nature* **1969**, *223*, 704-705.

(2) Wolfenden, R. *Acc. Chem. Res*. **1972**, *5*, 10-18.

(3) Wolfenden, R.; Kati, W. M. *Acc. Chem. Res.* **1991**, *24*, 209-215.

(4) Wolfenden, R. *Annu. Rev. Biophys. Bioeng*. **1976**, *5*, 271-306.

(5) Radzicka, A; Wolfenden, R. *Methods Enzymol*. **1995**, *249*, 284-312.

(6) Schramm, V. L. *Annu. Rev. Biochemistry* **1998**, *67*, 693-720.

(7) Morrison, J. F.; Walsh, C. T. *Adv. Enzymol. Relat. Area Mol. Biol.* **1988,** *61*, 201-301.

(8) Mader, M. M.; Bartlett, P. A. *Chem. Rev*. **1997**, *97*, 1281-1301.

(9) Lolis, E.; Petsko, G. A. *Annu. Rev. Biochemistry* **1990**, *59*, 597-630.

(10) Lipscomb, WN; Strater, N. *Chem. Rev.* **1996**, *96*, 2375-2433.

(11) Xiang, S.; Short, S. A.; Wolfenden, R.; Carter, C. W., Jr. *Biochemistry* **1995**, *34*, 4516-4523.

(12) Betts, L.; Xiang, S.; Short, S. A.; Wolfenden, R.; Carter, C. W., Jr. *J. Mol. Biol.* **1994**, *235*, 635-656.

(13) Carlow, D. C.; Short, S. A.; Wolfenden, R. *Biochemistry* **1996**, *35*, 948-954.

(14) Carlow, D. C.; Smith, A. A.; Yang, C. C.; Short, S. A.; Wolfenden, R. *Biochemistry* **1995**, *34*, 4220-4224.

(15) Smith, A. A.; Carlow, D. C.; Wolfenden, R.; Short, S. A. *Biochemistry* **1994**, *33*, 6468-6474.

(16) Snider, M. J.; Gaunitz, S.; Ridgway, C.; Short, S. A.; Wolfenden, R. *Biochemistry* **2000,** *39*, 9746-9753.

(17) Xiang, S.; Short, S. A.; Wolfenden, R.; Carter, C. W., Jr. *Biochemistry* **1996**, *35*, 1335-1341.

(18) Xiang, S.; Short, S. A.; Wolfenden, R.; Carter, C. W., Jr. *Biochemistry* **1997**, *36*, 4768-4774.

(19) Snider, M. J.; Wolfenden, R. *Biochemistry* **2001**, *40*, 11364-11371.

(20) Lewis, J. P.; Carter, C. W., Jr.; Hermans, J.; Pan, W.; Lee, T. S.; Yang, W. *J. Am. Chem. Soc*. **1998**, *120*, 5407-5410.

(21) Lewis, J. P.; Liu, S.; Lee, T. S.; Yang, W. *J. Comput. Phys*. **1999**, *151*, 242-263.

(22) Frick, L.; MacNeela, J. P.; Wolfenden, R. *Bioorg. Chem*. **1987**, *15*, 100-108.

(23) Frick, L.; Yang, C.; Marquez, V. E.; Wolfenden, R. *Biochemistry* **1989**, *28*, 9423-9430.

(24) Snider, M. J.; Reinhardt, L.; Wolfenden, R.; Cleland, W. W. *Biochemistry* **2002**, *41*, 415-421.

(25) (a) Cui, Q. ; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569-585.

(b) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, *24*, 565-581.

(26) For a recent review of the QM/MM methods and the summary of the applications, see: Gao, J; Truhlar, D. G. *Annu. Rev. Phys. Chem*. **2002**, *53*, 467-505.

(27) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.

(28) Guo, H.; Cui, Q.; Lipscomb, W.L.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9032-9037.

(29) Guo, H.; Cui, Q.; Lipscomb, W.L.; Karplus, M. *Angew. Chem. Int. Ed.* **2003**, *42*, 1508-1511.

(30) Mackerell, A. D., Jr.; Bashford, D.; Bellott, M.; Bunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E. III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; W.-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586-3616.

(31) Jorgensen, W. L. *J. Am. Chem. Soc*. **1981**, *103*, 335-340.

(32) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys*. **1996**, *105*, 1902-1921.

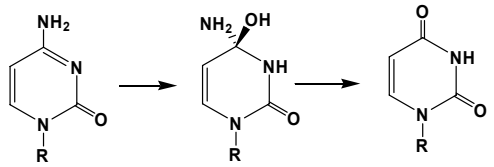
(33) Brooks, C. L. III; Brunger, A.; Karplus, M. *Biopolymers* **1985**, *24*, 843-865.

(34) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett*. **1974**, *28*, 578.

(35) Kumar, M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem*. **1992,** *13*, 1011-1021.

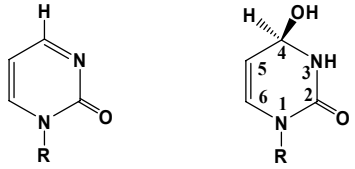
Appendix

A



cytidine tetrahedral intermediate uridine

B



ZEB ZEB-H_2O

Figure III-1. Reaction catalyzed by CDA and TSA of this reaction

(A) The hydrolytic deamination of cytidine to uridine catalyzed cytidine deaminase via a tetrahedral intermediate. **(B)** Pyrimidin-2-one ribonucleoside (zebularine or ZEB) and zebularine 3, 4-hydrate (ZEB-H_2O). ZEB-H_2O resembles the tetrahedral intermediate in (A) and the hydration process seems to be catalyzed by CDA by a mechanism similar to the formation of the tetrahedral intermediate during the CDA-catalyzed reaction.

Figure III-2. Simulation results of CDA-ZEB-H_2O complex

(A) Left: An average active-site structure obtained from 2,000 frames (or 100ps) of the molecular dynamics simulations of the CDA-ZEB-H_2O complex. The further increase of the number of frames has little effect on the structure. Distances are in Å. The 4-OH group of ZEB-H_2O is coordinated to the zinc ion (2.0Å) and hydrogen bonded to $O_{\varepsilon 1}$ of Glu104 (1.5Å). The C_4-O_4 bond has an average distance of about 1.6Å. $O_{\varepsilon 2}$ of Glu104 receives a hydrogen bond from the 3-NH group (1.6Å). The additional ligands to zinc include Cys129, Cys132 and His102. Glu91 and Asn89 interact with the 3'-OH group. **Right**: The X-ray structure from Ref. 11. **(B)** The motion of ZEB-H_2O in the active site of CDA as a function of time monitored by the C_4-O_4 distance (Blue) and C_3'-O_3' distance (Red). The distance fluctuations of the C_4-O_4 bond are significantly larger than those of the C_3'-O_3' bond. Also, the C_4-O_4 bond undergoes transient bond-breaking and making on a picoseconds scale during the simulations; the bond distance approaches or exceeds 2.0Å about 2-3% of the times. **(C)** The free energy (potential of mean force) as a function of the C_4-O_4 distance. There is a second minimum near r(C_4-O_4) = 2Å, corresponding the transient structures with the relatively long C_4-O_4 distances observed in Figure III-2A. The free energy difference of the two minima is about 2.5kcal/mol; the structure with the shorter C4-O4 distance is more stable in the wild-type enzyme. **(D)** The fluctuations of the $O_{\varepsilon 1}$ (Glu104)...H(O_4) (Red) and $O_{\varepsilon 2}$ (Glu104)...H_3(N_3) (Blue) hydrogen bond distances as functions of time. The average distance for the $O_{\varepsilon 1}$...H(O_4) hydrogen bond (1.6 Å) is shorter than that for the $O_{\varepsilon 2}$...H_3(N_3) hydrogen bond (1.75Å), suggesting that the former is more stable. **(E)** The average distances for the $O_{\varepsilon 1}$...H(O_4) (Red), $O_{\varepsilon 2}$...H_3(N_3) (Green) and Zn...O_4 (Blue) interactions as functions of the C_4-O_4 distance. These average distances are obtained from the trajectories of the umbrella sampling, and the corresponding free energy information is given in Figure III-2C (See Methods). The distance that undergoes the most significant change is the $O_{\varepsilon 1}$...H(O_4) hydrogen bond. This hydrogen bond is rather strong (based on the hydrogen bond distance) when the C_4-O_4 covalent bond is intact, but rather weak when the C_4-O_4 bond is broken. The distance between O_4 and Zn decrease (from 1.95Å to 1.81Å) with increasing the C_4-O_4 distance (from 1.5Å to 2.3Å).

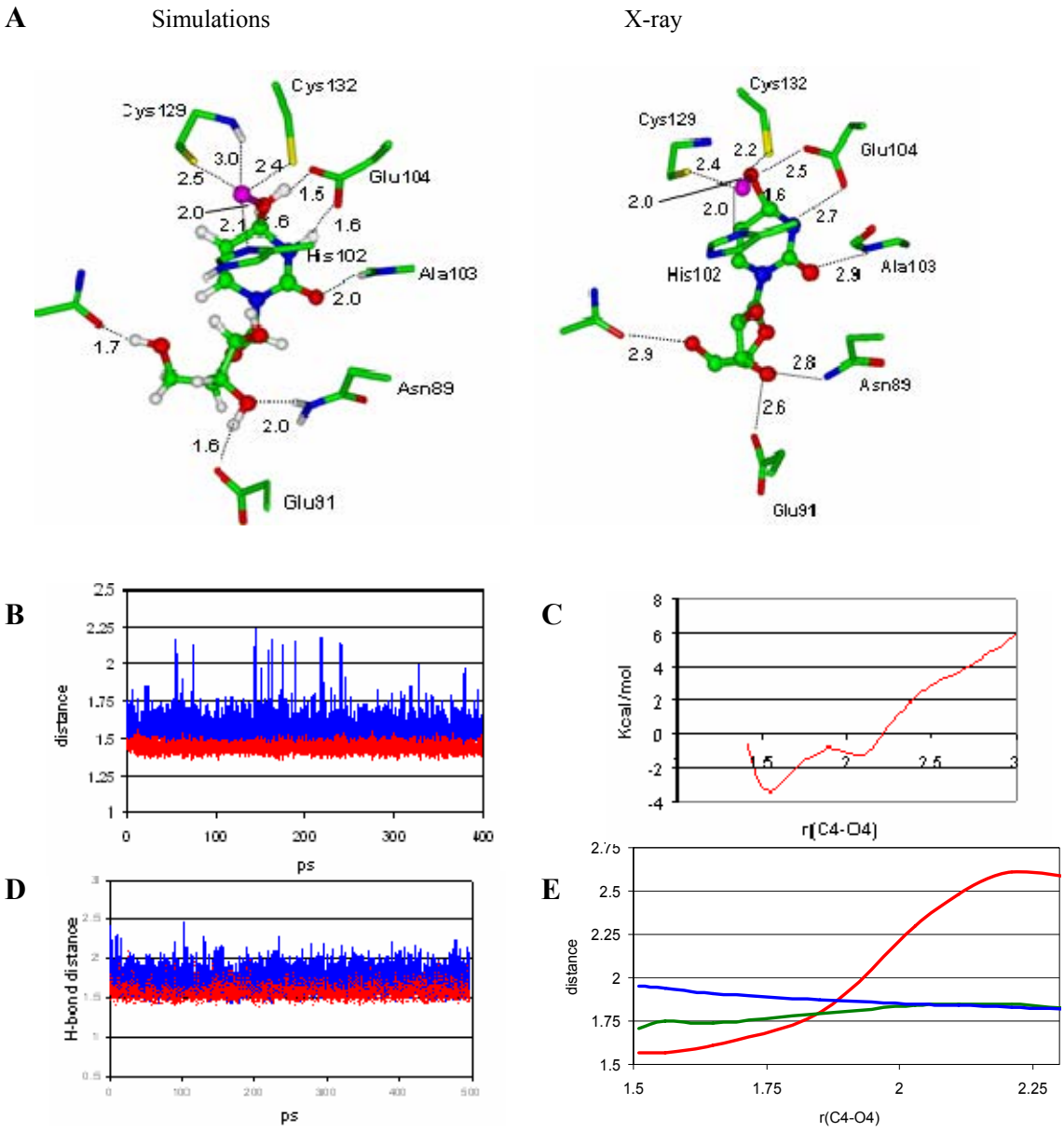


Figure III-2, continued.

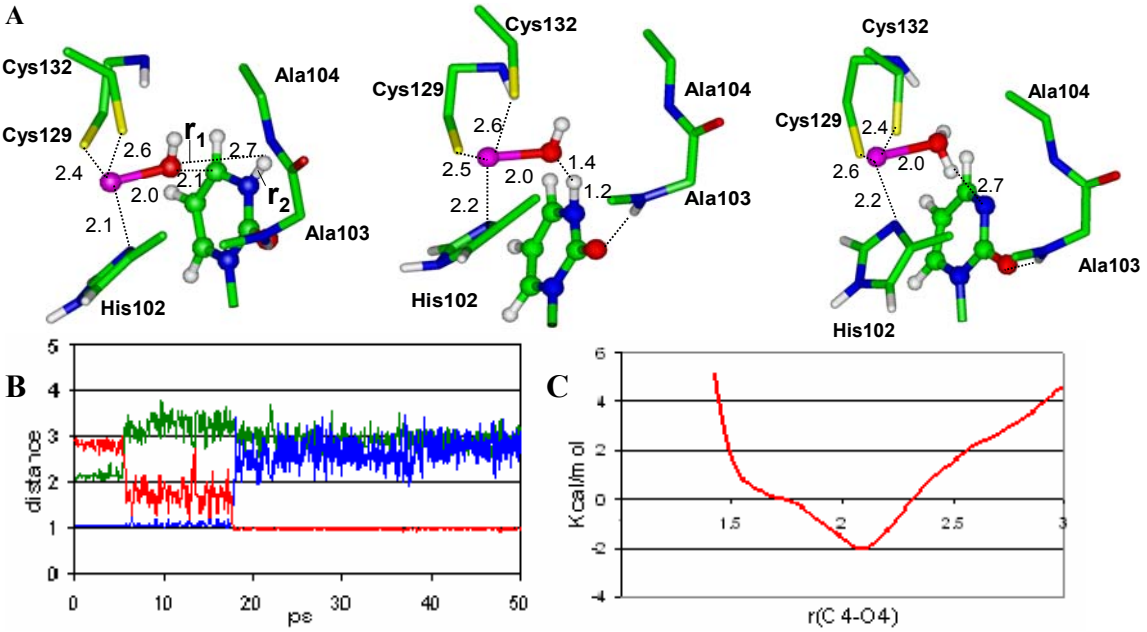


Figure III-3. Simulation results of E104A-ZEB-H_2O complex

(A) The dehydration of ZEB-H_2O into ZEB and a water molecule observed during the QM/MM MD simulations of the E104A-ZEB-H_2O complex. ***Left.*** The early stage of the simulations. The 4-OH group exists as the zinc hydroxide ion and is on the top of C_4 with a C_4-O_4 distance of about 2.5-3.0Å. ***Middle.*** The next stage of the dehydration. The zinc activated hydroxide oxygen interacts with the 3-NH group and shares the proton with N_3. ***Right.*** The final stage of the dehydration. The hydroxide ion has picked up the proton from N_3 and changed to water. It moved back to the initial position on the top of C_4. The umbrella sampling simulations were performed using the difference between $r_1(O_4...H_3)$ and $r_2(N_3-H_3)$ (See Figure III-3A) as the reaction coordinate (i.e., r_1 - r_2). The free energy decreases by about 20 kcal/mol as the proton transfers from 3-N to the hydroxide ion. **(B)** The dehydration of ZEB-H_2O is monitored by the O_4-H_3 (Red), N_3-H_3 (Blue) and C_4-O_4 (Green) distances. The O_4-H_3 distance changes from a distance of 3Å to 1.5Å (i.e., corresponding to the change from the structure on the left to the one in the middle in Figure III-3A). It then changes to 1Å (i.e., corresponding to the change to the structure on the right in Figure III-3A), and at the same time the N_3-H_3 bond is broken with the corresponding distance increases to about 3Å. **(C)** The free energy (potential of mean force) as a function of the C_4-O_4 distance in the E104A mutant. The mutation of Glu104 to Ala shifted the minimum of the free energy curve along the C_4-O_4 bond to the other location at $r(C_4-O_4) \sim 2.2$Å identified earlier (Figure III-2C).

Chapter IV

Stabilization of a Transition-State Analogue at the Active Site of Yeast Cytosine Deaminase through a General Acid-Base Mechanism

Qin Xu, Haobo Guo, Andrey Gorin and Hong Guo

Revised from *J. Phys. Chem. B* **2007**, *111*, 6501-6506

Abstract

The formation of transition-state analog at the active site of yeast cytosine deaminase (yCD) is investigated by the quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) and free energy simulations. It was thought previously that the binding of the inhibitor pyrimidin-2-one leads to the formation of 4-[R]-hydroxyl-3,4-dihydropyrimidine (DHP) at the active of cytosine deaminase. However, the computational study presented here suggests that the nucleophilic attack by Zn-hydroxide on the inhibitor is concerted with the two proton transfers from Glu64 to N_3 of the inhibitor and from Zn-OH to Glu64, respectively. This leads to the formation of a stable alkoxide-like TSA at the active site instead of DHP. It is demonstrated that such concerted process allows the TSA to be stabilized through a general acid-base mechanism.

Abbreviations

yCD, yeast cytosine deaminase; TSA, transition state analogues; QM/MM MD simulations, quantum mechanical/molecular mechanical molecular dynamics simulations; SCC-DFTB, the self-consistent charge density functional tight-binding method; SD, stochastic boundary.

Introduction

General acid-base transition-state stabilization[1-4] is one of the most important strategies that enzymes use to catalyze chemical reactions. Recent computational studies[5-8] suggested that enzymes might use the general acid-base mechanism to promote the formation of the transition state analog (or tetrahedral intermediate analog) and enhance their binding affinities at the active sites. One of the cases that have been studied is the binding of zebularine (ZEB) to cytidine deaminase (CDA) in the formation of a near-perfect transition state analog (TSA). CDA catalyzes the hydrolytic deamination of cytidine to uridine, and its interaction with the TSA inhibitor has become a paradigm for studying the tight TSA binding by enzymes. It was previously believed that the strong electrostatic interactions with the TSA were responsible for the tight binding. However, the quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) and free energy simulations[8] showed that the nucleophilic attack by the Zn-hydroxide on C_4 of ZEB was concerted with two proton transfers from the Zn-hydroxide to Glu104 and from Glu104 to ZEB, respectively. It was predicted from the computational study that this concerted process allows the TSA, an alkoxide-like inhibitor instead of zebularine 3,4-hydrate, to be stabilized through a mechanism that is similar to the transition-state stabilization in the general acid-base catalysis. It would therefore be of interest to determine the generality of the stabilization of TSAs or tetrahedral intermediate analogs through the general acid-base mechanism and whether the similar enhancement of binding also exists in other enzymes.

In this chapter, we study the formation of TSA in yeast cytosine deaminase (yCD). yCD catalyzes the deamination of cytosine into uracil with a catalytic mechanism that is believed to be similar to the cytidine deamination by CDA. We try to determine whether the stabilization of TSA through the general-acid-base mechanism observed in the case of CDA in the earlier study also exists for yCD. Cytosine deaminase is one of the enzymes that are important for the anticancer research. In addition to catalyzing the deamination of cytosine to uracil (Figure IV-1A), it also catalyzes the conversion of 5-fluorocytosine (5-FC) to 5-fluorouracil (5-FU), a compound that is widely used as an anti-cancer drug.[9] The ability of 5-FU to inhibit the DNA synthesis makes it a potent chemotherapeutic agent for cancer treatments. However, the existence of high gastrointestinal and hematological toxicities causes a serious problem.[10] Fortunately, the gene-directed enzyme prodrug therapy (GDEPT) can help to lower this risk by utilizing the less toxic 5-FC as a prodrug.[11-12] The CD/5-FC system of GDEPT has been proved useful in killing tumor cells with minimum toxic effect of 5-FU,[13-20] and it was suggest that the cytosine deaminase from *S. cerevisiae* (yCD) is more applicable than *E. coli* cytosine deaminase (bCD).[21] Understanding of the origin of the transition state stabilization in the yCD-catalyzed reactions is therefore of considerable interest.

The high resolution crystallographic structures of yCDs have been determined with and without an inhibitor.[22] The inhibitor 2-hydroxypyrimidine (2-HP) or its keto form pyrimidin-2-one that was used in the crystallographic studies is believed to be hydrated to 4-[R]-hydroxyl-3, 4-dihydropyrimidine (DHP) (see Figure IV-1B) in a similar process as the hydration of ZEB to form zebularine 3, 4-hydrate in CDA. In our

earlier work on CDA, [8] it was demonstrated that zebularine 3, 4-hydrate may in fact be unstable in the active site and that a proton transfer from the Zn-hydroxide group to Glu104 (the corresponding residue in yCD is Glu64) during the nucleophilic attack could be related to the very high affinity of TSA. Thus, question remains as to whether the similar general acid-base mechanism might be involved in the binding of 2-HD or pyrimidin-2-one by yCD to form TSA. Unfortunately, the location of the proton between the TSA and Glu64 cannot be clearly identified based on the experimental structures even at the rather high resolutions (e.g., 1.14 and 1.6Å for yCD). Although computational studies have been performed on yCD,[23] these previous calculations have mainly been focused on the reaction mechanism and the QM/MM MD and free energy simulations have not been performed for understanding the possible role of the general acid-base mechanism in the formation of TSA. We have therefore performed the QM/MM MD and free energy simulations on the yCD complex in this paper and examined the process of the interchange between TSA and pyrimidin-2-one in the active site. It is shown that, similar to the case of CDA, the nucleophilic attack by the Zn-hydroxide group on C_4 of pyrimidin-2-one was found to be concerted with the proton transfers from the Zn-hydroxide to $O_{\varepsilon 1}$ of Glu64 and from $O_{\varepsilon 2}$ of Glu64 to pyrimidin-2-one, respectively. The concerted process of the nucleophilic attack and the proton transfers leads to the stabilization of the TSA, also an alkoxide-like inhibitor, through the general acid-base catalysis. Our study of the process of the inhibitor binding in yCD presented here may provide not only important insights into the role of general acid-base catalysis in the TSA

stabilization, but also a better understanding of the origin of the transition state stabilization in the enzyme-catalyzed reaction.

Methods

A fast semi-empirical density-functional approach (SCC-DFTB)[24] implemented in the CHARMM program[25] was used for the QM/MM molecular dynamics and free energy simulations. The SCC-DFTB approach has been extensively tested against high-level quantum mechanical methods for a range of active site models containing Zn.[24] It was found that the SCC-DFTB approach reproduced structural and energetic properties rather reliably. Additional tests[8] were performed in the earlier study of CDA using several high-level QM/MM approaches. The QM methods used in the additional tests include B3LYP/6-31G(d,p) and HF/6-31G(d,p). It was found that the results from the SCC-DFTB/MM calculations were similar to those based on the high-level QM methods, consistent with the study[24] on the active site models. For detailed discussions of the testing results, see Ref. 8 and 24.

The initial coordinates for the simulations were obtained from the crystal structures of the yCD-TSA complexes (1UAQ at 1.60Å[23] and 1P6O at 1.14Å[22]), which have the transition state analog bound at the active site. *The TSA inhibitor was initially assumed to be DHP* as *generally believed, and this assumption was found to be incorrect (see below).* The MD simulations were performed on the yCD-DHP complex. For each of the two X-ray structures, two models of the yCD complex were constructed. The first model (referred here as the simple model) has zinc, DHP and the side chain of Glu64

treated by the QM method and the rest of the system treated by MM using the all-hydrogen CHARMM force field (PARAM22).[26] With the QM regions expand to the side chains of Thr86, Cys91 and Cys94, the second model was generated and is referred here as the bonded model. In this bonded model, several force constants for the bond length and bond angles related to the zinc cation and its ligands were added in order to reproduce the structure parameters of the original X-ray structures; these force constants are given in Table IV-S1. The link-atom approach[27] in the CHARMM program was used to separate the QM and MM regions, and the modified TIP3P water model[28, 29] was applied for the solvent molecules. The initial model systems that were generated from the X-ray structures containing DHP were found to be unstable; i.e., the proton transfer from O_4 to $O_{\varepsilon 1}$ of Glu64 occurred during the QM (SCC-DFTB)/MM energy minimization, leading to a stable alkoxide-like inhibitor at the active site. The resulting structures with the alkoxide-like inhibitor were found to be stable during the QM (SCC-DFTB)/MM MD simulations. The similar observation was also made during energy minimizations with use of the HCTH/6-31G** and B3LYP/DZVP treatments for the QM region. *Thus, the simulations suggest that TSA is not DHP, but the corresponding alkoxide-like species.* The high-level QM/MM approaches are extremely time-consuming for the MD and free energy simulations, and they were therefore not used for the MD and free energy simulations in the present study. Nevertheless, the similar results obtained based on the energy minimizations already suggested that DHP might not be stable in the yCD complex. This situation is similar to that observed for zebularine 3, 4-hydrate in CDA.[8]

The stochastic boundary molecular-dynamics method[30] was used for the QM/MM MD simulations. The system was separated into a reaction zone and a reservoir region, which was deleted; the reaction zone was further divided into the reaction region and the buffer region. The reference point for partitioning the system was chosen as the N_1 atom of TSA. The reaction region was a sphere with radius 16Å, and the buffer region has R equal to 16Å $\leq$ R $\leq$ 18Å. Inside the reaction region, the atoms were propagated by molecular dynamics, whereas atoms in the buffer region were propagated by Langevin dynamics. The friction constants for the Langevin dynamics were 250ps^{-1} for the protein atoms and 62ps^{-1} for the water molecules. A 1-fs time step was used for integration of the equations of motion, and the results of the simulations were saved on every 50fs. In initiating the runs, 500 steps of minimization using the steepest descent method were performed for the protein and solvent atoms. Then energy minimizations were performed for the entire stochastic boundary system with the Adapted Basis Newton Raphson method. The temperature of the system was gradually increased from 50K to 300K (20ps) and equilibrated at 300K (30ps). The simulations were performed on the resulting systems for several hundred picoseconds to 1ns.

The umbrella sampling method[31] implemented in the CHARMM program along with the Weighted Histogram Analysis Method (WHAM)[32] was used in the present study and applied to determine the changes of the free energies (potentials of mean force) for the processes of the nucleophilic attack of the Zn-hydroxide on C_4 (monitored by the distance between C_4 and O_4). The free energy curves for the simple and bonded models (see above) led to the same conclusion, and only the results from the simpler model based

on 1UAQ are presented. The harmonic potentials with a force constant of 50-1000 kcal/mol-Å^2 were used for the umbrella sampling of the nucleophilic attack. Twenty windows were used for the nucleophilic attack by the Zn-hydroxide. For each window, 70ps simulations (20ps equilibration and 50ps production run) were performed with the origin of the harmonic potential, $d_0(C_4\text{-}O_4)$, changed from 1.4Å to 3.0Å leading to a dataset of 1,000,000 (total) for the free energy analysis. To estimate the effect of the proton transfer from the Zn-hydroxide to Glu64 on the stability of the TSA, the free energy profile was also obtained for the nucleophilic attack step with the O_4-H bond fixed by SHAKE algorithm[33] at 1.1Å. As a result, Glu64 can still protonate N_3 but is unable to accept the proton from O_4.

Results and Discussions

To examine whether the methods of the QM/MM MD simulations used in the present study are reasonable, the average structures of the active site obtained from the simulations are compared with one of the structures determined by X-ray crystallography in Figure IV-2 (PDB ID=1UAQ). The average structures from the simple and bonded models (see Method) based on the initial coordinates of 1UAQ are shown in Figure IV-2B and 2C, respectively, while the average structure of the simple model based on 1P6O is plotted in Figure IV-2D. All of these average structures are rather close to the X-ray structures and have the similar hydrogen-bond network and critical distances. It is interesting to note, however, that in all the average structures Glu-64 has already accepted the proton from Zn-O_4 and becomes protonated during the energy

minimizations and MD simulations, leading to the alkoxide-like TSA inhibitor. This is in contrast to the general belief that 4-[R]-hydroxyl-3, 4-dihydropyrimidine (DHP) was formed in the active site even though the proton is invisible in the experimental structures. Thus, the QM/MM results indicate that the structure of the yCD-DHP complex is likely to be unstable and the TSA may exist as an alkoxide-like inhibitor at the active site (as shown in Fig. 2B-D). This suggestion is supported by the additional high-level QM/MM calculations (see Methods).

Figure IV-3 monitors the fluctuations of certain distances at the active site of the yCD complex with the alkoxide-like TSA, including R(C_4-O_4), R(Zn-O_4), R(N_3-H_3,), R($O_{\varepsilon 2}$ …H_3), R($O_{\varepsilon 1}$-H) and R(O_4…H). Figure IV-3 shows that the C_4-O_4 bond is stable with a bond distance of about 1.5Å (red line in the top panel). Consistent with the average structures in Figure IV-2B-D, both the Glu64 and N_3 are protonated with R($O_{\varepsilon 1}$-H) and R(N_3-H_3) around 1Å. The $O_{\varepsilon 2}$…H_3 (N_3) and O_4…H distances fluctuate around 2Å, indicating the formation of two stable hydrogen bonds between Glu64 and the ligand.

