Simulating Biochemical Physics with Computers

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Pinsker Yen-lin Lin

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Prof. Jiali Gao, Advisor

July, 2010

Acknowledgements

I would never have been able to finish my dissertation without the guidance of my advisor, help from friends, and support from my family.

I would like to express my deepest gratitude to my advisor, Prof. Jiali Gao, for his excellent guidance, caring, patience, providing me with an excellent atmosphere for doing research and financially supporting my study. I would like to thank Prof. Carmay Lim, who showed me the entrance of the research field to let me experience practical issues beyond the textbooks. I would also like to thank Prof. Sanford Lipsky, Prof. J Ilja Siepmann, Prof. Darrin York, and Prof. Gianluigi Veglia for the past several years and helping me to develop my background in physical chemistry and biochemistry. The members of my final defense committee, Prof. Christopher Cramer, Prof. David Blank, and Prof. David Thomas, have generously given their time and expertise to better my work. I thank them for their contribution and their good-natured support.

I am grateful to many persons, who were always willing to help and give his/her best suggestions, especially Prof. Shuhua Ma, Dr. Kin-Yiu Wong, Prof. Dan Major, Dr. Alessandro Cembran and Yan Zhou. I must acknowledge as well Dr. Yao Fan, Dr. Peng Bao, Dr. Jorge Mews Estevez, Dr. Wangshen Xie, Qiang Gao, Lei Shi, Jaebeom Han, Peng Zhang, and other workers in the laboratory of Prof. Gao for their friendship. Special thanks goes to Yan Zhou. It would have been a lonely lab without her.

Especially, I need to express my deep gratitude and appreciation to Harris Liu, whose support makes my dream come true. I couldn't go this far without him. I would also like to thank Zheng Chang, who gives me strength to face difficulties.

Finally, and most importantly, I offer my regards and blessings to my beloved family – my parents and my brothers. They were always supporting me and encouraging me with their best wishes during the completion of my Ph.D. study. A special thanks goes to the greatest woman – my Mother. She is my energy and my courage.

Minneapolis, July 2010
Pinsker Yen-lin Lin

To a remarkable person, my Mother…

Abstract

This dissertation is composed of three parts. The first part is to argue the solvent effects on the solvatochromic shift of the $n \rightarrow \pi^*$ excitation of acetone in ambient and supercritical water fluid using a hybrid QM−CI/MM potential in MC simulations. The solute is described by the AM1 approach and water molecules are treated classically. Specially, the spontaneous polarization of the solvent due to the excitation of the solute was considered. The solvent effects on the blue shift of acetone in water fluids at various temperatures and solvent densities are examined.

The second part is to investigate the role of dopa decarboxylase (DDC) in the catalysis of converting anti-Parkinson drug L-dopa into dopamine. By means of combined QM/MM potentials in MD simulations, we first analyze the factors contributing to the tautomeric equilibrium of an intramolecular proton transfer in the external PLP−L-dopa aldimine (the Michaelis complex). How the intrinsic properties, solvent effects as well as the enzyme environment control the shift of the equilibrium is discussed. Afterward, the free energy profiles for the decarboxylations of the external aldimines both in water and in DDC are calculated. The contributions of DDC to the rate enhancement of the reaction are elucidated. The reaction mechanism of L-dopa decarboxylation in DDC is proposed.

The third part is to study the structural dynamics of lysine-specific demethylase (LSD1) in complex with CoREST and protein-substrate interactions of LSD1 with histone H3 tail. MD simulations of LSD1•CoREST complex bound to a 16 a.a. of the N-terminal H3-tail peptide (H3-p16) were carried out using NAMD to study the conformational flexibility of the protein complex, especially the substantial oscillation of the TOWER domain. In addition, the simulations reveal some important protein-peptide and peptide-peptide interactions between LSD1 and H3-p16 that are absent in the crystal structure.

Contents

List of Tables

List of Figures

List of Schemes

Preface

This dissertation contains previously published work and is reproduced with permission.

Chapter 2:
"Solvatochromic Shifts of the $n \rightarrow \pi^*$ Transition of Acetone from Steam Vapor to Ambient Aqueous Solution: A Combined Configuration Interaction QM/MM Simulation Study Incorporating Solvent Polarization", Yen-lin Lin and Jiali Gao, *J. Chem. Theory Comput.,* **2007**, *3*, 1484–1493. © 2007 American Chemical Society.

Chapter 3:
"Internal Proton Transfer in the External Pyridoxal 5´-Phosphate Schiff Base in Dopa Decarboxylase", Yen-lin Lin and Jiali Gao, *Biochemistry* **2010**, *49*, 84–94. © 2010 American Chemical Society.

Appendix D:
"A Non-Orthogonal Block-Localized Effective Hamiltonian Approach for Chemical and Enzymatic Reactions", Alessandro Cembran, Apirak Payaka, Yen-lin Lin, Wangshen Xie, Yirong Mo, Lingchun Song, and Jiali Gao, *J. Chem. Theory Comput.*, 2010, *6*, 2242–2251. © 2010 American Chemical Society.

"Potential Using an initio Molecular Orbital Theory and Density Functional Theory", Lingchun Song, Jaebeom Han, Yen-lin Lin, Wangshen Xie, and Jiali Gao, *J. Phys. Chem. A* **2009**, *113*, 11656–11664. © 2009 American Chemical Society.

"Kinetic Isotope Effects from Hybrid Classical and Quantum Path Integral Computations" in ***Kinetic Isotoped Effects in Enzyme Catalysis***, Jiali Gao, Kin-Yiu Wong, Dan T. Major, Alessandro Cembran, Lingchun Song, Yen-lin Lin, Yao Fan, and Shuhua Ma, Eds. R. Alleman and M. Scruton, RSC Biomolecular Sciences, 2009, pp 105–131.

Chapter 1

Introduction

The enormous catalytic power of enzymes to accelerate the rates of chemical reactions has been a mystery for a long time, and remains to be a key fundamental question in biochemistry. How do enzymes enhance the rates of sophisticated chemical transformations and exquisite specificity toward substrates? The three-dimensional structure of an enzyme is the most important factor in achieving this remarkable catalytic power, and extensive experiments and computer simulations have provided abundant evidence. Nevertheless, the dynamic fluctuations and internal motions of enzymes are also fundamentally important for their biological functions and have been a focus of many experimental and computational enzymologists (*1.1–1.5*).

With the improvement of computer speed and the development of novel algorithms, many novel computational tools have become available for studying the mechanism and dynamics of enzymatic reactions (*1.2*, *1.6–1.8*). In particular, the work presented in this thesis makes use a broad range of computational methods, including high-level ab initio and density functional theory for small molecules and their reactions in the gas phase, molecular dynamics simulations employing classical force fields for large biomolecules in water solution, and combined quantum mechanical and molecular mechanical (QM/MM) models for enzyme-catalyzed processes. Of critical importance for successful computer simulations of enzymes is the potential energy function or a force field that describes intermolecular and intramolecular interactions.

The currently available methods may be divided into two main categories: empirical force fields, which is computationally efficient, and quantum mechanical models such as semiempirical and ab initio wave function theory and density functional electronic theory, which have increased computational demands. Each category has its own advantages and disadvantages and each method is still continuously being improved. For example, a significant effort is being spent to incorporate explicit polarization terms into the empirical force fields, while methods are being developed to use quantum mechanical models directly as a force field for macromolecular simulations. For the purpose of modeling enzymatic catalysis, in which only a small portion of the system requires the explicit treatment of the electronic structure using a quantum mechanical representation, it is natural to combine quantum mechanics with molecular mechanics, or QM/MM, in which a

quantum mechanical method is used to describe the active site to simulate the bond formation and cleavage processes, and the computationally efficient molecular mechanics is used to model the remainder of the system.

Over the past decade, computational procedures for studying enzyme mechanisms have become well established. First, it entails the determination of the potential of mean force, or the free energy reaction profile, along a pre-defined reaction coordinate in molecular dynamics simulations using a combined QM/MM potential. Although many studies are still reported with the use of energy minimization techniques, it is likely that these calculations do not capture the true origin of enzyme catalysis because the contributions and effects of protein dynamic fluctuations are ignored, and these fluctuations are known to be critical to the function and properties of enzymes (*1.5*, *1.8*, *1.9*). Having established a possible reaction mechanism, the second step in computational study of enzyme catalysis is to explore the specific contributions of enzyme dynamics using the techniques of activated molecular dynamics simulations (*1.5*, *1.10*, *1.11*). Dynamic factors in enzymatic mechanism may broadly include the classical transmission recrossing factor of transition state theory, quantum mechanical tunneling, and non-equilibrium dynamic fluctuations of the enzyme-solvent system. Clearly, not all factors are necessarily equally important for a given enzymatic process, and computational details may differ depending on the specific questions to be investigated. Based on these procedures, I will summarize in Chapters 3–5 in this dissertation results that help to enhance our understanding of the underlying catalytic principles and the reaction mechanism of

the enzyme dopa decarboxylase, which is involved in the treatment of Parkinson disease and catalyzes the decarboxylation reaction of aromatic amino acids.

Molecular dynamics simulation is instrumental in changing the view of proteins as relatively rigid structures in the early days when crystal structures of proteins and enzymes became available. Now, the study of protein dynamics and its relationship with function has become the central focus of experimental and computational investigations. Undoubtedly, internal motions of a protein play an integral role in determining its biologically function, and dynamics simulations can provide a detailed picture of thermal fluctuations and the associated spatial and temporal scales. Lysine specific demethylase-1 (LSD1) has captured enormous attention in the past 10 years since its discovery due to the association with gene expression through its capability of reversibly removing protein lysine modifications (*1.12−1.18*). In fact, the function of demethylases is a central component of the so-called histone code hypothesis in the understanding of epigenetics. Here, by sampling the conformational space of the LSD1•CoREST complex bound to the target histone 3 tail, the results of nanosecond MD simulations in water provide insights into the complex's behavior in aqueous solution, establishing a foundation for further investigations on the mechanisms and specific interactions in the enzymatic catalysis.

Many chemical and biological reactions involve electronically excited states, such as the photoisomerization of retinal in rhodopsin (*1.19−1.21*). Thus, it is of importance to model the excited state events in biological systems in order to gain an understanding of the underlying biophysical interactions. Our group has developed configuration interaction (CI)-QM/MM techniques for modeling excited states of

4

chromophores in solutions and in proteins. In Chapter 2, the solvatochromic shifts of a probe molecule acetone were studied in the entire range aqueous solvation, from steam vapor to supercritical conditions to ambient water. The effect of instantaneous solvent polarization, using a polarizable force field for the solvent, was taken into account by following the solute electronic excitation, providing a prelude for studying excited states in proteins.

Key results in each chapter are briefly summarized below:

- Chapter 2: In this chapter, a hybrid quantum mechanical and molecular mechanical potential is used in Monte Carlo simulations to examine the solvent effects on the electronic excitation energy of the $n \rightarrow \pi^*$ transition of acetone in ambient and supercritical water fluid, which covers a temperature range from 25 to 500 °C and pressures from 1 to 2763 atm. In this study, the chromophore acetone molecule is represented by the Austin Model 1 (AM1) Hamiltonian and the water molecules are treated classically. Two sets of calculations are performed. The first involves the TIP4P model for water, and the second employs a polarizable model, POL2, for the solvent. The first calculation yields the excitation energy by using a static ground-state solvent charge distribution obtained from QM-CI/MM calculations. The latter takes into account the effect of solvent polarization following the solute electronic excitation. The trend of the computed $n \rightarrow \pi^*$ blue shifts for acetone at different water fluids are in good agreement with experimental results. The Monte Carlo simulations of acetone in the supercritical, near supercritical, dense-liquid and ambient water fluids reveal that the solvatochromic shifts are

dominated by the electrostatic interactions between acetone and water molecules during the solute excitation. Additionally, the solvent charge redistribution following the solute electronic excitation has a small contribution ($0 \sim -37$ cm^{-1}) to the total solvatochromic shift and decreases linearly with water density. Both the solvatochromic shift and solvent polarization correction are more obvious in the ambient water than in the supercritical water because the solvent stabilization of the ground state over the excited state is more significant in the former condition.

- Chapter 3: This chapter documents combined quantum mechanical and molecular mechanical simulations of dopa decarboxylase that have been carried out to elucidate the factors contributing to the tautomeric equilibrium of the intramolecular proton transfer in the external PLP−L-dopa Schiff base. The presence of a carboxylate anion on the α-carbon of the Schiff base stabilizes the zwitterions and shifts the equilibrium in favor of the oxoenamino tautomer (protonated Schiff base). Moreover, protonation of the PLP pyridine nitrogen further drives the equilibrium towards the oxoenamino direction. On the other hand, solvent effects favors the hydroxyimino configuration, although the equilibrium favors the oxoenamino isomer with a methyl group as the substituent on the imino nitrogen. In dopa decarboxylase, the hydroxyimino form of the PLP(H+)−L-dopa Schiff base is predicted to be the predominant form with a relative free energy of −1.3 kcal/mol over that of the oxoenamino isomer. Both Asp271 and Lys303 stabilize the hydroxyimino

configuration through hydrogen-bonding interactions with the pyridine nitrogen of the PLP and the imino nitrogen of the Schiff base, respectively. Interestingly, Thr246 plays a double role in the intramolecular proton transfer process, in which it initially donates a hydrogen bond to the phenolate oxygen in the oxoenamino configuration, and then switches to a hydrogen bond acceptor from the phenolic hydroxyl group in the hydroxyimino tautomer.

- Chapter 4: This chapter describes a study of the decarboxylation reaction of L-dopa in dopa decarboxylase using a combined quantum mechanical and molecular mechanical potential in molecular dynamics (MD) simulations to elucidate the catalytic effect of the enzyme. It shows that pyridoxal 5′-phosphase (PLP) cofactor intrinsically accelerates the reaction rate by about 11 orders of magnitude relative to the uncatalyzed reaction in aqueous solution. The participation of PLP in the enzymatic decarboxylation also provides stabilization on the product intermediate of the reaction. In addition, dopa decarboxylase isolates the reaction center from water solvent by way of forming a solvent-occluded cavity with its catalytic loop to prevent solvent from stabilizing the carboxylate group in the Michaelis complex in the decarboxylation and thus further accelerates the reaction rate by about 7 orders of magnitude. Furthermore, the external hydroxyimine PLP aldimine is predicted to be preferred over the oxoenamine isomer in the enzymatic reaction by having a faster reaction rate and producing a more stable carbanion intermediate.

- Chapter 5: This chapter exposes the nuclear quantum effects on the decarboxylation reaction of L-dopa in dopa decarboxylase using a path-integral free-energy perturbation method. The predicted primary ^{13}C-kinetic isotope effects of the carboxyl carbon atom for the enzymatic decarboxylation reaction are in agreement with experiments on this class of enzymes, although experimental studies of kinetic isotope effects on this particular enzyme has not been performed. It turns out that the nuclear quantum effects contribute to catalysis in dopa decarboxylase, although the dominant effect in rate acceleration by the enzyme is the lowering of the quasiclassical free energy barrier in the solvent-excluded active site of the enzyme. In addition, hydroxyimine and oxoenamine PLP−L-dopa aldimines exhibit differences in kinetic isotope effects for the decarboxylation reaction that may be traced to different intramolecular bonds at the different transition states of the reactions with different PLP-aldimine tautomers.

- Chapter 6: Lysine-specific demethylase 1 (LSD1) is an enzyme that catalyzes the demethylation process on mono- and dimethylated lysine residues, which has been implicated in gene activation or repression depending on the interacting protein factors. An important question is the dynamical flexibility of the protein complex between LSD1 and the auxiliary protein CoREST, which recognizes specific DNA sequences around the histone proteins. In this chapter, by means of molecular dynamics simulations, the motion of the LSD1•CoREST protein-protein complex, where the active cavity of LSD1 is bound to a 16-a.a. substrate of N-terminal histone H3 tail (H3-p16) are

8

investigated. We report the conformational fluctuations of the protein complex in aqueous solution over a 20-ns simulation. The substantial flexibility of the TOWER domain of LSD1 observed in this work indicates that the domain tolerates a limited range of conformational oscillation. In addition, by compared with the crystal structure, the conformational rearrangement of the H3-p16 peptide in the catalytic cavity was examined with the presence of explicit water solvent. The simulations reveal some important protein-peptide and peptide-peptide interactions between LSD1 and the H3-p16 tail peptide.

The thesis also includes three Appendix chapters, documenting research results that have been published in the literature. Appendix A documents the supporting information of Chapters 2, listing the numerical data of the computed solvatochromic shift (Δv) and the corresponding reduced densities (ρ_r) in the $n \rightarrow \pi^*$ excitation of acetone in water fluid at various temperatures. Appendix B reports the supporting information of Chapters 3, including the fully optimized structures and atomic geometries of oxoenamine and hydroxyimine tautomers, the energies and free energies for the studied tautomerization reactions in the gas phase, as well as the free energies and solution free energies in aqueous solution. Appendix C records the free energy profiles for the decarboxylation reaction of L-dopa in dopa decarboxylase with different loop conformation involving in the enzymatic catalysis. Appendix D lists the computed KIEs of individual block-averages for the decarboxylation reactions of PLP−L-dopa in the enzyme dopa decarboxylase as well as in water solution.

Appendix D includes my other publications that I have contributed as a part of collaboration during my doctoral studies.

Chapter 2

Supercritical Solvation with Explicit Polarization: Towards an Understanding of Solvent Effects and the Solvatochromic Shifts of Acetone from Steam Vapor to Ambient Aqueous Solution

2.1 Introduction

The unusual properties of supercritical water (SCW) fluid have attracted considerable interests (*2.1–2.6*) as a viable medium for green chemical oxidations (*2.7–2.9*). At or near supercritical conditions, organic species and molecular oxygen are completely miscible (*2.2–2.7, 2.10, 2.11*), whereas electrolytes are nearly insoluble (*2.7, 2.12*). Thus, it offers a tremendous opportunity to develop alternative technologies for the destruction of chemical warfare agents and organic wastes by complete oxidation (*2.7, 2.9*). Another advantage of performing chemical oxidations

in supercritical fluid water (T_c = 374 °C, P_c = 217.7 atm, and ρ_c = 3.22 g·cm^{-3}) or near supercritical conditions is that the reaction conditions can be optimized by varying the density of the medium with a change in pressure. The variations of the fluid density, ranging from steam vapor to dense aqueous solution also provide an interesting medium for investigating solute-solvent interactions (*2.13–2.15*). The study presented in this Chapter is aimed at an understanding of the change of solute and solvent interactions over the entire spectra of solvent density and the unusual behaviors of solvation near supercritical conditions. We focus our study on the continuous change of solvatochromic shifts of the chromophore acetone as a probe solute in fluid water by using a combined quantum mechanical configuration interaction and molecular mechanical potential in statistical mechanical Monte Carlo simulations.

Solvatochromic shifts of organic chromophores have been used extensively as a probe to investigate solute-solvent interactions in solution (*2.16–25*). Based on the change in electronic absorption spectra of organic dye molecules, solvent polarity scales have been established including the popular $E_T(30)$ scale based on Reichardt's betaine dye (*2.16, 2.26, 2.27*). One group of chromophores, containing carbonyl, thiocarbonyl, and azo functional groups, are often used, which have characteristically weak $n \rightarrow \pi^*$ absorption bands (*2.16*). Typically, a blue shift in the absorption spectrum is observed in going from a low dielectric solvent to a more polar medium, although dispersion red-shifts are also found in nonpolar solvents such as carbon tetrachloride and hexane (*2.16, 2.28–2.30*).

Continuum solvation models coupled with electronic structure calculations have been widely used to model solvatochromic shifts (*2.18, 2.20, 2.31–2.34*). The

ZINDO program and its associated methods developed by Zerner have been applied to a variety of chromophores with remarkable success (2.18). Cossi and Barone evaluated the $n \rightarrow \pi^*$ transition of acetone in various polar and nonpolar solvents using the polarizable continuum solvation model (2.31), while a number of other groups have also studied this system using different techniques ($2.20, 2.33-2.40$). Although excellent agreement with experiment can be obtained, a shortcoming of the continuum solvation approach is a lack of treating specific hydrogen bonding interactions. Zerner showed that only when one or two explicit water molecules are included, would the computed spectral shifts for a series of pyrimidine and pyrazine compounds be in accord with experiment. On the other hand, combined QM/MM simulations even at the level of configuration interaction with only single excitations (CIS) can yield reasonable results (2.23). Of course, the latter computations are much more time-consuming as it requires configurational averaging over millions of solvent configurations. Avoiding explicit electronic structure calculations, Warshel and coworkers used the partial charges derived for the ground and excited states along with an atom-centered polarizable dipole model to determine the solvent effects on vertical excitation energy (2.41). This approach has been used by Levy et al. (2.42) and by Kollman et al. (2.43) in the analysis of excited state energy relaxation. Previously, our group developed a combined QM-CI/MM approach in Monte Carlo simulations, and applied it to a number of systems, including acetone in a variety of solvents ($2.22-2.24, 2.44, 2.45$). Later, the method was extended by incorporating a consistent treatment of the instantaneous electronic polarization between the solute and solvent in response to solute excitation ($2.23, 2.46$). Thompson and Schenter also

13

presented a combined QM-CI/MM-pol model that includes polarization effects in the MM region and have applied it to study both the ground and excited states (*2.47, 2.48*). In addition, Olivares del Valle *et al.* presented a strategy using the mean-field approximation combining QM/MM method to calculate the solvent shift of acetone in the ambient water (*2.40*). Combined QM/MM strategy has also been implemented in CASSCF calculations (*2.25*).

On the experimental side, Bennett and Johnston carried out a most comprehensive study, and measured the entire range of solvatochromic shifts of the $n \rightarrow \pi^*$ absorption band of acetone in vapor, fluid and liquid water (*2.1*). The experimental results showed that the spectral shifts can be divided into three regions. First, there is an initial phase of rapid increase in spectral shift, relative to the excitation energy of the isolated chromophore acetone in the gas phase, in the low density steam region. This is followed by a plateau region near supercritical fluid conditions. Finally, as the fluid density increases toward the ambient value, the absorption energy increases quickly again. The existence of a plateau region near supercritical conditions has been proposed as a feature of supercritical fluid solvation due to solvent clustering (*2.1, 2.49, 2.50*). In a separate study, Takebayashi et al., who utilized NMR spectroscopy and Monte Carlo simulations, found similar features, which were attributed to the variations in solute-solvent hydrogen bonding as the temperature and water density changes (*2.51, 2.52*). In the same vein, a number of molecular dynamics and Monte Carlos simulations have been reported, primarily focusing on solute-solvent interactions at or near the supercritical fluid region at a few selected states (*2.15, 2.53–2.55*). These studies provided support to solvent

clustering at supercritical conditions. Recently a classical Monte Carlo simulation of acetone in water followed by cluster calculations with semiempirical and time-dependent density functional theory has been reported at the supercritical point (*2.37*), the change of solvatochromic shifts of an organic chromophore in the entire range of solvent densities has not been demonstrated computationally.

In this work, we aim to assess the solvent effects on the $n \rightarrow \pi^*$ blue shift of acetone in the full region from steam vapor to supercritical conditions to ambient water. The computed $n \rightarrow \pi^*$ solvatochromic shifts of acetone in water fluids at various temperatures and solvent densities are compared with experimental values (*2.1*). To evaluate the contributions of different molecular interactions to the acetone $n \rightarrow \pi^*$ blue shift in these fluid states, a decomposition analysis of the energies was computed based on the method our group developed previously (*2.23, 2.46*). To this end, statistical Monte Carlo simulations using a hybrid quantum-mechanical–configuration interaction and molecular mechanical (QM-CI/MM) method have been carried out to explore the solvent effects in electronic spectroscopy. The effects of the solvent polarization in response to the solute electronic excitation is evaluated by using a polarizable MM solvent model (*2.23, 2.46, 2.56*). The results of the calculations reveal the factors governing the solvatochromic shifts of acetone at different water densities and temperatures, where a polarization correlation term from the instantaneous polarization of the solvent molecules following the solute excitation was also estimated. In the following, we first present the theoretical background and computational details. This is followed by

results and discussion. Finally, the main findings of this work are summarized in the conclusion.

2.2 Theoretical Methods

We use a combined quantum mechanical–configuration interaction and molecular mechanical (QM-CI/MM) potential in statistical mechanical Monte Carlo simulations to investigate the solvatochromic shifts in the $n \rightarrow \pi^*$ transition of acetone in fluid water (*2.22, 2.23*). In this hybrid system, the solute chromophore (*i.e.* acetone) is treated by a CI wave function and the solvent molecules are represented classically by empirical potential functions (*2.22, 2.47*). Except a few cases, most applications of combined QM/MM potentials make use of effective pairwise potentials for the MM region, in which the partial atomic charges on the solvent atoms are fixed at the same values both for the ground and excite states of the solute. This fixed-charge approach ignores the instantaneous charge polarization of the solvent due to solute electronic excitation, *i.e.* the interaction of the QM and solvent-induced dipoles was not involved, and the change in solvent configurations. In the present work, we also use a polarizable solvent model for water. Thus, the mutual "QM" solute and "MM" solvent polarization interactions are explicitly treated (*2.23, 2.46–2.48*). This is of particular interest in the present study because we examine the solvation of acetone by fluid water that covers the entire range of solvent densities, ranging from steam vapor to supercritical fluid to the dense liquid at the ambient condition. Previously we have implemented a polarizable combined QM/MM method for the study of electronic absorption in polar solvent (*2.23*) and the approach is the

16

same as that described by Thompson and Schenter (*2.47, 2.48*). Here, we study solvation effects in supercritical fluids by including the instantaneous polarization of solvent molecules in response to the solute excited-state wave function.

Energy Decomposition. The total ground-state energy of the QM/MM hybrid system with the utilization of a polarizable solvent model can be written as follows (*2.23, 2.46, 2.48*),

$$E_{tot}^g = \left\langle \Phi_{CI}^g \left| \hat{H}_X^o + \hat{H}_{Xs}^{stat}(\{q_s\}) + \hat{H}_{Xs}^{pol}(\{\boldsymbol{\mu}_s^g\}) \right| \Phi_{CI}^g \right\rangle + E_{Xs}^{vdW}$$
$$+ E_{ss}^{pair} - \frac{1}{2}\sum_s \boldsymbol{\mu}_s^g \cdot \boldsymbol{F}_s^o + \frac{1}{2}\sum_s \boldsymbol{\mu}_s^g \cdot \boldsymbol{F}_s^{qm}(\{\Phi_{CI}^g\}) \tag{2.1}$$

where the superscript "g" signifies quantities for the solute in the ground state, Φ_{CI}^g is the ground state CI wave function of the solute, which is the HF result at the CIS level, $\hat{H}_X^o$ is the Hamiltonian of the isolated solute (*X*), $\hat{H}_{Xs}^{stat}(\{q_s\})$ is the electrostatic interaction Hamiltonian between the QM system and the MM permanent charges (q_s), and $\hat{H}_{Xs}^{pol}(\{\boldsymbol{\mu}_s^g\})$ is the interaction between the QM solute and the MM induced dipoles ($\boldsymbol{\mu}_s^g$) in the solute ground state. The remaining terms do not involve the electronic degrees of freedom, except the last term due to the fact that energy terms are not additive in a polarizable force field and a combined QM/MM potential with explicit polarization in the MM model (*2.46*). In eq. (2.1), E_{Xs}^{vdW} is the van der Waals interaction between the solute and solvent atoms, E_{ss}^{pair} is the solvent pair interaction consisting of both Lennard-Jones and Coulomb terms, $\boldsymbol{F}_s^o$ is the static electrostatic field from the MM system, and $\boldsymbol{F}_s^{qm}(\{\Phi_{CI}^g\})$ is the electrostatic field generated by the solute wave function. The induced dipoles of MM atoms *s* ($\boldsymbol{\mu}_s^g$) in eq. (2.1) are determined self-consistently by an iterative procedure using eq. (2.2) (*2.46, 2.48*),

17

$$\boldsymbol{\mu}_s^g = \alpha_s \left[\boldsymbol{F}_s^o + \boldsymbol{F}_s^{qm}(\{\Phi_{CI}^g\}) + \sum_{t \neq s} \nabla_s \nabla_t \left(\frac{1}{r_{st}}\right) \cdot \boldsymbol{\mu}_t^g \right] \qquad (2.2)$$

where the subscripts (s,t) refer to MM atoms, α_s is the atomic polarizability of atom s, r_{st} is the distance between atoms s and t, and $\boldsymbol{\mu}_t^g$ ($t \neq s$) are induced dipoles of all other solvent atoms. The value of $\{\boldsymbol{\mu}_s^g\}$ is a function of the permanent charges of the MM atom s, all the other solvent-induced dipoles ($\boldsymbol{\mu}_t^g$, $t \neq s$) in the MM region, and the instantaneous external field from the QM system, $\boldsymbol{F}_s^{qm}(\{\Phi_{CI}^g\})$, which is derived from the molecular wave function of the solute. Since Φ_{CI}^g and $\{\boldsymbol{\mu}_s^g\}$ are mutually dependent on each other, they must be solved self-consistently. We have employed a triple-iterative self-consistent field (TSCF) procedure to achieve the convergences of both the solute wave function and the solvent induced dipole, and of the overall mutually polarized system (*2.46, 2.48*). The computational details have been described in refs 2.46 and 2.48. In short, we first use a set of induced solvent dipoles, which are kept frozen, along with the solvent permanent point charges to optimize the solute wave function. Then, the electric field of the solute molecule is included in eq. (2.2) to optimize the solute induced dipoles $\{\boldsymbol{\mu}_s^g\}$. The new set of $\{\boldsymbol{\mu}_s^g\}$ is again used to obtain an updated Φ_{CI}^g. This process continues until the total energy of the entire system in eq. (2.1) is fully converged.

For the excited state of the solute, a similar energy expression can be obtained (*2.46, 2.48*):

$$\begin{aligned} E_{tot}^e &= \left\langle \Phi_{CI}^e \middle| \hat{H}_X^o + \hat{H}_{Xs}^{stat}(\{q_s\}) + \hat{H}_{Xs}^{pol}(\{\boldsymbol{\mu}_s^e\}) \middle| \Phi_{CI}^e \right\rangle + E_{Xs}^{vdW} \\ &+ E_{ss}^{pair} - \frac{1}{2} \sum_s \boldsymbol{\mu}_s^e \cdot \boldsymbol{F}_s^o + \frac{1}{2} \sum_s \boldsymbol{\mu}_s^e \cdot \boldsymbol{F}_s^{qm}(\{\Phi_{CI}^e\}) \end{aligned} \qquad (2.3)$$

18

where the superscript "*e*" indicates excited state quantities, Φ_{CI}^e is an excited state CI wave function of the molecules in the QM region (in this study, it is the first singlet excited state), and $\boldsymbol{\mu}_s^e$ is the induced dipole of the solvent atom *s* in the MM region optimized in response to the presence of the QM solute in its electronically excited state. In eq. (2.3), the solvent polarization is assumed to be instantaneous in response to the solute electronic excitation.

In general, a similar triple-iterative SCF procedure as that for the ground state can be used, but this is very time consuming to optimize the excited state wave function. Fortunately, it is typically not necessary. In the work reported here, a simplified procedure is adopted to solve the coupled QM- and MM-SCF calculations in eq. (2.3) (*2.46, 2.48*): we use the excited-state electric field of the solute, determined by the optimized ground-state reference wave function, to determine the $\{\boldsymbol{\mu}_s^e\}$ dipoles. Thus, we do not further optimize Φ_{CI}^e. This is based on the Franck-Condon principle that the solvent and solute nuclei remain fixed in the Franck-Condon transition and the solvent's configuration can be approximately by that in the ground state (*2.28*). The small perturbation of $\{\boldsymbol{\mu}_s^e\}$ by optimized Φ_{CI}^e is ignored because this is of third-order effects (induction of the solute excited state by the induced solvent dipoles in the presence of the excited solute, over the mutual induction effects in the ground state). Consequently, the QM/MM polarization term in eq. (2.3) could be approximately defined as follows,

$$\left\langle \Phi_{CI}^e \right| \hat{H}_{Xs}^{pol}(\{\boldsymbol{\mu}_s^e\}) \left| \Phi_{CI}^e \right\rangle \approx \left\langle \Phi_{CI}^e \right| \hat{H}_{Xs}^{pol}(\{\boldsymbol{\mu}_s^g\}) \left| \Phi_{CI}^e \right\rangle + \left\langle \Phi_{CI}^e \right| \hat{H}_{Xs}^{pol}(\{\Delta\boldsymbol{\mu}_s\}) \left| \Phi_{CI}^e \right\rangle \quad (2.4)$$

where $\Delta\mu_s = \mu_s^e - \mu_s^g$. In eq. (2.4), the last term can be identically expressed classically for the interaction between the solvent-induced dipole with the QM electric field.

$$\left\langle \Phi_{CI}^e \left| \hat{H}_{Xs}^{pol}(\{\Delta\mu_s\}) \right| \Phi_{CI}^e \right\rangle = -\sum_s \Delta\mu_s \cdot F_s^{qm}(\{\Phi_{CI}^e\}) \tag{2.5}$$

We employ eq. (2.4) and (2.5) to rewrite eq. (2.3) as

$$
\begin{aligned}
E_{tot}^e &= \left\langle \Phi_{CI}^e \left| \hat{H}_X^o + \hat{H}_{Xs}^{stat}(\{q_s\}) + \hat{H}_{Xs}^{pol}(\{\mu_s^g\}) \right| \Phi_{CI}^e \right\rangle + E_{Xs}^{vdW} \\
&\quad + E_{ss}^{pair} - \frac{1}{2}\sum_s \mu_s^e \cdot F_s^o + \frac{1}{2}\sum_s \mu_s^e \cdot F_s^{qm}(\{\Phi_{CI}^e\}) - \sum_s \Delta\mu_s \cdot F_s^{qm}(\{\Phi_{CI}^e\})
\end{aligned}
\tag{2.6}
$$

The difference between eqs. (2.3) and (2.6) is that the former involves fully iterative QM-CI and solvent-polarization TSCF calculations, whereas the latter requires just the MM-SCF iteration to obtain μ_s^e. In eq. (2.6), excited-state energies in the CI calculations are determined by using the ground-state, solvent-induced dipoles (*2.46, 2.48*). Therefore, the transition energy of the solute from the ground state to the excited state in solution can be obtained by subtracting eq. (2.1) from eq. (2.6),

$$
\begin{aligned}
\Delta E_{tot}^{g\to e} &= \left\langle \Phi_{CI}^e \left| \hat{H}_X^o + \hat{H}_{Xs}^{stat}(\{q_s\}) + \hat{H}_{Xs}^{pol}(\{\mu_s^g\}) \right| \Phi_{CI}^e \right\rangle \\
&\quad - \left\langle \Phi_{CI}^g \left| \hat{H}_X^o + \hat{H}_{Xs}^{stat}(\{q_s\}) + \hat{H}_{Xs}^{pol}(\{\mu_s^g\}) \right| \Phi_{CI}^g \right\rangle \\
&\quad - \frac{1}{2}\sum_s \Delta\mu_s \cdot F_s^o + \frac{1}{2}\sum_s \left[\mu_s^e \cdot F_s^{qm}(\{\Phi_{CI}^e\}) - \mu_s^g \cdot F_s^{qm}(\{\Phi_{CI}^e\}) \right] \\
&\quad - \sum_s \Delta\mu_s \cdot F_s^{qm}(\{\Phi_{CI}^e\})
\end{aligned}
\tag{2.7}
$$

A further approximation of eq. (2.7) is that we assume that the van der Waals terms for the solute in the ground state and the excited state are the same. Implicitly, this means that we ignore the dispersion effects between solute and solvent in absorption spectral calculations (*2.29, 2.30*).

Explicit Simulation Studies. The excitation energy of a chromophore in solution as defined by eq. (2.7) can be partitioned into two components as follows (*2.46, 2.57*),

$$\Delta E_{tot}^{g \to e} = \Delta E_{stat}^{g \to e} + \Delta E_{pol}^{g \to e} \tag{2.8}$$

where

$$\Delta E_{stat}^{g \to e} = \left\langle \Phi_{CI}^{e} \middle| \hat{H}_X^o + \hat{H}_{Xs}^{stat}(\{q_s\}) + \hat{H}_{Xs}^{pol}(\{\boldsymbol{\mu}_s^g\}) \middle| \Phi_{CI}^{e} \right\rangle$$
$$- \left\langle \Phi_{CI}^{g} \middle| \hat{H}_X^o + \hat{H}_{Xs}^{stat}(\{q_s\}) + \hat{H}_{Xs}^{pol}(\{\boldsymbol{\mu}_s^g\}) \middle| \Phi_{CI}^{g} \right\rangle \tag{2.9}$$

and

$$\Delta E_{pol}^{g \to e} = -\frac{1}{2}\sum_s \Delta \boldsymbol{\mu}_s \cdot \boldsymbol{F}_s^o + \frac{1}{2}\sum_s \left[\boldsymbol{\mu}_s^e \cdot \boldsymbol{F}_s^{qm}(\{\Phi_{CI}^e\}) - \boldsymbol{\mu}_s^g \cdot \boldsymbol{F}_s^{qm}(\{\Phi_{CI}^g\}) \right]$$
$$- \sum_s \Delta \boldsymbol{\mu}_s \cdot \boldsymbol{F}_s^{qm}(\{\Phi_{CI}^e\}) \tag{2.10}$$

In eq. (2.9), $\Delta E_{stat}^{g \to e}$ represents the vertical excitation energy of the solute in the presence of the total electric field of the solvent that is equilibrated to the ground-state charge distribution of the solute. The remaining contributing terms in eq. (2.10), $\Delta E_{pol}^{g \to e}$, indicates the correlation effects (mutual many-body polarization to be distinguished from electron correlation effects in quantum chemistry) resulting from the instantaneous polarization of the solvent molecules by the solute excitation. The solvatochromic shift, Δv, is defined as the difference in excitation energies of the chromophore in solution and in the gas phase:

$$\Delta v = \left\langle \Delta E_{tot}^{g \to e} \right\rangle - \Delta E_{gas}^{g \to e} \tag{2.11}$$

21

where the bracket specifies an ensemble average configurations sampled by carrying out statistical mechanical the Monte Carlo or molecular dynamics simulations. Making use of the energy partition in eq. (2.7), we can formally separate the overall solvatochromic shift into two terms: (1) the spectral shift due to the solvent potential equilibrated to the ground state of the solute, $\langle \Delta\Delta E_{stat}^{g\to e} \rangle$, and (2) the subsequent energy change of the solvent dipole due to the solute electronic excitation. Thus,

$$\Delta\upsilon = \langle \Delta\Delta E_{stat}^{g\to e} \rangle + \Delta E_{pol}^{g\to e} \tag{2.12}$$

where

$$\langle \Delta\Delta E_{stat}^{g\to e} \rangle = \langle \Delta E_{stat}^{g\to e} \rangle - \Delta E_{gas}^{g\to e} \tag{2.13}$$

As described in previous works (*2.23, 2.46, 2.48*), the $\langle \Delta\Delta E_{stat}^{g\to e} \rangle$ term can be further decomposed into two components,

$$\langle \Delta\Delta E_{stat}^{g\to e} \rangle = \langle \Delta E_{Xs}^{g\to e} \rangle + \langle \Delta\Delta E_{X}^{g\to e} \rangle \tag{2.14}$$

where $\Delta E_{Xs}^{g\to e}$ describes the energy change of the solute-solvent interaction due to the solute electronic excitation, and $\Delta\Delta E_{X}^{g\to e}$ depicts the difference between the excitation energy of the solute in the gas phase ($\Delta E_{X,gas}^{g\to e}$) and that in solution ($\Delta E_{X}^{g\to e}$), due to, among other factors, the change in solute geometry at an instantaneous configuration and the polarization of the solute wave function:

$$\langle \Delta\Delta E_{X}^{g\to e} \rangle = \langle \Delta E_{X}^{g\to e} \rangle - \Delta E_{X,gas}^{g\to e} \tag{2.15}$$

The energy decomposition scheme of eqs. (2.7) and (2.12) provides us a convenient, approximate procedure for estimating the instantaneously mutual polarization effects upon solute electronic excitation. First, we carry out Monte Carlo

or molecular dynamics simulations using an effective, pairwise potential for the solvent such as the four-point charge TIP4P models. Since polarization effects for the ground-state configurations have been included in the potential in an average sense, on average, the computed excitation energy using such a non-polarizable model corresponds to the energy difference of eq. (2.7), which is written for a polarizable solvent model, averaged over the Monte Carlo trajectories. Then, we switch the solvent potential to a polarizable model, and use the configurations generated in the first step that employs a non-polarizable, effective potential to determine the ensemble average of the effects (or energy contribution) of instantaneous polarization of the solvent in response to the solute excitation. This average yields the energy terms in eq. (2.10).