To investigate the free energy change associated with the formation of TSA, the potential of mean force was determined as a function of the C_4-O_4 distance using the umbrella sampling method and WHAM program. Figure IV-4 (top) shows that the free energy of the complex with the alkoxide-like TSA at the active site (i.e., at position **B**) is more than 2kcal/mol lower than the substrate analog complex at position **A** (i.e., the yCD-pyrimidin-2-one complex). The barrier between **A** to **B** is somehow higher than that in CDA obtained in the earlier simulations.[8] Three average active-site structures at **A**, **B** and **B'** are plotted in Figure IV-4 (bottom); here **B**' is the free energy of the DHP

complex with the proton transfer from Zn-hydroxide to Glu64 prevented by use of the SHAKE algorithm.[33] The yCD-pyrimidin-2-one complex before the nucleophilic attack (i.e., with C_4-O_4 distance at about 2.4 Å) was found to be stable during the 500ps QM/MM MD simulations, although Glu64 underwent some rotation. At position **B**, the transition state analog is formed with the average bond length of C_4-O_4 about 1.5 Å and the hydrogen of the zinc hydroxide is transferred to Glu64, as discussed earlier. As can be seen from Figure IV-4, the shapes of the two free energy curves are essentially the same in the area of $d_0(C_4\text{-}O_4) \geq 2$Å. This result is expected, as the O_4-H bond is intact in the early stage of the nucleophilic attack and the corresponding structures along the reaction coordinate in the two cases are similar. Figure IV-4 also shows that the free energy after the nucleophilic attack is much higher for the curve with the constraint (i.e., at **B'**) than that without (i.e., at **B)**. The major difference for the two free energy simulations is that the ligand in the complex at **B'** is DHP, while it is the alkoxide-like TSA at **B**. Thus, it appears that a major stabilization of the alkoxide-like TSA complex in the active site compared to that of DHP arises from the proton transfer from the Zn-hydroxide to Glu64.

The H_3 proton transfer is monitored by the changes of distances $R(N_3\text{-}H_3)$ (blue) and $R(O_{\varepsilon 2}\text{-}H_3)$ (red) in Figure IV-5A based on the trajectories from the free energy simulations as $d_0(C_4\text{-}O_4)$ (the reaction coordinate in the free energy simulations) changes from 1.95Å (upper panel) to 1.85Å (lower panel). Accordingly, the H proton transfer is monitored by the changes of $R(O_4\text{-H})$ (blue) and $R(O_{\varepsilon 1}\text{-H})$ (red) in Figure IV-5B. Figure IV-5A (Figure IV-5B) shows that the positions of the blue and red lines are reversed from the upper panel to the lower panel, indicating that H_3(H) is transferred near the transition

state of the nucleophilic attack that occurs between $d_0(C_4\text{-}O_4)$ = 1.95Å and 1.85Å (see Figure IV-4). Figure 6 shows the average distances of $r(O_4\ldots H)$ (blue), $r(O_{\varepsilon 1}\text{-}H)$ (magenta), $r(O_{\varepsilon 2}\ldots H_3)$ (red) and $r(N_3\text{-}H_3)$ (yellow) as a function of $d_0(C_4\text{-}O_4)$ from the umbrella sampling. It is of interest to note that all the curves seem to cross with each other near the transition state, suggesting a concerted process of the nucleophilic attack and proton transfers as shown by the solid arrow from I to IV in Figure IV-1C.

It is believed that the binding of pyrimidin-2-one to CD leads to the formation of 4-[R]-hydroxyl-3,4-dihydropyrimidine (DHP) through the nucleophilic attack of Zn-hydroxide on the inhibitor. However, the proton on the C_4-OH group is invisible in the X-ray structures. Here we showed from the QM/MM MD simulations that a stable DHP would not exist in the active site of yCD and that the corresponding alkoxide-like structure would be more stable for which the proton on the C_4-OH group is transferred to Glu64. This alkoxide-like TSA complex cannot be distinguishable from that of DHP in the X-ray structures. Therefore, our prediction is still consistent with the experimental structures. The free energy simulations for the nucleophilic attack process showed that the proton transfer from Zn-hydroxide to Glu64 plays an important role in stabilizing the TSA complex and the DHP complex is even less stable than the pyrimidin-2-one complex (i.e., before the nucleophilic attack).

Conclusions

The formation of transition-state analog at the active site of yCD was investigated by the QM/MM MD and free energy simulations. Similar to our earlier simulations on

the binding of ZEB to CDA,[8] DHP was found to be unstable at the active site of yCD. It was suggested that the nucleophilic attack by the Zn-hydroxide group on C_4 of the inhibitor may be concerted with two proton transfer processes (from Glu64 to N_3 and from O_4 to Glu64, respectively), leading to the formation of a stable alkoxide-like TSA at the active site. Such concerted process seems to allow the TSA to be stabilized through a general acid-base mechanism. For some of the nucleophilic attack processes (e.g., by Zn-OH^- or Zn-OH_2) in enzyme-catalyzed reactions, the general acid-base mechanisms are not invoked to explain the efficiency of the reactions. It would be of interest to examine whether the tetrahedral intermediates resulting from the nucleophilic attacks could also be stabilized by the general acid-base mechanisms as we demonstrated for the inhibitor binding in CDA and CD.

Acknowledgements

We thank Prof. Martin Karplus for a gift of the CHARMM program. This work was supported in part by the Center of Excellence for Structural Biology and UT-ORNL Science Alliance, the University of Tennessee, and the Petroleum Research Fund.

Supporting Information

There are supporting information for the constrained parameters and the force constants for the bonded model (Table IV-S1) with the atomic labels used in the text (Figure IV-S1), comparison of the performances of the SCC-DFTB and B3LYP/6-31G(d,p) methods (Figure IV-S2), different PMF curves of the nucleophilic attack

starting from different initial structures with the energy correction based on energetic differences between SCC-DFTB and B3LYP/6-31G(d,p) (Figure IV-S3), and the PMF curve for the proton transfer between the Zn hydroxide and Glu64 (Figure IV-S4).

References

(1) Fersht, A. Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding, 2nd ed., W. H. Freeman and Company, New York, 1999.

(2) Jencks, W. P. *Chem. Rev.* **1972**, *72*, 705-718.

(3) Kirby, A. J. *Acc. Chem. Res.* **1997**, *30*, 290-296.

(4) Schowen, K. B.; Limbach, H.-H.; Denisov, G. S.; Schowen, R. L. *Biochim. Biophys. Acta* **2000,** *1458*, 43-62.

(5) Guo, H. B.; Wlodawer, A.; Guo, H. *J. Am. Chem. Soc.* **2005**, *127*, 15662-15663.

(6) Xu, Q.; Guo, H.; Wlodawer, A.; Guo, H. *J. Am. Chem. Soc*. **2006**, *128*, 5994-5995.

(7) Guo, H. B.; Wlodawer, A.; Nakayama, T.; Xu, Q.; Guo, H. *Biochemistry*, **2006**, *45*, 9129-9137.

(8) Guo, H. B.; Rao, N.; Xu, Q.; Guo, H. *J. Am. Chem. Soc.* **2005**, *127*, 3191-3197.

(9) Morris, S. M. *Mutat. Res*. **1993**, *297*, 39–51.

(10) Vauthey, J. N.; Marsh Rde, W.; Cendan, J. C.; Chu, N. M.; Copeland, E. M. *Br. J. Surg.* **1996**, *83*, 447–455.

(11) Aghi, M.; Hochberg, F.; Breakefield, X. O. *J. Gene Med.* **2000**, *2*, 148-164.

(12) Greco, O.; Dachs, G. U. *J. Cell. Physiol.* **2001**, *187*, 22-36.

(13) Austin, E. A.; Huber, B. E. *Mol. Pharmacol.* **1993**, *43*, 380–387.

(14) Freytag, S. O.; Khil, M.; Stricker, H.; Peabody, J.; Menon, M.; DePeralta-Venturina, M.; Nafziger, D.; Pegg, J.; Paielli, D.; Brown, S.; Barton, K.; Lu, M.; Aguilar-Cordova, E.; Kum, J. H. *Cancer Res.* **2002**, *62*, 4968–4976.

(15) Hirschowitz, E. A.; Ohwada, A.; Pascal, W. R.; Russi, T. J.; Crystal, R. G. *Hum. Gene Ther*. **1995**, *6*, 1055–1063.

(16) Huber, B. E.; Austin, E. A.; Good, S. S.; Knick, V. C.; Tibbels, S.; Richards, C. A. *Cancer Res.* **1993**, *53*, 4619–4626.

(17) Hwang, L. H. *J. Biomed. Science* **2006**, *13*, 453-468.

(18) Liu, Y.; Ye, T.; Maynard, J.; Akbulut, H.; Deisseroth, A. *Cancer Gene Ther.* **2006**, *13*, 346-356.

(19) Mullen, C. A.; Kilstrup, M.; Blaese, R. M. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 33–37.

(20) Warmann, S. W.; Armeanu, S.; Frank, H.; Buck, H.; Graepler, F.; Lemken, M. L.; Heitmann, H.; Seitz, G.; Lauer, U. M.; Bitzer, M.; Fuchs, J. *Pediatr. Surg. Int.* **2006**, *22*, 16-23.

(21) Kievit, E.; Bershad, E.; Ng, E.; Sethna, P.; Dev, I.; Lawrence, T. S.; Rehemtulla, A. *Cancer Res*. **1999**, *59*, 1417–1421.

(22) (a) Ireton, G. C.; Black, M. E. *Structure* **2003**, *11*, 961-972; (b) Ko, T. P.; Lin, J. J.; Hu, C. Y.; Hsu, Y. H.; Wang, A. H. J.; Liaw, S. H. *J. Biol. Chem.* **2003**, *278*, 19111-19117.

(23) (a) Sklenak, S.; Yao, L.; Cukier, R. I.; Yan H. *J. Am. Chem. Soc.* **2004**, *126*, 14879-4889; (b) Yao, L.; Sklenak, S.; Yan, H.; Cukier, R. I. *J. Phys. Chem. B* **2005**, *109*, 7500-7510.

(24) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem*. **2003**, *24*, 565-581.

(25) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.

(26) Mackerell, A. D., Jr.; Bashford, D.; Bellott, M.; Bunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E. III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; W.-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586-3616.

(27) Amara, P.; Field, M. J. *Theor Chem Acc* **2003**, *109*, 43-52.

(28) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335-340.

(29) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902-1921.

(30) Brooks, C. L., III; Brunger, A.; Karplus, M. *Biopolymers* **1985**, *24*, 843-865.

(31) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578-581.

(32) Kumar, M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011-1021.

(33) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.

Appendix

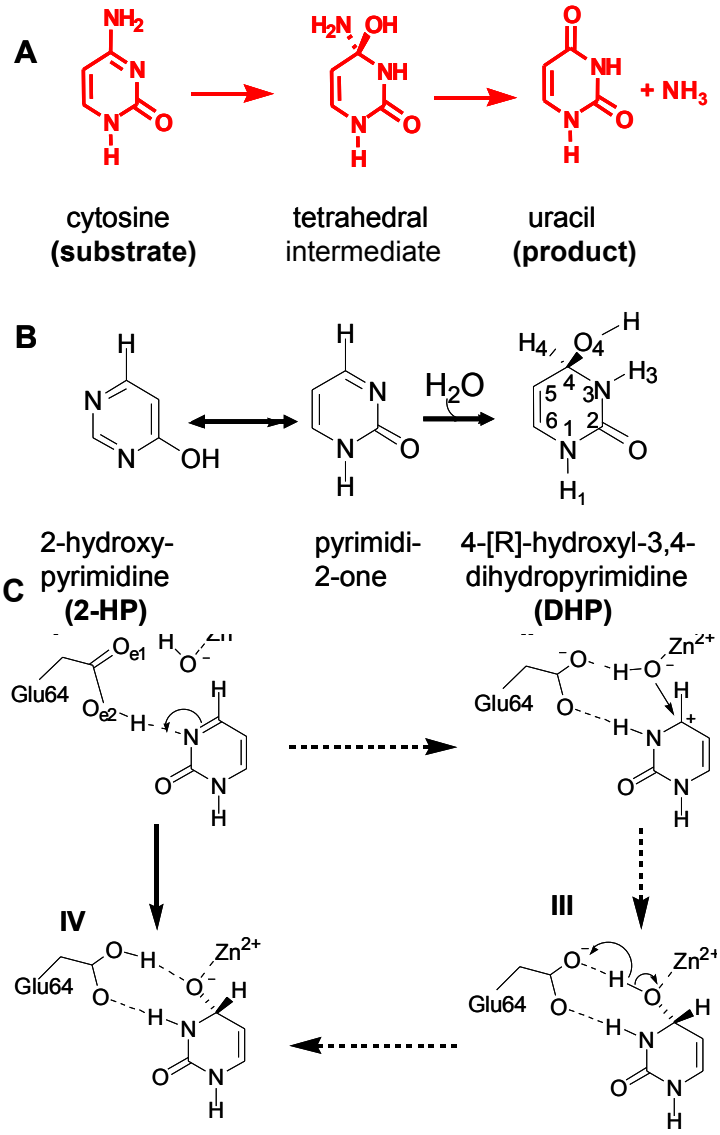


Figure VI-1. The reaction catalyzed by yCD and the formation of TSA in yCD

(A) The deamination of cytosine to uracil catalyzed by yCD. A tetrahedral intermediate is believed to be formed. **(B)** 2-hydroxypyrimidine (2-HP) or its tautomeric keto form pyrimidin-2-one is hydrated by yCD and 4-[R]-hydroxyl-3,4-dihydropyrimidine (DHP) is believed to be formed at the active site. **(C)** The possible steps for the TSA formation in yCD which may be concerted or step-wise. I→II: protonation of N_3; II→III: nucleophilic attack of Zn-hydroxide on C_4; III→IV: the proton transfer from O_4 to $O_{\varepsilon 1}$ (Glu64). A concerted I→IV process is suggested from our QM/MM MD and free energy simulations.

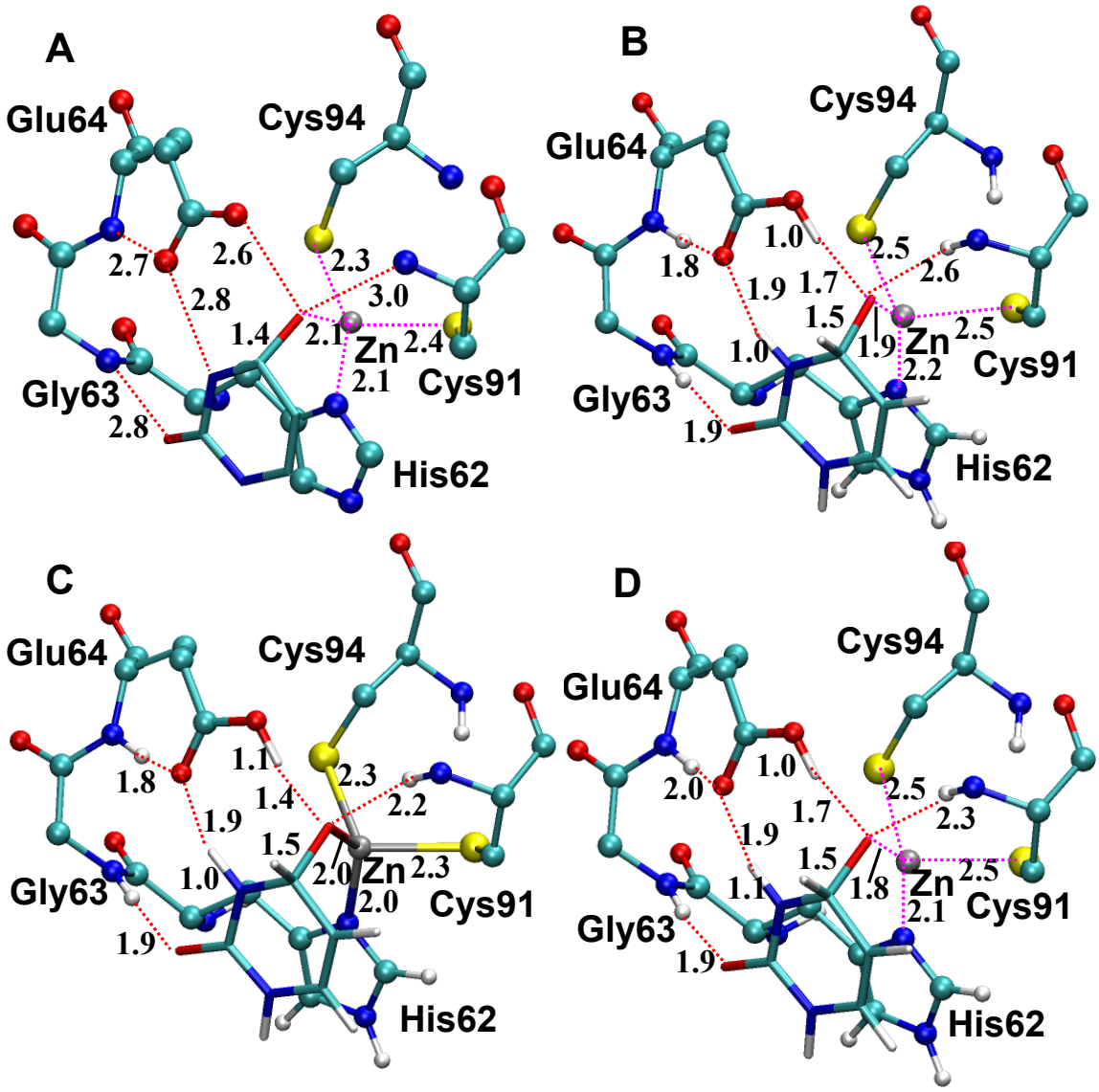


Figure IV-2. Comparison of yCD-TSA complexes between the average structures from the simulations and the X-ray structure

The hydrogen bond distances are between non-hydrogen atoms in the X-ray structure but not in the average structures. The ligations with zinc are also indicated. **(A)** The X-ray structure of 1UAQ. **(B)** The average structure of the simple model (see Methods) from the simulations of the yCD complex based on 1UAQ. **(C)** The average structure of the bonded model based on 1UAQ. **(D)** The average structure of the simple model based on the X-ray structure of 1P6O.

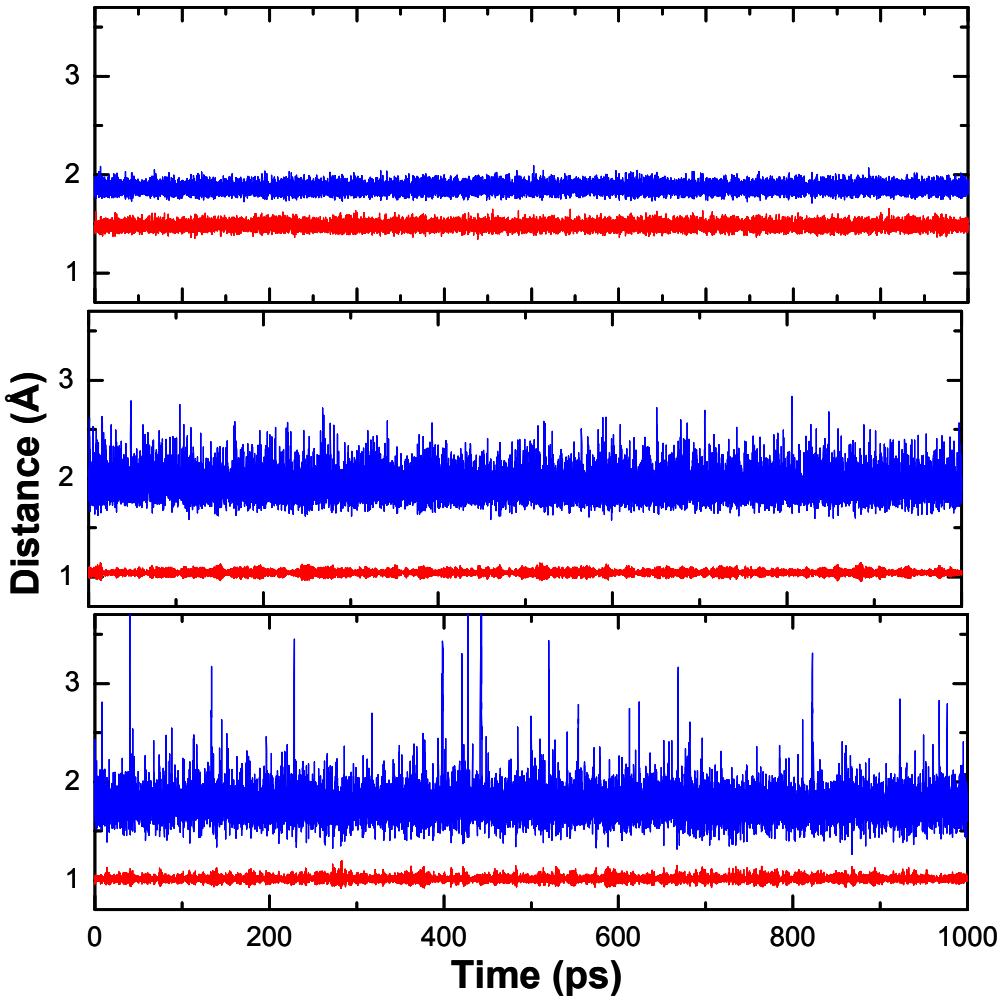


Figure IV-3. The distance fluctuations in simulations of yCD-TSA complex

The distance fluctuations were obtained from 1 ns simulations of the simple model of the yCD-DHP complex based on the structure of 1UAQ. **Top panel**: R(C_4-O_4) (red line) and R(Zn-O_4) (blue line); **Middle panel:** R(N_3-H_3) (red line) and R($O_{\varepsilon 2}$…H_3) (blue line); **Bottom panel**: R($O_{\varepsilon 1}$-H) (red line) and R(O_4…H) (blue line). See Figure IV-1B for the atom designations.

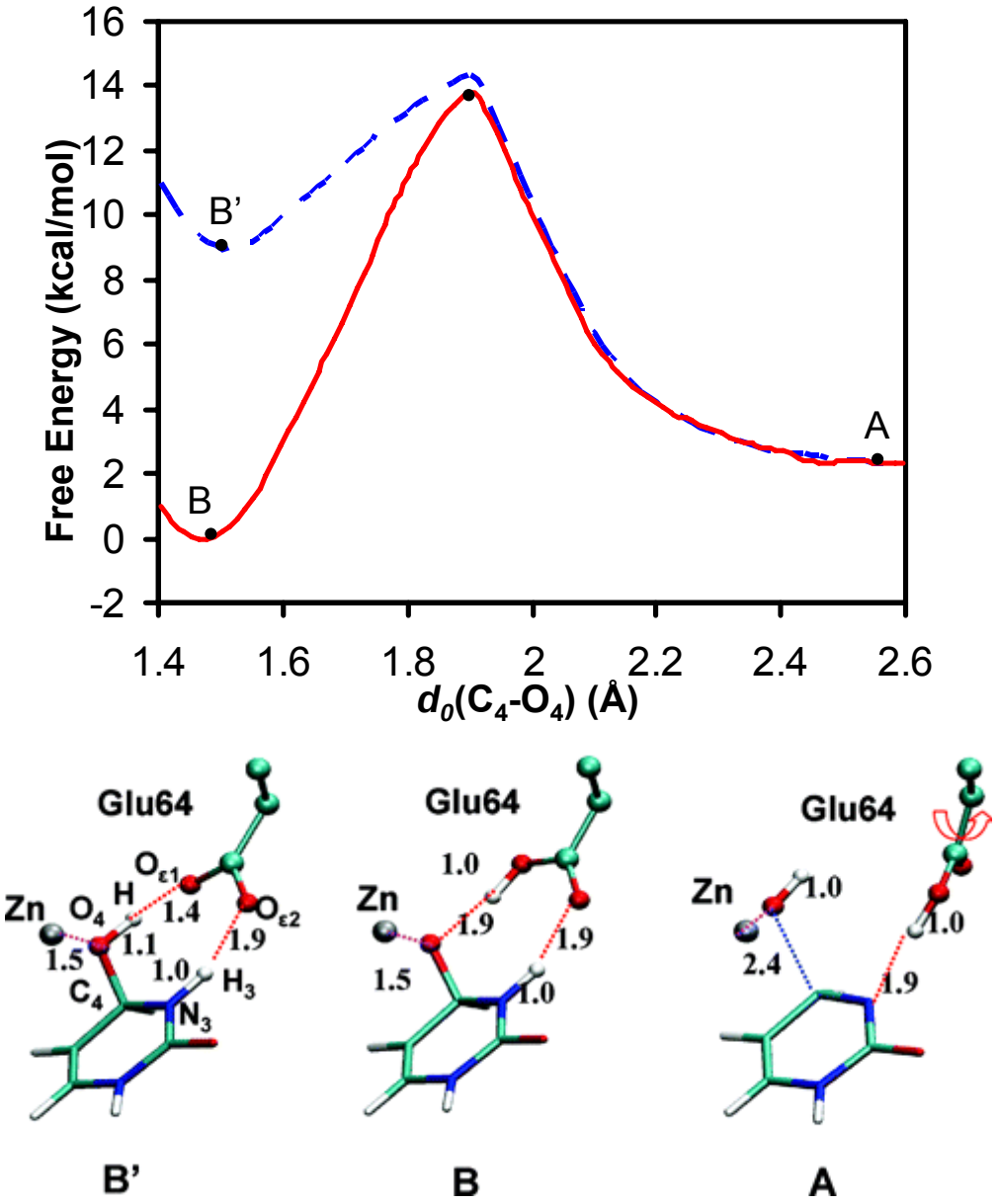


Figure IV-4. The nucleophilic attack in the yCD-TSA complex

Top: The potentials of mean force for the nucleophilic attack with (blue and dashed line) and without (red and solid line) the constraint on the O_4-H bond (see Methods). The labels A and B (B') indicate the different stages before and after the nucleophilic attack. The corresponding active-site structures are given in the figures below. **Bottom:** The active site structures at the different stages of the nucleophilic attack. **(A)** Before the nucleophilic attack. The arrow indicates that the side chain of Glu64 is flexible. **(B)** The formation of TSA. **(B')** The formation of DHP with addition of a constraint on the O_4-H bond.

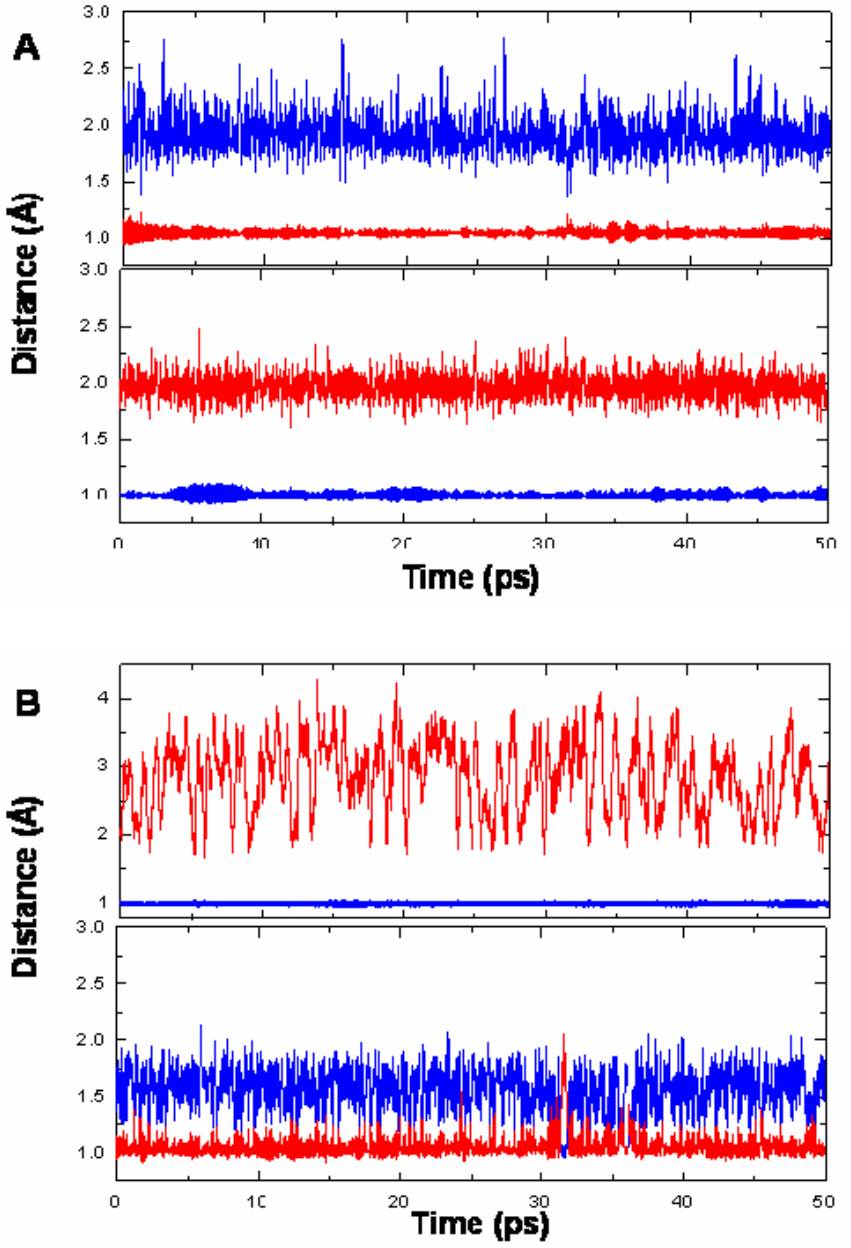


Figure IV-5. The proton transfers monitored by the changes of the distances

(A) The H_3 proton transfer monitored by the distances N_3-H_3 (blue) and $O_{\varepsilon 2}$-H_3 (red) before [upper panel with R(C_4-O_4)=1.95 Å] and after [lower panel with R(C_4-O_4)=1.85 Å] the transition state of nucleophilic attack; **(B)** The H (O_4) proton transfer monitored by the distances O_4-H (blue) and $O_{\varepsilon 1}$-H (red) before (upper panel) and after (lower panel) the transition state of nucleophilic attack.