2.3   Computational Details

All QM/MM calculations in statistical mechanical Monte Carlo simulations were performed using the MCQUB program (*2.58, 2.59*), in which the quantum mechanical energies were calculated using the MOPAC program (*2.60*). Monte Carlo simulations were carried out for a cubic box containing 396 water molecules and one acetone molecule with periodic boundary conditions. The isothermal isobaric (NPT) ensemble was employed at temperatures of 25, 50, 100, 200, 300, 400, 450, and 500 ºC and pressures in the range of 1 to 2763 atm. These results in bulk conditions of a reduced density (ρ_r) range from 0.05 to 3.10. A total of 29 unique conditions were included with various temperature and pressure conditions. The size has been shown to sufficiently describe thermodynamic and spectroscopic properties of solutes in

SCW and the ambient water, especially at high temperature regions where the fluid density is low (*2.13, 2.15, 2.51–2.55*). The intermolecular interaction among water molecules were spherically truncated at 9 Å. The spherical cutoff distances of the solvent-solute interaction employed in these calculations were about one-half of the edge of each unit box, ranging from 10.07–41.94 Å. This is reasonable since the solute molecule is not charged or having significant charge separations. Nevertheless, it might be advised to include long-range electrostatic effects in these simulations since the ability of solvent dielectric screening effects may be different in such a large density range. In all QM-CI/MM calculations, the acetone structure was held rigid at the AM1 geometry optimized in the gas phase (*2.61*). The electronic excited-state calculations were performed by configuration interaction that includes a total of 100 configurations from an active space of 6 electrons in 5 orbitals and these combinations have been shown to yield excellent results for acetone even though the model was not originally developed for spectroscopy (*2.22, 2.30*). The van der Waals parameters for the QM atoms were determined in a previous study (*2.62*).

Two separate calculations were involved. First, the combined QM-CI/MM potential with the pairwise four-point charge TIP4P (*2.63*) water model was utilized to yield the average values for $\left\langle \Delta E_{stat}^{g \to e} \right\rangle$. The TIP4P model has been verified to adequately describe the properties of SCW for the present purposes (*2.15, 2.54*). In particular, a series of Monte Carlo simulations have been carried out at 400 $^\circ$C and pressures ranging from 350 to 2000 atm. By analysis of reduced parameters, it was suggested that the TIP4P model may slightly underestimate the supercritical temperature by 30 to 50 degrees. In the second set of QM-CI/MM simulations, the

polarizable POL2 (*2.56*) model was adopted for the MM solvent to give the polarization correlation energy, $\Delta E_{pol}^{g \to e}$. In this step, only the single-point energies were evaluated based on the configurations generated in the first set of simulation. Each Monte Carlo simulation in the first computational step involves at least 4×10^6 configurations of equilibration, followed by 4×10^6 configurations for data averaging. The Owicki-Scheraga preferential sampling technique was used to enhance the statistics near the solute, such that solvent moves are made proportional to $1/(R^2 + W)$, where, $W = 350$ Å (*2.64*). The averages for $\Delta E_{pol}^{g \to e}$ in the latter calculations were equilibrated for at least 4×10^6 configurations, followed by single-point energy evaluations with a total of 50 structures to obtain the instantaneous polarization response by the solvent. Note that all spectral shifts correspond to Franck-Condon excitation, in which solute and solvent electronic polarization are assumed to be instantaneous in the excited state at the solvent nucleus positions equilibrated to the ground-state electronic structures.

2.4   Results and Discussion

Solvatochromic Shifts. The total solvatochromic shift ($\Delta \upsilon$) for $n \to \pi^*$ excitation of acetone in water fluid calculated by the QM-CI/MM method is plotted as a function of reduced density (ρ_r) of the fluid in the range from 0.02 to 3.11 (Figure 2.1). The reduced density of 3.11 corresponds to simulations at 25 ˚C and 1 atm. The theoretical results show that the initial increase in the reduced density (ρ_r) from 0.02−0.7 is accompanied by a rapidly rising blue shift in $\Delta \upsilon$. This is followed by

a slowly rising plateau region in the reduced density range of 0.7 to 1.5. In the third stage, the increase of Δv becomes markedly steeper at higher reduced densities from 1.5 to 3.1. The trend of Δv obtained in the calculations is in excellent agreement with the experimental data reported by Bennett and Johnston (*2.1*). The three distinctive regions in Figure 2.1 can be categorized as (1) the *gaseous steam phase* corresponding to a temperature of 500 K and pressures of 49 → 454 atm used in the Monte Carlo simulation, (2) the *supercritical fluid region* (T = 400–500 K, and P = 454–987 atm), and (3) the *dense-liquid phase* (T = 25–400 K, and P = 1–2763 atm). The plateau in the UV-absorption energy in the SCW region has been attributed to the effect of solvent clustering near the solute, which plays an important role in determining the chemical reactivity of organic solute in SCW (*2.65–2.67*). The experimental observation is nicely reproduced here (*2.1*), and we shall present structural analysis in a following section.
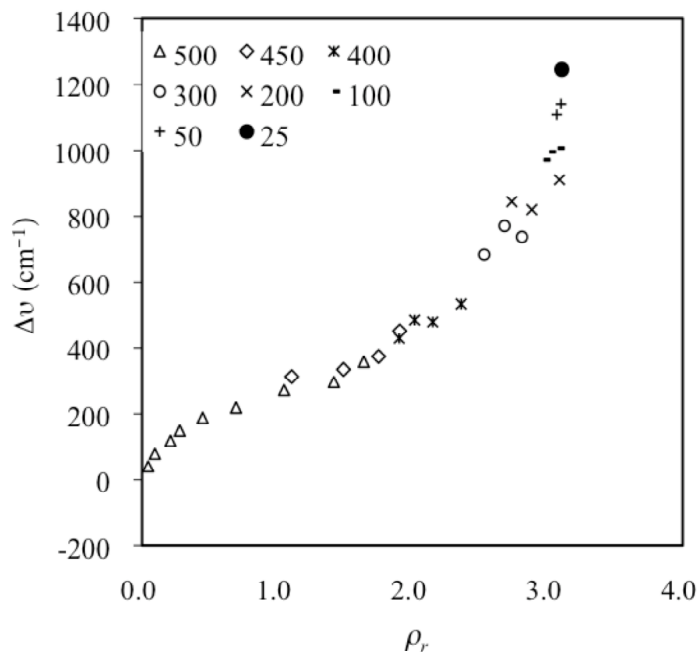
Figure 2.1
Computed solvatochromic shift (Δv) in the $n \rightarrow \pi^*$ excitation of acetone in water fluid as a function of reduced density (ρ_r). Numerical data are listed in Table A1 in Appendix A.

26

The quality of the present study is best illustrated by the computed $n \rightarrow \pi^*$ spectral shift (Δv) for acetone in ambient water, which is 1245 cm^{-1}. For comparison, the experimental value is 1560 cm^{-1} (*2.28, 2.68*), and the difference corresponds to an energy difference of only 0.9 kcal/mol. In an early study, the computed spectral shift is somewhat greater at 1690 cm^{-1} (*2.22*). Solvatochromic shifts of acetone in various solvents have been extensively investigated in ambient conditions (*2.22, 2.30, 2.32, 2.39, 2.48, 2.69–2.72*). It provides a prototypical system for studying solvation effects. The excellent agreement between the theoretical results and the experimental data indicates that the Monte Carlo simulations combined with the AM1 Hamiltonian for the QM atoms employed in the present work are adequate for analyses of the solvation structure and solvation energies of an organic solute in water fluids spanning the entire density ranges from vapor to supercritical fluids to dense liquid.

Energy Decompositions. To gain insight into the origin of the observed absorption spectral shifts and the possibility of solvent clustering near supercritical fluid conditions, we decomposed the total solvatochromic shifts into specific terms. If there is stable solvent cluster formation near the supercritical point, one would expect to find a relatively large and invariant condition from the $\Delta E_{pol}^{g \rightarrow e}$ term because the solvent polarization effects depend on the size of the cluster. To conveniently analyze the solvatochromic shift (Δv) of the acetone excitation in water fluid, the water reduced density (ρ_r) in this work is divided into four regions: the vapor phase ($\rho_r < 0.7$), the supercritical and near supercritical fluid region ($0.7 < \rho_r < 1.9$), the dense-liquid region ($\rho_r > 1.9$), and the ambient water state ($\rho_r = 3.11$) (*2.1*). In these four

regions, the values of Δv obtained from the hybrid QM-CI/MM calculations are in the range of 48–219 cm^{-1}, 273–452 cm^{-1}, 485–1142 cm^{-1}, and 1245 cm^{-1}, respectively (Figure 2.1).

Δv can be decomposed into $\Delta\Delta E_{stat}^{g\to e}$ and $\Delta E_{pol}^{g\to e}$ terms (eq. (2.11)). The first term represents the electrostatic stabilization of the ground state over the excited state due to solvation, of which the excitation energy in solution is obtained using the solvent charge distribution in the ground state of the solute. The second term is the polarization correlation energy due to the instantaneous solvent polarization following the solute electronic excitation. The decomposition results of Δv show that in the vapor, supercritical and near supercritical fluid, dense-liquid, and the ambient water regions, $\Delta\Delta E_{stat}^{g\to e}$ contribute to the blue shifts Δv by 40–226 cm^{-1}, 282–468 cm^{-1}, 500–1169 cm^{-1}, and 1282 cm^{-1}, respectively (Figure 2.2). On the other hand, $\Delta E_{pol}^{g\to e}$ contributes a small red shift to Δv in ranges of –0.02 to –7 cm^{-1}, –8 to –16 cm^{-1}, –15 to –35 cm^{-1}, and –37 cm^{-1}, respectively (Figure 2.3). Clearly, inclusion of the solvent instantaneous polarization effects leads to stabilization of the electronic excited state, giving rise to a red shift in the absorption energy, and the effect increases as the solvent density increase. However, it only makes a small correction to the total solvatochromic shift, suggesting that that the energy input required to reorient solvent dipoles following the solute excitation is small.
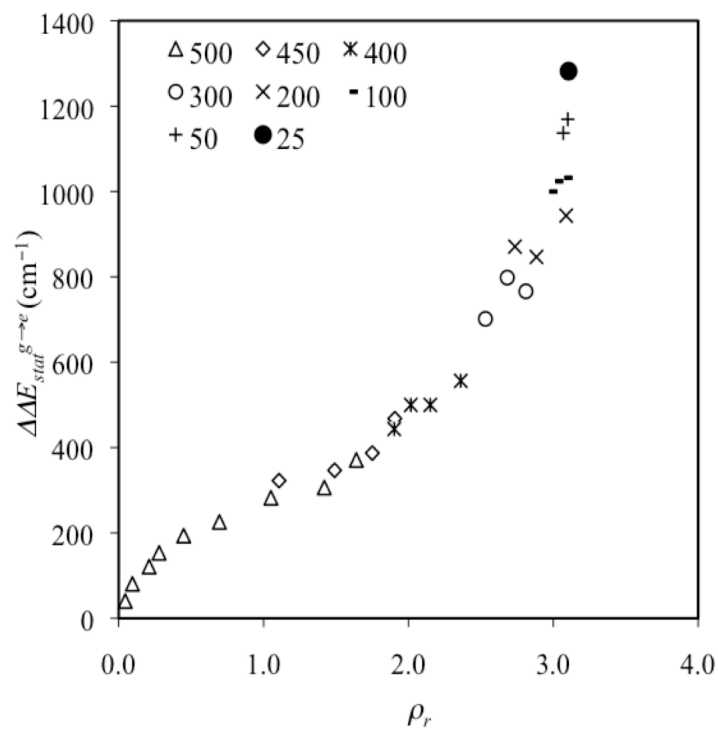
Figure 2.2   Electrostatic stabilization energy ($\Delta\Delta E_{stat}^{g\rightarrow e}$) due to solvent permanent

charge distribution for the $n \rightarrow \pi^*$ excitation of acetone in water fluid as a function

of fluid reduced density (ρ_r)
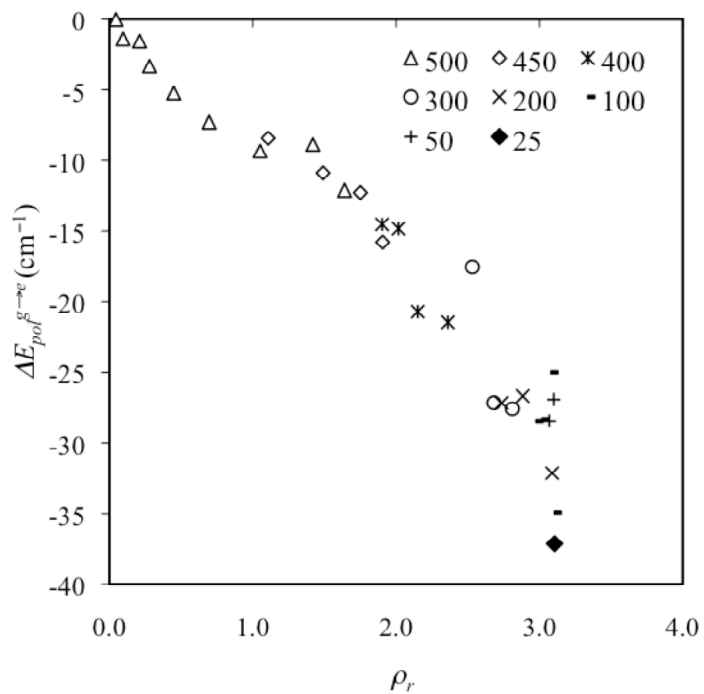
Figure 2.3 Computed solvent polarization contributions ($\Delta E_{pol}^{g \to e}$) to the overall spectral shifts for the $n \to \pi^*$ excitation of acetone in water fluid as a function of fluid reduced density (ρ_r)

To further understand the solute-solvent interactions and the solute intrinsic energy contributing to the electrostatic stabilization over the excitation due to solvation ($\Delta\Delta E_{stat}^{g\rightarrow e}$), the $\Delta\Delta E_{stat}^{g\rightarrow e}$ term is further separated into $\Delta E_{Xs}^{g\rightarrow e}$ and $\Delta\Delta E_{X}^{g\rightarrow e}$ two terms using eq. (2.14). The $\Delta E_{Xs}^{g\rightarrow e}$ represents the energy change of the solute-solvent interaction due to different solute charge distributions in the ground state and excited state, and $\Delta\Delta E_{X}^{g\rightarrow e}$ is the change of the intrinsic excitation energy of the solute in solution. In the vapor, supercritical and near supercritical, dense-liquid and ambient states, the energy component of $\Delta E_{Xs}^{g\rightarrow e}$ is the dominant component of the total solvatochromic shift, and the computed values are 23–167 cm^{-1}, 256–394 cm^{-1}, 484–1110 cm^{-1}, and 1253 cm^{-1}, respectively (Figure 2.4). Together with the solvent polarization correction, $\Delta E_{pol}^{g\rightarrow e}$, we find that the observed spectral shifts nearly come entirely from the difference in solute-solvent interaction energy between the excited and the ground states, comprising 95% of the total Δv. Surprisingly, the intrinsic excitation energy of the solute does not change significantly relatively to the gas phase value, with the computed $\Delta\Delta E_{X}^{g\rightarrow e}$ less than 50 cm^{-1} in all density ranges (Figure 2.4). Evidently, the polarization of the solute wave function does not affect the energy gap between the ground state and the $n \rightarrow \pi^{*}$ excited state. The results of the energy decompositions for the solvatochromic shift of the acetone $n \rightarrow \pi^{*}$ excitation reveal that the electrostatic stabilization from the solute-solvent interaction in the ground state over that in the excited state ($\Delta E_{Xs}^{g\rightarrow e}$) primarily dominates the spectra blue shift. Furthermore, $\Delta E_{Xs}^{g\rightarrow e}$ increases continuously without a plateau

behavior in SCW region, although it shows a clear transition in that region, leading to

a rapid increase as the solvent density further increases (Figure 2.4).
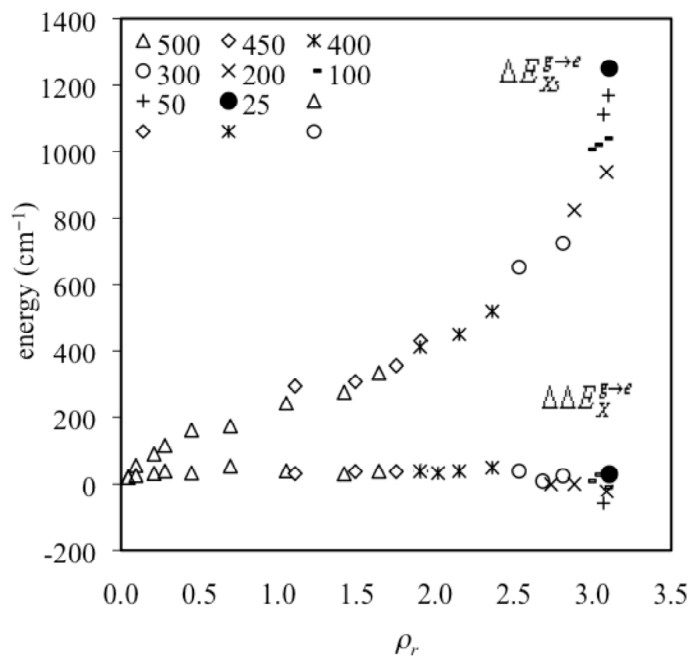
Figure 2.4   Decomposition of the permanent electrostatic energy term in Figure

2.2 into the change in net solute-solvent interaction energy ($\Delta E_{Xs}^{g\rightarrow e}$) and the intrinsic

excitation energy of the solute ($\Delta\Delta E_{X}^{g\rightarrow e}$) for the $n \rightarrow \pi^*$ transition of acetone in

water fluid.

Dipole Moment vs. Spectra Shift. A further measure of the molecular polarization is provided by calculating the ground-state and excited-state induced dipole moments of acetone due to solvation (Table 2.1). The calculated average and induced dipole moments continuously increase as the fluid density increases, whereas the values of $\langle \mu^g \rangle - \langle \mu^e \rangle$ are relatively small. Thus, although the ground state is more strongly solvated than the excited state, the computed dipoles in Table 2.1 show that specific solute-solvent interactions are critically important in molecular solvation and the overall dipole moment of a molecule is not a direct indication of its strength of solvation.

Table 2.1   Theoretical Results of Ground-state Dipole Moment in Water Fluid ($(\langle \mu^g \rangle)^{a,d}$, Ground-state Induced Dipole Moment ($\Delta \mu_{ind}^g)^{b,d}$, Excited-state Dipole Moment in Water Fluid ($(\langle \mu^e \rangle)^d$, and Excited-state Induced Dipole Moment ($\Delta \mu_{ind}^e)^{c,d}$ for Acetone in the Supercritical ($\rho_r < 0.7$ g·cm^{-3}), Near-critical ($0.7 < \rho_r < 1.9$ g·cm^{-3}), Dense-liquid Regions ($\rho_r > 1.9$ g·cm^{-3}), and Ambient Water ($\rho_r = 3.1$ g·cm^{-3})

region	$\langle \mu^g \rangle$	$\Delta \mu_{ind}^g$	$\langle \mu^e \rangle$	$\Delta \mu_{ind}^e$
supercritical	2.94–3.17	0.02–0.25	2.89–3.06	0.02–0.19
near-critical	3.21–3.41	0.29–0.49	3.15–3.36	0.28–0.49
dense-liquid	3.42–3.98	0.51–1.06	3.34–3.95	0.47–1.08
ambient water	4.07	1.15	3.87	1.00

[a] Ensemble average of AM1 ground-state dipole moment in water fluid. [b] $\Delta \mu_{ind}^g = \langle \mu^g \rangle - \mu_{gas}^g$, where μ_{gas}^g, a value of 2.91 D, is the ground-state dipole moment of acetone using the optimized AM1 geometry in the gas phase. [c] $\Delta \mu_{ind}^e = \langle \mu^e \rangle - \mu_{gas}^e$, where μ_{gas}^e, a value of 2.87 D, is the excited-state dipole moment of acetone in the gas phase. [d] Unit in Debye.
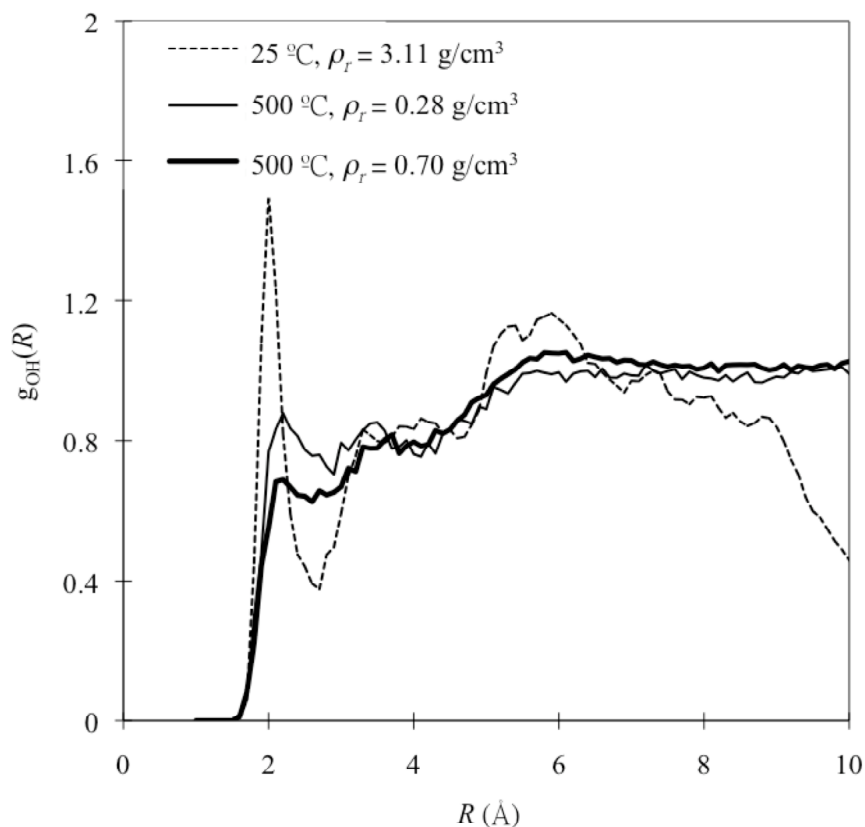
Figure 2.5 Computed radial-distribution functions for the acetone oxygen and water hydrogen ($g_{OH}(R)$) in ambient water (dashed line) and in the supercritical water states of 500 °C (solid lines).

Solvent Structure vs. Spectral Shift. The structural interpretation of energy component analyses is confirmed by examining the radial distribution functions (rdfs) between the solute and solvent. In particular, we focus on the acetone oxygen (O) and the water hydrogen (Hw) rdf, $g_{OH}(R)$, which gives the probability of finding a water hydrogen atom (Hw) at a distance R from the acetone oxygen (O). Figure 2.5 shows

34

the rdfs obtained in the ambient water ($T = 25$ °C, $\rho_r = 3.11$) and a state just above the supercritical conditions of 500 °C at a reduced density of 0.70. Table 2.2 presents the positions of the first peaks of the rdfs and the coordination number of water molecules in the first solvation layer for acetone in these four water conditions. In the ambient water, the position of the first solvation peak is well-defined appearing at 2.0 Å, but the second peak is less structured. The calculated solvent structure in the ambient water is similar to that observed by Thompson (*2.48*), by Takebayashi *et al.* (*2.51, 2.52*), and by Martin *et al* (*2.40*). In contrast, there is no well-defined first peak of $g_{OH}(R)$ in the supercritical conditions (Figure 2.5), which is also in accord with the results obtained by Takebayashi et al (*2.51, 2.52*).

Table 2.2   Computed Positions of the First Peaks (r_1) in Radial Distribution Functions (rdfs) and Coordination Numbers of Water Molecules in the First Solvation Shell of Acetone (N_{H_2O}) in the Supercritical, Near-critical, Dense-liquid, and Ambient Water fluids

	r_1 (Å)	N_{H_2O}
supercritical region	2.1–2.3	0.08–0.72
near-critical region	2.1–2.3	0.61–1.44
dense-liquid region	2.0–2.2	1.22–2.51
ambient water	2.0	2.90

The positions of the first peak of $g_{OH}(R)$ in the vapor, supercritical and near supercritical, and dense-liquid water fluids were shifted to longer distances in comparison with that in ambient water by 0.1–0.3 Å, 0.1–0.3 Å, and 0.0–0.2 Å, respectively (Table 2.2). Furthermore, the coordination numbers in the first solvation layer about the oxygen of acetone (N_{H_2O}) are 0.08–0.72, 0.61–1.44, and 1.22–2.51 in the corresponding conditions (Table 2.2 and Figure 2.6d). These results show that the average number of N_{H_2O} is an increasing function of water density, which is consistent with the finding by Takebayashi et al (*2.51, 2.52*). It is worthy to note that the trend of the coordination numbers of water (N_{H_2O}) in the first layer around acetone reflects the total solvatochromic shifts (Δv) in the $n \rightarrow \pi^*$ excitation of acetone with a linear correlation of $r^2 = 0.91$ (Figure 2.6a). The solvent polarization correction ($\Delta E_{pol}^{g \rightarrow e}$) and the ground-state induced dipole ($\Delta \mu_{ind}^g$) also show linear correlations with N_{H_2O} (Figure 2.6b and 2.6c). Overall, it implies that the density-dependent N_{H_2O} of specific hydrogen-bond interactions between acetone and water molecules directly influences the magnitude of the solvatochromic shift, solvent polarization correction, and the induced dipole of acetone.

(a)
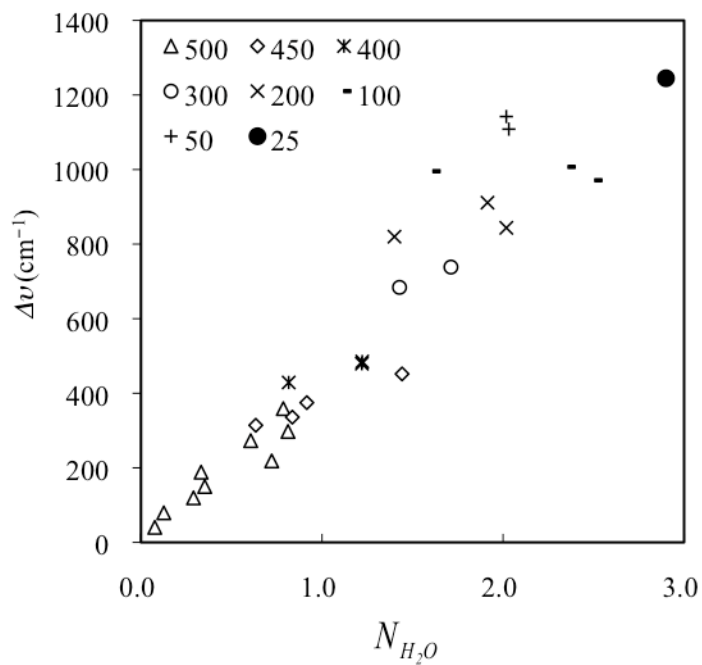
(b)
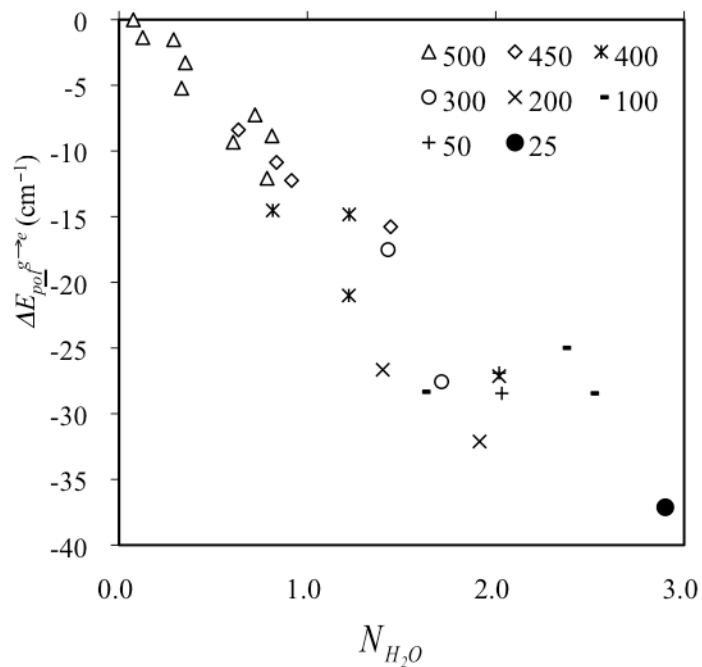
Figure 2.6 Correlations with the computed coordination number of water molecules in the first-solvation layer of acetone (N_{H_2O}) for (a) solvatochromic shift (Δv), (b) solvent polarization contribution ($\Delta E_{pol}^{g \to e}$)

37
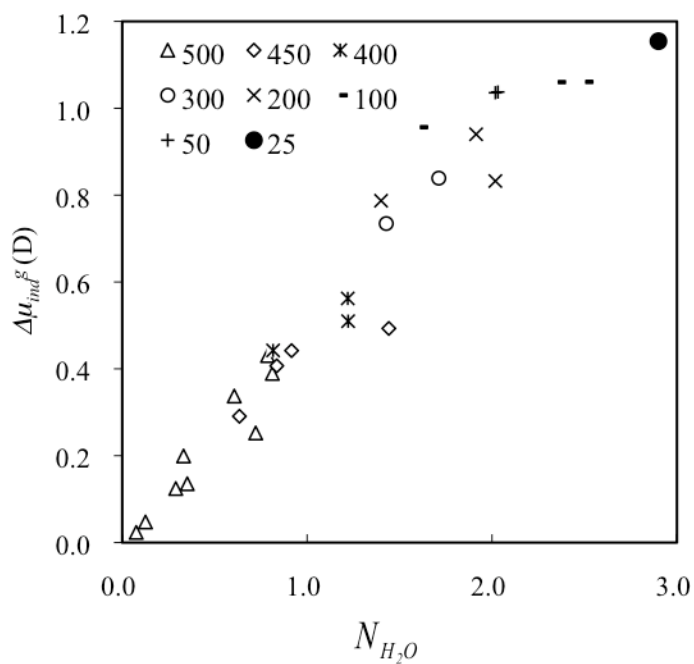
(d)
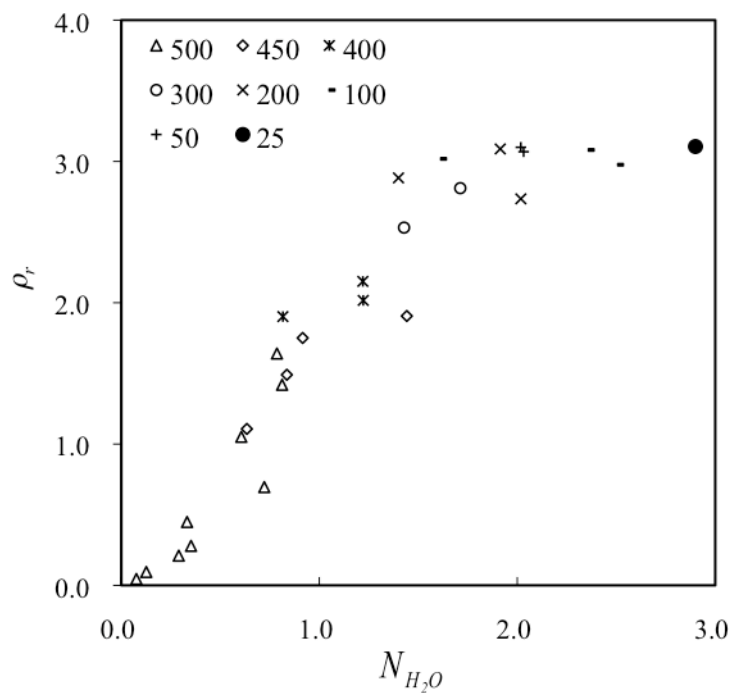
Figure 2.6 (cont.) Correlations with the computed coordination number of water molecules in the first-solvation layer of acetone (N_{H_2O}) for (c) ground-state induced dipole moment of acetone ($\Delta\mu_{ind}^g$), and (d) reduced density (ρ_r) of fluid.

2.5   Conclusions

Hybrid QM-CI/MM Monte Carlo simulations have been carried out to investigate the solvatochromic shifts for the acetone $n \rightarrow \pi^*$ excitation in the supercritical ($\rho_r < 0.7$), near-critical ($0.7 < \rho_r < 1.9$), dense-liquid ($\rho_r > 1.9$), and ambient water conditions. In the present work, the solvent polarization correlation following the solute electronic excitation was included. The computed $n \rightarrow \pi^*$ blue shift in ambient water (1245 cm^{-1}) is in reasonable agreement with the experimental value ($\Delta\upsilon_{exp} = 1560$ cm^{-1}) (*2.28, 2.68*). The trend of the solvatochromic shift as a function of water reduced density with the range from 0.05 to 3.11 was in accord with the experiment probed by the UV-visible absorption spectroscopy (*2.1*). The results of the energy decompositions show that the solvatochromic shifts in the supercritical, near-critical, dense-liquid, and ambient water fluids are mainly determined by the electrostatic interactions between acetone and water molecules during the solute excitation. Furthermore, the energy required to orient solvent molecules following the acetone excitation is quite small and decreases linearly with water density. The solvent-density dependent blue shift and the solvent polarization correction for the acetone $n \rightarrow \pi^*$ excitation in water fluid are governed by the induced dipole of acetone in the ground and excited states and the specific hydrogen-bond interactions between the oxygen of acetone and the hydrogen of water. In addition, both energy terms are more obvious in the ambient water than in the supercritical water because the solvent stabilization of the ground state over the excited state is more significant under the former conditions.

Chapter 3

Internal Proton Transfer in the External Pyridoxal 5′-Phosphate Schiff Base in Dopa Decarboxylase

3.1   Introduction

Pyridoxal 5′-phosphate (PLP), derived from vitamin B_6, is a versatile enzyme cofactor to facilitate many chemical transformations, including racemization, decarboxylation, and transamination reactions (*3.1*). One important yet still not fully resolved question is the tautomeric equilibrium of the Schiff base of PLP, which involves an intramolecular proton transfer between the covalent hydroxyimine and zwitterionic oxoenamine configurations (Scheme 3.1). Here, we use the term "covalent" and "zwitterionic" to emphasize the difference in electronic structure between the tautomers. This equilibrium is a major factor affecting the reactivity of the PLP Schiff base in the active site (*3.2*). To understand the role of PLP cofactors in enzyme catalysis, it is critical to elucidate the position of the bridging proton in PLP-

dependent enzymes (*3.3*). In this Chapter, I present computational results from combined quantum mechanical and molecular mechanical (QM/MM) simulations to elucidate the factors that influence the tautomeric equilibrium of the external aldimine Schiff base, both in water and in the active site of dopa decarboxylase.

keto-enol
tautomerism

oxoenamino (zwitterionic) tautomer   hydroxyimino (covalent) tautomer
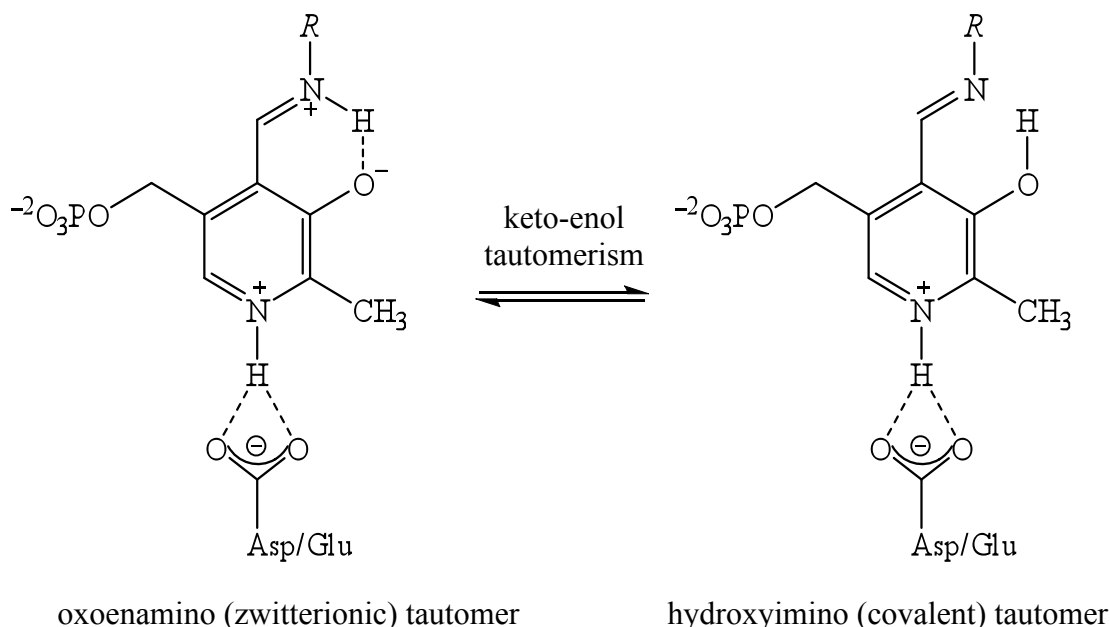
Scheme 3.1 Schematic view of the tautomeric equilibrium of an external PLP aldimine in PLP-dependent enzyme

Dopa decarboxylase (DDC, EC 4.1.1.28) is a PLP-dependent enzyme, which catalyzes the irreversible decarboxylation reaction of aromatic L-amino acid substrates, such as dopa, phenylalanine, and tryptophan. DDC plays an important role in the conversion of the anti-Parkinson drug L-dopa into dopamine. The X-ray crystal structure (*3.4*) shows that the PLP cofactor forms an *internal* Schiff base with Lys303 in the absence of the substrate. The internal Schiff base is converted into the *external*

PLP−L-dopa Schiff base, displacing Lys303 by the substrate L-dopa via a transaldimination process (*3.5, 3.6*). The resulting PLP−L-dopa aldimine is embedded in an extensive hydrogen-bond network in the enzyme (Figure 3.1) (*3.4*), in which the carboxylate side chain of Asp271 forms a salt bridge with the pyridine nitrogen of the PLP (*3.7*). The active-site residues, including Thr82, Ser149, Asn300, and His302, participate in hydrogen-bonding interactions with the phosphate group of the cofactor. Thr246 donates a hydrogen bond to the phenolic oxygen of the PLP, which plays a critical role in the hydroxyimine and oxoenamine tautomerization (*3.4, 3.7, 3.8*).
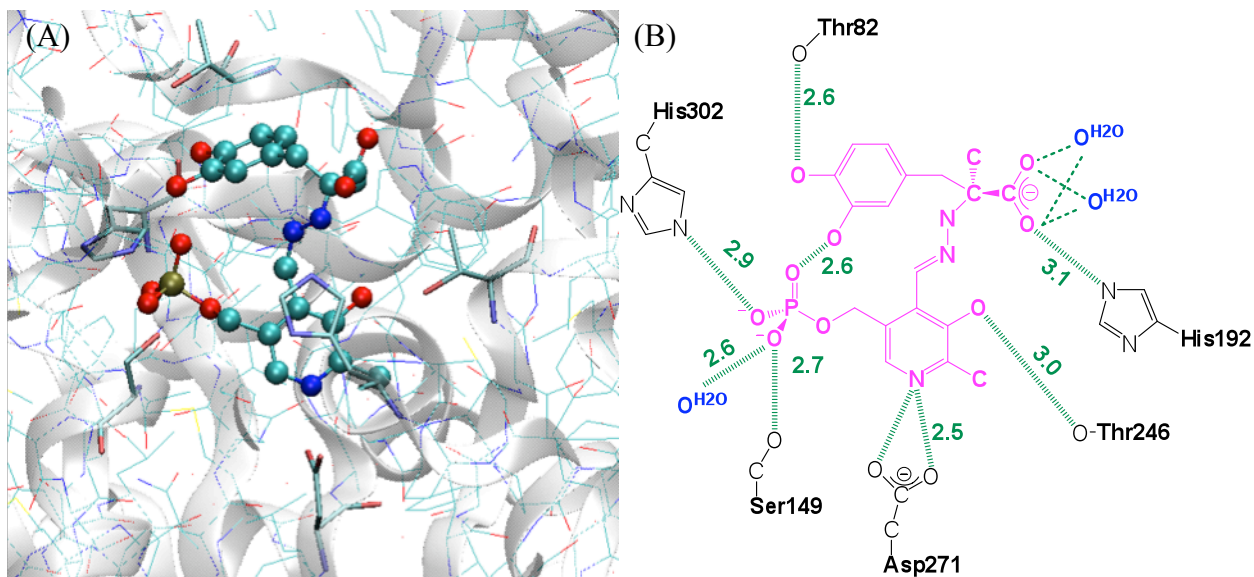
Figure 3.1   Partial view of the active center of hog kidney dopa decarboxylase in complex with external PLP−carbiDopa Schiff base (PDB entry: 1JS3) (*3.2*) (A) PLP−carbiDopa Schiff base is shown in ball and stick. (B) Detailed view of the hydrogen bond interactions (green dash lines), including structural water molecules (shown in blue) in the active site. The numbers indicate the hydrogen bonding distances found in the X-ray structure. PLP−carbiDopa Schiff base is colored in magenta.