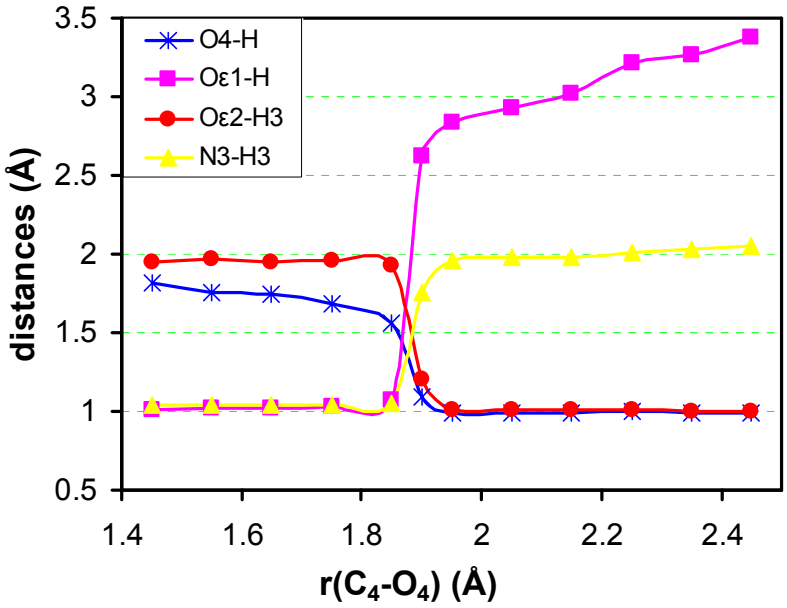


Figure IV-6. The changes of the distances during the nucleophilic attack

The average distances of $R(O_4\text{-}H)$ (blue), $R(O_{\varepsilon 1}\text{-}H)$ (magenta), $R(O_{\varepsilon 2}\text{-}H_3)$ (red) and $R(N_3\text{-}H_3)$ (yellow) as a function of the $C_4\text{-}O_4$ distance obtained from the free energy simulations.

Table IV-S1. Parameters for the zinc-bonded model

Top: bond-length parameters; Bottom: bond-angle parameters. See Figure IV-S1 for labeling.

	Force constant K_r (kcal/mol/Å^2)	r_0 (Å)
Zn-$N_{\delta 1}$	40	2.05

	Force constants K_θ (kcal/mol/rad^2)	θ_0 (°)
O_4-Zn-$N_{\delta 1}$	20	111.2
C_γ-$N_{\delta 1}$-Zn	20	129.9
$C_{\varepsilon 1}$-$N_{\delta 1}$-Zn	20	120.3
S (Cys91)-Zn-$N_{\delta 1}$	36	108.2
S (Cys94)-Zn-$N_{\delta 1}$	36	108.2

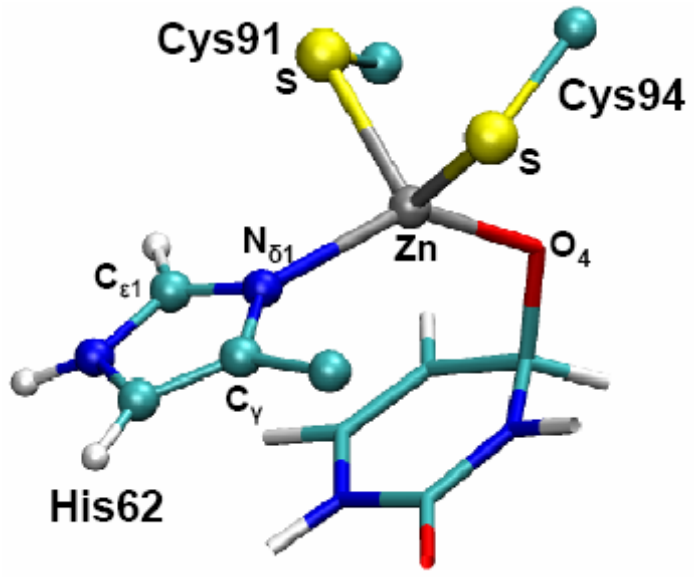


Figure IV-S1. The structure of zinc ligation in the bonded model

The transition state analogue is shown in stick while the residues of yCD are shown in ball and stick. The labeling of the atoms around the zinc atom are same to those in Table IV-S1.

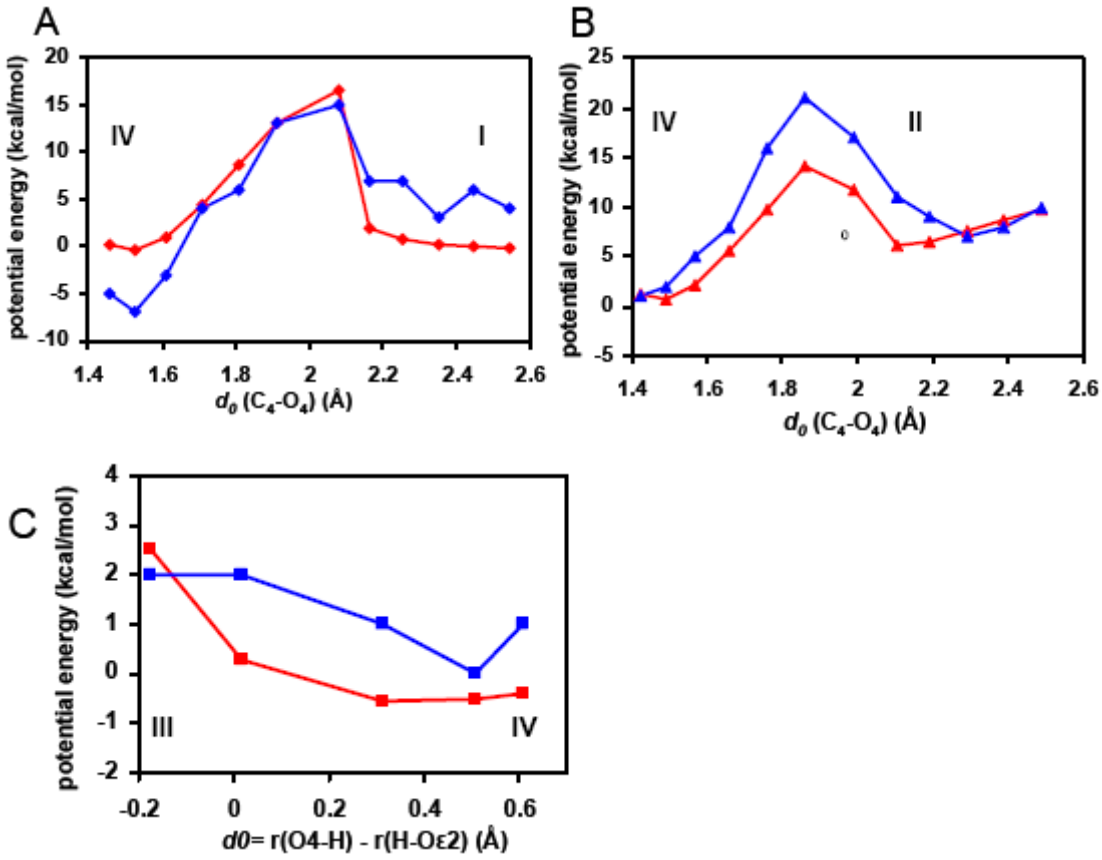


Figure IV-S2. Comparison of the performances of the SCC-DFTB and B3LYP/6-31G(d,p) methods

Comparison of the free energy changes between different complexes along the corresponding reaction coordinates from the SCC-DFTB (red) and B3LYP/6-31G** (blue) calculations, respectively. For the structures of I, II, III and IV, see Figure IV-1C. Energy minimizations were performed at the SCC-DFTB level and single point B3LYP/6-31G** calculations were undertaken at the SCC-DFTB geometries. The structure IV is the most stable structure in all the cases. **(A)** Between I and IV; **(B)** Between II and IV; **(C)** Between III and IV.

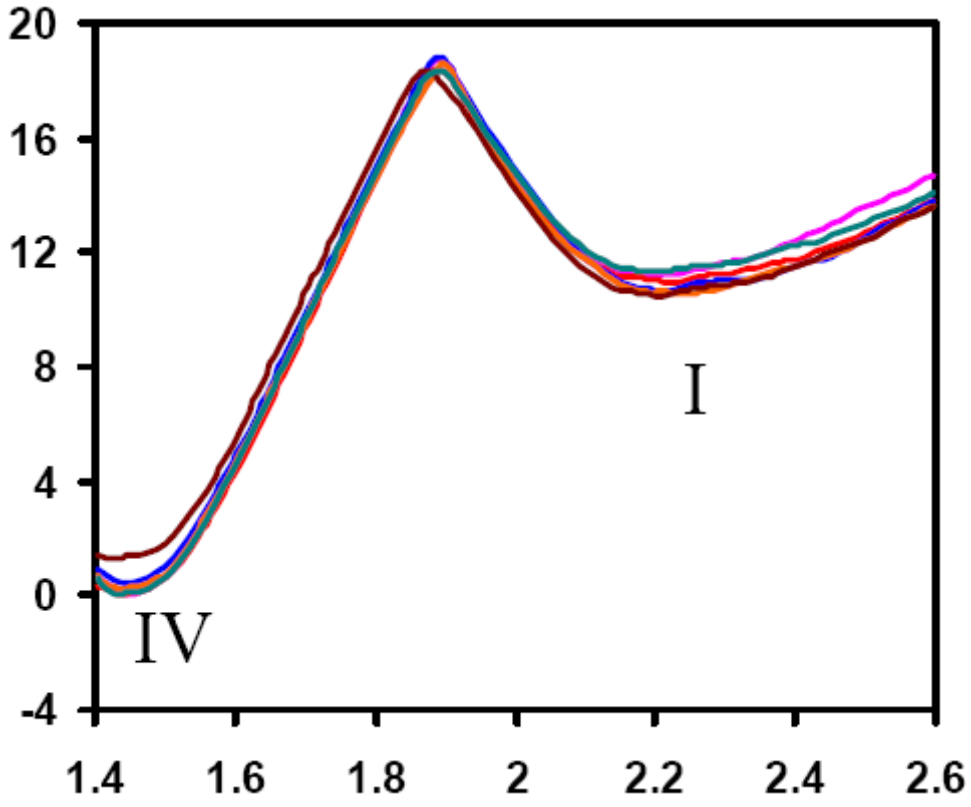


Figure IV-S3. Comparison of the free energy changes of the nucleophilic attack starting from different initial structures

The free energy (potential of mean force) curves are of umbrella sampling simulations starting from different initial structures obtained from the MD simulations (e.g., at 400ps, 450ps, 500ps, 550ps, 700ps and 800ps), including a correction based on the energy differences from the SCCDFT and B3LYP/6-31G** calculations on model systems. The results show that the relative free energy of the alkoxide-like TSA complex (IV) would be even lower and the complex would be relatively more stable compared to the complex I if a correction based on a high level QM method was introduced into the SCC-DFTB free energy curves (see Figure IV-4). This is because that SCC-DFTB underestimates the stability of the alkoxidelike TSA complex (see Figure IV-S2A). In addition, the similar free energy curves obtained suggest that the results are not sensitive to the initial structures used in the PMF simulations.

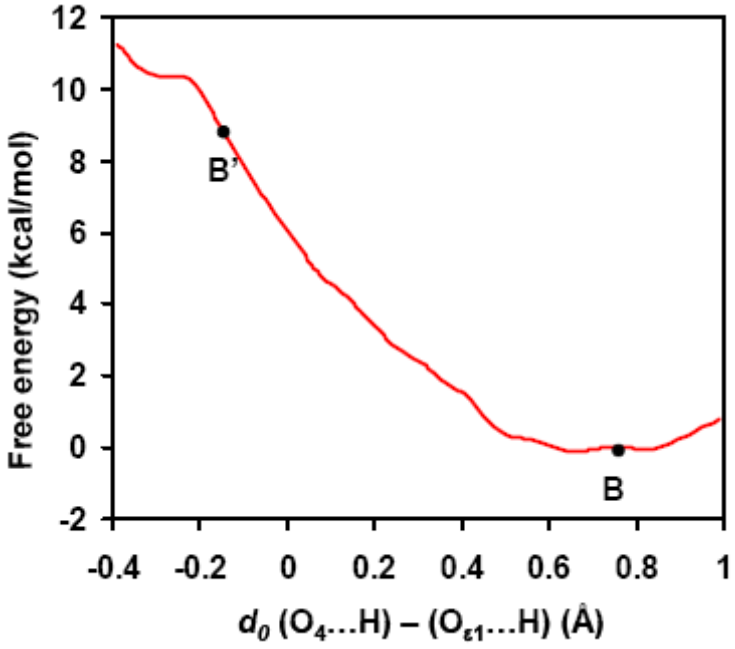


Figure IV-S4. The free energy changes for the proton transfer between O_4 and $O_{\varepsilon 1}$

During the QM(SCC-DFTB)/MM simulations for the transfer of H between O_4 and $O_{\varepsilon 1}$ (Figure IV-1C, between III and IV), a constraint has been added on the C_4-O_4 bond to prevent it from being broken. The free energy difference between B and B' is similar to that between B and B' obtained from the nucleophilic attack and reported in Figure IV-4.

Chapter V

The Importance of Dynamics in Substrate-Assisted Catalysis and Specificity

Qin Xu, Haobo Guo, Alexander Wlodawer, and Hong Guo

Revised from *J. Am. Chem. Soc.* **2006**, *128*, 5994-5995

Abstract

The QM(SCC-DFTB)/MM(CHARMM) molecular dynamics (MD) and free energy simulations have been performed on the nucleophilic attack by Ser278 of kumamolisin-As on a 5 amino acid peptide RPXaa*FR, in order to study the origin of the catalytic activity and specificity of this enzyme. The aspartic acid residue, Asp164 seems to act as a general acid catalyst to protonate and stabilize the tetrahedral intermediate, assisting in the nucleophilic attack. At the same time, a significant rotation of the P_1-His side chain was observed, with the formation of a salt bridge with the deprotonated Asp164, which may give help to the function of Asp164. The effect of P_1-His seems to be dynamic substrate-assisted catalysis (SAC).

Abbreviations:

SAC, substrate-assisted catalysis; QM/MM Quantum mechanical/molecular mechanical; MD molecular dynamics; TI, Tetrahedral intermediate; PMF, potential of mean force; SCC-DFTB, self consistent charge density-functional tight-binding; WHAM, weighted histogram analysis method.

Introduction

Substrate-assisted catalysis (SAC) is the process in which one or more functional groups from the substrate, in addition to those from the enzyme, contribute to the rate acceleration for the enzyme-catalyzed reaction. There is growing evidence that SAC may exist in many naturally occurring or engineered catalysts.[1-2] The existence of product-assisted catalysis has also been suggested recently.[3] Understanding the role of SAC in enzyme specificity is of considerable interest, as it may provide new insights into how substrates compete with each other for enzymes. Many of the previous studies of SAC[1-2] have focused on the catalytic effects resulting from the well-positioned substrate groups that can participate in SAC directly, without undergoing significant conformational changes. A fundamental question is whether some substrate groups which are not located at such ideal positions in enzyme-substrate complexes would be able to undergo conformational changes and participate in SAC as well. In this chapter, we report the results of computer simulations and show that the dynamics involving distant substrate groups may indeed play an important role in SAC. The importance of this type of SAC involving conformational changes of substrates might have been overlooked in some enzymes due in part to the dynamic nature of the effects that may not be well reflected in the X-ray structures.

The system used to demonstrate the importance of dynamics in substrate-assisted catalysis is kumamolisin-As, a member of a recently characterized sedolisin family of proteolytic enzymes.[4-7] The QM(SCC-DFTB)/MM(CHARMM) molecular dynamics (MD) and free energy simulations have been performed in this study.[8] Sedolisins have a

fold resembling that of subtilisin and a maximal activity at low pH (~3.9).[4-5] The defining features of kumamolisin-As and other members of this family are a unique catalytic triad, Ser278-Glu78-Asp82 (kumamolisin-As numbering), as well as the presence of an aspartic acid residue, Asp164, in the active site. Asp164 replaces Asn155 of subtilisin, a residue that creates the oxyanion hole.[6] Unlike Asn155 that provides hydrogen bonds to stabilize the oxyanion of the tetrahedral intermediate during the catalysis by subtilisin,[6] Asp164 seems to act as a general acid catalyst to protonate the tetrahedral intermediate and assist in the nucleophilic attack by Ser278 of kumamolisin-As.[7] The substrates used in the present work are RPXaa*FR (here * designates the scissile peptide bond and Xaa denotes the residue at the P_1 site). The formation of the tetrahedral intermediate (TI) is examined, as the efficiency of this step of the reaction is crucial for high substrate specificity of proteolytic enzymes.[6] The models were built using the X-ray structure of kumamolisin-As complexed with an inhibitor *N*-acetyl-isoleucyl-prolyl-phenylalaninal (AcIPF)[5b] and the substrate coordinates were obtained based on those from the corresponding residues in the Ser278Ala mutant of pro-kumamolisin.[5f] We demonstrate that there is a conformational transition of the His residue at the P_1 site for RPH*FR and that a salt bridge is formed between His and Asp164 as the system changes from the substrate to the TI complex. It is suggested that this conformational change and the formation of the salt bridge make it possible for the His residue to participate in SAC during the formation of the tetrahedral intermediate. Our suggestion for the existence of dynamic SAC is supported by the comparison of the free energy profiles of the TI formation for RPH*FR and RPF*FR.

Results and Discussions

The average structures of the active site of the kumamolisin-As complex during the nucleophilic attack by Ser278 on the RPH*FR substrate are given in Figure V-1. As is evident from Figure V-1A, Ser278 is well aligned for the nucleophilic attack on C in the substrate complex. Glu78 and Asp164 are well positioned to act as the general base and acid catalysts, respectively, to assist in the nucleophilic attack. His at the P_1 site forms two hydrogen bonds (H-bonds) to the C=O group of Pro at P_2 and the carboxylate of Asp179, respectively. Moreover, there is an H-bond between the backbone N-H group of the P_1 residue and Glu78. Figure V-1B shows that the interaction of the His sidechain with the C=O group is significantly weakened as the nucleophilic attack proceeds to the transition state, while the H-bond with Asp179 remains intact. The weakening of this interaction involving His seems to be triggered by the breaking of the H-bond between the backbone N-H group at P_1 and Glu78, as Glu78 accepts the proton from Ser278 and becomes uncharged during the nucleophilic attack. Thus, there appear to be the cooperative effects of hydrogen bonding[16] in the stabilization of the H-bond between the His sidechain and the P_2 C=O group in the substrate complex. Figure V-1C shows that the histidine sidechain rotated significantly (mainly around the C_β–C_γ bond) to interact with the unprotonated Asp164 residue during the formation of the TI complex. Since Asp164 is the general acid catalyst,[7] the formation of the salt bridge between His and Asp164 may make the TI complex more stable (see below).

The changes of free energy (potential of mean force) as functions of the reaction coordinate (ξ) for the nucleophilic attack by Ser278 on the RPH*FR and RPF*FR

substrates are given in Figure V-2. The free energy change for the RPH*FR complex in which the proton on Asp164 is fixed by the SHAKE algorithm[19] (to prevent Asp164 to act as the general acid) is also given. Figure V-2 shows that the free energy barrier of the TI formation is ~ 15 kcal/mol for RPH*FR, and this barrier increases to ~ 20 kcal/mol when the substrate is changed to RPF*FR. The major structural difference for the two substrates in the TI complexes is that His at the P_1 site in RPH*FR is able to form a salt bridge with Asp164 through a conformational transition (see above), and this salt bridge may help to stabilize the charge formation on Asp164 and make it a better general acid. In contrast, Phe at P_1 for RPF*FR is unable to interact with Asp164 and the stabilization of the TI through this mechanism is therefore impossible. It is of interest to note that the TI for RPH*FR becomes considerably less stable when the proton transfer from Asp164 to the ligand is prevented by using the SHAKE algorithm, supporting the suggestion that Asp164 acts as a general acid catalyst.[7]

The results of the simulations reported here are consistent with the available experimental data.[5] It has been shown that kumamolisin-As can cleave the –Pro–Xaa*Yaa– sequence and that this enzyme generally has higher specificity for the substrates with a positively charged residue at the P_1 site than for those without such a residue. For instance, it was found that the k_{cat} (K_M) value for a substrate containing the –Pro–Arg*Gly– sequence is ~ 65-fold (~ 6.5-fold) higher than for a substrate containing the –Pro–Gly*Gly– sequence, leading to 10-fold increase of k_{cat}/K_M.[5a] Moreover, a specificity profile analysis[5b] using a peptide library showed that kumamolisin-As generally displays higher specificity for the substrates with His, Lys or Arg at P_1 than for

the other substrates. Quantitative comparisons of the experimental data and theoretical results require determinations of the catalytic mechanism and rate limiting step which are poorly understood.

We suggest that the dynamics may play an important role in SAC and specificity of kumamolisin-As. The conformational change observed is triggered by the deprotonation/protonation events during the general acid-base catalysis. It is of interest to examine whether this would also occur for other enzymes.

Supporting Information: Computational Method

A fast semi-empirical density-functional approach (SCC-DFTB)[10] recently implemented in the CHARMM program[11] was used for quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) and free energy (potential of mean force or PMF) simulations. The initial coordinates of the enzyme were obtained from the crystal structure (PDB ID:1SIO. Resolution: 1.80 Å) of kumamolisin-As complexed with the inhibitor AcIPF (*N*-acetyl-isoleucylprolyl- phenylalaninal).[5b] The coordinates of the RPH*FR substrate were obtained from the crystal structure of the Ser278Ala mutant of pro-kumamolisin (PDB ID: 1T1E. Resolution 1.18 Å) which contains an activation peptide/linker running through the active site cleft. The catalytic domain of the Ser278Ala mutant of pro-kumamolisin has a virtually identical structure compared to the active kumamolisin.[5f] Kumamolisin-As and kumamolisin have a high sequence identity (92.7%), and their structures can be superimposed very well.[5b] The structures of the kumamolisin-As complex and the catalytic domain of the Ser278Ala mutant of pro-

kumamolisin were superposed, and the coordinates of the P_3-Arg169p—P_2-Pro170p—P_1-His171p—$P_{1'}$-Phe172p—$P_{2'}$-Arg173p peptide fragment from the linker were used to generate the structures of the RPH*FR substrate and the kumamolisin-As—RPH*FR complex; the backbone of P_3-Arg169p—P_2-Pro170p—P_1-His171p was superimposed with the inhibitor AcIPF very well. The structure of the kumamolisin-As—RPF*FR complex was generated simply by replacing His at P_1 by Phe; the Phe residue at the P_1 site of RPF*FR occupies essentially the same position as Phe of AcIPF in the X-ray structure of the kumamolisin-As—AcIPF complex.

The stochastic boundary molecular dynamics method[9] was used for QM(SCC-DFTB)/MM(CHARMM) simulations. The system was separated into a reaction zone and a reservoir region; the reaction zone was further divided into the reaction region and buffer region. The reaction region was a sphere with radius R of 16 Å, and the buffer region had R equal to 16 Å $\leq R \leq$ 18 Å. The reference center for partitioning the system was chosen to be the C atom of the His (Phe) carbonyl at P_1 of the substrate (Figure V-1A). The resulting system contains 3454 atoms (2522 enzyme atoms, 107 substrate atoms and 275 water molecules). The backbone of the substrates starting from the C atom of Pro at P_2 to the C_α atom of Phe at $P_{1'}$ was treated by QM along with the side chains of Glu32, Asp82, Glu78, Ser278 and Asp164. The rest of the system was treated by MM. The all-hydrogen potential function (PARAM22)[12] was used for MM atoms. The link atom approach[17, 18] was used to separate the QM and MM regions. A modified TIP3P water model[13] was employed for the solvent. The initial structure for the entire stochastic boundary system was optimized by adopted basis Newton-Raphson (ABNR) method.

The system was gradually heated from 50 to 300 K in 20 ps, and equilibrated at 300 K for 30 ps. A 1-fs time step was used for integration of the equations of motion, and the coordinates were saved every 50 fs for analyses. For the atoms in the buffer region, the Langevin Dynamics (LD) frictional constants were 250 ps^{-1} for the protein atoms and 62 ps^{-1} for the water molecules. The bonds involving hydrogen atoms in the MM region were fixed by the SHAKE algorithm.[19] The QM/MM MD simulations were performed for about 1 *ns* for the substrate complexes. The tetrahedral complexes were found to be unstable, and the C-O_γ bond was broken during the simulations, leading to the substrate complexes.

The umbrella sampling method[14] implemented in the CHARMM program along with the Weighted Histogram Analysis Method (WHAM) [15] was applied to determine the changes of the free energy (potential of mean force). The free energy curves for the wild-type enzyme (with and without the QM treatment of D164) complexed with the RPH*FR and RPF*FR substrates were obtained as functions of $\xi = r(\text{C-O}_\gamma)$ for the nucleophilic attack by Ser278 on the C atom of the substrate between His (Phe) at P_1 and Phe at $P_{1'}$ to form the tetrahedral intermediate. Harmonic biasing potentials were defined by the RESD module of CHARMM, with force constants in range of 100~800 $kcal \cdot mol^{-1} \cdot Å^{-2}$. The statistical error of the free energy simulations[20] was estimated to be about ± 0.6 kcal/mol.

Acknowledgment

We thank Professor Martin Karplus for a gift of the CHARMM program and an anonymous reviewer of the earlier paper for pointing out a possible role of dynamics in

aspartic protease catalysis. This work was supported in part by the Center of Excellence for Structural Biology, the University of Tennessee and the Petroleum Research Fund (HG) and in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research (AW).

References

(1) (a) Dall'Acqua, W.; Carter, P. *Protein Sci*., **2000**, 9, 1-9.

(b) Kosloff, M.; Selinger, Z. *Trends Biochem. Sci.* **2001**, *26*, 161-166.

(2) (a) Weinger, J. S.; Parnell, K.M.; Doner, S.; Green, R.; Strobel, S.A. *Nat. Struct. Mol. Biol*. **2004**, *11*, 1101-1106.

(b) Jogl G.; Tong, L. *Cell*, **2003**, *112*, 113-122.

(c) Ryan, M.; Liu, T.; Dahlquist, F. W.; Griffith, O. H. *Biochem*., **2001**, *40*, 9743-9750.

(d) Breldenbach, M. A.; Brunger, A. T. *Nature*, **2004**, *432*, 925-929.

(3) Fromme, J.C.; Bruner, S. D. ; Yang, W. ; Karplus, M. ; Verdine, G. L. *Nat. Struct. Mol. Biol*. **2003**, *10*, 204-210.

(4) (a) Wlodawer, A.; Li, M.; Gustchina, A.; Oyama, H.; Dunn, B. M.; Oda, K. *Acta Biochim. Pol*. **2003**, *50*, 81–102.

(b) Vandeputte-Rutten, L.; Gros, P. *Curr. Opin. Struct. Biol*. **2002**, *12*, 704–708.

(c) Powers, J. C.; Asgian, J. L.; Ekici, O. D.; James, K. E. *Chem. Rev*. **2002**, *102*, 4639–4750.

(d) Wlodawer, A. *Structure*, **2004**, *12*, 1117–1119.

(5) (a) Tsuruoka, N.; Nakayama, T.; Ashida, M.; Hemmi, H.; Nakao, M.; Minakata, H.; Oyama, H.; Oda, K.; Nishino, T. *Appl. Environ. Microbiol*. **2003**, *69*, 162-169.

(b) Wlodawer, A.; Li, M.; Gustchina, A.; Tsuruoka, N.; Ashida, M.; Minakata, H.; Oyama, H.; Oda, K.; Nishino, T.; Nakayama, T. *J. Biol. Chem*. **2004**, *279*, 21500–21510.

(c) Wlodawer, A.; Li, M.; Dauter, Z.; Gustchina, A.; Uchida, K.; Oyama, H.; Dunn, B. M.; Oda, K. *Nature Struct. Biol.* **2001**, *8*, 442–446.

(d) Wlodawer, A.; Li, M.; Gustchina, A.; Dauter, Z.; Uchida, K.; Oyama, H.; Goldfarb, N. E.; Dunn, B. M.; Oda, K. *Biochemistry*, **2001**, *40*, 15602–15611.

(e) Comellas-Bigler, M.; Fuentes-Prior, P.; Maskos, K.; Huber, R.; Oyama, H.; Uchida, K.; Dunn, B. M.; Oda, K.; Bode, W. *Structure*, **2002**, *10*, 865–876.

(f) Comellas-Bigler, M.; Maskos, K.; Huber, R.; Oyama, H.; Oda, K.; Bode, W. *Structure* **2004**, *12*, 1313–1323.

(6) (a) Hedstrom, L. *Chem. Rev*. **2002**, *102*, 4501–4523.

(b) Vandeputte-Rutten, L.; Gros, P. *Curr. Opin. Struct. Biol*. **2002**, *12*, 704–708.

(c) Powers, J. C.; Asgian, J. L.; Ekici, O. D.; James, K. E. *Chem. Rev*. **2002**, *102*, 4639–4750.

(d) Perona, J. J.; Craik, C. S. *J. Biol. Chem*., **1997**, *272*, 29987-29990.

(7) Guo, H. B.; Wlodawer, A.; Guo, H. *J. Am. Chem. Soc*. **2005**, *127*, 15662-15663.

(8) For details of the computational method, see the **supporting information** above.

(9) Brooks, C. L., III.; Brunger, A.; Karplus, M. *Biopolymers* **1985**, *24*, 843.

(10) Cui, Q.; Elstmer, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569–585.

(11) Brooks, B., R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem*. **1983**, *4*, 187–217.

(12) Mackerell, A. D., Jr.; Bashford, D.; Bellott, M.; Bunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy D.; Kuchnir, L.;

Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E. III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; W.-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586-3616.

(13) (a) Jorgensen, W. L. *J. Am. Chem. Soc*. **1981**, *103*, 335–340.