NMR, UV/vis, absorption, and fluorescence spectroscopic techniques showed that there is a keto-enol equilibrium, corresponding to an intramolecular proton transfer in model compounds for the internal and external aldimines (*3.3, 3.9–3.15*). Kinetic and spectroscopic studies of aromatic amino acid decarboxylases with and without the substrate or a substrate analog have been used to elucidate the physicochemical properties as well as the reaction mechanisms of the enzymatic processes (*3.16–3.18*). In the absence of substrate, PLP-dependent enzymes typically exhibit an absorption band in the range of 400–440 nm, corresponding to the oxoenamine configuration of the internal PLP–lysine aldimine (*3.16*). However, the absorption spectra of the internal PLP Schiff base of both rat liver and pig kidney DDC show a prominent absorption maximum at 335 nm, and a smaller absorption at 425 nm (*3.16, 3.19–3.22*). The former absorption has been attributed to the hydroxyimine aldimine tautomer (*3.16*). In addition, upon addition of the substrate, there is an increase in absorption at 425 nm, accompanied by a decrease at 335 nm. This spectrum is consistent with the L-dopa and enzyme complex companied by a shift of the tautomeric equilibrium towards the oxoenamino internal aldimine (*3.16*). Subsequently, a new absorbance at 380–390 nm (*3.16, 3.17*) occurs, which has been assigned to the external aldimine in oxoenamino conformation. It was proposed that the blue-shifted wavelength is due to a nonplanar structure of the oxoenamino PLP-aldimine cofactor. Alternatively, this new absorption band can be explained as the formation of an unprotonated-form of the external PLP Schiff base (*3.23*), although it is catalytically inert in aspartate aminotransferase (*3.24*).

Several experimental studies have been reported to rationalize the tautomeric equilibrium and the intramolecular O···H···N hydrogen bond of small Schiff bases. Examination of X-ray structures of model compounds and analysis of Schiff bases reveal that the enol form is the dominant configuration (*3.3, 3.4, 3.13, 3.25, 3.26*), whereas there is a small number of examples with the bridging hydrogen on the imino nitrogen (*3.25, 3.27*). Furthermore, UV−vis and solid-state solution NMR show that the tautomeric equilibrium of Schiff bases is influenced by many factors, such as the substituent on the imino group (*3.28−3.30*), local polarity around the intramolecular hydrogen bond (*3.28, 3.31*), solvent polarity (*3.3, 3.11, 3.31*), temperature (*3.11, 3.29*), and the protonation state of the pyridine nitrogen (*3.28, 3.30*). Finally, hydrogen-bonding interactions to the phenolic oxygen of Schiff bases can stabilize of the keto tautomer (*3.11, 3.31*).

Previous theoretical investigations (*3.15, 3.28, 3.31, 3.32−3.35*) focused on ab initio, DFT, and semiempirical calculations of Schiff bases in the gas phase with emphasis on substituent effects, intramolecular hydrogen bonds, and the photochromic properties (*3.15, 3.32−3.38*). Kiruba et al. investigated solvent effects on PLP derivatives (*3.39*). Bach et al. (*3.37*) modeled the gas-phase decarboxylation reaction of aminoformylacetic acid and the corresponding reaction in a cluster of six water molecules. Furthermore, the pyruvoyl-dependent enzymatic decarboxylation in histidine decarboxylase was investigated using ab initio and DFT methods (*3.38*). These calculations provided valuable insights into the electronic properties to the equilibrium and the effects of PLP cofactor on the decarboxylation.

Although a wealth of information has been accumulated on the hydroxyimine and oxoenamine tautomerism of PLP or PLP-analog Schiff bases, the location of the bridging proton of the intramolecular hydrogen bond in the enzyme active site remains elusive. Our goal here is to understand the tautomeric equilibrium of the external PLP−L-dopa aldimine in dopa decarboxylase. Molecular dynamics (MD) simulation with a combined QM/MM potential is used to model the enzyme complexed with the external PLP Schiff base. In the following, we first describe the computational details, followed by results and discussion.

3.2   Computational Details

Model for the External PLP Schiff base and Dopa Decarboxylase. The X-ray crystal structure of pig kidney DDC in a ternary complex with the PLP cofactor and carbiDopa inhibitor (PDB entry: 1JS3) (*3.4*) was used to generate the initial Michaelis complex structure containing the PLP−L-dopa external Schiff base. DDC is a homodimer, which consists of 486 amino acid residues in each subunit. There is only one active site and one PLP cofactor per dimer structure, located at the interface of the two monomers (*3.16, 3.20, 3.41*). The substrate, L-dopa, was constructed by modifying the structure of the carbiDopa inhibitor to yield the external Schiff base. The pK_a for the pyridine nitrogen of the external aldimine was recently established to be 5.8 in water due to the presence of the strongly electron withdrawing imino group in the protonated Schiff base (*3.3*). This is significantly more acidic than the aldehyde form of the PLP cofactor itself, which has a pK_a of 8.5 (*3.42, 3.43*). In the active site, the crystal structure reveals that there is strong interaction between the pyridine

nitrogen and the side chain of Asp271, a typical structural feature in PLP-dependent enzymes. However, the exact location of the proton, either on Asp271 or on pyridine is unknown. Given that the pK_a of the pyridine of the external aldimine is somewhat more basic than a carboxylate group (*3.3*), we have kept the dopa-external aldimine protonated, (PLP(H+)−L-dopa), throughout the simulation.

To study the intramolecular proton transfer reaction in the PLP(H+)−L-dopa Schiff base, we first built the structure of the oxoenamino tautomer. The residues in the missing loop (residue 328′−339′, where the prime indicates a residue from the second subunit) were modeled as follows. First, we reconstituted the missing loop in each monomer using the BUILD module in InsightII (Accelrys) (*3.44*). The phenol side chain of Tyr332′ forms a hydrogen bond with the carboxylate group of the substrate at a distance of 2.0 Å. The extended structure was subjected to 800 steps of the energy minimization using the adopted-basis Newton-Raphson method (*3.45*) (ABNR) with the side chain of Tyr332′ harmonically restrained, and all the other residues held fixed using the CHARMM program (*3.46*). The active site residue Lys303, which is released from the internal aldimine Schiff base, was set neutral, ready for the next reaction cycle. His192, His262, His348, His386, and His434 that form ion-pairs or are exposed to the solvent were protonated, and we treated the remaining titratable residues corresponding to ionization states at pH 7. Although recent progresses have been reported on loop optimization (*3.47*−*3.52*), the computational results indicate that the procedure described above is adequate for the present study, which in turn validates the selection of loop configurations in activity

studies. This reconstructed enzyme system complex with PLP(H+)–L-dopa Schiff base in the oxoenamino form was chosen as the initial structure in our simulations.

Potential Energy Function. We used a dual-level (*3.53–3.57*) combined QM/MM (*3.58–3.61*) potential to describe the intramolecular proton transfer reaction of the PLP(H+)–L-dopa Schiff base in dopa decarboxylase. There are 39 atoms treated quantum mechanically by the Austin Model 1 (AM1) (*3.62*) method (Figure 3.2). Previous calculations showed that the AM1 model can yield good geometrical results for the Schiff bases both in the ground and in the excited states in comparison with experiments and ab initio results (*3.36, 3.63*). To obtain accurate energetic results, we applied a dual-level approach (*3.53–3.57*), in which Density Functional Theory (DFT) calculations are used as the high-level results to replace the semiempirical energies for the isolated "QM" species, while QM-MM interactions are determined at the lower-level using the AM1/CHARMM potential. In hybrid QM/MM simulations, we placed the QM/MM boundary at the C5A position of the PLP(H+)–L-dopa Schiff base, which was treated with the generalized hybrid orbital (GHO) (*3.64–3.66*) method.
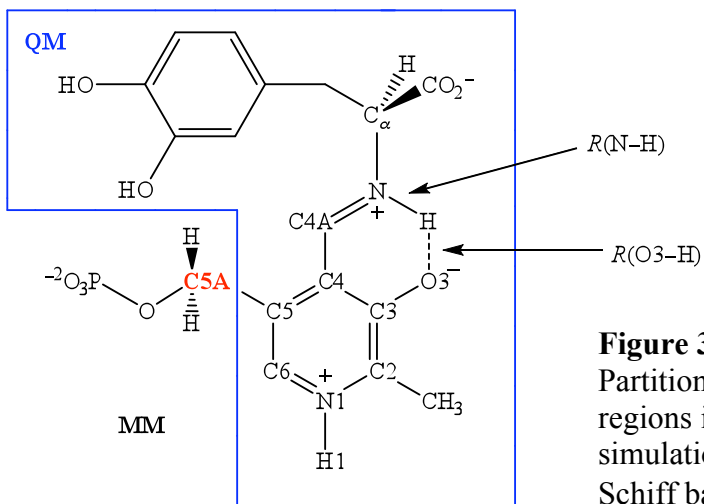
Figure 3.2
Partition of quantum and classical regions in combined QM/MM MD simulations for the PLP(H+)–L-Dopa Schiff base

47

(a1)

(a2)

(b1)

(b2)
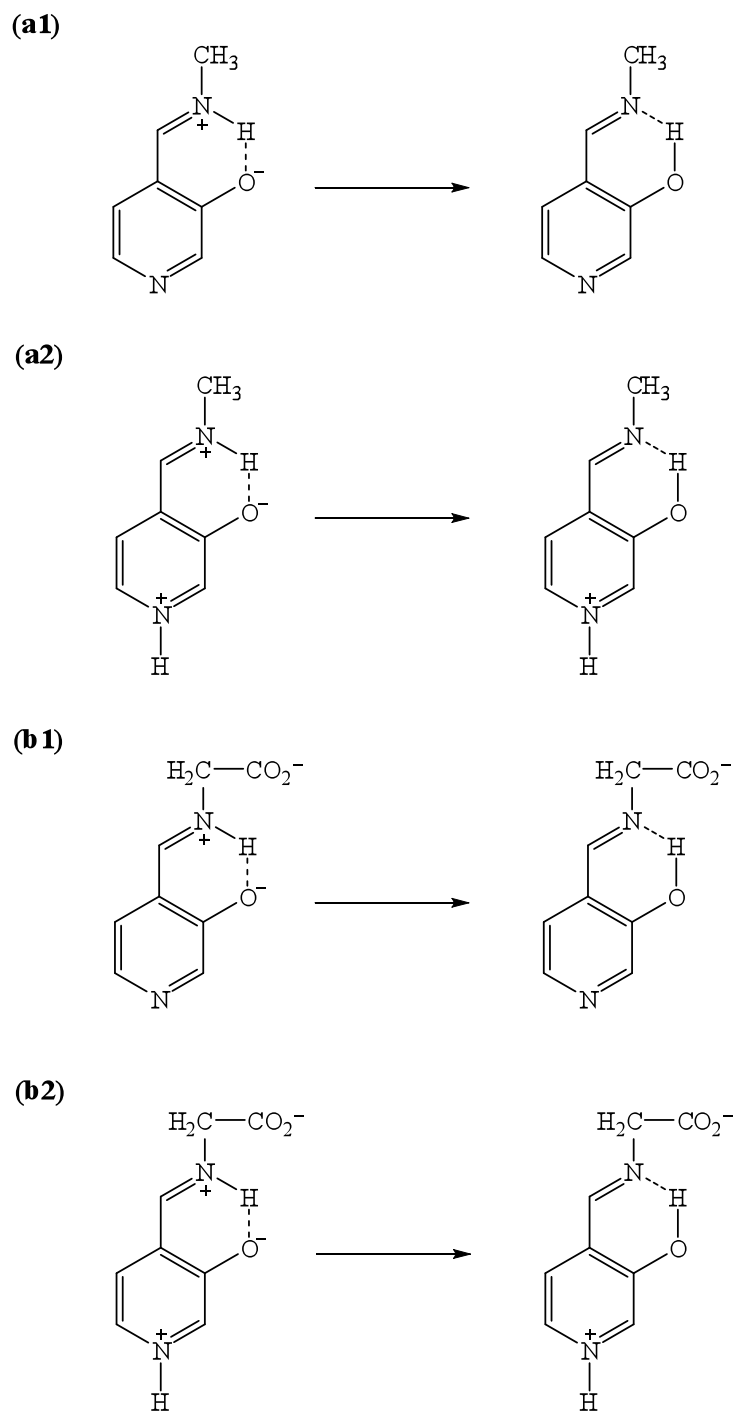
Scheme 3.2   Model Reactions for Intramolecular Proton Transfer Reaction in Aqueous Solution

(c1)
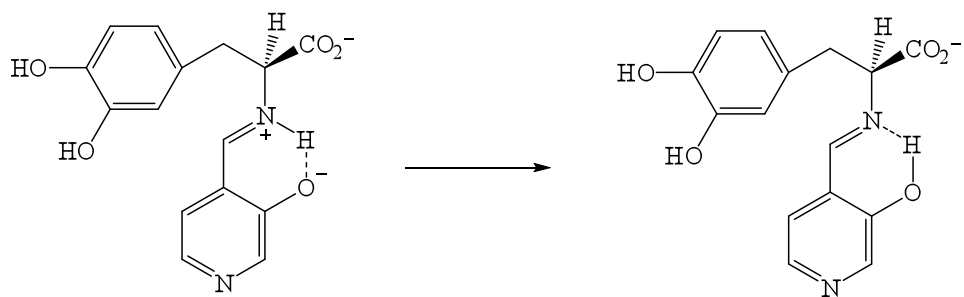
(c2)
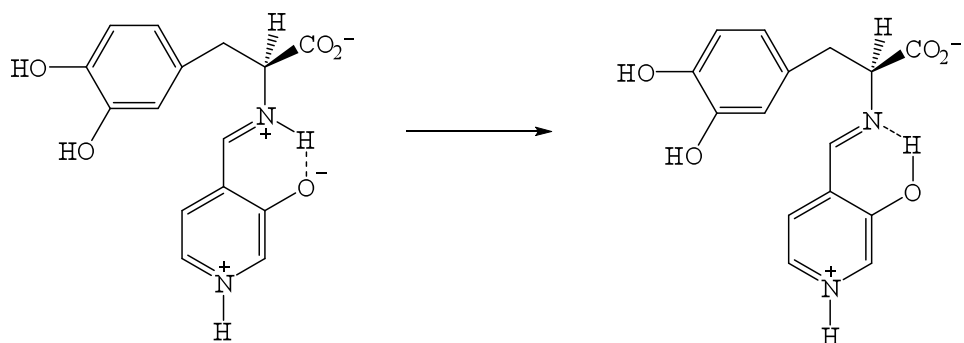
(d1)
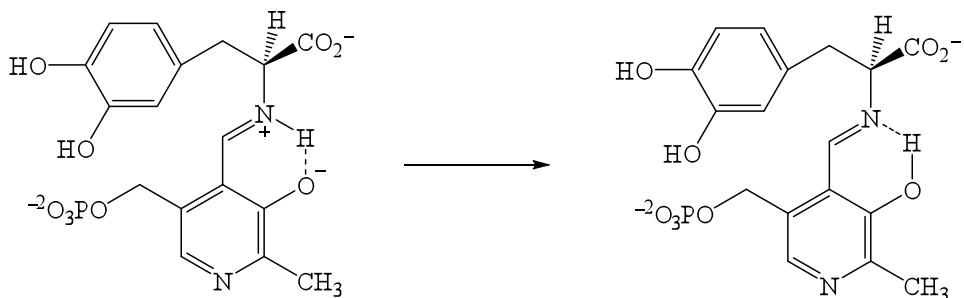
(d2)
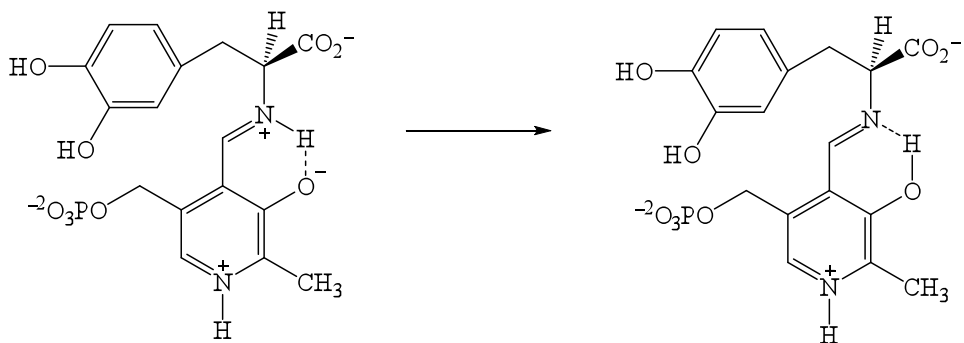
Scheme 3.2 (cont.) Model Reactions for Intramolecular Proton Transfer Reaction

in Aqueous Solution

49

All protein atoms and the phosphate group were modeled by the CHARMM22 all-atom force field (*3.67*), and water was represented by the three-point-charge TIP3P (*3.68*) model. To provide insight into solvation effects on the tautomeric equilibrium of the PLP Schiff base, we designed a total of eight model reactions, **a1/a2**, **b1/b2**, **c1/c2**, and **d1/d2** (Scheme 3.2), for free energy simulations in aqueous solution. For reactions **d1** and **d2**, the phosphate group was represented by the CHARMM22 force field (*3.67*) and the C5A atom was described as the GHO atom (*3.65, 3.66*).

Molecular Dynamics Simulations. Periodic boundary conditions along with the isothermal-isobaric (NPT) ensemble at 298.15 K and 1 atm were used, and long-range electrostatic effects are modeled using the particle-mesh Ewald method (PME−QM/MM) (*3.69, 3.70*). The leapfrog-Verlet integration scheme (*3.71*) was used in all simulations with a time step of 1 fs, and the non-bonded interaction list was updated on every 25 integration steps using a cutoff of 14 Å. The van der Waals energies were feathered to zero between 12 Å and 13 Å with a shift function. All bonds involving hydrogen atoms, except those in the QM region, were constrained to their equilibrium distances using the SHAKE algorithm (*3.72*) during all dynamics simulations.

The initial enzyme system described above was embedded in a cubic box of water molecules about $89 \times 89 \times 89$ Å^3. Water molecules at distances less than 2.8 Å from any non-hydrogen protein atoms, the PLP(H+)−L-dopa Schiff base, or crystallographic waters were removed. The resulting system has a net charge of zero

and no additional counterions were added. The final model of the enzyme system consists of 67518 atoms.

All water molecules were first minimized for 500 steps using the ABNR algorithm (*3.45*) to reduce initial close contacts, while the rest of the system was held constrained. This was followed by short simulations (5 ps) at 298.15 K to relax the water positions. Subsequently, we fixed the coordinates of all water molecules, the substrate PLP(H+)−L-dopa, and residues in the active site (Thr82, Ser149, His192, Thr246, Asp271, Asn300, His302, Lys303, Tyr332′, and Lys334′), and we optimized the rest of the system first by 500 steps of minimization to remove close crystallographic contacts. Then, the system was heated from 0 K to 298.15 K in 30 ps of molecular dynamics simulations. After the initial setup, the entire system was relaxed and equilibrated at 298.15 K for 250 ps under harmonic restraints on non-hydrogen atoms of the PLP(H+)−L-dopa Schiff base and the active-site residues listed above, and the force constants of these restraints were gradually changed from 50 to 5 kcal/mol/Å^2. This ensures the relative positions of the external PLP aldimine and key residues in the active site to be maintained close to the X-ray structure. The structure at this stage was further equilibrated using the combined QM/MM potential for 100 ps without restraints, and the resulting structure was adopted as the starting configuration to perform the potential of mean force (PMF) calculations for the intramolecular proton transfer reaction in the enzyme.

For the eight model reactions in water, the center-of-masses of each reactant (*i.e.*, oxoenamino tautomer) was placed at the center of a cubic water box. The lengths of edges for the cubic box for the smaller systems, **a** and **b**, are 30 Å and 45 Å

for reactions **c** and **d**. In all QM/MM free energy simulations, each system was initially heated from 0 K to 298.15 K within 30 ps, followed by an equilibration of least 60 ps with the solute restrained. Then, each system was equilibrated for an additional 250 ps in the absence of any restraint. The equilibrated structure of each model reaction was used for the following PMF calculations.

Free Energy Simulations. We carried out a series of umbrella sampling simulations to obtain the PMFs for the intramolecular proton transfer reaction of the PLP(H+)–L-dopa Schiff base in dopa decarboxylase enzyme, and for each of the model reactions in aqueous solution. For each reaction, the bridging hydrogen was transferred from the imino nitrogen to the phenolic oxygen, corresponding to the tautomerization reaction of oxoenamino isomer → hydroxyimino isomer. The reaction coordinate, as depicted in Figure 3.2, is defined as $z = R(N–H) – R(O3–H)$. We used a total of 13 to 18 simulation windows for the model reactions in water, and 29 windows for the enzymatic process. For the model reactions, each simulation window consisted of 50 ps for equilibration, followed by an additional 50 ps for averaging. For the enzymatic reaction, we performed 10 ps of equilibration, followed by 30 ps for averaging and trajectory collection for each window. Overall, a total of 1.55 ns of MD simulations for the enzyme system were performed. The weighted histogram analysis method (WHAM) (*3.73*) was used to analyze the probability density and to obtain the free energy profiles for the unbiased systems along the proton transfer reaction coordinate, z.

Interaction Energy Decomposition. The interaction energy decomposition method has been widely applied to enzymes (*3.55*, *3.74–3.81*), which is adopted in

52

the present study to probe electrostatic contributions from each residue on the tautomeric equilibrium of the external PLP(H+)−L-dopa Schiff base in the active site of DDC. We used 300 configurations to calculate the average interaction energies for each tautomer. For each configuration, we sequentially zeroed the MM charges of one residue and calculated the QM/MM energies in the order of the distance between the C_a atom of the residue and the bridging hydrogen of the Schiff base. The energy difference between the total QM energies that include (I) and exclude (I-1) residue I corresponds to the electrostatic interaction energy between residue I and the QM system (PLP(H+)−L-dopa Schiff base):

$$\Delta E_{elec}(I) = \left[E_{QM}(I) + E_{QM/MM}(I) \right] - \left[E_{QM}(I-1) + E_{QM/MM}(I-1) \right] \qquad (3.1)$$

where $E_{QM}(I)$ is the energy of the "QM" subsystem and $E_{QM/MM}(I)$ is the interaction energy between the QM region and the rest of the system in which residue I is included. The last two terms in eq (3.1) represent energies computed when residue I is excluded in the decomposition analysis (*3.55, 3.74–3.81*). Since the most interesting quantity is the differential electrostatic interaction energies in going from the oxoenamino to the hydroxyimino tautomer, we computed and reported the difference for this reaction in Figure 3.5:

$$\Delta\Delta E_{elec}(I) = \Delta E_{elec}^{enol}(I) - \Delta E_{elec}^{keto}(I). \qquad (3.2)$$

53

3.3   Results and Discussion

We have studied the intramolecular proton transfer reaction of PLP(H+)–L-dopa Schiff base in dopa decarboxylase using a combined QM/MM potential with MD simulations. We used a dual-level approach to compute the potential of mean force for the intramolecular proton transfer reaction (tautomeric equilibrium), in which the intrinsic (gas-phase) free energy change was determined using Gaussian03 (*3.82*) at the hybrid DFT method at the B3LYP/6-311+G(d,p) level (*3.83*) and the semiempirical QM(AM1)/MM model was adopted to obtain solvation effects as the lower-level model. For comparison, we also estimated the solvation effects by using the polarizable continuum model (PCM) (*3.84, 3.85*), which provides a validation of the QM/MM simulations. In what follows, we first examine the effects of the protonation state of the pyridine ring in the external Schiff base, and substituents on the imino nitrogen. Then, we compare solvation results from the dual-level QM/MM simulations with those from the PCM calculations. The potential of mean force for the intramolecular proton transfer reaction of the PLP(H+)–L-dopa substrate in dopa decarboxylase is examined, with special emphasis on individual residue contributions in the active site. Finally, we discuss biological implications arising from these results.

A. Hydroxyimine and Oxoenamine Tautomerism in the External PLP–Aldimine.   There are two main factors contributing to the tautomeric equilibrium of the PLP Schiff base: (a) the protonation state of the pyridine ring, and (b) the substituent on the imino nitrogen of the Schiff base. The pyridine nitrogen of the external aldimine substrate has a pK_a of about 5.8 in aqueous solution (*3.3*), which

is unprotonated under physiological conditions. However, in the active site of an enzyme, the pyridine nitrogen is assumed to be protonated due to stabilization through ion-pair interactions with a basic residue such as Glu and Asp or hydrogen bonded to polar residues such as Ser and Thr. In the present case of DDC enzyme, Asp271 is the counter ion of the pyridinium ion. In rare situations, such as in the active site of alanine racemase (*3.81, 3.86, 3.87*), the pyridine nitrogen is unprotonated and accepts a hydrogen bond from an arginine residue.

The protonation state of the pyridine nitrogen directly affects the tautomeric equilibrium between the hydroxyimino and oxoenamino isomers, which is reflected by the free energy difference between reactions **1** and **2** (Table 3.1). Protonation at the pyridine nitrogen shifts the intrinsic tautomeric equilibrium towards the oxoenamine direction with a free energy change of ca. 5 kcal/mol. In this case, the protonated pyridine ring acts as an electron sink, stabilizing the phenolate anion configuration through induction effects. Therefore, the positive charge of the protonated pyridine ring favors electrostatic stabilization of the zwitterionic configuration more than that of the covalent neutral tautomer.

Substituent effects at the imino site of the PLP Schiff base are revealed in the series depicted in Scheme 3.2. Reaction **b** introduces a carboxylate group, whereas reaction **c** models the full L-dopa substrate. For reaction **b**, the α-carboxylate group stabilizes the oxoenamino isomer by about 7 kcal/mol both in the unprotonated and the protonated pyridine systems in comparison with the parent compounds in reaction **a**. Similarly, the inclusion of the L-dopa group in reaction **c** have similar effects, indicating that the oxoenamino tautomer gains greater stabilization in the presence of

a carboxylate substituent at the imino nitrogen. The carboxylate effects can be attributed to ion-pair interactions between the carboxylate ion and the zwitterionic Schiff base (**b**−**d**).

Table 3.1   Computed Free Energies of Tautomerization Reaction (Oxoenamine → Hydroxyimine) for the Model Reactions Located in the Gas Phase and in Aqueous Solution, and for the Reaction in Dopa Decarboxylase at 298.15 K[a]

reaction	ΔG_{gas}	ΔG_{PCM}	$\Delta G_{PMF\copyright}$
a1	−4.4[b]	0.7[e]	1.4
a2	0.4[b]	1.9[e]	2.9
b1	2.3[b]	2.3[e]	−0.1
b2	7.4[c]	3.6[f]	−2.4
c1	0.9[b]	0.7[e]	−0.7
c2	7.0[c]	1.4[f]	−1.3
d1	5.0[d]	363.4[g]	−0.9
d2	7.1[c]	9.4[f]	−2.0
d2 in DDC	−−	−−	−1.3

[a] All energies are given in kcal/mol. [b] B3LYP/6-311+G(d,p)//B3LYP/6-311+G(d,p). [c] B3LYP/6-311+G(d,p)//HF/6-311+G(d,p). [d] B3LYP/6-311+G(d,p)//HF/6-31+G(d). [e] Solution free energies were calculated at HF/6-311+G(d,p) using full optimized structures at B3LYP/6-311+G(d,p) in the gas phase. [f] Solution free energies were calculated at HF/6-311+G(d,p) using full optimized structures at HF/6-311+G(d,p) in the gas phase. [g] Solution free energies were calculated at HF/6-311+G(d,p) using full optimized structures at HF/6-31+G(d) in the gas phase. [h] Correction of B3LYP/6-311+G(d,p) gas-phase energy has been made to the lower-level (AM1) energy, i.e., $\Delta G_{PMF'} = \Delta G_{PMF} - \Delta G_{gas}^{AM1} + \Delta G_{gas}^{B3LYP/6-311+G(d,p)//AM1}$.

The results presented in the second column of Table 3.1 are consistent with the recent experimental findings by Limbach, Toney and their coworkers (*3.30*, *3.31*). In that work, a series of aldimine Schiff bases and the corresponding acid-base adducts with carboxylic acids were examined using solid-state ^{1}H and ^{15}N NMR. An interesting coupling between intramolecular hydrogen bonding and the protonation state of the pyridine ring was observed; the bridging proton was found to be located on the phenolic oxygen in the crystal structure of the aldimine Schiff base that carries a methyl substituent on the imino nitrogen. This observation is in accord with the results of reaction **a1** (Scheme 3.2) in Table 3.1. In addition, protonation of the pyridine nitrogen shifts the bridging proton from the phenolic oxygen to the imino nitrogen, which is consistent with the present findings in comparison of reaction **2** with reaction **1**.

B. Solvation Effects on the Tautomeric Equilibrium of PLP Schiff Base. The free energy results from the potentials of mean force for the intramolecular proton transfer reactions in the hydroxyimine and oxoenamine tautomerization in aqueous solution, $\Delta G_{PMF'}^{keto \to enol}$, are summarized in Table 3.1. Calculated dipole moments in the gas phase (μ_{gas}), in aqueous solution (μ_{aq}), and the induced dipole moments from vacuum to water solution ($\Delta\mu_{ind}$) are listed in Table 3.2.

For the parent system, we found that solvation stabilizes the oxoenamino form of the PLP Schiff base by 5.8 and 2.5 kcal/mol for reactions **a1** and **a2**, respectively, relative to the gas phase reactions. In the unprotonated-pyridine system, the oxoenamino tautomer has a larger dipole moment than the hydroxyimino configuration, resulting in greater solvation effects in water (*3.11*). This is further

reflected by a larger induced dipole moment in the oxoenamino isomer ($\Delta\mu_{ind}^{oxoenamino}$ = 2.9 D) than that in the hydroxyimino tautomer ($\Delta\mu_{ind}^{hydroxyimino}$ = 1.73 D) for reaction **a1**. On the other hand, when the pyridine is protonated, solvent effects on the tautomeric equilibrium is less pronounced in **a2** than in **a1** since the dominant factor is the solvation of the cation. For comparison, the results from the PCM continuum model are in excellent agreement with explicit simulations (Table 3.1).

For reactions **b** and **c,** aqueous solvation strongly stabilizes the hydroxyimino tautomer by 2.4 and 1.6 kcal/mol, respectively, when the pyridine ring is unprotonated, and the effects are increased to 8−10 kcal/mol if pyridine is protonated. Similarly, in the full PLP−L-dopa Schiff base model in reaction **d**, the proton transfer from oxoenamine to hydroxyimine is strongly favored by −5.9 kcal/mol and −9.1 kcal/mol in water for **d1** and **d2**, respectively. The hydroxyimino form of the tautomer in reactions **b−d** can gain greater solvation stabilization without the interference of a neighboring zwitterion. Furthermore, we found that the induced dipole moments ($\Delta\mu_{ind}$) for the hydroxyimino tautomers are greater than that for the oxoenamino isomers (Table 3.2). In comparison with the results from the PCM calculations, the trends are in reasonable agreement with QM/MM simulations, except that in reaction **d2**. In the latter case, the dihydroxyphenyl group in the optimized structures of the hydroxyimino and oxoenamino tautomers in reaction **d2** is closer to the pyridine ring in the hydroxyimino conformation (see Appendix B), resulting in a smaller solvent-accessible surface. Consequently, the hydroxyimino structure is poorly solvated in comparison with the more exposed configurations sampled in explicit simulations.

Table 3.2   Computed Dipole Moments (Debye) in the Gas Phase (μ_{gas}) and in Aqueous Solution (μ_{aq}) using DFT, ab initio, and PCM Methods

reaction	tautomeric state	$\mu_{gas}{}^{a}$	$\mu_{aq}{}^{b}$	$\Delta\mu_{ind}{}^{c}$
a1	oxoenamine	7.29	10.17	2.88
	hydroxyimine	5.05	6.78	1.73
a2	oxoenamine	4.37	6.64	2.27
	hydroxyimine	5.77	8.28	2.51
b1	oxoenamine	8.20	9.71	1.51
	hydroxyimine	11.11	13.41	2.30
b2	oxoenamine	18.64	23.06	4.43
	hydroxyimine	22.13	27.80	5.66
c1 [b]	oxoenamine	7.15	8.71	1.56
	hydroxyimine	7.91	10.58	2.67
c2 [b]	oxoenamine	19.11	23.68	4.57
	hydroxyimine	20.93	26.81	5.88
d1 [b]	oxoenamine	12.45	16.41	3.96
	hydroxyimine	12.90	16.77	3.87
d2 [b]	oxoenamine	15.76	20.48	4.72
	hydroxyimine	18.18	23.28	5.11

[a] Dipole moments μ_{gas} were calculated at HF/6-311+G(d,p) with fully optimized geometries using the methods as described in the footnote of Table 3.1. [b] Dipole moments μ_{aq} were calculated using PCM method at HF/6-311+G(d,p) with fully optimized geometries using the same method as described in the footnote of Table 3.1. [c] Induced dipole moments, $\Delta\mu_{ind}$, is defined as $\Delta\mu_{ind} = \mu_{aq} - \mu_{gas}$.

On the experimental side, Sharif et al. studied the chemical shifts of the Schiff base of PLP-analogs in aqueous solution at different pH (*3.3*). It was found that the pK_a values of the ring nitrogen and the imino nitrogen are 5.8 and 11.4, respectively, for methylated Schiff base systems. The computational results are consistent with these findings in that the oxoenamino form of the aldimine in **a1** is predominant in aqueous solution both in the protonated and unprotonated forms of the pyridine ring (Table 3.1).

Listed in Table 3.3 are the average bond distances associated with the intramolecular hydrogen bond, i.e., between atoms O3–H (r_{O3-H}), N–H (r_{N-H}), and O3–N ($r_{O3\cdots N}$), and the dihedral angle of $\phi_{N-C4A-C4-C3}$ (Figure 3.2) depicting out-of-plane torsion of the Schiff base. For all reactions, the donor-acceptor distances ($r_{O3\cdots N}$) are shorter in zwitterion configurations than neutral hydrogen-bond pairs, thanks to electrostatic interactions. The intramolecular ion-pair interactions enforce the protonated Schiff base to stay in the planar configuration characterized by relatively small dihedral fluctuations in a range of −5° to 7°. On the other hand, the average dihedral angles ($\phi_{N-C4A-C4-C3}$) for the hydroxyimino isomer exhibit a greater degree of flexibility, ranging from ±17° to ±163°. Furthermore, in most cases, the average dihedral angles are far from planarity from the aromatic ring, suggesting that the PLP–Schiff base does not maintain strong intramolecular hydrogen bonds. In contrast, geometry optimizations using the PCM continuum model always result in planar structures, favoring intramolecular hydrogen bonds, a clear illustration of the difference between a single structural optimization and dynamics simulations at a given temperature.

60

Table 3.3 Averaged Bond Distances and Dihedral Angle for the Intramolecular O3···H···N Hydrogen Bonding Interaction in the Oxoenamino and Hydroxyimino Tautomers of the PLP Schiff Bases in Aqueous Solution [a,b] and in DDC Enzyme [a,c]

reaction	tautomeric state	r_{O3-H}	r_{N-H}	$r_{O3L\ N}$	$\phi_{N-C4A-C4-C3}$
a1:	oxoenamine	2.16 ± 0.11	1.01 ± 0.03	2.84 ± 0.08	1.6 ± 9.4
	hydroxyimine	0.98 ± 0.02	4.14 ± 0.56	3.99 ± 0.22	−66.2 ± 126.6
a2:	oxoenamine	2.20 ± 0.13	1.02 ± 0.03	2.86 ± 0.08	−4.6 ± 10.2
	hydroxyimine	0.98 ± 0.03	4.00 ± 0.61	4.02 ± 0.14	100.6 ± 111.9
b1:	oxoenamine	2.17 ± 0.10	1.01 ± 0.02	2.84 ± 0.07	7.4 ± 9.9
	hydroxyimine	0.98 ± 0.03	4.92 ± 0.24	4.03 ± 0.11	−5.4 ± 163.4
b2:	oxoenamine	2.20 ± 0.12	1.01 ± 0.02	2.85 ± 0.08	−3.2 ± 12.5
	hydroxyimine	0.98 ± 0.03	3.05 ± 0.73	3.23 ± 0.25	−59.7 ± 26.3
c1:	oxoenamine	2.18 ± 0.13	1.01 ± 0.03	2.85 ± 0.10	−3.2 ± 12.5
	hydroxyimine	0.97 ± 0.03	4.55 ± 0.58	4.05 ± 0.16	39.4 ± 148.8
c2:	oxoenamine	2.21 ± 0.13	1.01 ± 0.03	2.87 ± 0.08	1.9 ± 12.7
	hydroxyimine	0.98 ± 0.03	3.70 ± 0.36	2.94 ± 0.14	−41.6 ± 17.4
d1:	oxoenamine	2.15 ± 0.12	1.01 ± 0.03	2.80 ± 0.07	7.2 ± 10.6
	hydroxyimine	0.97 ± 0.03	2.63 ± 0.46	3.05 ± 0.17	46.9 ± 20.8
d2:	oxoenamine	2.17 ± 0.12	1.02 ± 0.03	2.81 ± 0.08	4.2 ± 17.0
	hydroxyimine	0.98 ± 0.02	2.55 ± 0.46	2.98 ± 0.16	−15.3 ± 44.2
d2 in DDC	oxoenamine	2.11 ± 0.11	1.03 ± 0.03	2.76 ± 0.07	−14.6 ± 10.3
	hydroxyimine	0.98 ± 0.03	3.29 ± 0.57	3.26 ± 0.31	79.8 ± 34.7

[a] All distances are given in angstroms and angles in degrees. Standard deviations in the average distances and dihedral angles have been shown for each value. [b] Averaged over 50 configurations of the last 50-ps MD trajectories. [c] Averaged over 200 configurations of the last 100-ps MD trajectories.