(b) Neria, E.; Fischer, S.; Karplus, M., *J. Chem. Phys*. **1996**, *105*, 1902–1921.

(14) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett*. **1974**, *28*, 578–581.

(15) Kumar, M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem*. **1992**, *13*, 1011-1021.

(16) Guo, H.; Salahub, D. R. *Angew. Chem. Int. Ed. Engl*. **1998**, *37*, 2985-2990.

(17) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720-1735.

(18) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700-733.

(19) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, 23, 327-341.

(20) Kobrak, M.N. *J. Comput. Chem.* **2003**, 24, 1437-1446.

Appendix

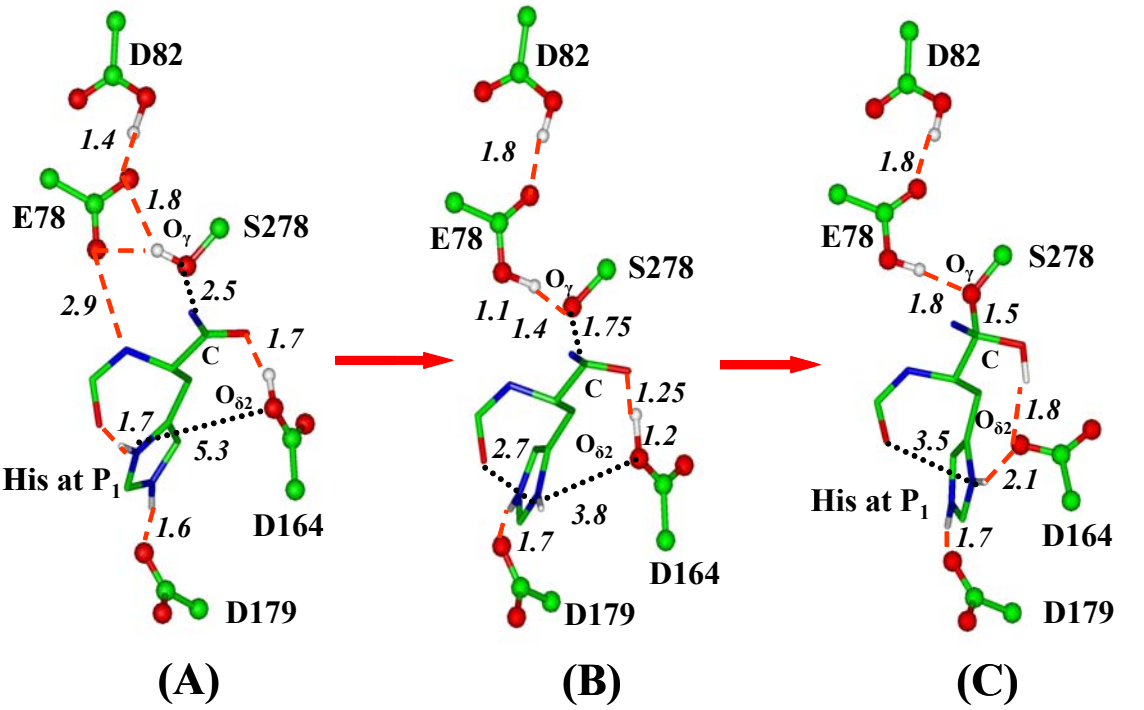


Figure V-1. Average structures of kumamolisin-As complex during the nucleophilic attack

Average structures of kumamolisin-As complex during the nucleophilic attack by Ser278 on substrate RPH*FR obtained from the QM/MM MD and free energy simulations. The enzyme is plotted in ball and stick and the ligand in stick. Hydrogen bonds (H-bonds) are indicated by red dashed lines. Ser278-Glu78-Asp82 is the catalytic triad. Only the side chain of the P1 (His) residue for the substrate is shown for clarity. **(A)** The substrate complex. His side chain at P_1 donates an H-bond to the C=O group of Pro at P_2, in addition to the carboxylate of Asp179. Glu78 accepts an H-bond from the backbone N-H group of the P_1 residue (the H-bond distance shown here is between the non-hydrogen atoms). **(B)** An average structure near the transition state of the nucleophilic attack. The interaction of the His sidechain with the C=O group is significantly weakened and the H-bond between Glu78 and the N-H group is broken. **(C)** The TI complex. His at P_1 has undergone a conformational change to interact with the unprotonated Asp164 residue and the H-bond to the C=O group observed in the substrate complex is broken.

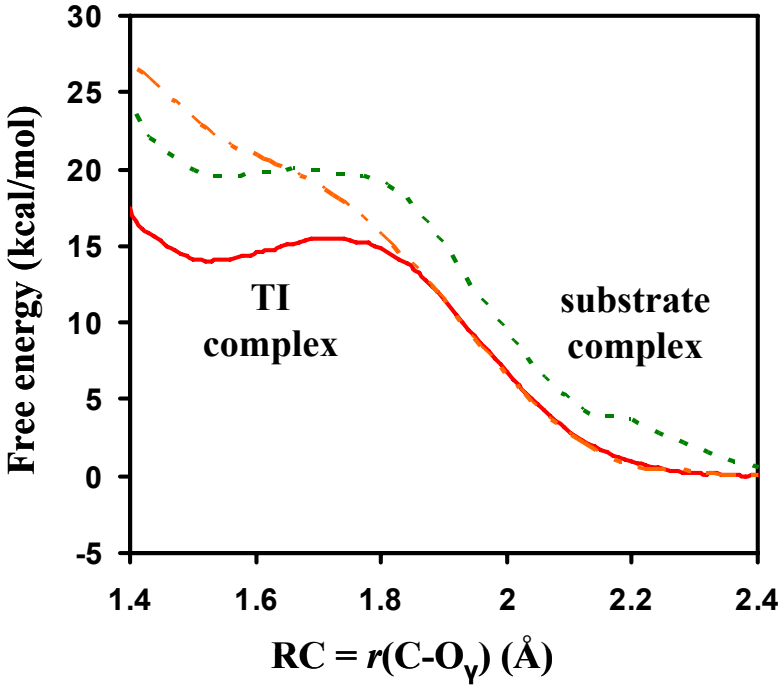


Figure V-2. The free energy changes of the nucleophilic attacks

The free energy changes from the substrate to TI complex as a function of $\xi = r(C–O_\gamma)$, the reaction coordinate for the nucleophilic attack. **Red solid line**: the RPH*FR substrate without use of the SHAKE algorithm; **Orange dot-dashed line**: the RPH*FR substrate with use of the SHAKE algorithm to fix the proton on Asp164; **Green dashed line**: the RPF*FR substrate.

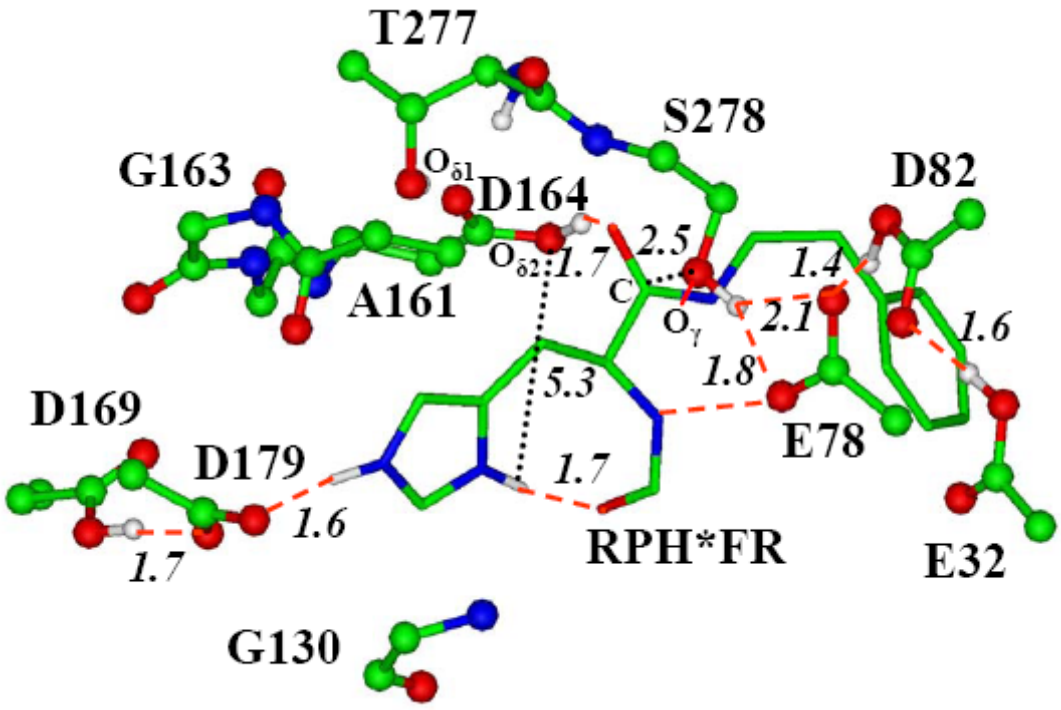


Figure V-S1. The average structure of the substrate complex for RPH*FR

The average structure of the substrate complex for RPH*FR obtained from MD simulations. The enzyme is plotted in ball and stick and the substrate in stick. Hydrogen bonds are indicated by red dashed lines. Ser278-Glu78-Asp82 is the catalytic triad. The S1 binding pocket is formed by Ala161, Gly163, Gly130, Thr277 and Asp179, as in the X-ray structure of the kumamolisin-As-AcIPF complex. Only the sidechains of the P_1 (His) and P_1' (Phe) residues of the substrate are shown for clarity.

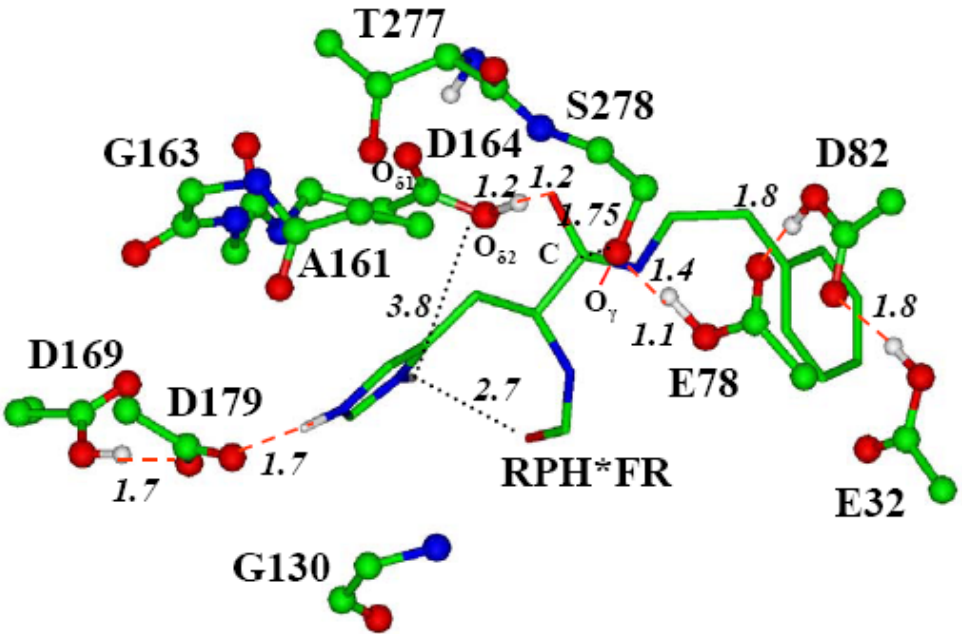


Figure V-S2. The average structure of the transition state complex for RPH*FR

The average structure at r(C-Oγ) = 1.75 Å (i.e., near the transition state) for the RPH*FR complex during the nucleophilic attack obtained from the free energy simulations.

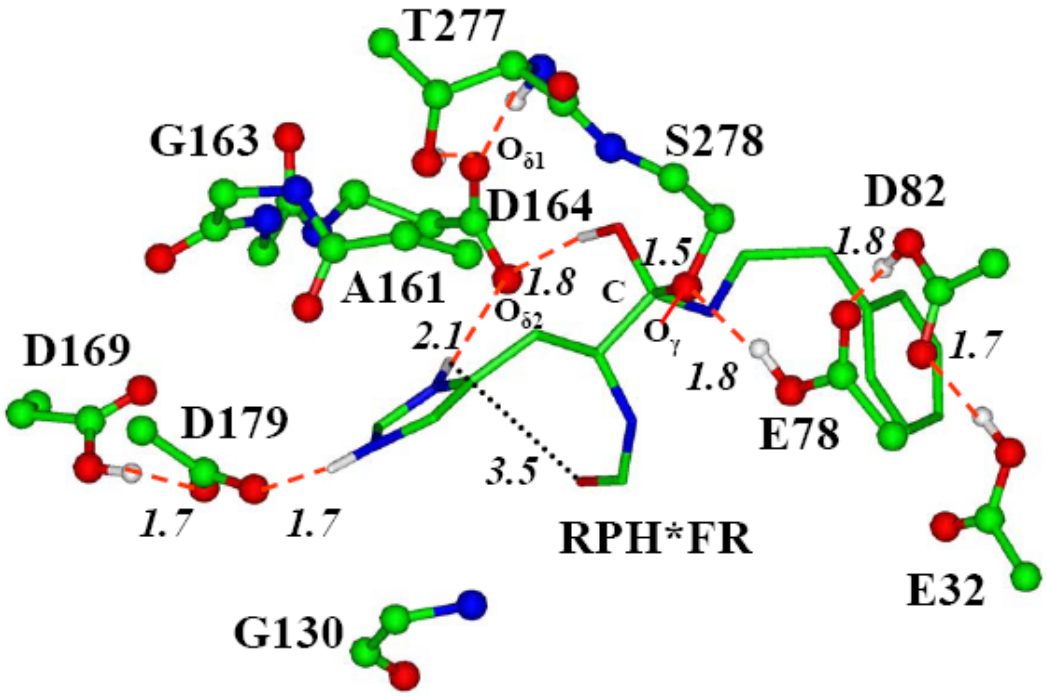


Figure V-S3. The average structure of the TI complex for RPH*FR

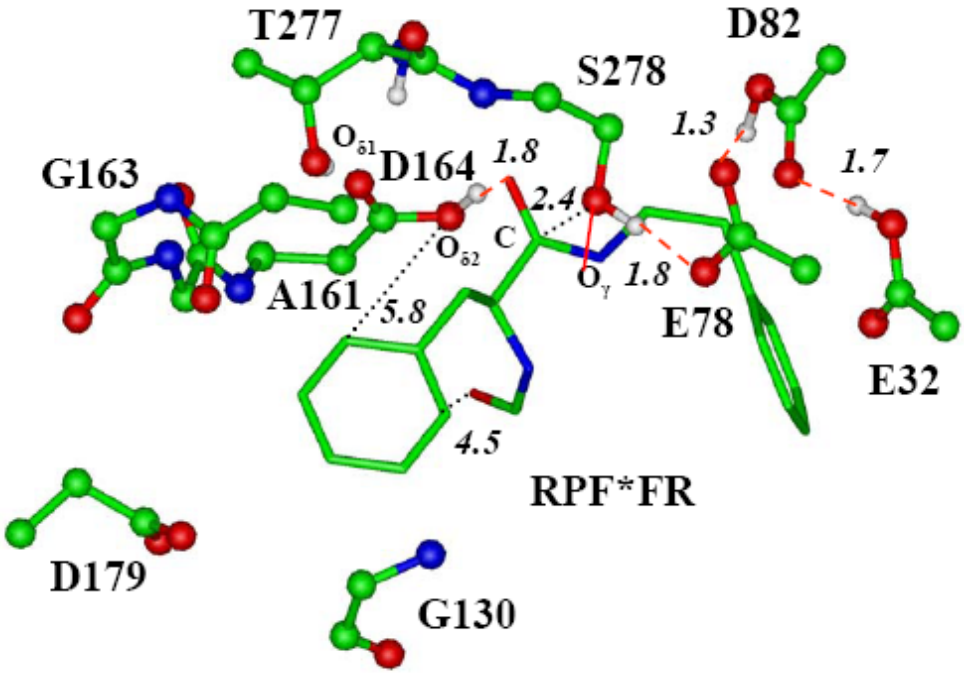


Figure V-S4. The average structure of the substrate complex for RPF*FR

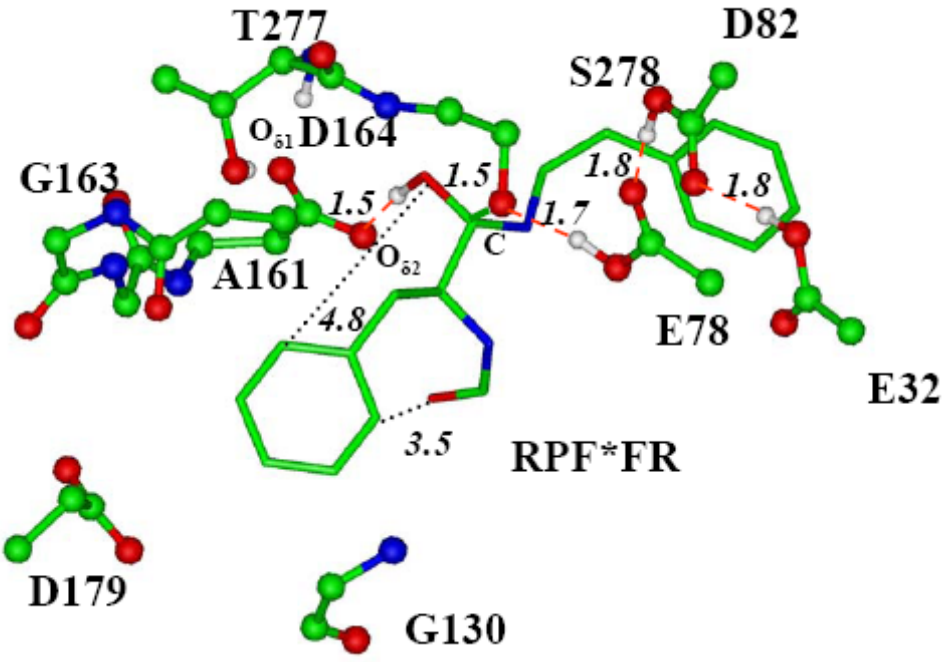


Figure V-S5. The average structure of the TI complex for RPF*FR

Chapter VI

The QM/MM molecular dynamics and free energy simulations of the acylation reaction catalyzed by the serine carboxyl peptidase kumamolisin-As

Qin Xu, Haobo Guo, Alexander Wlodawer, Toru Nakayama, and Hong Guo

Revised from *Biochemistry* **2007**, 46, 3784-3792

Abstract

Quantum mechanical/molecular mechanical molecular dynamics and free energy simulations are performed to study the acylation reaction catalyzed by kumamolisin-As, a serine-carboxyl peptidase, and to elucidate the catalytic mechanism and the origin of substrate specificity. It is demonstrated that the nucleophilic attack by the serine residue on the substrate may not be the rate limiting step for the acylation of the GPH*FF substrate. The present study also confirms the earlier suggestions that Asp164 acts as a general acid during the catalysis and that the electrostatic oxyanion-hole interactions may not be sufficient to lead a stable tetrahedral intermediate along the reaction pathway. Moreover, Asp164 is found to act as a general base during the formation of the acyl-enzyme from the tetrahedral intermediate. The role of dynamic substrate assisted catalysis (DSAC) involving His at the P_1 site of the substrate is examined for the acylation reaction. It is demonstrated that the bond breaking and making events at each stage of the reaction trigger a change of the position for the His sidechain and lead to the formation of the alternative hydrogen bonds. The back and forth movements of the His sidechain between the C=O group of Pro at P_2 and $O_{\delta 2}$ of Asp164 in a ping-pong-like mechanism and the formation of the alternative hydrogen bonds effectively lower the free energy barriers for both the nucleophilic attack and the acyl-enzyme formation and may therefore contribute to the relatively high activity of kumamolisin-As towards to the substrates with His at P_1 site.

Abbreviations:

DSAC, Dynamic substrate assisted catalysis; SAC, substrate-assisted catalysis; QM/MM Quantum mechanical/molecular mechanical; MD molecular dynamics; TI, Tetrahedral intermediate; PMF, potential of mean force; SCC-DFTB, self consistent charge density-functional tight-binding; TS, transition state; WHAM, weighted histogram analysis method.

Introduction

Kumamolsin-As belongs to the recently characterized family of serine-carboxyl peptidases (sedolisins) that were originally described by Murao, Oda and co-workers about 20 years ago. [1-3] Sedolisins are present in a wide variety of organisms, including archaea, bacteria, molds, slime molds (mixomycetes), amoebas, fishes, and mammals, and they are active at low pH and often high temperature. This family of enzymes seems to have been derived through divergent evolution from a common ancestor with classical serine peptidases; the structural studies [4-10] showed that sedolisins have a fold resembling that of subtilisins, although they are significantly larger (the mature catalytic domains contain approximately 375 amino-acid residues). One interesting question is why Nature needs another family of subtilisin-like peptidases. For bacterial sedolisins, one possible explanation is that some organisms might have to utilize such evolved serine-carboxyl peptidases for their survival, presumably under acidic environments and high temperature. The biological importance of sedolisins in mammals has already been demonstrated by the fact that mutations leading to the loss of the human enzyme CLN2[11-13], another member of the sedolisin family, result in a fatal neurodegenerative disease, classical late-infantile neuronal ceroid lipofuscinosis.[14]

The three-dimensional structures are available for three members of the sedolisin family, including sedolisin (also known as pepstatin-insensitive carboxyl proteinase or PSCP) [4-6], kumamolisin [7-8], and kumamolisin-As [9-10]. The defining features of these enzymes are a unique catalytic triad, Ser-Glu-Asp (Ser278-Glu78-Asp82 for kumamolisin-As; see Figure VI-1A), as well as the presence of an aspartic acid residue

(Asp164 for kumamolisin-As) that replaces Asn155 of subtilisin, a residue that creates the oxyanion hole in the classical subtilisin-like serine peptidase. The X-ray crystallographic and mutagenesis studies [4-10] demonstrated that the serine residue is the catalytic nucleophile for sedolisins, while the nearby Glu is likely to act as the general base to accept the proton from Ser and assist in the nucleophilic attack. Recent computational studies [15-17] further suggested that Asp164 may act as a general-acid catalyst to protonate the substrate and stabilize the tetrahedral intermediate (TI). This mechanism of the TI stabilization is similar to the one proposed for aspartic peptidases[18], but is different from that of serine peptidases, in which the electrostatic oxyanion-hole interactions are used instead [19]. Thus, although sedolisins might have evolved from a common precursor of classical serine peptidases, they seemed to have ended up with the use of some of the chemistry of aspartic peptidases for the catalysis. The existence of significant mechanistic similarities between sedolisins and both serine and aspartic peptidases makes the detailed understanding of this newly characterized family of enzymes extremely interesting.

Here we examine the acylation reaction catalyzed by kumamolisin-As of the GPH*FF substrate (* designates the scissile peptide bond), using quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) and free energy (potential of mean force) simulations. Although the enzyme has been a subject of previous computational studies [15-17], the earlier investigations have mainly focused on the inhibitor binding and the nucleophilic attack by Ser278. To the best of our knowledge, a detailed examination of the acylation reaction as well as the corresponding free energy

barriers has not been performed for this enzyme, or indeed for any sedolisins. Kumamolisin-As can cleave the –Pro–Xaa*Yaa– sequence (here Xaa and Yaa denote the residues at the P_1 and P_1' sites, respectively). One experimental observation that is of considerable interest is the fact that this enzyme generally has higher specificity for the substrates with a positively charged residue at the P_1 site than for those that do not carry such charge [3, 9]. For instance, it was found that the k_{cat} value for a substrate containing the –Pro–Arg*Gly sequence is ~ 65-fold higher than for a substrate containing the –Pro–Gly*Gly sequence and K_M ~ 6.5-fold higher, leading to 10-fold increase of k_{cat}/K_M (*3*). The higher specificity for the substrates with a positively charged residue at P_1 was also confirmed from specificity profile analysis using a peptide library [18-20]. It was suggested in our earlier computational study[16] that the dynamics involving His at P_1 triggered by the bond breaking and making events may play an important role in the stabilization of the tetrahedral intermediate and therefore contribute to the relatively high specificity for the substrates with His at P_1. This type of catalytic processes is normally referred to as substrate-assisted catalysis (SAC) [21], with one or more functional groups from the substrate, in addition to those from the enzyme, contributing to the rate acceleration for the enzyme-catalyzed reaction. However, many of the previous studies of SAC [22] have focused on the catalytic effects resulting from the well-positioned substrate groups that can participate in SAC directly, without undergoing significant conformational changes. A major difference in the case of kumamolisin-As is that His at the P_1 site is not located at such an ideal position in the enzyme-substrate complex, but is able to undergo a conformational change for the TI stabilization during the nucleophilic attack. We define

this type of SAC involving conformational changes of the substrates as dynamic SAC (DSAC), to distinguish it from the standard SAC for which the conformational changes of the substrate are not required. It is shown here that such DSAC involving His at P_1 also plays an important role for the acyl-enzyme formation. The importance of DSAC or similar types of catalytic processes involving the conformational changes of protein groups triggered by the bond breaking and making events might have been overlooked in some other enzymes, due in part to the dynamic nature of the effects that may not be well reflected in the X-ray structures.

Here we demonstrate that the nucleophilic attack by the serine residue on the substrate may not be the rate limiting step for the acylation reaction for kumamolisin-As, which is different from some previous calculations on classical serine peptidases [20]. The results support our earlier conclusion that the general acid mechanism involving Asp164 is crucial for the stabilization of the tetrahedral intermediate during the catalysis and that the electrostatic oxyanion-hole interactions may not be sufficient to generate a stable TI along the reaction pathway [15-17]. Moreover, Asp164 is found to act as a general base during the formation of the acyl-enzyme from the tetrahedral intermediate, and therefore plays multiple roles in the catalysis. The present study also provides important new insights into the role of DSAC. The previous suggestion [16] for the existence of DSAC was based on the results of the nucleophilic attack on a substrate (RPH*FR). Since the nucleophilic attack may not be rate limiting for the acylation (see above), a question remains as to whether the dynamics involving the His residue at P_1 would also play a role in stabilizing the transition state for the acyl-enzyme formation (so that an overall rate

enhancement for the acylation reaction could be achieved). In this paper, it is shown that the bond breaking and making events at the different stages of the acylation reaction trigger the back and forth movements of the His sidechain at P_1 between the C=O group of Pro at P_2 and $O_{\delta 2}$ of Asp164, leading to the formation of the alternative hydrogen bonds and the reduction of the free energy barriers for both the nucleophilic attack and formation of the acyl-enzyme. It is proposed that such motions and the formation of the alternative hydrogen bonds may contribute to the relatively high activity of kumamolisin-As towards the substrates with His at the P_1 site and lead to the reduction of the free energy barriers for the de-acylation reaction as well.

Methods

A fast semi-empirical density-functional method (SCC-DFTB) [23] implemented in the CHARMM32b program [24] was used in the present study for the QM/MM molecular dynamics (MD) and free energy (potential of mean force or PMF) simulations. The efficiency of the semi-empirical QM methods (such as SCC-DFTB) makes possible to sample millions of the structures and conformations for enzyme systems and to determine the free energy profiles of the enzyme-catalyzed reactions. These important tasks for understanding enzymes are not feasible with high-level first-principle *ab initio* methods. Moreover, recent testing studies and comparisons with high-level *ab initio* QM methods on model systems seem to indicate that the SCC-DFTB method offers certain advantages over some other semi-empirical QM methods [25-27]. However, it should be pointed out that care must be exercised in using the semi-empirical QM methods for the studies of

enzymes. Indeed, while the high-level QM/MM approaches based on energy-minimization may face the intrinsic problem for the lack of sampling sufficient conformational spaces, the accuracy of the semi-empirical QM/MM MD and free energy simulations may be limited by the approximations in their descriptions of the breaking and formation of chemical bonds as well as interactions. For instance, the deficiencies of the SCC-DFTB method used here could affect some of the conclusions reached from this study (even though the SCC-DFTB energies are not used directly to generate the profiles). Thus, the comparisons of performance of the SCC-DFTB method on certain models as well as on simplified enzyme systems with those from high-level *ab intio* calculations are of fundamental importance. This would allow us to obtain important insights into the question as to whether the conclusions obtained from the QM/MM MD and free energy simulations could still hold if high-level QM methods would be used.

Although the SCC-DFTB method has been tested on a large number of models and enzyme systems [25-27], additional comparisons with the results from high-level *ab initio* calculations for the systems related to the present study are necessary. In our earlier work [15], we compared the performance of the QM(SCC-DFTB)/MM simulations with the calculations based on the QM(B3LYP/6-31G**)/MM energy minimization for the process of the nucleophilic attack of Ser278 on an inhibitor *N*-acetyl-isoleucyl-prolyl-phenylalaninal (AcIPF) in the kumamolisin-As complex. We found that the results from the both methods are consistent with each other concerning the role of Asp164 and Glu78 as the general acid-base catalysts. There are two additional conclusions obtained from the present study that could also be affected by possible deficiencies of SCC-DFTB. One

conclusion is that the nucleophilic attack by Ser278 on the substrate may not be rate limiting for the acylation reaction and the other is the existence of dynamic substrate assisted catalysis (DSAC). The first conclusion could be affected if the SCC-DFTB method underestimates (overestimates) the activation barrier for the nucleophilic attack more (less) significantly than the activation barrier for the acyl-enzyme formation (i.e., the breaking of the peptide bond). In this case, the relatively low barrier for the nucleophilic attack might be simply due to the deficiencies of SCC-DFTB. We have compared the performance of SCC-DFTB with B3LYP/6-31G** on a small model system and on selected snapshots of the enzyme complex. It was found that the SCC-DFTB method may in fact underestimate the activation barrier for the acyl-enzyme formation more significantly than that for the nucleophilic attack. Thus, the first conclusion mentioned above is likely to be true even with the high level *ab initio* method. For the conclusion concerning the existence of DSAC, the high-level *ab initio* MD simulations could not be performed due to the reason that these simulations are extremely expensive. In order to confirm this conclusion, we changed the treatment of His at P_1 from SCC-DFTB to MM [29]. The same conformational transition involving His at P_1 was also observed, confirming our conclusion of the existence of DSAC.