Table 3.4 Selected Average Distances r_{A-B} (Å) between Oxoenamino or Hydroxyimino Tautomer of PLP(H+)–L-dopa Schiff Bases and Enzyme Residues in the Active Site of DDC

PLP(H+)-L-dopa	Residue	oxoenamino tautomer	hydroxyimino tautomer	X-ray
O3	Hγ1, Thr246	2.00 ± 0.13	3.41 ± 0.21	--
H	Oγ1, Thr246	3.61 ± 0.25	2.05 ± 0.29	--
O3	Oγ1, Thr246	2.83 ± 0.12	2.87 ± 0.17	3.03
H1	Oδ1, Oδ2, Asp271	1.99 ± 0.23	2.01 ± 0.22	--
N1	Oδ1, Oδ2, Asp271	2.89 ± 0.16	2.91 ± 0.17	2.78
C4A	Hε2, His192	3.95 ± 0.23	3.65 ± 0.27	3.99
C$_\alpha$	Hε2, His192	4.63 ± 0.23	5.19 ± 0.29	--
O^{CO2}	Hε2, His192	3.77 ± 0.28	4.28 ± 0.39	--
C4A	Nζ, Lys303	3.50 ± 0.18	4.01 ± 0.27	2.84
N	Hζ, Lys303	3.86 ± 0.27	3.00 ± 0.41	--
H	Nζ, Lys303	4.80 ± 0.26	5.73 ± 0.44	--
O^{CO2}	Hη, Tyr332′	1.86 ± 0.13	1.91 ± 0.16	--
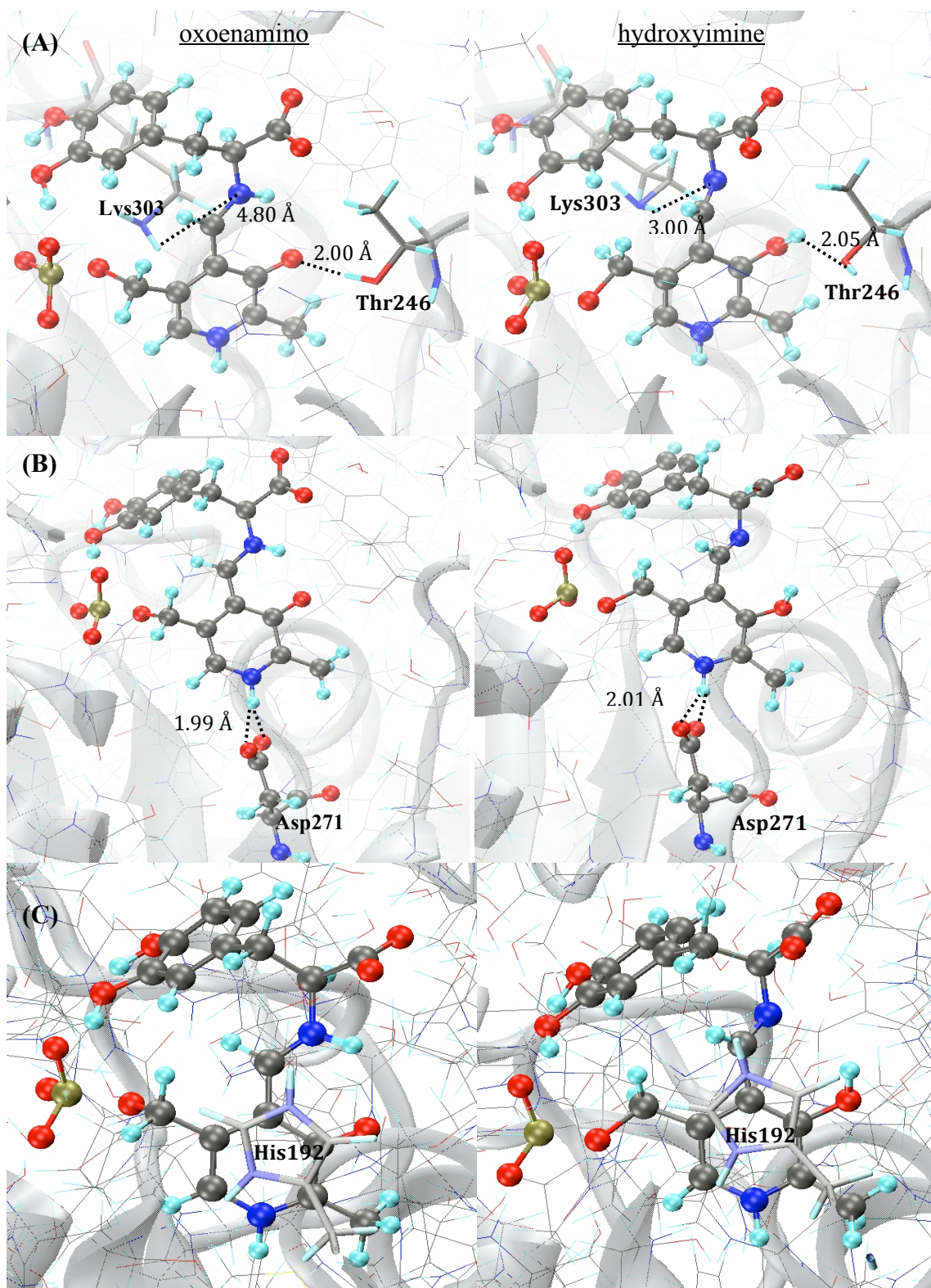
Figure 3.3 Snapshots of active-site pocket of dopa decarboxylase with oxoenamino and hydroxyimino PLP(H+)−L-dopa Schiff bases. PLP(H+)−L-dopa Schiff bases are displayed in ball-and-stick, and the specific amino acid residues: (A) Lys303 and Thr246, (B) Asp192, and (C) His192.

C. The PLP(H+)−L-dopa external Schiff Base in Dopa Decarboxylase.

The active site of dopa decarboxylase is located at the interface of two monomers, in which the PLP(H+)−L-dopa Schiff base is embedded in a network of hydrogen bonds coupled with hydrophobic interactions. The hydrogen-bonding network between the PLP external aldimine and active site residues is crucial in controlling the cofactor reactivity. In PLP-dependent enzymes, the phenolic oxygen of the PLP is stabilized by various hydrogen-bond donating residues, including tyrosine and asparagine in aspartate aminotrasferases. In DDC, a threonine residue (Thr246) is found in close proximity of the phenolic oxygen. For the oxoenamino tautomer, Thr246 donates a hydrogen bond to the phenolate anion, $^{PLP}O3^-\cdots H\gamma 1^{Thr246}$, at an average distance of 2.0 Å (Table 3.4 and Figure 3.3A). Interestingly, we found that Thr246 accepts a hydrogen bond from the phenolic hydroxyl group in the hydroxyimino isomer. The average distance for the interaction, $^{PLP}O3-H\cdots O\gamma 1^{Thr246}$, is 2.1 Å (Figure 3.3A). Thus, a conformational switch involving the hydroxyl group of Thr246 is accompanied with the proton transfer reaction. Moreover, the average $\phi_{N-C4A-C4-C3}$ dihedral angle changes from about $-15 \pm 10^{\circ}$ in the oxoenamino configuration to $80 \pm 35^{\circ}$ (Table 3.3) in the hydroxyimino isomer, indicating that the intramolecular hydrogen bond with the Schiff base is disrupted. In the latter isomer, the imino nitrogen of the external aldimine forms a weak hydrogen bond with the side chain of Lys303 at a distance of 3.0 Å (Figure 3.3A).

The pyridine nitrogen of the cofactor is protonated in DDC, forming an ion pair with Asp271, $^{PLP}N1^+-H\cdots O_\delta^{Asp271}$. A protonated pyridine ring is critical for catalysis and has been proposed to serve as an electron sink to stabilize the carbanionic

intermediate produced in the catalytic step. The average $H \cdots O_\delta^{Asp271}$ distance is about 2.0 Å from the present QM/MM simulations (Table 3.4 and Figure 3.3B), which is somewhat longer than values of 1.76 and 1.60 Å from X-ray based on heavy atom positions (*3.1*) and NMR (*3.30*) experiments. Previously, Hayashi et al. (*3.16, 3.89*) showed that both the protonated and unprotonated α-amino substrate L-dopa can bind to rat liver DDC to form the Michaelis complex, which is subsequently converted into the external aldimine. If the α-amino group is protonated, a basic residue is needed to remove the proton, and His192 can fulfill this role. Our MD simulations show that the imidazolium side chain of His192 is stacked above the pyridine ring of the PLP and interacts with the carboxylate group of the external aldimine at averaged distances of 3.8 Å and 4.3 Å (Table 3.4 and Figure 3.3C).

Sharif et al. showed that the intramolecular proton transfer is coupled with hydrogen bonding interactions at the pyridine ring. A protonated pyridine strongly favors the oxoenamino form of the Schiff base in the solid state and in polar aprotic solvents (*3.30, 3.31*). However, the intramolecular hydrogen bond is not coupled to the protonation state of pyridine ring in aqueous solution (*3.3*). This has been attributed to competing hydrogen-bonding interactions with water, causing the imine group to rotate out of the aromatic plane. The results in Table 3.1 show that solvation effects strongly favor the hydroxyimino configurations for reactions **b−d** when the pyridine nitrogen is protonated since the carboxylate anion and pyridinium cation are separated without the interference of an internal zwitterion pair. Turning to the DDC enzyme, Figure 3.4 shows that the covalent hydroxyimino tautomer is preferred by −1.3 kcal/mol over the oxoenamine configuration in the active site, and a similar

preference is found in aqueous solution for reactions **b** and **c** (Table 3.1). Note that the hydroxyimino configuration of the dopa external aldimine is not a local minimum using hybrid DFT method, although a free energy of about 7 kcal/mol can be estimated at the HF/6-31+G(d) geometry.
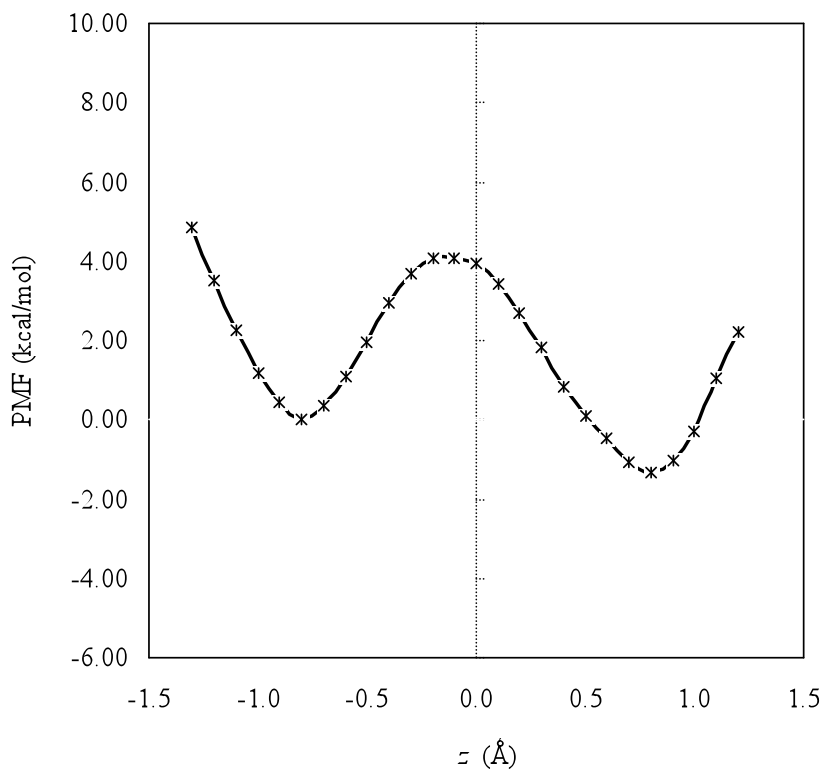
Figure 3.4   Computed dual-level potential of mean force for the tautomerization reaction (oxoenamine form → hydroxyimine form) of PLP(H+)−L-dopa Schiff base in dopa decarboxylase

The relative interaction energies of individual residues with the hydroxyimino and oxoenamino tautomers are depicted in Figure 3.5. Not surprisingly, Asp271 makes the largest differential electrostatic contributions to the PLP(H+)−L-dopa tautomeric equilibrium thanks to the ion-pair interactions with the protonated pyridine. However, in contrast to the findings for model compounds in the solid state and in polar aprotic solvents, in which intermolecular hydrogen bonds favor the oxoenamine configuration, Asp271 strongly stabilizes the hydroxyimino tautomer by nearly 9 kcal/mol. We attribute this finding to enhanced charge localization to favor ion-pair interactions in the covalent configuration of the Schiff base (Scheme 3.1). Lys303 also has strong stabilizing contributions to the neutral imino nitrogen of the cofactor, whereas the interactions are repulsive in the protonated Schiff base. Interestingly, His192 which is stacked over the pyridine ring interacts more favorably with the oxoenamino zwitterions configuration than with the hydroxyimino tautomer by 4.7 kcal/mol. Overall, the dual-level QM/MM simulations show that the tautomeric equilibrium of PLP(H+)−L-dopa Schiff base is strongly influenced by the hydrogen bonding network in the active site of DDC. Although intrinsically, the oxoenamino configuration is favored in the gas phase (Table 3.1) in all model reactions except **a1** (Scheme 3.2), electrostatic interactions involving Asp271 and Lys303 favor the equilibrium shift towards the hydroxyimine side. In addition, a hydrogen bond switch of Thr246 from a hydrogen bond donor to the phenolate ion in the oxoenamine isomer to a hydrogen bond acceptor from the phenolic hydroxyl group in the hydroxyimine configuration further helps stabilization of the latter in the intramolecular proton transfer process.
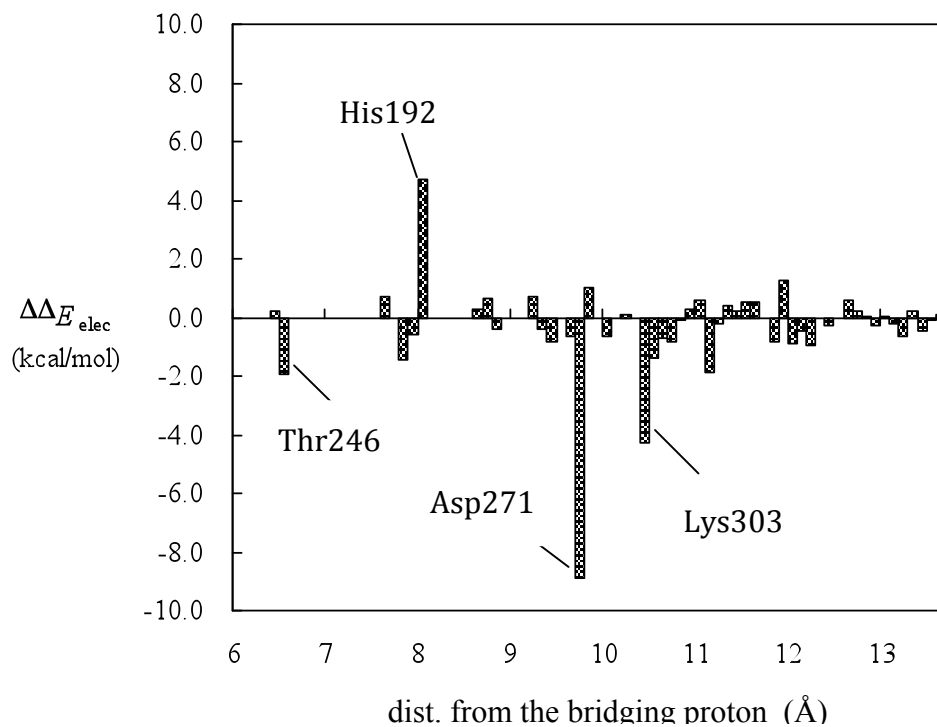
Figure 3.5   Residual dopa decarboxylase to the relative stabilization between the oxoenamino and hydroxyimino PLP(H+)–L-dopa Schiff base as a function of the distance between the C_α atom of the residue *I* and the bridging hydrogen of the intramolecular hydrogen bond in the dopa external aldimine. The residue numbers with contributions more than 2 kcal/mol are indicated.

Barboni et al. reported that the binding of L-dopa methyl ester (DopaOMe) in pig kidney DDC caused the appearance of a new absorbance at 390 nm (*3.17*). DopaOMe is an analog of L-dopa, which forms the external aldimine with the PLP cofactor, but is incapable of decarboxylation. Thus, it is used as a model for the Michaelis complex of the external aldimine. Hiyashi et al. reported a similar phenomenon in the binding of L-dopa to rat liver DDC, accompanied by the appearance of an absorption band at 380 nm (*3.16*). The species at 380–390-nm absorption was suggested to be the oxoenamine tautomer of the external PLP Schiff base; however, oxoenamino PLP Schiff bases typically have an absorption peak in the range of 400–440 nm (*3.16*–*3.19*, *3.21*, *3.22*). The blue shift of the absorption has been proposed to be either due to a nonplanar structure of the PLP Schiff base with the imine group orthogonal to the pyridine ring, or due to a special active-site environment and hydrogen bonds to PLP (*3.16*). Although the present simulations show that the imine group and the pyridine ring of the oxoenamino-form PLP(H+)–L-dopa Schiff base roughly maintain a coplanar configuration, there is a significant degree of conformational flexibility at an average torsional angle ($\phi_{N-C4A-C4-C3}$) of $-14.6 \pm 10.3^{\circ}$ (Table 3.3) in DDC. Thus, the blue-shifted absorption spectrum may be explained by the out-of-plane fluctuations about the imino moiety and the aromatic ring, which reduce the extent of π-electron delocalization. In solution, Toney and Limbach and their coworkers found that there is no cooperative interaction between the protonation state of the pyridine and the imino group in water, suggesting that the imino group prefers an out-of-plane conformation in favor of hydrogen bonding interactions with the solvent (*3.3*).

69

The intramolecular proton transfer of the Schiff base has a low free energy barrier of about 5 kcal/mol at the B3LYP:AM1/QM/MM dual-level of theory (Figure 3.4) with a preference for the hydroxyimine tautomer in the active site. The external PLP Schiff base is considered to be in an equilibrium mixture of both tautomeric configurations. Studies on the decarboxylation reaction of the PLP(H+)−L-dopa Schiff base in DDC (to be published) indicate that the hydroxyimine tautomer favors decarboxylation reaction over the reaction in the oxoenamine form. Thus, it is possible that the catalyzed decarboxylation reaction is coupled with a low-barrier intramolecular proton transfer in the PLP-dependent enzymes and responsible for the observed spectral shifts.

In summary, the presence of the hydroxyimino configuration in the external aldimine of the PLP and L-dopa complex has important implications to the enzymatic activity. First, it allows the formation of a relatively more stable enzyme-intermediate complex in decarboxylation reaction. Second, it can effectively lower the reaction barrier for enzymatic decarboxylation process by forming a more stable transition state. Further theoretical studies are being carried out to systematically investigate the decarboxylation mechanism in both tautomeric configurations of the PLP(H+)−L-dopa Schiff bases in dopa decarboxylase.

3.4   Conclusions

We have carried out combined QM/MM molecular dynamics simulations to study the oxoenamine and hydroxyimine tautomeric equilibrium of PLP(H+)−L-dopa Schiff base in DDC. Eight model reactions (Scheme 3.2) have been investigated in

the gas phase and in aqueous solution to elucidate the contributions of the intrinsic properties as well as solvent effects on the equilibrium. We employed a dual-level combined quantum mechanical and molecular mechanical (QM/MM) approach in molecular dynamics simulations, in which density functional theory was used to represent the intrinsic free energy change in the tautomeric equilibrium and the semiempirical AM1 model was adopted to describe QM and MM interactions.

Both substituent effects on the Schiff base imine group and the protonation state of the PLP pyridine ring affect the tautomeric equilibrium of the PLP Schiff bases. The presence of a carboxylate anion on the a-carbon of the imine group helps to stabilize the zwitterionic configuration and shifts the equilibrium is from the hydroxyimino tautomer to the oxoenamine isomer. Moreover, protonation of the PLP pyridine nitrogen further drives the equilibrium towards the oxoenamine direction. Solvent effects may favors either oxoenamino or hydroxyimino tautomer, depending on the electronic properties of the substituent on the Schiff base nitrogen. With a carboxylate group on the a-carbon, solvent effects favors the equilibrium in the direction of the hydroxyimino tautomer; however, but the equilibrium is driven towards the oxoenamino isomer when the substituent is a methyl group.