There are additional questions that need to be mentioned for the model building processes. The family of sedolisins is only active at low pH, and the models used in the present study (e.g., the information concerning the protonation states of Asp and Glu residues) were based on the computer models of the tetrahedral adduct complex obtained in the earlier studies [15-17], in which the models were manually modified by adding protons

on some ionizable residues (e.g., Asp and Glu), based on their local environments. In other words, the models were built to mimic the structures determined at low pH and to represent the active enzyme at low pH (i.e., not at neutral pH at which the enzyme is inactive). The average structure from the 1.2 ns QM/MM MD simulations for the tetrahedral adduct complex was then performed to test whether the experimental X-structure could be well reflected from the models that we built. If the interactions in the experimental structure could not be well represented by a model, the structure would be expected to deviate significantly from the experimental structure during the simulations. After several tests, we were able to generate the model for which the average structure from the MD simulations was rather close to the experimental structure obtained at low pH, suggesting that the interactions in the tetrahedral adduct complex are well represented and that the model should be reasonable. For instance, the distances between the non-hydrogen atoms for the key hydrogen bonds in the active site from the simulations are very close to those in the experimental structure, with an average deviation of only about 0.1 Å. We also tested other cases and noticed that we could not reproduce the X-structures from the simulations if we used different (presumably incorrect) protonation states for certain residues.

The initial coordinates for the enzyme were taken from the crystal structure (PDB ID: 1SIO, resolution 1.8 Å) of kumamolisin-As complexed with the inhibitor AcIPF (*N*-acetyl-isoleucyl-prolyl-phenylalaninal) [9-10]. The substrate used in the present study, GPH*FF, contains the key residues of the internally quenched fluorogenic (IQF) substrate [NMA-MGPH*FFPK-(DNP)DRDR] used in the earlier experimental study [9-10],

although it is considerably smaller. One of the key questions for performing the simulations on kumamolisin-As is how to obtain the coordinates for the substrate (tetrahedral intermediate) in the enzyme complex. Fortunately, an X-ray structure for the S278A mutant of pro-kumamolisin (PDB ID: 1T1E. Resolution 1.18 Å) [8] is available that makes generating the enzyme-complex possible. This X-ray structure contains an activation peptide/linker running through the active site cleft. The catalytic domain of the Ser278Ala mutant of pro-kumamolisin has a virtually identical structure compared to the active kumamolisin. Kumamolisin-As and kumamolisin share high sequence identity (92.7%) and their structures superimpose very well. The structures of the kumamolisin-As-AcIPF complex and the catalytic domain of the S278A mutant of pro-kumamolisin were first superposed, and the coordinates of the P_3-Arg169p—P_2-Pro170p—P_1-His171p—P_1'-Phe172p—P_2'-Arg173p peptide fragment from the linker were used to generate the coordinates of the GPH*FF substrate, with P_3-Arg169p replaced by Gly and P_2'-Arg173p replaced by Phe. The coordinates of kumamolisin-As and the GPH*FF peptide were then combined to generate the putative kumamolisin-As substrate complex used to initiate the computational studies.

The stochastic boundary MD method [28] was used for the QM(SCC-DFTB)/MM simulations. The system was separated into a reaction zone and a reservoir region; the reaction zone was further divided into the reaction region and the buffer region. The reaction region was a sphere with radius R of 16 Å, and the buffer region had R equal to 16 Å $\leq R \leq$ 18 Å. The reference center for partitioning the system was chosen to be the carbonyl carbon atom (C) of the His residue at the P_1 site (see Figure VI-1). The side

chains of Glu32, Asp82, Glu78, Ser278, and Asp164, as well as a part of the substrate [the backbone atoms of P_2-Pro (C=O), P_1-His, P_1'-Phe (C_α-NH) and the sidechain of P_1-His] were treated by QM and the rest of the system by MM. The all-hydrogen potential function (PARAM22) [29] was used for MM atoms. The link-atom approach [30] available in the CHARMM program [24] was used to separate the QM and MM regions. A modified TIP3P water model [31, 32] was employed for the solvent, and explicit water spheres covering the whole region of interest were used to solvate the system using the standard procedure in the CHARMM program. The resulting system contains 3431 atoms (2630 enzyme atoms and 267 water molecules, including 79 crystal water molecules). The initial structure for the entire stochastic boundary system was optimized by adopted basis Newton-Raphson (ABNR) method. The system was gradually heated from 50 to 310 K in 20 ps and equilibrated at 300 K for 80 ps. A 1-fs time step was used for integration of the equations of motion, and coordinates were saved every 10 fs for analyses. The LD frictional constants were 250 ps^{-1} for the protein atoms and 62 ps^{-1} for the water molecules. The bonds involving hydrogen atoms in the MM region were fixed by the SHAKE algorithm (*33*). 500 *ps* QM/MM MD simulations were performed for the substrate, tetrahedral intermediate and acyl-enzyme complexes of the wild-type enzyme; the initial structures of the tetrahedral intermediate and acyl-enzyme complexes were obtained from the free energy simulations (see below). The fluctuations of some important distances related to the proton motions were monitored with the coordinates obtained from the trajectories of the 500 *ps* simulations.

The umbrella sampling method [34] implemented in the CHARMM program and Weighted Histogram Analysis method (WHAM) [35] were applied to determine the change of the free energy (PMF) for the acylation reaction in wild-type. The reaction coordinate [ξ = R(C-N) – R(C…O_γ)] used in the study is the difference between the bond distance of the scissile peptide bond [R(C-N)] and the R[O_γ(Ser278)…C] distance of the nucleophilic attack (see Figure VI-1A), the two key distances for the change of the substrate to the acyl-enzyme. The determination of multi-dimensional free energy maps involving all the coordinates would be too time-consuming. Our previous study [17] indicated that the one-dimensional free energy simulations with the suitable reaction coordinate would be sufficient to determine the key energetic properties for the reaction. The free energy profile for the wild-type enzyme, with the proton fixed on Asp164 using the SHAKE algorithm [30], was also obtained. 18-22 windows along the reaction coordinate were used between the substrate complex and acyl-enzyme. The force constants that were used are in the range of 100-800 kcal·Mol^{-1}·$Å^{-2}$. Similar free energy profiles and the same conclusions were obtained from three different sets of PMF calculations with the MD simulations ranging from 40 ps (20 ps equilibration and 20 ps production run) to 100 ps (50 ps equilibration and 50 ps production run) for each window.

Results and Discussion

Structural and dynamic properties of the substrate complex, tetrahedral intermediate, and acyl-enzyme

The average active-site structure for the kumamolisin-As-substrate (GPH*FF) complex obtained from the QM/MM MD simulations is given in Figure VI-1A. Ser278 is the nucleophile that attacks the carbonyl carbon atom (C) of the substrate, while Glu78 and Asp164 act as the general base and acid catalysts, respectively [4-10, 15-17]. Two additional residues, Glu32 and Trp129, interact with Asp82 of the catalytic triad through hydrogen bonding networks. These two residues may play an important role in maintaining the structural integrity of the active site. Indeed, the previous experimental study (*8*) on kumamolisin showed that Glu32→Ala and Trp129→Ala mutations affect the active-site conformation, leading to a 95% decrease in the activity. The backbone N-H group and sidechain of Thr277 interact with Asp164, and it was suggested in the earlier computational study [15] that these interactions might play a role in stabilizing negative charge formation on Asp164 during the nucleophilic attack. The motions of the three protons that may be important for catalysis (i.e., those between Asp164 and substrate, Ser278 and Glu78, and Glu78 and Asp82) are monitored from the trajectories of the MD simulations of the substrate complex and are plotted in Figure VI-1B. Figure VI-1B shows that, for the substrate complex, these protons are located on Asp164, Ser278 and Asp82, respectively, consistent with the average structure shown in Figure VI-1A. The fluctuation of the H-O(D82) distance is larger than that for a normal H-O bond [e.g., comparing the fluctuations with those of H-O(D164) and H-O(S278)], indicating that the

corresponding H-O(D82) covalent bond is weakened as a result of the strong hydrogen bonding interaction with Glu78 (the distance of the hydrogen bond between Glu78 and Asp82 is only 1.5 Å).

The average structure for the tetrahedral intermediate obtained from the simulations is shown in Figure VI-1C. In this structure, the Ser278 residue has already attacked the substrate and transferred its proton to Glu78. The proton on Asp164 has transferred to the substrate during the formation of the tetrahedral intermediate. Thus, Glu78 and Asp164 played the roles of the general base and acid catalysts, respectively, during the nucleophilic attack. Preventing the proton transfer from Asp164 to the substrate increases the free energy for the formation of the tetrahedral intermediate significantly (see below), highlighting the importance of the role of Asp164 as the general acid catalyst. A similar discussion can be made for Glu78 [17]. It is of interest to note from Figure VI-1D (bottom panel) that the H-O(D82) bond does not undergo the large fluctuations (as observed in the substrate complex) in the tetrahedral intermediate; the corresponding H-bond distance to Glu78 is also longer (~1.9 Å). The average structure of the acyl-enzyme is given in Figure VI-1E, and the fluctuations of the distances related to the proton motions are plotted in Figure VI-1F. Figures VI-1E and VI-1F show that Glu78 has given up its proton to the nitrogen atom (N) of Phe at P_1', leading to the breaking of the C-N peptide bond between P_1-His and P_1'-Phe and the formation of the acyl-enzyme. Asp164 becomes protonated again and therefore acted as a general base during the acyl-enzyme formation. It is interesting to note that a low-barrier hydrogen bond is formed between Glu78 and Asp82 in the acyl-enzyme (Figure VI-1F,

bottom panel); the proton is almost equally distributed on Glu78 and Asp82 during the MD simulations.

Free energy profiles of the acylation

The changes of free energy (potential of mean force) as functions of the reaction coordinate (ξ) for the acylation reaction in kumamolisin-As and D164A mutant are given in Figure VI-2A. The free energy change in which the proton on Asp164 is fixed by the SHAKE algorithm [30] (to prevent Asp164 acting as the general acid and base catalyst for the nucleophilic attack and acyl-enzyme formation, respectively) is also given. It should be pointed out the free energy profiles in Figure VI-2A were based on the one-dimensional free-energy simulations. For the step-wise bond breaking and making events this may not pose a serious problem. However, for the concerted processes some information that would be contained in multi-dimensional free-energy profiles may be absent in the one-dimensional profiles. Indeed, there are several proton transfers during the acylation reaction, and some of these proton transfers may be coupled to the rearrangement of the system (i.e., the nucleophilic attack and the formation of the acyl-enzyme). For instance, our earlier two-dimensional free energy simulations on the nucleophilic attack of Ser-278 on AcIPF in kumamolisin-As showed that there are important free-energy relationships between the nucleophilic attack and the proton transfers. Such free energy relationships between the nucleophilic attack (acyl-enzyme formation) and proton transfers are absent in the one-dimensional free energy profiles (Figure VI-2A). However, the previous study [17] also indicates that one-dimensional free

energy simulations with the suitable reaction coordinate should be sufficient to determine the key energetic properties for the acylation reaction.

Figure VI-2A shows that the nucleophilic attack by the serine residue on the substrate is not rate limiting for the acylation reaction, although the free energy barriers for the nucleophilic attack (TS1) and formation of the acyl-enzyme (TS2) do not differ significantly, with the barrier at TS2 slightly higher. The result obtained for kumamolisin-As is different from the previous calculations on trypsin [20] which suggested that the nucleophilic attack is rate limiting for the acylation in the serine peptidase. It is of interest to note in Figure VI-2A that preventing the proton transfer away from Asp164 leads to an increase of the free energy barrier of the acylation reaction by about 10 kcal/mol and the tetrahedral intermediate along the reaction pathway disappeared as a result of fixing the proton on Asp164. Thus, the electrostatic oxyanion-hole interaction involving the protonated Asp164 is unlikely to be sufficient for generating a stable tetrahedral intermediate during catalysis. Figure VI-2A also shows that the replacement of Asp164 by Ala increases the free energy barrier considerably. The results of the simulations reported here support the earlier proposal [15] for the existence of the general-acid mechanism involving Asp164 in the stabilization of the tetrahedral intermediate and are consistent with the available mutagenesis studies [7-10]. For instance, the D164A mutant of kumamolisin did not show any measurable proteolytic activity [7]. For the D164N mutant of kumamolisin-As, a low but appreciable level of catalytic activity (1.3%) was found [10], corresponding to an increase of the activation barrier of about 2.5 kcal/mol. It should be pointed out, however, that care must be exercised in making direct

comparisons of the computational results with experimental data. For instance, the acylation reaction has not been proven to be rate-limiting and the processes of the deacylation, substrate binding, and product release may need to be considered as well. Furthermore, structural analyses of the unliganded D164N mutant showed that, although the catalytic triad is intact, the mutated residue (Asn164) unexpectedly makes a hydrogen bond with the side chain of Ser278 [10]. The consequence of this structural change and the way by which the side chain of Asn164 stabilizes the transition state in the reaction catalyzed by the D164N mutant are still unknown.

The average active-site structures near the transition states for the nucleophilic attack (TS1) and the formation of the acyl-enzyme (TS2) are given in Figure VI-2B. Figure VI-2B (left) shows that there is little progress of the proton transfer from Asp164 to the substrate before the system reaches TS1. The result is consistent with the earlier suggestion that such proton transfer mainly occurs at a later stage of the nucleophilic attack [15, 17]. The proton transfer from Ser278 to Glu78 seems to be more synchronous with the nucleophilic attack, although two-dimensional free energy maps are required for a detailed understanding of the relationship between the nucleophilic attack and proton transfer [17]. Figure VI-2B (right) shows that the proton transfer from Glu78 to the nitrogen atom (N) of Phe at P_1' has been basically completed at TS2.

Dynamic substrate assisted catalysis (DSAC)

Kumamolisin-As was found to have a higher specificity for the substrates with a positively charged residue (such as His or Arg) at the P_1 site than for those they do not [2, 3,

[9]. Interestingly, the linkers of pro-kumamolisin-As and pro-kumamolisin (-P_3-Arg169p—P_2-Pro170p—P_1-His171p—P_1'-Phe172p—P_2'-Arg173p-) also have a positively charged His residue at P_1 [7-10]; the coordinates of this peptide fragment from pro-kumamolisin have been used to generate the substrate complex used in this study (see above). It was found in our earlier simulations [16] that the His sidechain at P_1 rotated significantly during transition from the substrate to the tetrahedral intermediate. This movement was triggered by the deprotonation/protonation events and resulted in the interaction of His at P_1 with the unprotonated Asp164 making the tetrahedral intermediate more stable. We suggested that the dynamics involving the His residue at the P_1 site may play an important role in the catalysis and contribute to the higher specificity for the substrates with His at P_1. These results are confirmed here with a different substrate. Indeed, while His at P_1 forms a hydrogen bond with the C=O' group of Pro at P_2 in the substrate complex (Figure VI-1A), it interacts with Asp164 in the tetrahedral intermediate (Figure VI-1C).

The previous proposal for the existence of DSAC in the kumamolisin-As-catalyzed reaction was based on the results of the nucleophilic attack by Ser278 on a substrate. However, as demonstrated earlier, the nucleophilic attack may not be the rate limiting step for the acylation. Thus, a question remains as to whether the His residue at P_1 would be able to stabilize the transition state for the acyl-enzyme formation (TS2) as well, and therefore lead to an overall rate enhancement for the acylation reaction. The structure shown in Figure VI-1E suggests that this may be indeed the case, as His at P_1 moves back to its original position (i.e., to that found in the substrate complex) to interact

with the C=O' group of Pro at P_2 in the acyl-enzyme. Figure VI-2B (right) shows that the motion of the His residue from the position in the tetrahedral intermediate to that in the acyl-enzyme is almost completed as the system reaches TS2, the transition state for the acyl-enzyme formation. This is in contrast to the case of the TI formation, where His at P_1 undergoes the conformational change mainly after the system passes TS1 (see above).

Figure VI-3 shows a plot of the free energy profiles for His at P_1 moving between the C=O' group of Pro at P_2 and $O_{\delta 2}$ of Asp164 in the substrate complex, tetrahedral intermediate, and acyl-enzyme; the distance between the $H_{\delta 1}$ hydrogen atom of the His sidechain (i.e., H indicated in Figures VI-1A, VI-1C and VI-1E) and $O_{\delta 2}$ of Asp164 is used as the reaction coordinate (RC). For the substrate complex, the free energy for the conformations for which the His residue forms a hydrogen bond with the C=O' group (RC ~ 4.2Å) is considerably lower than the free energy of those with the $O_{\delta 2}$(Asp164)…H(His) hydrogen bond; the conformations in the former case are therefore more stable. This result is consistent with the average structure shown in Figure VI-1A where the His residue interacts with C=O'. For the TI complex, the conformations with the $O_{\delta 2}$(Asp164)…H(His) hydrogen bond become more stable, as indicated by the relatively low free energy at RC=1.7Å. The conformations with the $O_{\delta 2}$(Asp164)…H(His) hydrogen bond become less stable again after the system changes to the acyl-enzyme. Therefore, the His residue at P_1 moves between the C=O' group of Pro at P_2 and $O_{\delta 2}$ of Asp164. Each movement is triggered by the bond breaking and making events at a given stage of the reaction, as well as the corresponding changes of the interactions with the His residue. The back and forth movements of the His sidechain

between the C=O' group of Pro at P_2 and $O_{\delta 2}$ of Asp164 in a ping-pong-like mechanism and the formation of the alternative hydrogen bonds effectively lower the free energy barriers for the both nucleophilic attack and the acyl-enzyme formation. It is possible that this mechanism also works for the deacylation reaction. In other words, the water attack on the acyl-enzyme and the formation of the product might also be helped by the motions of the His residue in the same ways as what happened for the nucleophilic attack by Ser278 and the formation of the acyl-enzyme, respectively.

Conclusions

The QM/MM MD and free energy simulations demonstrated that the nucleophilic attack by the serine residue on the substrate may not be the rate limiting step for the acylation of the GPH*FF substrate. The results given here confirm the earlier suggestions that Asp164 acts as a general acid during the nucleophilic attack by Ser278. It was also shown that this residue acts as a general base during the formation of the acyl-enzyme from the tetrahedral intermediate. The role of dynamic substrate assisted catalysis (DSAC) involving His at the P_1 site was examined, and it was shown that the bond breaking and making events at each stage of the reaction trigger a change of the position for the His sidechain and lead to the formation of alternative hydrogen bonds. The back and forth movements of the His sidechain between the C=O' group of Pro at P_2 and $O_{\delta 2}$ Asp164 and the formation of the corresponding hydrogen bonds effectively lower the free energy barriers for the both nucleophilic attack and the acyl-enzyme formation and leads to a relatively high activity.

Acknowledgment

We thank Professor Martin Karplus for a gift of the CHARMM program.

References

(1) Oda, K.; Sugitani, M.; Fukuhara, K.; Murao, S. *Biophys. Acta* **1987**, *923*, 463-469.

(2) Wlodawer, A.; Li, M.; Gustchina, A.; Oyama, H.; Dunn B. M.; Oda, K. *Acta Biochim. Pol*. **2003**, *50*, 81-102.

(3) Tsuruoka, N.; Nakayama, T.; Ashida, M.; Hemmi, H.; Nakao, M.; Minakata, H.; Oyama, H.; Oda, K.; Nishino. T. *Appl. Envioron. Microbiol.* **2003**, *69*, 162-169.

(4) Wlodawer, A.; Li, M.; Gustchina, A.; Uchika, K.; Oyama, H.; Dunn B. M.; Oda. K. *Nat. Struct. Biol.* **2001**, *8*, 442-446.

(5) Wlodawer, A.; Li, M.; Gustchina, A.; Oyama, H.; Oda, K.; Beyer, B. B.; Clemente, J.; Dunn, B. M. *Biochem. Biophys. Res. Commun.* **2004**, *314*, 638-645.

(6) Wlodawer, A.; Li, M.; Gustchina, A.; Dauter, Z.; Uchida, K.; Oyama, H.; Goldfar, N. E.; Dunn B. M.; Oda, K. *Biochemistry* **2001**, *40*, 15602-15611.

(7) Commellas-Bigler, M., Fuentess-Prior, P., Maskos, K., Huber, R., Oyama, H., Uchida, K.; Dunn, B. M., Oda, K., and Bode, W. *Structure* **2002**, *10*, 865-876.

(8) Commellas-Bigler, M.; Maskos, K.; Huber, R.; Oyama, H.; Oda, K.; Bode, W. *Structure* **2004**, *12*, 1313-1323.

(9) Wlodawer, A.; Li, M.; Gustchina, A.; Tsuruoka, N.; Ashida, M.; Minakata, H.; Oyama, H.; Oda, K.; Nishino, T.; Nakayama, T. *J. Biol. Chem*. **2004**, *279*, 21500-21510.

(10) Okubo, A.; Li, M.; Ashida, M.; Oyama, H.; Gustchina, A.; Oda, K.; Dunn, B. M.; Wlodawer, A.; Nakayama, T. *FEBS J*. **2006**, *273*, 2563-2576.

(11) Rawlings, N. D.; Barrett, A. J. *Biochim. Biophys. Acta.* **1999**, *1429*, 496-500.

(12) Lin, L.; Sohar, I.; Lackland, H.; Lobel, P. *J. Biol. Chem.* **2001**, *276*, 2249–2255.

(13) Walus, M.; Kida, E.; Wisniewski, K. E.; Golabek, A. A. *FEBS Lett*. **2005**, *579*, 1383-1388.

(14) Sleat, D. E.; Donnelly, R. J.; Lackland, H.; Liu, C. G.; Sohar, I.; Pullarkat, R. K.; Lobel, P. *Science* **1997**, *277*, 1802–1805.

(15) Guo, H. B.; Wlodawer, A.; Guo, H. *J. Am. Chem. Soc*. **2005**, *127*, 15662–15663.

(16) Xu, Q.; Guo, H. B.; Wlodawer, A.; Guo, H. *J. Am. Chem. Soc*. **2006**, *128*, 5994–5995.

(17) Guo, H. B.; Wlodawer, A.; Nakayama, T.; Xu, Q.; Guo, H. *Biochemistry* **2006**, *45*, 9129–9137.

(18) Dunn, B. M. *Chem. Rev.* **2002**, *102*, 4431-4458.

(19) Hedstrom, L. *Chem. Rev*. **2002**, *102*, 4501-4523.

(20) Ishida, T.; Kato, S. *J. Am. Chem. Soc*. **2003**, *125*, 12035–12048.

(21) Carter, P.; Wells, J. A. *Science*, **1987**, *237*, 394-399.

(22) Dall'Acqua, W.; Carter, P. *Protein Sci*. **2000**, *9*, 1-9.

(23)Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim T.; Karplus, M. *J. Phys. Chem. B* **2001**, *110*, 6458-6469.

(24) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.

(25) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H. B.; Ghosh, N.; Prat-Resina, X.; Konig, P.; Li, G. H.; Xu, D. G.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458-6469.

(26) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus M. *J. Comput. Chem.* **2003**, *24*, 565-581.

(27) Konig, P. H.; Hoffmann, M.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. B* **2005**, *109,* 9082-9095.

(28) Brooks, C. L. III; Brunger, A.; Karplus, M. *Biopolymers* **1985**, *24*, 843-865.

(29)MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Bunbrack R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher III, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586-3616.

(30) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700-733.

(31) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

(32) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902-1921.

(33) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys*. **1977**, *23*, 327-341.

(34) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578-581.

(35) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman P. A.; Rosenberg, J. M. *J. Comput. Chem*. **1992**, *13*, 1011-1021.

Appendix

Figure VI-1. The average active-site structures of the kumamolsin-As-substrate (GPH*FF) complex, tetrahedral intermediate and acyl-enzyme obtained from the simulations, and the motions of the three key protons involved in the catalysis.

The initial structures of the simulations for the tetrahedral intermediate and acyl-enzyme were based on the structures from the free energy simulations (see below). **(A)** The average structure of the substrate complex. The enzyme is plotted in ball and stick and the substrate in stick. Hydrogen bonds (H-bonds) are indicated by red dashed lines. Only the sidechain of the P_1 (His) residue of the substrate is shown for clarity. The catalytic triad contains the Ser278, Glu78 and Asp82 residues. The His side chain at P_1 donates an H-bond to the C=O' group of Pro at P_2, in addition to the carboxylate of Asp179 (not shown). The backbone N-H group of the P_1 residue donates a hydrogen bond to Glu78. **(B)** Fluctuations of the distances (in Å) related to proton motions (vibrations) between Asp164 and substrate, Ser278 and Glu78, and Glu78 and Asp82 in the substrate complex. Top panel: fluctuations of the H-O(D164) and H…O(substrate) distances as function of time. Red, R[H…O(substrate)]; blue, R[H-O(D164)]. The proton is on Asp164 and is involved in an H-bond with the C=O group of the substrate (see also Figure VI-1A). Middle panel: the fluctuations of the H-O(S278) and H…O(E78) distances as function of time. Red, R[H…O(E78)]; blue, R[H-O(S278)]. The proton is on Ser278 and is involved in an H-bond with Glu78. Bottom panel: the fluctuations of the H…O(E78) and H-O(D82) distances as function of time. Red, R[H-O(D82)]; blue, R[H…O(E78)]. The H-bond formed between Glu78 and Asp82 has a short distance of ~1.5 Å, and the proton is mainly located on Asp82. **(C)** The average structure of the tetrahedral intermediate after the nucleophilic attack by Ser278 on the substrate. The His sidechain at P_1 now interacts with $O_{\delta 2}$ of Asp164 and may, therefore, play a role in stabilizing the charge formation on Asp164. The H-bond between the backbone N-H group of the P_1 residue and Glu78 becomes significantly weaker due to the protonation of Glu78. **(D)** Fluctuations of the distances related to proton motions (vibrations) in the tetrahedral intermediate complex. Top panel: the fluctuations of the H…O(D164) and H-O(tetrahedral intermediate) distances as functions of time. Red,

R[H-O(tetrahedral intermediate)]; blue, R[H…O(D164)]. The proton is on the tetrahedral intermediate and is involved in an H-bond with Asp164 (see Figure VI-1C). Middle panel: the fluctuations of the H…O(S278) and H-O(E78) distances as function of time. Red, R[H-O(E78)]; blue, R[H…O(S278)]. The proton is on Glu78 and is involved in an H-bond with Ser278. Bottom panel: the fluctuations of the H…O(E78) and H-O(D82) distances as function of time. Red, R[H-O(D82)]; blue, R[H…O(E78)]. **(E)** The average structure of the acyl-enzyme with the C-N bond broken. **(F)** Fluctuations of the distances related to proton motions (vibrations) in the acy-enzyme. Top panel: the fluctuations of the H-O(D164) and H…O distances as function of time. Red, R[H…O]; blue, R[H-O(D164)]. Middle panel: the fluctuations of the H…O(S278) and H…O(E78) distances as function of time. Red, R[H…O(E78)]; blue, R[H…O(S278)]. Bottom panel: the fluctuations of the H…O(E78) and H…O(D82) distances as function of time. Red, R[H…O(D82)]; blue, R[H…O(E78)].

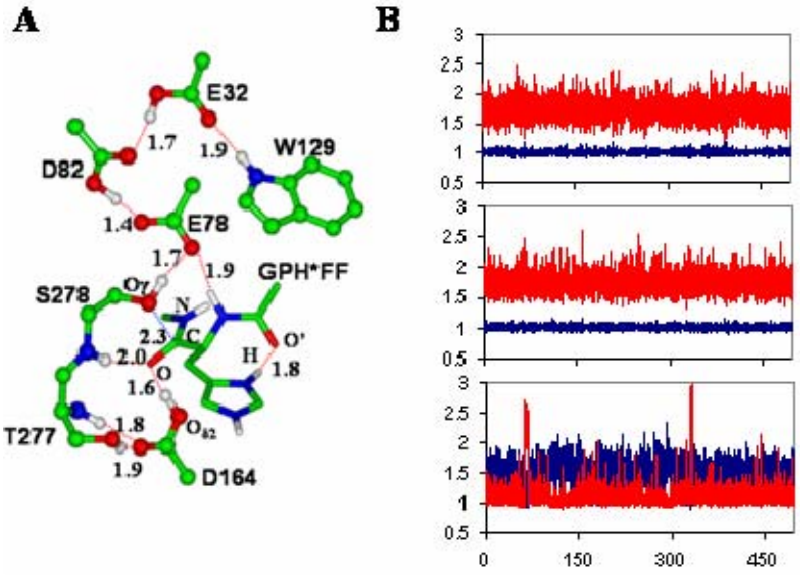


Figure VI-1, continued.

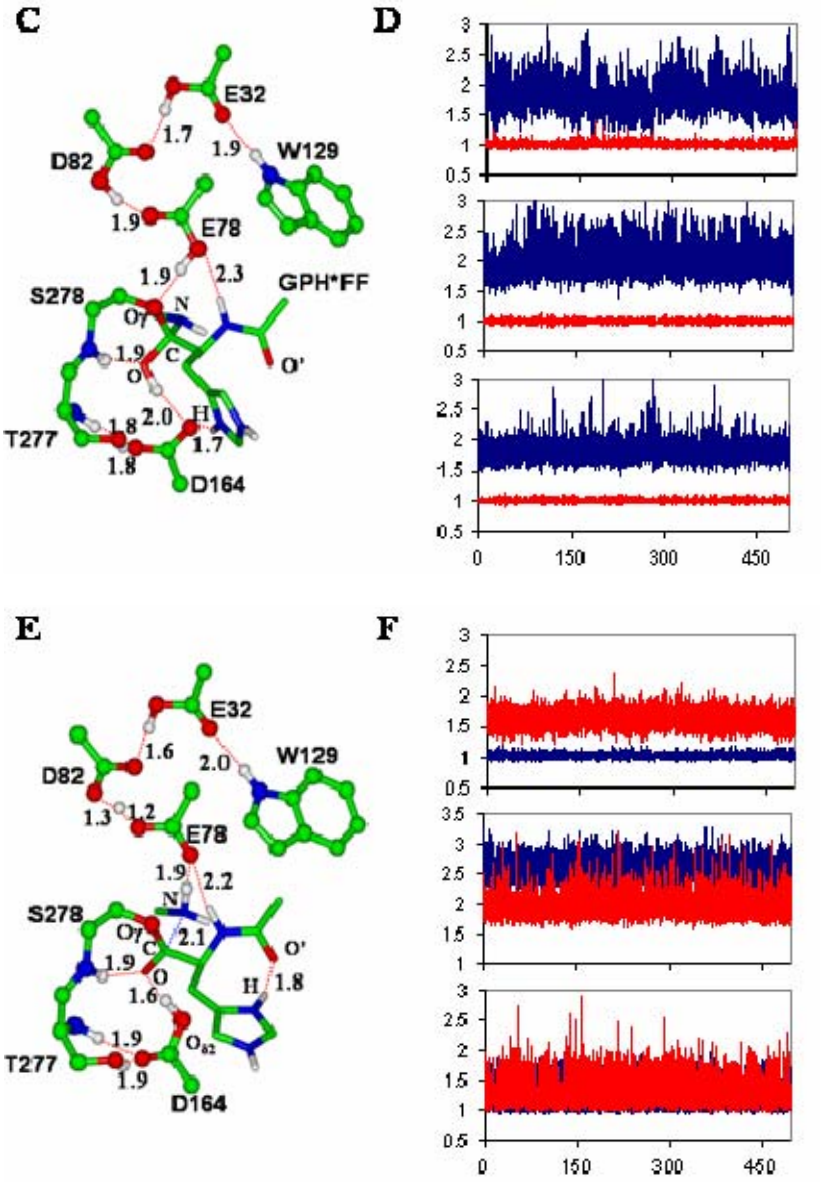


Figure VI-1, continued.

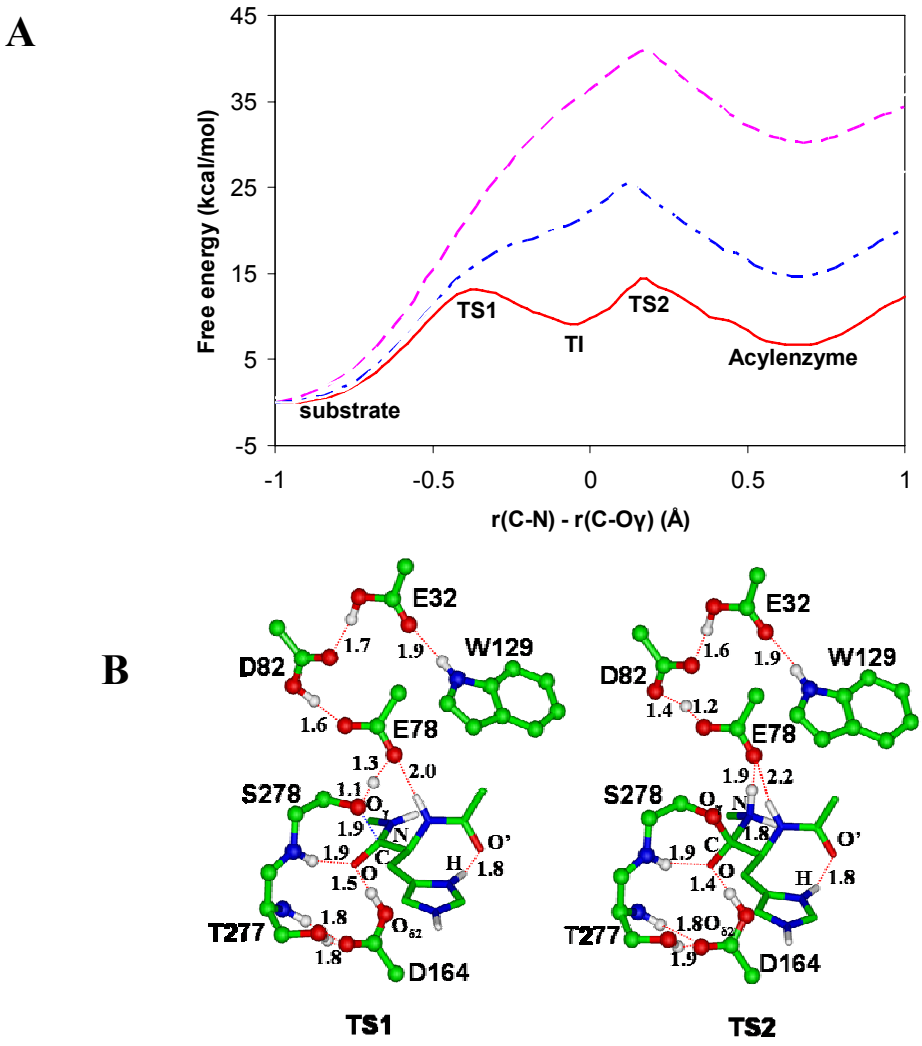
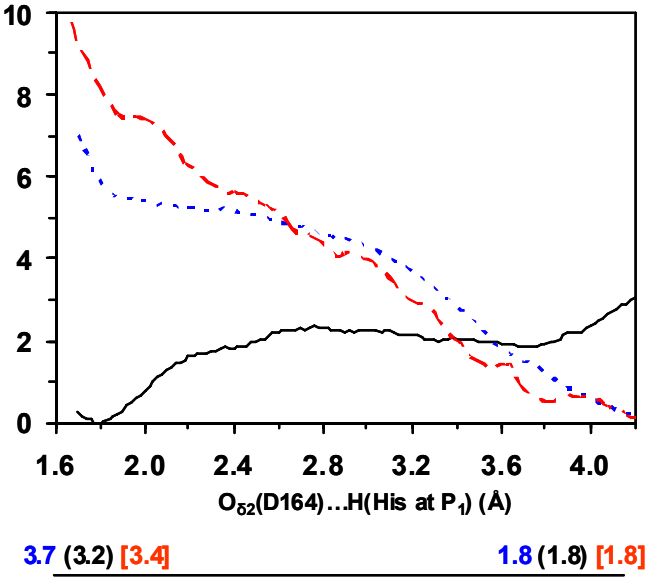


Figure VI-2. The free energy profiles of the acylation and the average structures at the transition states for the nucleophilic attack and the formation of the acyl-enzyme

(A) The free energy profiles of the acylation reaction for the wild-type kumamolisin-As and D164A mutant as a function of the reaction coordinate [ξ = R(C-N) – R(C…O_γ)] obtained from the QM/MM free energy (potential of mean force) simulations. **Red solid line**: wild-type; **blue dot-dashed line**: wild-type with the proton fixed on Asp164 with the SHAKE algorithm; **pink dashed line**: D164A mutant. The average structures for the substrate complex, tetrahedral intermediate and acyl-enzyme for wild-type are given in Figure VI-1A, VI-1C and VI-1E, respectively. **(B)** The average active-site structures for wild-type near the transition states for the nucleophilic attack (TS1) and the formation of the acyl-enzyme (TS2), respectively, obtained from the free energy simulations.



Average distances between H(P_1) and O'(P_2) for the substrate complex (blue), TI (black, in parenthesis), and acyl-enzyme (red, in square parenthesis).

Figure VI-3. The free energy changes for the shifts of P_1 His side chain

The free energy change as the side chain of His residue at P_1 moves between $O_{\delta 2}$ of Asp164 and O' of the C=O' group of Pro at P_2 in the substrate complex, tetrahedral intermediate and acyl-enzyme of kumamolisin-As (see Figures VI-1A, VI-1C and VI-1E for average structures). The reaction coordinate is the distance between the hydrogen atom (H) of the His residue at P_1 and $O_{\delta 2}$ of Asp164. The corresponding average distances between H and O' are given below. For instance, when the reaction coordinate is 4.2 Å for the substrate complex, the average distance for *r*(H…O') is about 1.8 Å based on the trajectory of the corresponding window (i.e., H forms a hydrogen bond with the C=O' group of Pro at P_2). **Blue dotted line**: the free energy change in the substrate complex. For the substrate complex, the conformations for which His at P_1 forms a hydrogen bond with the C=O' group of Pro at P_2 (i.e., the structure shown in Figure VI-1A) are considerable more stable (by about 5-6 kcal/mol) than the conformations for which the His residue interacts with $O_{\delta 2}$ of Asp164. **Black solid line**: the free energy change in the tetrahedral intermediate. For the tetrahedral intermediate, the conformations for which the His at P_1 forms a hydrogen bond with $O_{\delta 2}$ of Asp164 (i.e., the structure shown in Figure VI-1C) are more stable (by about 2 kcal/mol) than the conformations for which His interacts with C=O group of Pro at P_2. **Red dashed line**: the free energy change in the acyl-enzyme. The conformations for which the His residue forms a hydrogen bond with the C=O' group of Pro at P_2 (i.e., the structure shown in Figure VI-1E) are considerable more stable than the conformations for which His at P_1 interacts with $O_{\delta 2}$ of Asp164.

Chapter VII

A Triplet Code for Writing Epigenetic Marks: Application of QM/MM MD and Free Energy Simulations to Understand Product Specificity of Protein Lysine Methyltransferases DIM-5 and SET7/9

Qin Xu, Hao-Bo Guo, Xiaodong Cheng, Jeremy Smith, and Hong Guo

ABSTRACT

Histone lysine methylation catalyzed by protein lysine methyltransferases (PKMTs) can govern many important biological processes, including heterochromatin formation, X-chromosome inactivation, and transcriptional silencing and activation. Biological consequences of histone lysine methylation may differ depending on whether the lysine residue is mono-, di- or tri-methylated. Therefore, the ability of different PKMTs to generate different methylation states for the lysine residue, which is termed as product specificity, adds more complexity to the histone code and provides an additional layer of regulatory control. Understanding the origin of the product specificity and developing suitable computational models for its prediction are of fundamental importance. Here quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) simulations are performed on the reactant complexes for the 1^{st}, 2^{nd} and 3^{rd} methyl transfers from AdoMet to the target lysine/methyl lysine in the *N. crassa* DIM-5 (a trimethylase) and human SET7/9 as well as their mutants to establish the correlation between the formation of the reactive configurations for the methyl transfers and product specificity. In addition, the free energy profiles for the methyl transfer processes are determined. The results provide important insights into the origin of product specificity for PKMTs. Moreover, they support a hypothesis that the triplet code consisting of the difference in free energy barriers for the three methyl transfers can be used as important parameters in the determination of how the epigenetic marks of lysine methylation are written by the enzymes. Other important results from the simulations are also reported and discussed.

Introduction

The nucleosome is the fundamental building block of eukaryotic chromatin, and histone proteins play an important role in packaging DNA into nucleosome.[1] The tails of the histone proteins are subject to a variety of post-translational covalent modifications. These modifications form the so-called epigenetic histone code read by other proteins, which leads to distinct downstream events in the regulation of chromatin structure and gene expression.[2] One such modification is histone lysine methylation catalyzed by protein lysine methyltransferases (PKMTs).[3] Previous studies have shown that histone lysine methylation can govern many important biological processes, including heterochromatin formation, X-chromosome inactivation, and transcriptional silencing and activation. (For reviews, see Ref. 4) Several lysine residues on histone proteins have been identified to be the sites of methylation, including histone H3 Lys-4 (or simply H3-K4), H3-K9, H3-K27, H3-K36, H3-K79 and H4-K20. The methylated lysine residues can then serve as the docking sites for methyl-lysine binding domains and effector proteins to mediate the downstream events in the regulation of chromatin structure and gene expression.[4] There are two known enzyme families involved in the lysine methylation, the SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain proteins and Dot1p, and these two families have entirely different structural scaffolding. A number of crystal structures for the complexes of the SET domain PKMTs (including SET7/9, SET8, Clr4, DIM-5, vSET and pLSMT) have been determined.[5-6] These structures showed that substrate and AdoMet are located on the opposite surfaces, with the target lysine side chain accessing

the active site and the cofactor through a narrow channel. The crystal structures for the non-SET domain enzymes, yeast Dot1p and human Dot1L, have also been determined.[7]

In addition to the differences in their selection of the sites for methylation (substrate specificity), PKMTs may also differ in their ability to transfer one, two or three methyl groups from S-adenosyl-L-methionine (AdoMet) molecule(s) to the ε-amino group of the target lysine. This is because lysine residue has three protons, and each of them may be replaced by a methyl group (Figure VII-1A). For instance, DIM-5 of *N. crassa* (to be studied in this paper) generates primarily trimethyl H3 Lys-9, which marks chromatin regions for DNA methylation.[6] Another PKMT, which has been a subject of previous computational investigations,[8-10] is human SET7/9 that generates exclusively mono-methyl H3 Lys-4. The ability of different PKMTs to direct different degrees of methylation is called product specificity, and the enzymes can therefore be classified as mono-methylases, di-methylases or tri-methylases. For the SET domain di-methylases and tri-methylases, the methylation processes are believed to proceed processively without the release of the intermediates from the active sites (Figure VII-1B).[6b, 14] Biological consequences of histone lysine methylation may be quite different depending on whether the lysine residue is mono-, di- or tri-methylated. Therefore, the product specificity adds more complexity to the histone code and provides an additional layer of regulatory control. Thus, determination of the origin of the product specificity and development of suitable computational models/techniques for its prediction are of fundamental importance. This type of research may help to develop strategies for manipulating signal properties and turning epigenetic mechanisms of gene regulation into

tools of therapeutic intervention in diseases. Nevertheless, our understanding of the factors that control the product specificity is still lacking.

Structural analyses of the SET-domain PKMT complexes led to the suggestion that the geometry and shape at the bottom of the lysine access channel might be one of the key factors for controlling the product specificity for these enzymes (for reviews, see Ref. 11). For instance, the crystal structure[5b] of SET7/9 complexed with *S*-adenosyl-L-homocysteine (AdoHcy) and a histone H3 methyl-lysine-4 (H3-K4me) peptide showed that the rotation around the C-N bond of the methyl-lysine side chain could be prohibited at the active site,[5b, 11d] making the formation of the reactive configuration for addition of the second methyl group more difficult. This suggestion is consistent with the fact that SET7/9 is a mono-methylase.[5-6] Additional experimental studies were performed to pinpoint the active-site amino acid residue(s) that are responsible for product specificity. It should be pointed out that the key question for understanding product specificity is why for mono-methylases only one methyl group can be added to the target lysine, while for tri-methylases (di-methylases) multiple methyl additions can be achieved. In other words, the main focus is to identify the residue(s) and factors that control the *methylation state* of the product. This question is different from the question of identifying the residues and factors that are crucial for the general catalysis (i.e., regardless whether it is mono-methylation, di-methylation or tri-methylation). One proposal that is of considerable interest is the existence of a tyrosine/phenylalanine switch that controls the product specificity for some of the SET-domain PKMTs. Comparison of the active sites of SET7/9 and DIM-5 showed that a single amino acid residue occupies a structurally

similar position in both enzymes (F281 of DIM-5 and Y305 of SET7/9) and is in proximity of the ε-amino group of the target lysine.[6b] It was hypothesized that the Y305 hydroxyl in SET7/9 might be the source of steric hindrance that limits the further methylation by the enzyme.[6b] This hypothesis has been supported by the experimental studies. Indeed, it was shown that DIM-5 could be converted from a K-9 tri-methylase to a K-9 mono/di-methylase by the F281Y mutation, while the SET7/9 Y305F mutant generated dimethylated instead of monomethylated K-4.[6b] In each case, the substrate specificity (K-9 *vs.* K-4) and the overall reaction rate were not changed as a result of the mutation. The further supports for the existence of the tyrosine/phenylalanine switch for product specificity came from the studies on human SET8 (a histone H4-K20 mono-methylase) [5e] and G9a (a predominant H3-K9 PKMT that directs euchromatic mono- and di-methylation).[6e] Although previous structural studies have provided considerable insights, there remain many questions that need to be addressed in order to understand the origin of the product specificity. For instance, the role of the steric hindrance at the bottom of the lysine access channel in controlling product specificity has been called into question recently.[9b] Hu *et al*.[9b] argued that since the methyl group is very small and protein structures are very dynamic, there is a question as to how the steric hindrance could be imposed (see below for further discussions). Nevertheless, the functions of many enzymes depend on their remarkable sensitivity to subtle changes in the structures of the target molecules and reaction transition states,[12] and these powers of binding discrimination are a result of the combination and balance of different interactions in the enzyme complexes.[13] This is probably true for the product specificity as well, but an

intriguing question is exactly how the existence and absence of a single hydroxyl group on one residue can lead to such large effects on the reaction product.

Molecular dynamics (MD) simulations have been applied to study structural and energetic consequences of different interactions in the enzyme complexes and to understand the origin of product specificity. One approach that has been used for PKMTs [8-9] is to examine the reactant complexes prior to each of the methyl transfers and determine the relationship between the product specificity and the populations of the reactive configurations from which the methyl transfers would occur efficiently. Since the lone pair of electrons on N_ζ of the target lysine/methyl lysine is the site for the methyl addition, the structures with good alignments of the methyl group of AdoMet with the electron lone pair on N_ξ are expected to lead to an efficient methyl addition (see Figure VII-1C), whereas the structures with poor alignments would make the methyl transfer less likely, leading to the potential termination of the methylation process. The previous simulations[8-9] have showed that while there was a large population of the reactive structures with relatively short r(C_M…N_ζ) distances (~2.9 Å) and small θ angles (close to 0°) for the first methyl transfer in wild-type SET7/9, the population of such well-aligned reactive structures became very small for the second methyl transfer. The results from the simulations are therefore consistent with the fact that SET7/9 is a mono-methylase and can only catalyze the first methyl addition to H3-K4. Interestingly, the tyrosine/phenylalanine switch hypothesis proposed earlier was also confirmed from the simulations on SET7/9 and its Y305F mutant.[8] Indeed, it was shown[8] that the active site of Y305F was able to adopt the reactive structures with short *r*(C_M…N_ζ) distances and

small θ angles both for the first and second methyl transfers, consistent with the experimental observation that Y305F is a di-methylase.[6b] Although there is a good correlation between the formation of reactive configurations and product specificity, there may exist different interpretations. Hu *et al*.[9b] argued recently that the main controlling factor for the lacking of di-methylation activity for SET7/9 was not the steric hindrance which resulted in the disruption of the formation of the reactive conformation. They showed that the lone pair of electrons on N_ξ of the methyl lysine formed an $N_\xi \cdots H{-}O$ hydrogen bond in the reactant complex for the 2nd methyl transfer from the MD simulations, which was connected with solvent water molecules on the protein surface through a water chain. It was argued that extra energy would be required to break this hydrogen bond and to push a water molecule out of the water channel during the 2nd methyl transfer in SET7/9 and that it was this energetic cost that made SET7/9 a mono-methylase.[9b] Additional simulations on other PKMT systems are still necessary to confirm this proposal.

Although understanding the connection between the structural properties of the reactant complexes and product specificity is of importance, determining the energetic properties for the methyl transfers, including the activation barriers at the transition states, may provide some key insights into the origin of product specificity. Such energetic information can be obtained through the use of quantum mechanical/molecular mechanical (QM/MM) free energy (potential of mean force) simulations. Since the main focus is to understand and predict the methylation states of the products generated by different PKMTs (see above), it is important to establish the relationship between the

relative activation barriers for different methyl transfer processes with the product specificity. In our earlier study,[8] it was demonstrated that for wild-type SET7/9 the free-energy barrier for the 2nd methyl transfer from AdoMet to H3-K4me was 5 kcal/mol higher than that of the 1st methyl transfer from AdoMet to H3-K4. The higher free-energy barrier for the 2nd methyl transfer means that the dimethylation would be a much slower process and the product of mono-methylation would presumably be released before the second methyl could be added. The results of the simulations therefore agreed with the findings that SET7/9 is a mono-methylase. For Y305F, the free-energy simulations demonstrated that, unlike wild type (WT), the both free-energy barriers for the 1st and 2nd methyl transfers were rather low, consistent with the experimental observations that Y305F is a dimethylase and supporting the tyrosine/phenylalanine switch hypothesis.[6b]

In addition to the interpretations of the product specificity based on the methyl transfer processes (including those based on the populations of the reactive configurations and the free energy profiles for the methyl transfers mentioned above), Zhang and Bruice[10] suggested that the product specificity was controlled by the process of the deprotonation of the positively charged methyl lysine generated from the earlier methylation step (see Figure VII-1B). This deprotonation is absolutely required, because the basic form of the methyl lysine is the nucleophile and the lone pair of electrons on its N_ζ is the site where an additional methyl group can be added. They proposed that only water molecules in water channels observed from their MD simulations could play the role in the deprotonation, and the formation of the water channel therefore had to precede each of the next methyl additions.[10] Thus, for mono-methylases (di-methylaes) the

failure to form a water channel required for the deprotonation of Lys-$N_\zeta(Me)H_2^+$ (Lys-$N_\zeta(Me)_2H^+$) was believed to prevent further methylation by the enzymes, leading to the mono-methyl (di-methyl) lysine product.[10] One of the key assumptions in this proposal is that for the reactions catalyzed by tri-methylases the PKMT·Lys-$N_\zeta(Me)H_2^+$·AdoMet and PKMT·Lys-$N_\zeta(Me)_2H^+$·AdoMet complexes must be formed during the processive methylation processes; for di-methylases PKMT·Lys-$N_\zeta(Me)H_2^+$·AdoMet must be formed. Indeed, the results of the simulations by Zhang and Bruice[10] showed that the water channels that were observed in PKMT·Lys-$N_\zeta(Me)H_2^+$·AdoMet and PKMT·Lys-$N_\zeta(Me)_2H^+$·AdoMet could not form in PKMT·Lys-$N_\zeta(Me)H_2^+$·AdoHcy, PKMT·Lys-$N_\zeta(Me)_2H^+$·AdoHcy, PKMT·Lys-$N_\zeta(Me)H_2^+$ or PKMT·Lys-$N_\zeta(Me)_2H^+$. Nevertheless, to the best of our knowledge there is no experimental evidence to support the formation of PKMT·Lys-$N_\zeta(Me)H_2^+$·AdoMet and PKMT·Lys-$N_\zeta(Me)_2H^+$·AdoMet during the methylation processes, and previous studies[8,14] proposed that the deprotonation of methyl lysine might actually occur after dissociation of AdoHcy and before binding of the next AdoMet molecule (i.e., starting from PKMT·Lys-$N_\zeta(Me)H_2^+$ or PKMT·Lys-$N_\zeta(Me)_2H^+$). Moreover, one would expect potential strong repulsions between the two positively charged groups, AdoMet and Lys-$N_\zeta(Me)_2H^+$ (AdoMet and Lys-$N_\zeta(Me)H_2^+$), in the PKMT·Lys-$N_\zeta(Me)_2H^+$·AdoMet (PKMT·Lys-$N_\zeta(Me)H_2^+$·AdoMet) complex. Therefore, additional experimental and computational work is necessary to determine whether the PKMT·Lys-$N_\zeta(Me)H_2^+$·AdoMet and PKMT·Lys-$N_\zeta(Me)_2H^+$·AdoMet complexes exist or not. Furthermore, there seem to be other inconsistencies in this proposal. For instance, the

simulations predicted that the Y245F mutant was a di-methylase, [10a] while previous experimental data showed that Y245F is a weak mono-methylase.[5b]

In this article, the results of quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) and free energy simulations are reported for DIM-5; some unpublished data for SET7/9 are given as well. DIM-5 from *N. crassa* is a SUV39-type H3K9 tri-methylase that is essential for DNA methylation in vivo. The existence of a large body of experimental data[6] for the wild type and mutated DIM-5 allows us to test a variety of proposals through computational investigations. To the best of our knowledge, this is the first time that the simulation results are reported for DIM-5. Our results on DIM-5 suggest that the triplet code consisting of the difference in free energy barriers for the three methyl transfers from AdoMet to the target lysine/methyl lysine may determine how the epigenetic marks of lysine methylation are written by the enzyme. The origin for the existence of the tyrosine/phenylalanine switch is discussed, and detailed interpretations are also provided.

Methods

Quantum mechanical/molecular mechanical (QM/MM) free energy (potential of mean force) simulations were applied to the studies of the 1st, 2nd and 3rd methyl transfers catalyzed by wild type DIM-5, F281Y mutant of DIM-5, wild type SET7/9 and Y305F mutant of SET7/9, using the CHARMM program[15]. AdoMet/AdoHcy and lysine/methyl-lysine side chains of the substrates and products were treated by QM and the rest of the system by MM. The link-atom approach[18] available in the CHARMM program was

applied to covalent bonds on the boundary of the QM and MM regions. The SCC-DFTB[16] method was used for the QM description in the present study; high level *ab initio* methods (e.g., B3LYP and MP2) are too time-consuming to be used for MD and free energy simulations. Based-on an iterative energy minimization by adopted basis Newton-Raphson (ABNR) and the reaction coordinate driving approach,[9d] the energetic profiles using the SCC-DFTB and B3LYP/6-31G** methods for the description of the methyl transfer in a small model system were first compared.[8] Such comparisons are necessary and allow us to understand the performance of the semi-empirical method for the systems under investigations. (See Ref. 8 for additional discussions.)

The all-hydrogen potential function (PARAM22)[17] was used for the MM atoms. A modified TIP3P water model[19] was employed for the solvent. The stochastic boundary molecular dynamics method[20] was used for the QM/MM MD and free energy simulations. The system was separated into a reaction zone and a reservoir region; the reaction zone was further divided into the reaction region and buffer region. The reaction region was a sphere with radius r of 22 Å, and the buffer region had r equal to 20 Å $\leq r \leq$ 22 Å. The reference center for partitioning the system was chosen to be the C_δ atom of the target lysine residue, Lys9 of the histone H3 peptide or the methylated Lys9. The resulting systems contain around 5500 atoms, including about 800-900 water molecules.

The initial coordinates for the reactant complexes of the 1st, 2nd and 3rd methyl transfers were mainly based on the crystallographic dimmer complex (PDB code: 1PEG) of DIM-5 containing AdoHcy and a short H3K9 peptide.[6b] For the mono-methylation, a methyl group was manually added to AdoHcy to change it into AdoMet. By comparing

with the structure of SET7/9 (PDB code: 1O9S)[5b] and LSMT (pea Rubisco large subunit methyltransferase, PDB code: 1OZV)[5h], a crystal water was docked to the active site based on the X-ray structure of SET7/9[5b] with superposition and further optimization using MOE.[24] This procedure led to the reactant complex for mono-methylation containing DIM-5, AdoMet and the H3-K9 peptide. For the study of the second methyl transfer from AdoMet to the mono-methylated Lys-9 (H3-K9me), one methyl group was manually added to AdoHcy to change it into AdoMet and one to H3K9 peptide (based on the X-ray structure of SET7/9 that contains a methyl-lysine), respectively, along with the water molecule (see above). The methyl group added on H3K9 takes the place of the hydrogen close to the water molecule, since the other position forms a more stable hydrogen bond with Tyr-178 and has less space to accommodate the methyl group. This leads to the reactant complex of di-methylation. Two methyl groups were added to H3K9 and one was added to AdoHcy for the reactant complex for the 3rd methyl transfer (along with the water). The F281Y mutant was generated simply by changing Phe-281 in wild-type to Tyr with the reactant complexes generated using the similar procedures. The systems of SET7/9 and its Y305F mutant for the 3rd methyl transfer were generated in a similar way as in the case of DIM-5, using the structures in our earlier simulations on SET7/9 [8].

The initial structures for the entire stochastic boundary system were optimized by steepest descent (SD) method and adopted basis Newton-Raphson (ABNR) method. 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. 1 ns QM/MM MD simulations were carried out for each of the reactant and product complexes of the 1st, 2nd and 3rd methyl transfers after 500-ps of equilibration was performed. As discussed in the previous studies and mentioned earlier in this paper, the S_N2 methyl transfer from AdoMet to K9, K9me or K9$(me)_2$ would presumably be more efficient if the S-CH_3 group of AdoMet is well aligned with the lone pair of electrons of N_ξ in the reactant complex with a small θ angle and relatively short C_M-N_ξ distance.[8-9, 12-14, 24] Therefore, during the MD simulations, we monitored the distributions of $r(C_M\text{-}N_\xi)$ and θ, which was defined as the angel between the direction of the C_M-S_δ bond ($\boldsymbol{r}_2$) and the direction of the electron lone pair ($\boldsymbol{r}_1$), (see Figure VII-1C). This allows us to estimate the free energy contribution to the activation barrier due to the formation of the reactive substrate complexes (or near attack conformers). The free energy curves as functions of $r(C_M\text{-}N_\xi)$ and θ in the reactant complexes were generated from the probability densities $\rho[r(C_M\text{-}N_\xi)]$ and $\rho(\theta)$ based on the trajectories of the 1-ns MD simulations and following formula:[21]

$$W(\xi) = - k_B T \, ln \, \rho(\xi).$$

Here k_B is the Boltzmann's constant, T is the temperature (283.15 K), and ξ is $r(C_M\text{-}N_\xi)$ or θ.

The umbrella sampling method[22] implemented in the CHARMM program along with the Weighted Histogram Analysis Method (WHAM) [23] 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-K9me or H3-K9$(me)_2$ in each case. The reaction coordinate is defined as a linear combination of $r(C_M\text{-}N_\xi)$ and

$r(C_M\text{-}S_\delta)$ [i.e., $R = r(C_M\text{-}S_\delta) - r(C_M\text{-}N_\xi)$].)]. Each methylation process was described by twenty windows, in which 50 ps productive runs were performed after 20 ps equilibrations. The force constants of the harmonic biasing potentials used in the PMF simulations were 50 to 500 kcal·mol^{-1}·Å^{-2}. The free energy profiles for the reverse methyl-transfer processes (methyl transfer from methylated lysine to AdoHcy) were also performed afterward, most of which are almost identical to the corresponding free energy profiles for the methylation process (methyl transfer from AdoMet to lysine/methyl-lysine/dimethyl-lysine).

Results

Methyl transfers in wild type DIM-5

The average structures of the active sites of the reactant complexes of wild type DIM-5 for the 1st, 2nd, and 3rd methyl transfers are compared in Figure VII-2A-C. It can be seen from Figure VII-2A-C that the lone pair of electrons on N_ζ of the target lysine/methyl lysine and the methyl group of AdoMet are well aligned in all the three structures. An average distance of 3.0Å between N_ξ and C_M were obtained for these three reactant complexes. The oxygen atoms of the carbonyls of Val203, Arg238, Ile240 and that of the Tyr283 side chain interact with the methyl group of AdoMet, and it has been proposed that such C-H…O hydrogen bonds may play a role in the catalysis by fixing the methyl group toward the lone pair of electrons on N_ξ.

The average structure of the reactant complex for the 1st methyl transfer (Figure VII-2A) is similar to the X-ray structure of DIM-5 (PDB code: 1PEG) [6b]; the RMSD was

calculated to be 0.6Å for all the atoms within 5Å from the reaction center (C_M). For this structure, the position and the orientation of the ε-amino group seems to be fixed by its hydrogen bond with Tyr178, indicating an important role of Tyr178 in the catalysis.[6] The activity of DIM-5 was indeed found to be complete lost and dramatically reduced as a result of Y178→V and F mutation, respectively.[6b] There is a water molecule (W1) in Figure 2A-C. This water molecule was built manually by comparing DIM-5 with the structures of SET7/9 (PDB code: 1O9S) [5b] and LSMT (PDB code: 1OZV)[5h], as the crystal structure of DIM-5 did not show this water molecule. Figure 2A-C shows that one of the important differences in the three average structures of reactant complexes is the position of W1. As shown in Figure VII-2A, in the reactant complex of DIM-5·Lys-$N_\zeta H_2$·AdoMet W1 forms hydrogen bonds with the carbonyl groups of Ile240 and Thr237 (not shown) and the ε-amino group of H3K9 (similar to the case of SET7/9). These three hydrogen bonds fix the position of W1 at the active site and may contribute to the stabilization of the position of H3K9 as well. In the reactant complex of DIM-5·Lys-N_ζ(Me)H·AdoMet, one hydrogen on N_ξ still forms a stable hydrogen bond with Tyr178, while the other one has been replaced by a methyl group. The water molecule W1 lost its hydrogen bonds with both the lysine ε-amino group and the carbonyl of Ile240. W1 moves away from its position in DIM-5·Lys-$N_\zeta H_2$·AdoMet with an average distance between W1 oxygen and N_ξ increased from 3.2Å to 4.9Å. In the reactant complex of DIM-5·Lys-$N_\zeta(Me)_2$·AdoMet, the ε-amino group of the target lysine has both of its two hydrogen atoms replaced by methyl groups. It is therefore unable to hydrogen-bond to Tyr178, and the methyl groups seem to form hydrophobic interactions with the aromatic

rings of Tyr178, Tyr283 and Phe281. These hydrophobic interactions may help to fix the position of the ε-amino group so that the lone pair of electrons on N_ζ can be well aligned with the methyl group of AdoMet for the efficient methyl transfer. W1 has been pushed away from the active site (about 8Å to N_ζ) and forms new hydrogen bonds with the carbonyl groups of Asp176 and Leu265.

Consistent to the average structures of the three reactant complexes, the two dimensional plots of $r(C_M \ldots N_\zeta)$ and θ distributions (see the definitions of the two parameters in **method**) exhibit similar patterns. All of the three reactant complexes have $r(C_M \ldots N_\zeta)$ concentrated in 2.5 to 3.5Å and θ concentrated in 0 to 60 degree. (Figure VII-2D-F) These patterns suggest that the active site in each case adopts a reactive structure with a good alignment of the methyl group with the electron lone pair on N_ζ. The results therefore show that the wild type DIM-5 may efficiently transfer all the 1st, 2nd, and 3rd methyl groups from AdoMet to lysine 9, which is consistent to the experimental result that wild type DIM-5 mainly produces trimethyl-lysine. The product specificity of wild type DIM-5 as a trimethylase was further confirmed by the free energy simulations (see below).

Methyl transfers in F281Y mutant of DIM-5

For the reactant complexes of DIM-5 F281Y mutant, the average structure of the active site for the 3rd methyl transfer is very different from those for the 1st, 2nd transfers (Figure VII-3A-C). In the reactant complexes for the 1st and 2nd methyl transfers, the positions of the target H3K9 and H3K9me relative to the methyl group of AdoMet are

similar to those in the WT complexes. The average distances between N_ζ and C_M are both about 3.0Å, with the lone pair of electrons on N_ζ well aligned to the methyl group to be transferred. Consistent to the average structures, the two dimensional plots of $r(C_M\ldots N_\zeta)$ and θ distributions both have the distributions concentrated in the area of 2.5 to 3.5Å and 0 to 70 degree (Figure VII-3D, E). In contrast, the reactant complex for the 3rd methyl transfer has an much larger average distance of $r(C_M\ldots N_\zeta)$ as 4.6Å (Figure VII-3C). The amino group of H3K9(me)$_2$ undergoes large fluctuations in both the position and the orientation. The two methyl groups on N_ζ rotate constantly with no stable interactions formed with surrounding residues. The two dimensional plots of $r(C_M\ldots N_\zeta)$ and θ distributions have a much dispersed pattern, with $r(C_M\ldots N_\zeta)$ ranging from 3 to 6Å and θ to all the angles (Figure VII-3F).

Although the reactant complexes for 1st and 2nd methyl transfers have similar $r(C_M\ldots N_\zeta)$ and θ distributions, the stabilization of the amino groups seems a little different in the two complexes. Similar to the stabilization of H3K9 in the wild type DIM-5, one hydrogen atom on N_ζ forms stable hydrogen bonds with the side chain of Tyr178; the other hydrogen may form a hydrogen bond with the side chain of Tyr281 or the oxygen of W1. For the reactant complex of the 2nd methyl transfer, the reactant complex has 90% chance to have a structure in which H3K9me forms a hydrogen bond with oxygen of Val203 carbonyl, instead of to Tyr178 (which was observed in the reactant complex of WT-DIM-5·Lys-N_ζ(Me)H·AdoMet). The average distance between N_ζ and the W1 oxygen is increased from 3.6Å in the 1st methyl transfer reactant complex to 4.5Å in the 2nd methyl transfer reactant complex.

The 3rd methyl transfers in wild type and Y305F mutant of SET7/9

The correlation between the product specificity and the reactant complex structures are also observed in our simulations on the wild type and Y305F mutant of SET7/9, a mono-methylase. The wild type SET7/9 reactant complex only has reactive structures for the 1st methyl transfer. The two dimensional plot has $r(C_M \ldots N_\zeta)$ and θ concentrated within 2.5 to 3.5Å and 0 to 60 degree, respectively.[8] Whereas in the reactant complex for 2nd methyl transfer, the methylated ε-amino group of H3K9me is quite flexible, corresponding to a dispersed distribution of $r(C_M \ldots N_\zeta)$ and θ between 3 to 5.5Å and 0 to 120 degree, respectively.[8] The reactant complex for the 3rd methyl transfer has lower flexibility in the position and orientation of H3K9$(me)_2$ group, but it is "stabilized" at an unfavorable conformation for the methyl transfer. The average distance of $r(C_M \ldots N_\zeta)$ is as long as 4.3Å, and the lone pair of electrons on N_ζ is almost perpendicular to the direction of the methyl group to be transferred. (Figure VII-4A) Correspondingly, the two dimensional plot shows a concentrated pattern with $r(C_M \ldots N_\zeta)$ between 3.5 to 5Å and θ around 90 degree. (Figure VII-4C)

In the simulations of SET7/9 Y305F mutant, the reactant complexes for both the 1st methyl transfer and the 2nd methyl transfer have a reactive average structure similar to the one for 1st methyl transfer in wild type SET7/9. The average distances $r(C_M \ldots N_\zeta)$ are 2.9 and 3.2Å, respectively, and the lone pair of electrons on N_ξ is well aligned toward the methyl group to be transferred. Correspondingly, the free energy simulations obtained comparable low barriers as about 15kcal/mol for both the 1st and the 2nd methyl

transfers.[8] However, the structure of the reactant complex for the 3rd methyl transfer is quite distinct. Although the conformation of the ε-amino group of H3K9(me)$_2$ is relative stable, the methyl group of AdoMet undergoes a movement deviating from the active site. The distance of $r(C_M \ldots N_\zeta)$ fluctuates from 4.5 to 7Å, averaged at 5.6Å. And the lone pair of electrons on N_ξ points away from the methyl group, resulting in the angle θ as 90 to 150 degree. This result is consistent to the mass spectrometry analysis with only dimethylated product obtained for Y305F mutant.[6b]

The relative free energy barriers for methyl transfers: a triplet code for writing epigenetic marks

An important factor that may control the product specificities of PKMTs is the relative efficiencies of the 1st, 2nd, and 3rd methyl transfers. To reflect such relative efficiencies of PKMTs for the 1st, 2nd, and 3rd methyl transfers, we designed a triplet code (0, $\Delta_{2\text{-}1}$, $\Delta_{3\text{-}1}$) to present the differences in free energy barriers for the methyl transfers. The first parameter of the triplet code is the free energy barrier for the 1st methyl transfer and set to be the zero. The second ($\Delta_{2\text{-}1}$) and the third ($\Delta_{3\text{-}1}$) numbers of the triplet code are the difference in free energy barriers for the 2nd and the 3rd methyl transfers compared to the barrier for the 1st methyl transfer, respectively. In this way, the triplet code can clearly reflect the different efficiency of a PKMT to transfer the 1st, 2nd, and 3rd methyl groups to the target lysine/methyl lysine.

Since the free energy barrier for the 3rd methyl transfer in wild type DIM-5 is 1kcal/mol lower than those for the 1st and 2nd methyl transfers (which have the same

barriers), the triplet code for the three methyl transfers can be written as (0, 0, -1) for wild type DIM-5 (Figure VII-5A). This triplet code suggests a comparable or even higher activity for the 2nd and 3rd methyl transfers than the 1st methyl transfer, consistent with the product specificity of DIM-5 as a trimethylase.[6b] At the same time, with the free energy barriers of the 1st, 2nd, and 3rd methyl transfers in F281 mutant of Dim-5 as 15, 18 and 20 kcal/mol, the triplet code for DIM-5 F281 mutant can be written as (0, 3, 5), suggesting increased free energy barriers and decreased activities in 2nd and 3rd methyl transfers (Figure VII-5B). The 3 kcal/mol difference for the 2nd methyl transfer compared to that for the 1st methyl transfer may not prohibit the production of dimethylated lysine, but would slow down the production. The 5kcal/mol difference for the 3rd methyl transfer is expected to make the trimethylated product unlikely, consistent with experimental observations.[6b] Similarly, the free energy barriers of the methyl transfer in SET7/9 are 18, 23 and 26 kcal/mol, and the triplet code can be written as (0, 5, 8), indicating increased free energy barriers for 2nd and 3rd methyl transfers too. However, the 5 kcal/mol difference for 2nd methyl transfer and 8 kcal/mol difference for 3rd methyl transfer make it very difficult for the further methylation, consistent with the fact that SET7/9 is a mono-methylase (Figure VII-5C).

Discussions

Earlier QM/MM simulations has confirmed the methyl transfers in PKMTs to be an in-line S_N2 nucleophilic substitution.[8, 9a, 10] The different structures of the reactant complex may contribute to the different energetic barriers. As shown in Figure VII-2,

VII-3 and VII-4, the alignment of the lone pair electrons on N_ζ to the methyl group of AdoMet can be reflected by the two dimension plot of $r(C_M \ldots N_\zeta)$ and θ distributions. A short distance of $r(C_M \ldots N_\zeta)$ about 3.0Å and a small angle of θ lower than 60 degree generally suggest a good alignment that may facilitate the methyl transfer efficiently. In contrast, a long distance of $r(C_M \ldots N_\zeta)$ and a dispersed distribution of θ may suggest a unfavorable alignment for the methyl transfer. Thus, the average structures of reactant complexes should be related to the efficiencies of the methyl transfers and the product specificity of the PKMT.

Another way to compare the efficiencies of the different methyl transfers is to determine their free energy barriers. We proposed a triplet code here that gives the difference in free energy barriers for the three methyl transfers in the enzyme. We found that the triplet code (0, 0, -1) obtained from simulations on DIM-5 is consistent to the experimental result of Tyr/Phe switch for the product specificity of DIM-5. In other words, the further methylations are comparable or even easier than the 1st methylation. Considering that the multi-methyl transfers by the SET domain PKMTs is a processive instead of a distributive process (which has been suggested by evidences from the catalytic kinetics and product analyses[14]), the products would expect to be the trimethylated final product, instead of the monomethylated or dimethylated intermediates. This is consistent with the fact that DIM-5 is a tri-methylase. In contrast, our simulation results are also consistent to the experimental result that the Phe→Tyr mutation at residue 281 switches DIM-5 from a tri-methylase into a mono-/di- methylase. The concentrated distributions of $r(C_M \ldots N_\zeta)$ and θ in the reactant complex for 1st and 2nd

methyl transfers suggested the potential activity for F281Y mutant to transfer one or two methyl groups to H3K9, whereas the dispersed distributions of $r(C_M \ldots N_\zeta)$ and θ in a broad area suggested an extra cost in energy to fix the electron lone pair on N_ζ and methyl group of AdoMet in the reactive configurations. (Figure VII-3) Correspondingly, the free energy simulations resulted in a triplet code for DIM-5 F281Y mutant of (0, 3, 5) (Figure VII-5B). As indicated by the second number in the triplet code, the barrier for the 2nd methyl transfer is about 3 kcal/mol higher than that of the 1st methyl transfer. This may not entirely prevent the production of dimethylated lysine, but can lower the activity for 2nd methyl transfer significantly, consistent with experimental observations.[6b] The third number of the triplet code indicates that over 5 kca/mol more energy is needed for the 3rd methyl transfer. Therefore, an even lower activity for the 3rd methyl transfer is expected, which is also consistent with the experimental result.[6b]

The correlation of the product specificity with the reactant complexes structures and the triplet code for the free energy barrier differences is also observed in the simulations of another PKMT, the monomethylase SET7/9. Consistent with the product specificity of wild type SET7/9, the two dimensional plot of $r(C_M \ldots N_\zeta)$ and θ distributions only suggest a favorable structure for the S_N2 reaction in the reactant complex for 1st methyl transfer.[8] The structure in the reactant complex for the 2nd methyl transfer generates a quite dispersed distribution in $r(C_M \ldots N_\zeta)$ and θ,[8] and the plot of $r(C_M \ldots N_\zeta)$ and θ in the reactant complex for the 3rd methyl transfer suggest the ε-amino group to be stabilized at unfavorable position for the methyl transfer (Figure VII-4). Correspondingly, the triplet code (0, 5, 8) suggests over 5 kcal/mol additional energy is

needed for the 2nd methyl transfer and over 8kcal/mol additional energy is required for the 3rd methyl transfer. The Y305F mutant of SET7/9 switches the product specificity of wild-type into a dimethylase as demonstrated experimentally.[6b] Consistent with this, the average structures of reactant complexes for the 1st and the 2nd methyl transfers have favorable reactive configurations,[8] and the reactant complex for the 3rd methyl transfer has the dispersed and unfavorable distributions of $r(C_M \ldots N_\zeta)$ and θ (Figure VII-4). In summary, both the $r(C_M \ldots N_\zeta)$ and θ distributions and the triplet code for the difference in free energy barriers can be used to understand and predict the product specificity for SET7/9 as well.

Conclusions

The quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) simulations on the 1st, 2nd and 3rd methyl transfers from AdoMet to the target lysine/methyl lysine in the *N. crassa* DIM-5 (a trimethylase) and human SET7/9 as well as their mutants suggest the existence of correlations between product specificity and the population of $r(C_M \ldots N_\zeta)$ and θ distributions and the relative free energy barriers for methyl transfers. The difference in free energy barriers for methyl transfers are summarized into a triplet code, which can pretty well explain the product specificity switch by F281Y mutation in DIM-5 and Y305F mutant in SET7/9.

References

(1) (a) Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997**, *389*, 251-260.

(b) Watson, J. D.; Baker, T. A.; Bell, S. P.; Gann, A.; Levine, M.; Losick, R. *Molecular Biology of the Gene*, 5th edition. CSHL Press, **2004**.

(2) (a) Strahl, B. D.; Allis, C. D. *Nature*, **2000**, 403, 41-45.

(b)Turner, B. M. *Cell* **2002**, *111*, 285-291.

(c) Lachner, M.; O'Carroll, D.; Rea, S.; Mechtler, K.; Jenuwein, T. *Nature* **2001**,410, 116-120.

(d) Bannister, A. J.; Zegerman, P.; Partridge, J. F.; Miska, E. A.; Thomas, J. O.; Allshire, R. C.; Kouzarides, T. *Nature* **2001**, 410, 120-124.

(e) Jacobs, S. A; Taverna, S. D.; Zhang, Y.; Briggs, S. D.; Li, J.; Eissenberg, J. C.; Allis, C. D.; Khorasanizadeh, S. *EMBO J.* **2001**, 20, 5232-5241.

(f) Taverna, S. D.; Li, H; Ruthenburg, A. J.; Allis C. D.; Patel D. J. *Nat. Struct. Biol.* **2007**, 14, 1025-1040.

(3) (a) Rea, S.; Eisenhaber, F.; O'Carroll, D.; Strahl, B. D.; Sun, Z. W.; Schmid, M.; Opravil, S.; Mechtler, K.; Ponting, C. P.; Allis, C. D.; Jenuwein, T. *Nature* **2000**, *406*, 593-599.

(b) Nakayama, J.; Judd, C. R.; Strahl, B. D.; Allis, C. D.; Grewal, S. I. S. *Science* **2001**, 292, 110-113.

(c) Tamaru, H; Selker, E. R. *Nature* **2001**, 414, 277-283.

(4) (a) Lall, S. *Nat. Struct. Mol. Biol.* **2007**, *14*, 1110-1115.

(b) Luger, K. *Curr. Opin. Genet. Dev.* **2003**, *13*, 127-135.

(c) Bannister, A. J.; Schneider, R.; Kouzarides, T. *Cell* **2002**, *109*, 801-806.

(d) Khorasanizadeh, S. *Cell* **2004**, *116*, 259-272.

(e) Sanders, S. L.; Portoso, M.; Mata, J.; Bahler, J.; Allshire, R. C.; Kouzarides, T. *Cell* **2004**, *119*, 603-614.

(f) Chuikov, S.; Kurash, J. K.; Wilson, J. R.; Xiao, B.; Justin, N.; Ivanov, G. S.; McKinney, K.; Tempst, P.; Prives, C.; Gamblin, S. J.; Barlev, N. A.; Reinberg, D. *Nature* **2004**, *432*, 353-360.

(g) Nishioka, K.; Chuikov, S.; Sarma, K.; Erdjument-Bromage, H.; Allis, C. D.; Tempst, P.; Reinberg, D. *Genes & Dev.* **2002**, *16*, 479-489.

(h) Rice, J. C.; Briggs, S. D.; Ueberheide, B.; Barber, C. M.; Shabanowitz, J.; Hunt, D. F.; Shinkai, Y.; Allis, C. D. *Mol. Cell* **2003**, *12*, 1591-1598.

(i) Kouskouti, A.; Scheer, E.; Staub, A.; Tora, L.; Talianidis, I. *Mol. Cell* **2004**, *14*, 175-182.

(j) Ruthenburg, A. J.; Allis, C. D.; Wysocka, J. *Mol. Cell* **2007**, *25*, 15-30.

(k) Jenuwein, T. *FEBS J.* **2006**, *273*, 3121-3135.

(l) Martin, C.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 838-849.

(m) Nightingale, K. P.; O'Neill, L. P.; Turner, B. M. *Curr. Opin. Genet. Dev.* **2006**, *16*, 125-136.

(n) Peters, A. H.; Schubeler, D. *Curr. Opin. Cell Biol.* **2005**, *17*, 230-238.

(o) Goldberg, A. D.; Allis, C. D.; Bernstein, E. *Cell* **2007**, *128*, 635-638.

(p) Su, I. H.; Tarakhovsky, A. *Curr. Opin. Immunol.* **2006**, *18*, 152-157.

(q) Turner, B. M. *Nat. Struct. Mol. Biol.* **2005**, *12*, 110-112.

(5) (a) Wilson, J. R.; Jing, C.; Walker, P. A.; Martin, S. R.; Howell, S. A.; Blackburn, G. M.; Gamblin, S. J.; Xiao, B. *Cell* **2002**, *111*, 105-115.

(b) Xiao, B.; Jing, C.; Wilson, J. R.; Walker, P. A.; Vasisht, N.; Kelly, G.; Howell, S.; Taylor, I. A.; Blackburn, G. M.; Gamblin, S. J. *Nature* **2003**, *421*, 652-656.

(c) Kwon, T.; Chang, J. H.; Kwak, E.; Lee, C. W.; Joachimiak, A.; Kim, Y. C.; Lee, J.; Cho, Y. *EMBO J.* **2003**, *22*, 292-303.

(d) Couture, J. F.; Collazo, E.; Hauk, G.; Trievel, R. C. *Nat. Struct. Mol. Biol.* **2006**, *13*, 140-146.

(e) Couture, J. F.; Collazo, E.; Brunzelle, J. S.; Trievel, R. C. *Genes & Dev.* **2005**, *19*, 1455-1465.

(f) Xiao, B.; Jing, C.; Kelly, G.; Walker, P. A.; Muskett, F. W.; Frenkiel, T. A.; Martin, S. R.; Sarma, K.; Reinberg, D.; Gamblin, S. J.; Wilson, J. R. *Genes & Dev.* **2005**, *19*, 1444-1454.

(g) Manzur, K. L.; Farooq, A.; Zeng, L.; Plotnikova, O.; Koch, A. W.; Sachchidanand; Zhou, M. M. *Nat. Struct. Biol.* **2003**, *10*, 187-196.

(h) Trievel, R. C.; Flynn, E. M.; Houtz, R. L.; Hurley, J. H. *Nat. Struct. Biol.* **2003**, *10*, 545-552.

(i) Trievel R. C.; Beach B.M.; Dirk L. M.; Houtz R. L.; Hurley J. H. *Cell* **2002**, *111*, 91-103.

(j) Couture, J. F.; Hauk, G.; Thompson, M. J.; Blackburn, G. M.; Trievel, R. C. *J. Biol. Chem.* **2006**, *281*, 19280-19287.

(k) Jacobs, S. A.; Harp, J. M.; Devarakonda, S.; Kim, Y.; Rastinejad, F.; Khorasanizadeh, S. *Nat. Struct. Biol.* **2002**, *9*, 833-838.

(l) Min, J.; Zhang, X; Cheng, X. D.; Grewal, S.; Xu, R. *Nat. Struct. Biol.* **2002**, *9*, 828–832.

(6) (a). Zhang, X.; Tamaru, H.; Khan, S. I.; Horton, J. R.; Keefe, L. J.; Selker, E. U.; Cheng, X. *Cell* **2002**, *111*, 117-127.

(b) Zhang, X.; Yang, Z.; Khan, S. I.; Horton, J. R.; Tamaru, H.; Selker, E. U.; Cheng, X. *Mol. Cell* **2003**, *12*, 177-185.

(c) Tamaru, H.; Zhang, X.; McMillen, D.; Singh, P. B.; Nakayama, J.; Grewal, S. I.; Allis, C. D.; Cheng, X.; Selker, E. U. *Nat. Genet.* **2003**, *34*, 75–79.

(d) Rathert, P.; Zhang, X.; Freund, C.; Cheng, X.; Jeltsch, A. *Chem. & Biol.* **2008**, *15*, 5-11.

(e) Collins, R. E.; Tachibana, M.; Tamaru, H.; Smith, K. M.; Jia, D.; Zhang, X.; Selker, E. U.; Shinkai, Y.; Cheng, X. *J. Biol. Chem.* **2005**, *280,* 5563–5570.

(7) (a) Min, J.; Feng, Q.; Li, Z.; Zhang, Y.; Xu, R. M. *Cell* **2003**, *112*, 711–723.

(b) Sawada, K.; Yang, Z.; Horton, J. R.; Collins, R. E.; Zhang, X.; Cheng, X. D. *J. Biol. Chem.* **2004**, *279*, 43296–43306.

(8) Guo, H-B.; Guo, H. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 8797-8802.

(9) (a) Hu, P.; Zhang, Y. *J. Am. Chem. Soc.* **2006**, *128*, 1272-1278.

(b) Hu, P.; Wang, S.; Zhang, Y. *J. Am. Chem. Soc* **2008**, *130*, 3806-3813.

(c) Wang, S.; Hu, P.; Zhang, Y. *J. Phys. Chem. B* **2007**, *111*, 3758-3764.

(d) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, 112, 3483-3492.

(10) (a) Zhang X.; Bruice T. C. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 5728-5732.

(b) Zhang X.; Bruice T. C. *Biochem.* **2008**, *47*, 2743-2748.

(c) Zhang X.; Bruice T. C. *Biochem.* **2007**, *46*, 14838-14844.

(d) Zhang X.; Bruice T. C. *Biochem.* **2007**, *46*, 9743-9751.

(e) Zhang X.; Bruice T. C. *Biochem.* **2007**, *46*, 5505-5514.

(11) For recent reviews, see (a) Cheng, X.; Zhang, X. *Mutat Res* **2007**, *618*, 102-115.

(b) Dillon, S. C.; Zhang, X.; Trievel, R. C.; Cheng, X. *Genome Biol.* **2005**, *6*, 227.

(c) Cheng, X.; Collins, R. E.; Zhang, X. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 267-294.

(d) Xiao, B.; Wilson, J. R.; Gamblin, S. J. *Curr. Opin. Struct. Biol.* **2003**, *13*, 699-705.

(e) Couture, J. F.; Trievel, R. C. *Curr. Opin. Struct. Biol.* **2006**, *16*, 753-760.

(12) Wolfenden, R.; Kati, W. M. *Acc. Chem. Res.* **1991**, *24*, 209-215.

(13) (a) Xu, Q.; Guo, H-B.; Wlodawer, A.; and Guo, H. *J. Am. Chem. Soc.* **2006**, *128*, 5994-5995.

(b) Xu, Q.; Guo, H-B.; Wlodawer, A.; Nakayama, T.; Guo, H. *Biochem.* **2007**, 46, 3784-3792.

(c) Guo, H-B.; Rao, N.; Xu, Q.; Guo, H. *J. Am. Chem. Soc.* **2005**, *127*, 3191-3197.

(d) Guo, H.; Cui, Q.; Lipscomb, W. N.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9032–9037.

(14) Dirk, L. M.; Flynn, E. M.; Diwtzel, K.; Couture, J.; Trievel, R.; Houtz, R. L. *Biochem.* **2007**, *46*, 3905–3915.

(15) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.

(16) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569-585.

(17) Mackerell, A. D., Jr.; Bashford, D.; Bellott, M.; Bunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E. III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; W.-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586-3616.

(18) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700-733.

(19) (a) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

(b) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902-1921.

(20) Brooks C. L., III; Brunger A.; Karplus, M. *Biopolymers* **1985**, *24*, 843-865.

(21) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics. John Wiley & Sons: New York, **1988**.

(22) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578-581.

(23) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011-1021.

(24) MOE, Chemical Computing Group Inc., 1010 Sherbrooke Street West, Suite 910, Montreal, Quebec, Canada H3A 2R7.

Appendix

Figure VII-1. The methyl transfers in PKMTs

(A) Lysine residue has three protons, and each of them may be replaced by a methyl group. Therefore, the lysine residue may be mono-, di- or tri-methylated by PKMTs. **(B)** Hypothetical processive methylation process catalyzed by the tri-methylase DIM-5. The basic form of the lysine or methyl lysine is the nucleophile, and the lone pair of electrons on its N_ζ is the site where additional methyl group can be added. Therefore, the further deprotonation of Lys-$N_\zeta(Me)H_2^+$ (Lys-$N_\zeta(Me)_2H^+$) is required before the 2nd (3rd) methyl group can be transferred from AdoMet to the methyl lysine (dimethyl lysine). Two of the possible ways by which the enzymes may control the product specificity are: (1) limiting the further methylation through the increase of the activation barrier[8-9] of methyl transfers and (2) stopping the deprotonation process.[10] **(C)** The definition of the structural parameters for monitoring the relative orientation of AdoMet and H3K9me [H3K9 and H3K9$(me)_2$] in the reactant complex. The efficiency of the methyl transfer may be related to the distributions of $r(C_M\ldots N_\zeta)$ 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 $R = r(C_M\ldots S_\delta) - r(C_M\ldots N_\zeta)$.

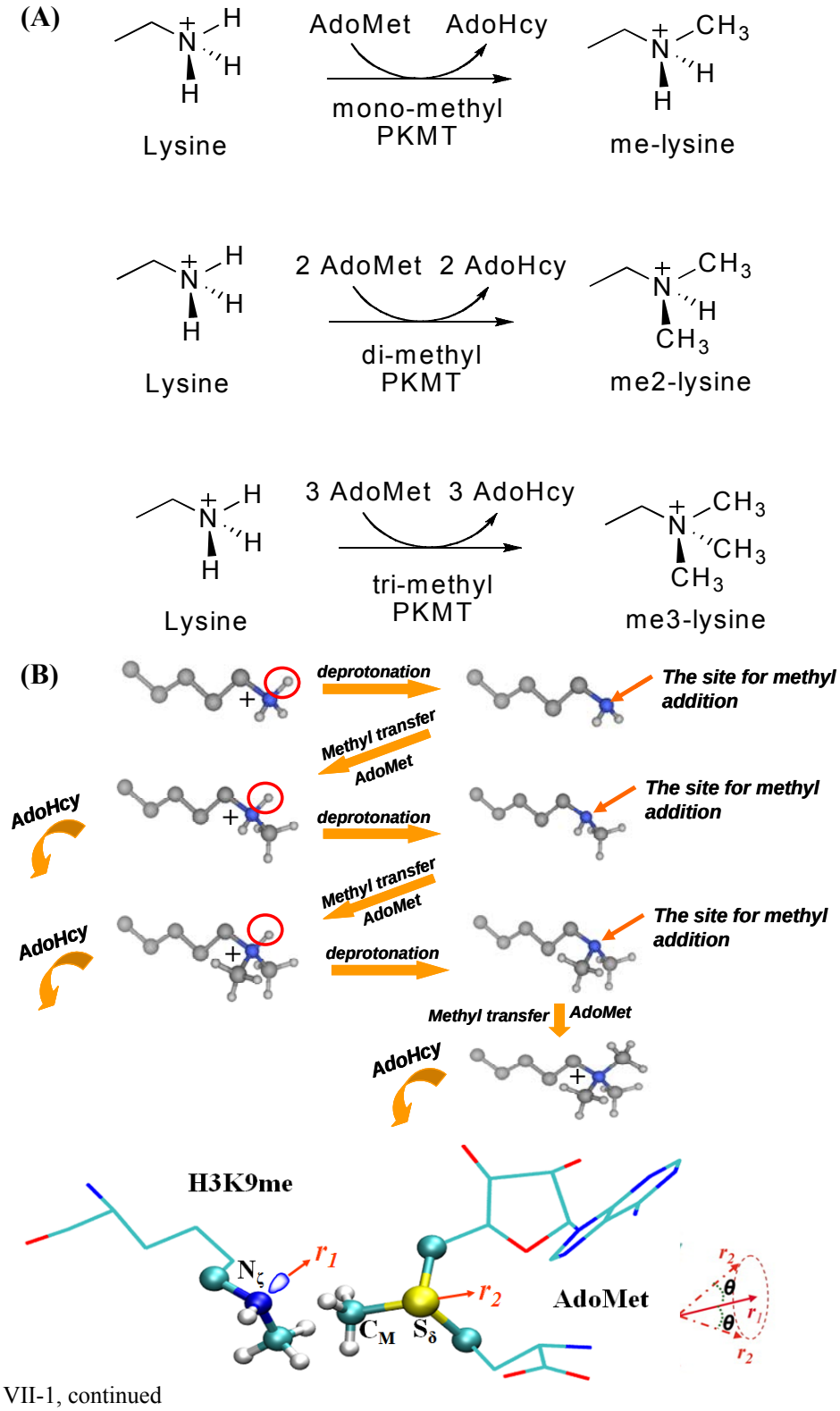


Figure VII-1, continued

Figure VII-2. The average structures and $r(C_M \ldots N_\zeta)/\theta$ distributions of the reactant complexes for methyl transfers in wild type DIM-5

(A) The average active-site structure of the reactant complex for the 1st methyl transfer in DIM-5 (DIM-5 complexed with AdoMet and histone H3K9 peptide) obtained from the QM/MM MD simulations. DIM-5 is shown in balls and sticks, and AdoMet and the H3K9 sidechain are in sticks. Hydrogen bonds are indicated by red dotted lines, and carbon-oxygen hydrogen bonds involving the methyl group are shown in green dotted lines. The distances related to the reactant coordinate are also shown. **(B)** The average structure of the reactant complex for the 2nd methyl transfer (DIM-5 complexed with AdoMet and H3H9me). The hydrogen atoms of the methyl group on H3K9me are not shown for clarity. **(C)** The average structure of the reactant complex for the 3rd methyl transfer [DIM-5 complexed with AdoMet and H3K9$(me)_2$]. The hydrogen atoms of the methyl groups on H3K9$(me)_2$ are not shown for clarity. **(D)** Two dimensional plot of $r(C_M \ldots N_\zeta)$ and θ distributions from the 1ns QM/MM MD simulations of the reactant complex for the 1st methyl transfer in DIM-5. **(E)** Two dimensional plot of $r(C_M \ldots N_\zeta)$ and θ distributions of the reactant complex for the 2nd methyl transfer. **(F)** Two dimensional plot of $r(C_M \ldots N_\zeta)$ and θ distributions of the reactant complex for the 3rd methyl transfer.

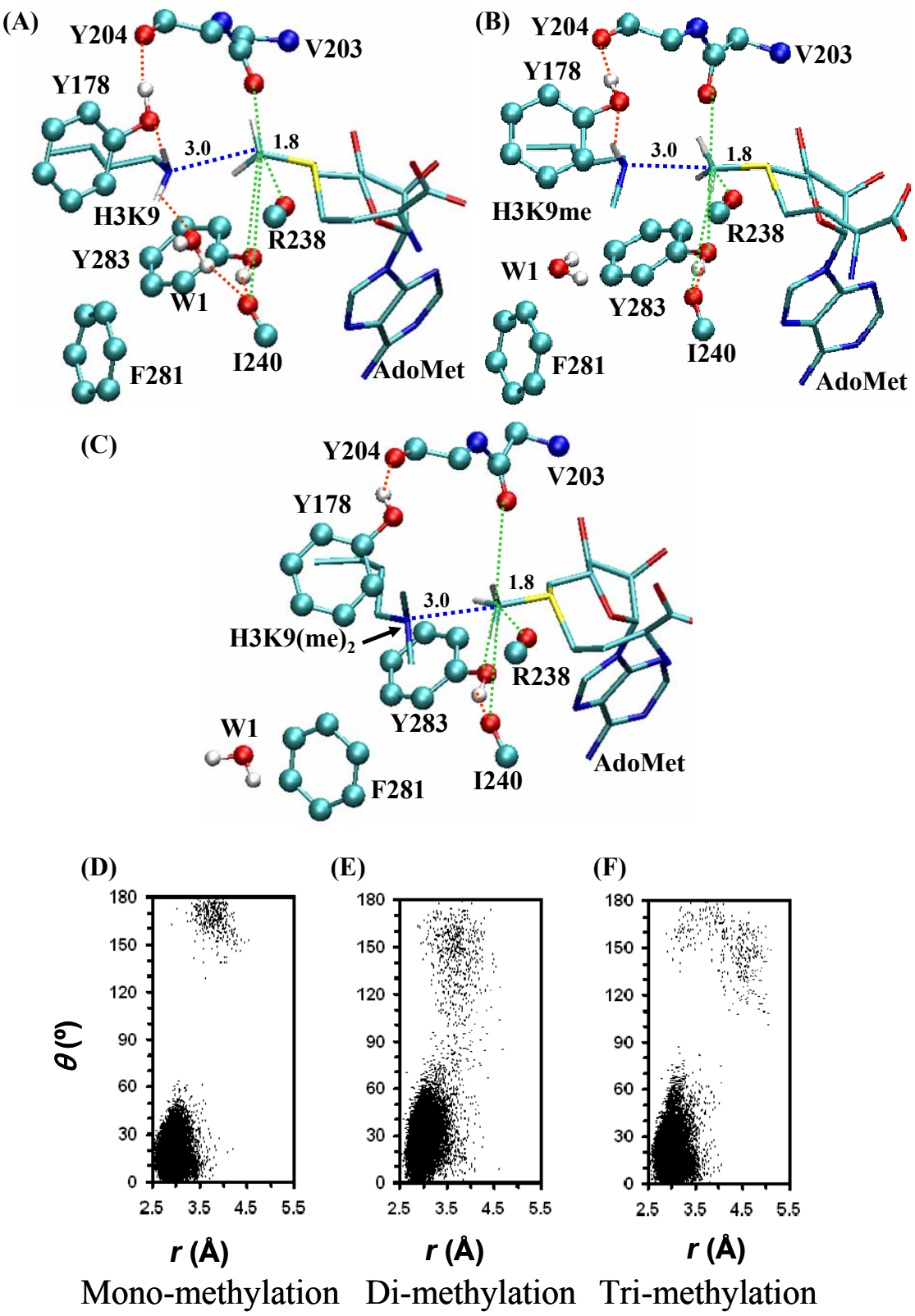


Figure VII-2, continued.

Figure VII-3. The average structures and $r(C_M \ldots N_\zeta)/\theta$ distributions of the reactant complexes for methyl transfers in DIM-5 F281Y mutant

(A) The average active-site structure of the reactant complex for the 1st methyl transfer in F281Y (F281Y complexed with AdoMet and histone H3K9 peptide). **(B)** The average structure of the reactant complex for the 2nd methyl transfer in F281Y (F281Y complexed with AdoMet and H3H9me). The hydrogen atoms of the methyl group on H3K9me are not shown for clarity. **(C)** The average structure of the reactant complex for the 3rd methyl transfer in F281Y [F281Y complexed with AdoMet and H3K9$(me)_2$]. The hydrogen atoms of the methyl groups on H3K9$(me)_2$ are not shown for clarity. **(D)** Two dimensional plot of $r(C_M \ldots N_\zeta)$ and θ distributions of the reactant complex for the 1st methyl transfer in F281Y. **(E)** Two dimensional plot of $r(C_M \ldots N_\zeta)$ and θ distributions of the reactant complex for the 2nd methyl transfer. **(F)** Two dimensional plot of $r(C_M \ldots N_\zeta)$ and θ distributions of the reactant complex for the 3rd methyl transfer.

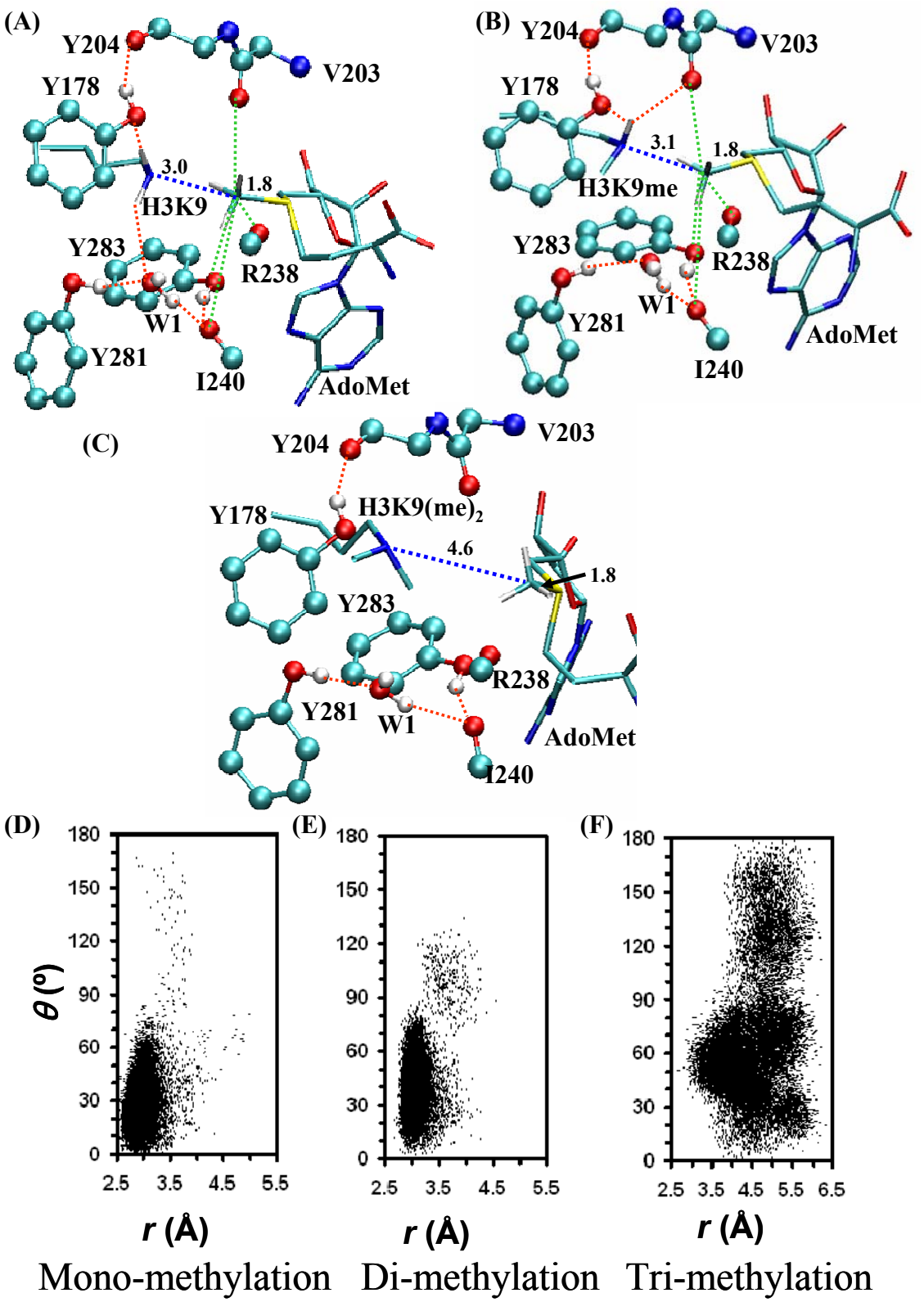


Figure VII-3, continued.

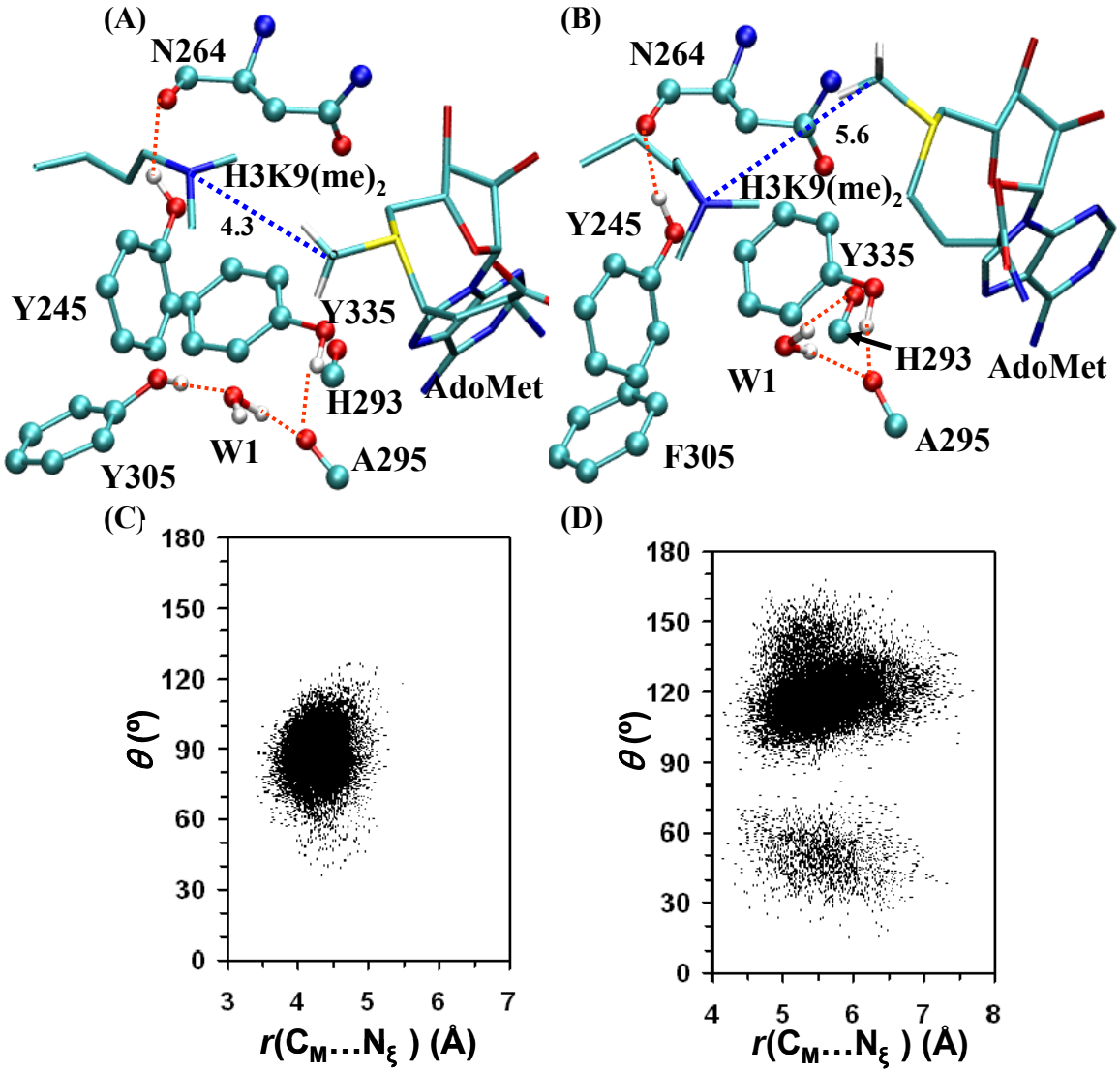


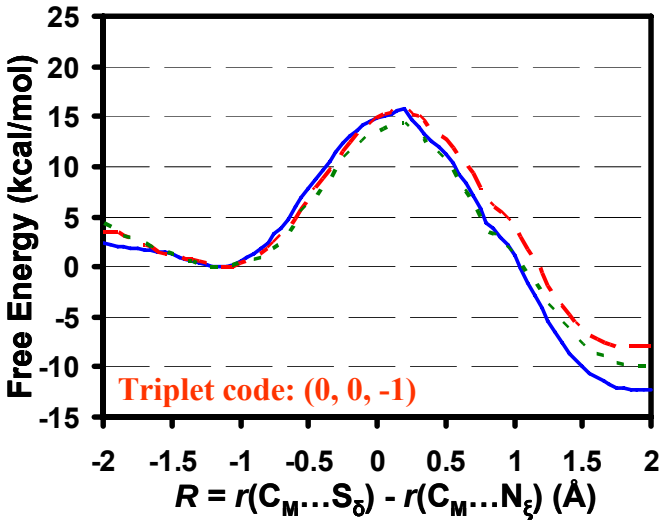
Figure VII-4 The average structures and $r(C_M…N_\zeta)/\theta$ distributions of the reactant complexes for the 3rd methyl transfers in the wild type and Y305F mutant of SET7/9

(A) The average active-site structure of the reactant complex for the 3rd methyl transfer in wild type SET7/9 [SET7/9 complexed with AdoMet and H3K4(me)$_2$] **(B)** The average active-site structure of the reactant complex for the 3rd methyl transfer in the Y305F mutant of SET7/9 [Y305F complexed with AdoMet and H3K4(me)$_2$]. **(C)** Two dimensional plot of $r(C_M…N_\zeta)$ and θ distributions in the reactant complex for the 3rd methyl transfer of wild type SET7/9. **(D)** Two dimensional plot of $r(C_M…N_\zeta)$ and θ distributions in the reactant complex for the 3rd methyl transfer of Y305F mutant.

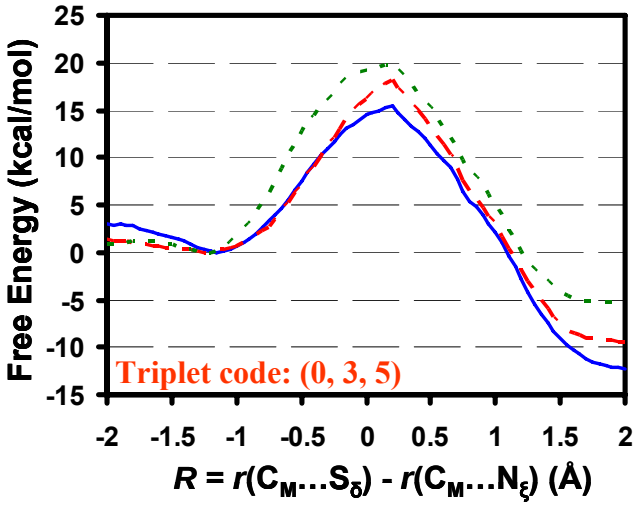
Figure VII-5. The free energy changes for the methyl transfers in the wild type and the F281Y mutant of DIM-5, as well as in wild type SET7/9

(A) The free energy (potential of mean force) changes for the 1st, 2nd and 3rd methyl transfers from AdoMet to H3-K9, H3-K9me and H3-K9(me)$_2$, respectively, as functions of the reaction coordinate [$R = r(C_M \ldots S_\delta) - r(C_M \ldots N_\zeta)$] in wild-type DIM-5. The 1st methyl transfer: blue and solid line; the 2nd methyl transfer: red and dashed line; the 3rd methyl transfer: green and dotted line. The relative free energy barriers can be represented by a triplet code ($0, \Delta_{2\text{-}1}, \Delta_{3\text{-}1}$) = (0, 0, -1). Here $\Delta_{2\text{-}1}$ and $\Delta_{3\text{-}1}$ are the changes of the free energy barriers from the 1st to the 2nd methyl transfer (about 0 kcal/mol) and from the 1st to the 3rd methyl transfer (about -1kcal/mol), respectively. The free energy barrier for the 1st methyl transfer is default to be zero in this triplet. **(B)** The free energy changes for the 1st, 2nd and 3rd methyl transfers as functions of the reaction coordinate in the F281Y mutant. The 1st methyl transfer: blue and solid line; the 2nd methyl transfer: red and dashed line; the 3rd methyl transfer: green and dotted line. The triplet code for the relative difference in activation barriers is ($0, \Delta_{2\text{-}1}, \Delta_{3\text{-}1}$) = (0, 3, 5). **(C)** The free energy changes for the 1st, 2nd and 3rd methyl transfers as functions of the reaction coordinate in wild-type SET7/9. The 1st methyl transfer: black and solid line; the 2nd methyl transfer: red and dotted line; the 3rd methyl transfer: blue and dashed line. The triplet code for the relative difference in activation barriers is ($0, \Delta_{2\text{-}1}, \Delta_{3\text{-}1}$) = (0, 5, 8).

(A)



(B)



(C)

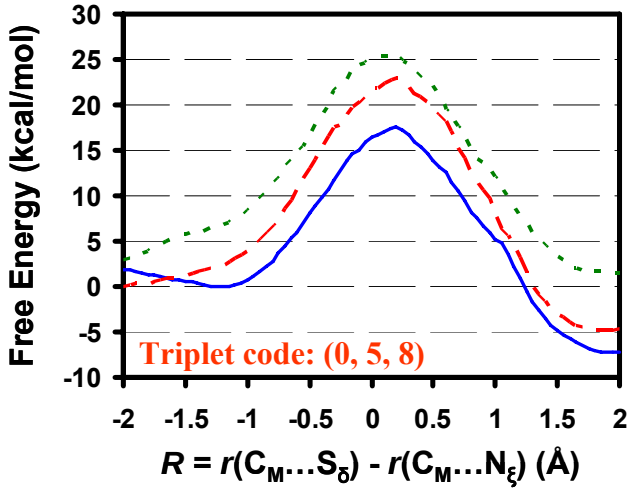


Figure VII-5, continued.

Chapter VIII

Conclusions and future directions

The dynamic features of enzymes in solution have been gradually recognized as an important factor for enzyme activities and specificities. However, the dynamic motions such as proton transfers, side chain movements and conformational changes of subdomains can not be easily reflected in the high resolution structures determined by X-ray crystallography. With the developments in NMR and many other methods, the conformational ensembles of enzyme structure can be observed at atomic level, but the changes of interactions and energies during the dynamic motions are still difficult to determine. Complementary to conventional structural analysis, computational simulations have the advantage of monitoring the dynamic motions in proteins or enzymes. More importantly, the energetic changes related to these motions can be calculated using computer programs.

In this dissertation, the quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) simulations, as well as free energy simulations using the umbrella sampling method and the weighted histogram analysis method (WHAM) program, were applied to the studies on the bindings of transition state analogue (TSA) to cytidine deaminase (CDA) and yeast cytosine deaminase (yCD), the catalytic mechanism of kumamolisin-As, as well as the product specificities of two protein lysine methyltransferases (PKMTs), DIM-5 and SET7/9.

In the simulations of the TSA bindings of zebularine 3, 4-hydrate (ZEB-H_2O) to CDA and of 4-[R]-hydroxyl-3,4-dihydropyrimidine (DHP) to yCD, proton transfers were observed between the TSA and a catalytic Glu residue (Glu104 in CDA and Glu64 in yCD). The absence of this critical Glu residue destabilizes the TSA at the active site. The

TSA of CDA, ZEB-H_2O was rapidly dehydrated at the active site of E104A mutant. In addition, the stabilization of the tetrahedral TSA by the Glu residue seems to be attributed to a proton transfer from the TSA to the Glu, instead of a hydrogen bond between the hydroxyl and the negative carboxylate of the Glu residue. In both CDA and yCD, free energy simulations estimated about 8kcal/mol lower energy for the tetrahedral TSA complex with this proton transferred than the one with the proton not transferred.

The general acid-base mechanism was also observed in the stabilization of the tetrahedral intermediate by Asp164 during the acylation of kumamolisin-As. Asp164 acts as a general acid to protonate the negative oxygen of the tetrahedral intermediate formed by the nucleophilic attack of Ser278. Then this residue acts as a general base and accepts the proton back from the tetrahedral intermediate during the formation of the acyl-enzyme. In addition to the stabilization by Asp164, Glue78 aslo acts as a general acid-base to mediate the proton transfer from Ser278 to the nitrogen of the scissile peptide bond. Moreover, dynamic substrate-assisted catalysis (DSAC) involving the His of the substrate at the P_1 site was proposed to be triggered by the bond breaking and making events during the acylation, expecially the protonation/deprotonation of the Asp164 carboxylate. It was suggested that the back and forth movements of the His sidechain between the hydrogen bond with carbonyl of Pro at P_2 site and the salt bridge with $O_{\delta 2}$ of Asp164 may contribute to the transition state stabilizations and substrate specificity of kumamolisin-As.

The origin of the product specificities of PKMTs was studied by comparison of QM/MM MD simulations on the first, second and third methyl transfers in the

trimethylase DIM5 and the monomethylase SET7/9, as well as in two of their mutant forms of the enzymes. The product specificities of the enzymes can be well explained by the population distributions of $r(C_M \ldots N_\zeta)$ and the angle θ (the distance for N_ζ to attack the methyl group and the angle between the orientation of the electron lone pair on N_ζ relative to the C_M-S_δ bond) in the dynamic conformational ensemble of the well-aligned reactant structures, as well as the relative free energy barriers for the methyl transfers. Based on the relative diffierences in the free energy barriers for the three methyl transfers, a triplet code was proposed in the determination of product specificities of PKMTs.

In summary, our results of QM/MM MD simulations on different enzymes (CDA, yCD, kumamolisin-As and PKMTs) support the importance of dynamics in enzymatic catalysis. In the first three cases, we reported the importance of the proton transfers to lower the reaction barriers and to stabilize the tetrahedral intermediates or transition state analogues at the active sites. This is in contrast with the previous proposals based on the structural analyses of these enzymes, in which the hydrogen bonding interactions were the key reason for the stabilizations. In the case of kumamolisin-As, we also proposed DSAC, in which the conformational changes of the substrate groups and formation of alternative interactions may lower the energetic barriers at different stages of the reaction cycle. This is an important development for traditional SAC studies, which focus on the contributions from well positioned substrate groups without involving conformational changes. In the case of PKMTs, we designed a triplet code reflecting the differences in the energetic barriers for the methyl transfers catalyzed by PKMTs and used this triplet code to explain the product specificities of PKMTs and their mutants.

Based on our current simulation conditions, future studies in QM/MM MD simulations may still focus on reactions in local regions. For examples, we may further study the binding of other TSAs on CDA, such as 3,4,5,6-tetrahydrouridine (THU), so as to confirm the importance of the proton transfer to Glu. We are also interested in the role of the critical Glu in catalytic processes of cytidine deamination in CDA and cytosine deamination in yCD, which was suggested as a general acid-base proton shuttle both during the nucleophilic attack and in the formation of product. We may continue our studies on kumamolisin-As in the process of deacylation and attemp to determine the rate limiting step of the reaction cycle. The collaboration between the proton transfers is of great interest also. We may further our studies on PKMTs in many aspects. For instance, the Tyr245 mutants of SET7/9 involve another residue critical to the methyltransferases activity of PKMTs. The Y245F mutant of SET7/9 and the corresponding Y178F mutant of DIM-5 both have dramatically lowered activities but their product specificities are not changed. However, it is quite interesting that the Y245A mutant of SET7/9 has higher efficiencies for the second and third methyl transfers compared to the first methyl transfer. Monitoring of the reactant complexes before the first, second and third methyl transfers in the Y245A mutant of SET7/9 may help to explain this difference in efficiency of methyl transfers. The product specificities of other PKMTs, such as the monomethylase SET8 may also be simulated. What is more, we are also interested in studies of the deprotonation, such as the possible role of Tyr335 as suggested in our earlier works (Ref. 8 of chapter VII). Finally, we would like to expand our computational simulations to the studies of other enzymatic reactions and catalysis.

VITA

Qin Xu was bron in Beijing, China in 1979. He attended Peking University in 1997 and graduated with a Bachelor of Science in 2001. He began his graduate study in 2002 at the University of Tennessee, Knoxville. He defended his dissertation in Aug., 2008.