The present dual-level QM/MM MD simulations show that the hydroxyimino configuration of the PLP(H+)−L-dopa Schiff base is preferred over the oxoenamino isomer in the active site of DDC. The computed free energy barrier is about 5 kcal/mol from B3LYP:AM1//QM/MM simulations, suggesting that both tautomeric structures can be rapidly interconverted in the enzyme. Specific hydrogen bonding and electrostatic interactions between the aldimine Schiff base and active site residues

play an important role in the keto-enol equilibrium in PLP-dependent enzymes. In DDC, Asp271 forms an ion-pair with the protonated pyridine nitrogen, and Lys303 helps to stabilize the covalent configuration of the Schiff base nitrogen in the hydroxyimino tautomer. Importantly, Thr246 is found to play a double role of hydrogen bond donor to the phenolate ion in the oxoenamino configuration, and a hydrogen bond acceptor from the phenolic moiety of the PLP Schiff base in the oxoenamino tautomer during the intramolecular proton transfer in the external dopa-PLP aldimine complex.

Supporting Information: Fully optimized structures and atomic geometries of oxoenamino and hydroxyimino tautomers in reactions **a–d** are shown in Figure B1 and Table B1 in Appendix B. Energies and free energies in the gas phase, solvation free energies and solution free energies in aqueous solution for the tautomerization reactions of **a–d** are listed in Table B2.

Chapter 4

Catalyzing Decarboxylation of L-dopa in Dopa Decarboxylase

4.1　Introduction

Dopa decarboxylase (DDC, EC 4.1.1.28) is a pyridoxal 5′-phosphate (PLP) dependent enzyme that catalyzes the irreversible decarboxylation reaction of aromatic L-amino acid substrates, such as dopa, phenylalanine, and tryptophan (*4.1–4.3*). DDC plays an important role in the conversion of the anti-Parkinson drug 3,4-dihydroxyphenylalanine (L-dopa) or 5-hydroxytryptophan into dopamine or serotonin (Scheme 4.1). PLP is one of the most versatile enzyme cofactors, which assists a variety of chemical transformations such as amino acid decarboxylation, racemization, transamination, elimination, and aldol condensation reactions. Thus, it is important to elucidate the factors that contribute to the stereochemical control of

the enzymatic processes by a single cofactor. Furthermore, it is particularly interesting to understand the enormous catalytical power of PLP-dependent decarboxylases, which accelerates the rate of decarboxylation reactions by as large as 18 orders of magnitude over that of the uncatalyzed spontaneous amino acid decarboxylation reaction in aqueous solution (*4.4−4.10*). In this Chapter, I present results on the mechanism and the origin of catalysis in the decarboxylation reaction of L-dopa by dopa decarboxylase, making use of combined quantum mechanical and molecular mechanical (QM/MM) simulations.
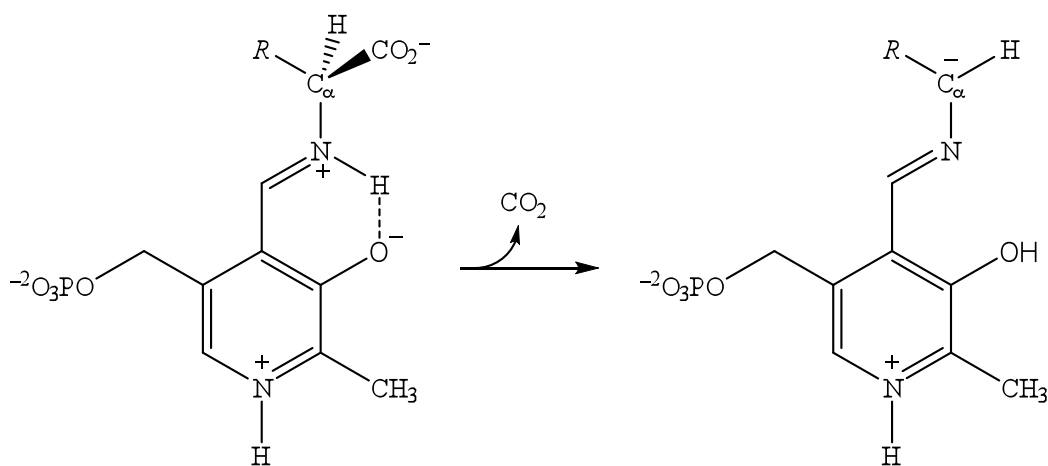
Scheme 4.1   Illustration of the Decarboxylation Reaction of L-amino Acid Substrate Catalyzed by Dopa Decarboxylase

X-ray crystal structures of the DDC–PLP enzyme binary complex and a covalent bound inhibitor ternary complex have been determined, which serve as the starting structures in the study. The active site of DDC is located at the homodimer interface, and a flexible loop, essential to catalysis, closes the active site upon binding of the substrate L-dopa. However, the loop structure was disordered in the crystal

structures, and our initial aim was to evaluate loop configurations through studies of the enzymatic reaction mechanism. Our hypothesis is that the best representative loop configurations are those that can yield a reaction mechanism and the associated free energy barrier for the decarboxylation reaction in good agreement with experimental data. Since the difficulty of evaluating re-constructed structural elements in the active site of enzyme structures that are not fully resolved in structural determination or coordinates that are generated from homology modeling is often encountered in enzyme kinetics and mechanism modeling, the strategy presented in this study, which makes use of combined QM/MM simulation techniques to evaluate enzyme activity using the reconstructed structures, can be generally used for structural validations. Our goal is to illustrate this approach in the study of dopa decarboxylase catalysis.

The X-ray crystal structures show that, in the absence of the substrate, Lys303 through its side chain ε-amino group forms an internal aldimine complex with the PLP cofactor via a Schiff base linkage (*4.3*). Binding of the substrate leads to the conversion of the internal PLP−Lys303 Schiff base into the non-covalent, external PLP-substrate complex by a transaldimine process (*4.4, 4.7*). Several active site residues can be identified to form a hydrogen-bonding network that stabilizes the external aldimine. The hydroxyl group of Thr246 interacts with the phenolic group of PLP, while the carboxylate group of Asp271 forms a salt bridge with the pyridine nitrogen of the cofactor that is typically protonated in PLP-dependent enzymes (*4.11−4.17*). The side chain of His192, a highly conserved residue among other group II PLP-dependent enzymes (*4.18*) to which dopa decarboxylase belongs, has stacking interactions with the pyridine ring of PLP and has been proposed to act as a proton

donor at the C_α atom of the intermediate after the decarboxylation step. Thr82, Ser149, Asn300, and His302 participate in the stabilization of the PLP phosphate group. However, the active site configuration in all DDC X-ray crystal structures is incomplete in that an 11-residue loop (residues 328–339), which has been proposed to cap the substrate in the binding cavity (*4.19*), is disordered, preventing a full structural characterization of the catalytic mechanism of the enzyme. .

This mobile loop is located at the dimer interface, and is possibly extended to the catalytic cavity of the neighbor subunit (*4.19*). Cleavage of the dynamic loop by proteases results in the nicked enzyme retaining a similar conformation identical to the native enzyme but the catalytic activity is abolished (*4.6, 4.20*). Kagamiyama et al. constructed a model of the external PLP–L-dopa aldimine complex structure in the enzyme active site and also proposed a structure of this mobile loop using the program MOE to explain the instability of apo-form dopa decarboxylase (*4.19*). In this model, the loop of one monomer lays on the entrance the catalytic cavity of the other subunit and occludes the active site from the solvent. The hydrophobic side chains of Met328, Pro330, Val331, Tyr332, and Leu333 were placed at the opening of the active site. Moreover, several conserved residues on the loop (*4.21*), including Tyr332 and Leu333, have been found to be essential in the catalytic reaction based on site-directed mutagenesis (*4.6, 4.9, 4.21*). The substitution of Tyr332 with a structure-analogue phenylalanine or alanine residue results in a catalytically inactive enzyme. These observations illustrated the importance of the dynamic loop in the catalytic mechanism and in the stabilization of the enzyme (*4.22*).

76

NMR and optical spectroscopic techniques have shown that PLP aldimine adduct exits a keto-enol tautomeric equilibrium between the Schiff base and phenolic hydroxyl group corresponding to an intramolecular proton transfer (*4.23–4.30*). The structure of PLP aldimine thus can present in two protonation states: the covalent hydroxyimine (enol-form, **1** in Figure 4.1) and the zwitterionic oxoenamine (keto-form, **2** in Figure 4.1). The two tautomeric configurations have different spectroscopic signatures with an absorption band in the range of 330–350 nm (*4.2, 4.31, 4.32*) and 400–440 nm, respectively. However, spectroscopic studies of dopa decarboxylase showed that the external PLP aldimine exhibited an unusual absorbance at 380–390 nm (*4.4, 4.33, 4.34*). Similar result was observed for rat histidine decarboxylase (HDC) (*4.35*), a group II PLP-dependent decarboxylase (*4.36*). This dominated 380-nm absorbing species was proposed to represent an unprotonated structure (**3** in Figure 4.1) that has a typical absorption at 360–380 nm (*4.37*) but is catalytically inert in aspartate aminotransferase (*4.38*). This absorbance was in another way assigned to the oxoenamino Schiff base, and the blue shift in wavelength was explained due to the unusual structure of the external aldimine or the contribution of the enzyme environment. In addition, combined quantum mechanical and molecular mechanical (QM/MM) molecular dynamics (MD) simulations of the tautomeric equilibration in PLP–L-dopa complex in the active site of dopa decarboxylase have been carried out, showing that the hydroxyimino configuration is preferred in the active site (*4.17*).
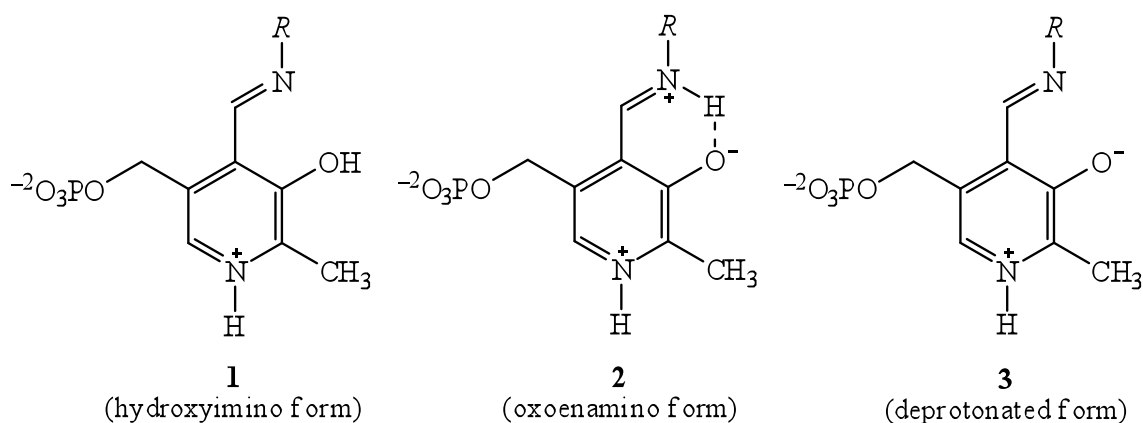
Figure 4.1 Schematic diagram of hydroxyimino, oxoenamino, and deprotonated PLP aldimines

There have been extensive computational studies of the decarboxylation reactions. Bach et al. modeled the decarboxylation reaction of alanine catalyzed by a PLP analogue in the gas phase using ab initio and DFT calculations as well as in a low dielectric medium using the polarizable continuum model (*4.14*). They emphasized the roles of the Asp/Glu residue (which is hydrogen bonded to the pyridine nitrogen of PLP) and the Arg residue (which form an ion pair interaction with the carboxylate group of substrates in certain PLP-dependent enzymes) in a cluster of two water molecules. Brill and Li investigated the gas-phase decarboxylation mechanisms of amino acids without water and with one explicit water molecule at the B3LYP/6-31G level (*4.10*). Toney studied the nonenzymatic and enzymatic PLP-dependent decarboxylation of 2-aminoisobutyrate (AIB), including the considerations of different protonation states of the external PLP

aldimine (*4.39*). It was found that the hydroxyimino aldimine with neutral pyridine ring is the most stable in the active site of dialkylglycine decarboxylase, and a concerted decarboxylation/proton transfer mechanism was proposed (*4.39*). Truhlar and Paneth et al. presented a comprehensive study of the wild type and E274A mutant ornithine decarboxylase (ODC) with ornithine and lysine using a combined QM/MM potential employing the AM1 Hamiltonian to treat the decarboxylation reaction (*4.40*). They applied ensemble-averaged variational transition state theory (EA-VTST) with quantized vibrations (*4.75*) to compute the kinetic isotope effects (KIEs) for the ODC-catalyzed decarboxylation reactions, and a good agreement with experimental results was found. Notably, the PLP external aldimine used in their calculations was the hydroxyimino tautomer, which has been shown to be lower in free energy than the oxoenamino form in the present DDC enzyme (see Chapter 3) (*4.17*).

In the chapter, we employ umbrella sampling MD simulations with a combined QM/MM potential to obtain the free energy profile for the decarboxylation reaction of L-dopa catalyzed by dopa decarboxylase. The enzymatic reaction is compared with the corresponding model reactions in aqueous solution to evaluate the contributions of PLP cofactor and the effects of enzyme stabilization to lower the free energy barrier. In addition, the role of the dynamic loop in the decarboxylation reaction is addressed by comparing the free energy profiles of the reaction using different loop structures. Moreover, the effects of the tautomeric equilibrium of the PLP cofactor on the decarboxylation reaction of L-dopa are investigated. Finally, a mechanism for the L-dopa decarboxylation in dopa decarboxylase is proposed.

4.2   Computational Details

To investigate various factors contributing to the enzymatic process of the DDC-catalyzed dopa decarboxylation reaction, we examined three decarboxylation reactions in the active site in this work. The first two models are designed to probe the effects of the keto-enol equilibrium of the PLP cofactor, resulting in the external PLP−L-dopa aldimine is in the hydroxyimino (enol) configuration or the oxoenamino aldimine (keto) isomer in the Michaelis complex. In both cases, the dynamic loop is in the closed form that occludes the active site from the solvent. The *d*ecarboxylation reaction corresponding to the *c*losed loop conformation with the *P*LP *h*ydroxyimino tautomer is labeled as DCPH, and the *de*carboxylation in the closed loop conformation with the *P*LP *o*xoenamino isomer is designated as DCPO throughout this chapter. The third model represents the decarboxylation of the hydroxyimino PLP aldimine catalyzed by the enzyme in which the loop configuration is open at the active site, and this reaction is denoted by DOPH. In the open-loop structure, we only considered the enol (hydroxyimine) configuration because it yields a lower free energy barrier in the closed-loop configuration (*vide infra*), and the results on the DOPH system are sufficient to differentiate the effects of loop conformation on enzyme catalysis. In all cases, the X-ray structure of the dimeric DDC−carbiDopa inhibitor complex (PDB code: 1JS3, 2.25 Å resolution) (*4.3*) was adopted to build the Michaelis complex model, and the procedure or constructing the three model systems has been thoroughly described in Chapter 3 (*4.17*).

We used a combined QM/MM potential (*4.41−4.43*) to describe the decarboxylation reactions of the L-dopa under three different conditions,

corresponding to the DCPH, DCPO, and DOPH complexes. In each model system, there are 39 atoms (Figure 4.2) that are represented quantum mechanically with the standard Austin Model 1 (AM1) (*4.44*) method. Although semiempirical models can have large errors for chemical reactions, as it turns out, the AM1 model performs exceptionally well for decarboxylation reactions in general, including the present L-dopa decarboxylation (*4.76–4.78*). AM1 was also successfully used to model the decarboxylation by L-ornithine decarboxylase, and good agreement with experiments was obtained in the computed KIEs. Consequently, it is not necessary to further adjust the original parameters of the QM model. The C5A atom of the PLP–L-dopa Schiff base is chosen as a QM/MM boundary atom treated by the generalized hybrid orbitals (GHO) method (*4.45, 4.46*). All protein atoms, and the phosphate group of the PLP are represented by the CHARMM22 all-atom force field (*4.47*) (MM22), whereas the solvent water molecules are described by the three point charge TIP3P (*4.48*) model.
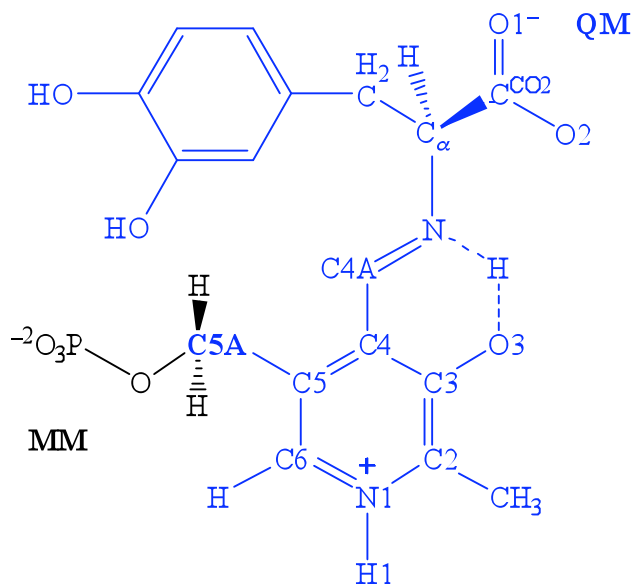
Figure 4.2

Partition of quantum and classical regions in the combined QM/MM simulations for the decarboxylation reactions in the present study, except reaction aqDOPA

In addition, to elucidate the importance of the PLP cofactor in DDC catalysis, we have investigated the uncatalyzed, spontaneous decarboxylation of L-dopa in water (denoted as aqDOPA) and the PLP-catalyzed L-dopa decarboxylation reactions in the hydroxyimino tautomeric configuraiton (aqPH) and in the oxoenamino form (aqPO), both of which are performed in the absence of the enzyme in water.

For each enzyme system, the Michaelis complex structure constructed as describe in Chapter 3 is initially fully embedded in a $93 \times 93 \times 93$ Å^3 TIP3P water box, and water molecules that are within 2.8 A of any protein atoms are removed from the system. The net charge of the enzyme system is neutral, and no counter ions are needed. A total of 75513 atoms are included in the loop-closed DCPH and DCPO systems, and 80649 atoms are in the loop-open DOPH system. For each of the uncatalyzed model reactions, aqDOPA, aqPH, and aqPO, the center of mass of the solute was first placed at the origin of the coordinate, which is the center of a rectangular water box of $30 \times 45 \times 45$ Å. No additional counter ions were included in aqDOPA system, and two sodium ions (Na$^+$) were added to neutralize the aqPH, and aqPO systems (*4.17*).

A series of umbrella sampling simulations were performed to compute the potentials of mean force (PMFs) for the decarboxylation reactions of L-dopa in dopa decarboxylase as well as for the model reactions in aqueous solution. The reaction coordination, z, of the decarboxylation processes is defined as $z \equiv R(C_\alpha - C^{CO_2^-})$, where $R(C_\alpha - C^{CO_2^-})$ is the distance between the C$_\alpha$ atom and the carbon atom of the

leaving group CO_2^-. We used a total of 13 windows for the enzymatic reactions and 12 windows for the model reactions. For the enzymatic simulations, each window consisted of 30-ps equilibration, followed by additional 30 ps for averaging and trajectory collection. For the model reactions, each window was equilibrated for 50 ps and subsequently sampled for another 50 ps. The weighted histogram analysis method (*4.49, 4.50*) (WHAM) was used to analyze the probability density and to obtain the free energy profile for the decarboxylation reactions of the unbiased systems along the reaction coordinate.

4.3   Results and Discussion

We first present results to understand the origin of the catalytic power of the enzyme by comparing the enzymatic reaction with the corresponding reaction in aqueous solution. Then, the contributions of the active site residues on the chemical reaction are emphasized, and the role of the dynamic loop in catalysis is examined. This is followed by analyzing the effects of the tautomeric structures of the PLP cofactor. Finally, we present a mechanism for the L-dopa decarboxylation reaction in dopa decarboxylase.

Catalytic Effects of Dopa Decarboxylase. Key results are depicted in Figure 4.3, which displays the computed potentials of mean force for a series of L-dopa decarboxylation reactions in dopa decarboxylase enzyme solvated in water as well as in complex with the PLP catalyst in aqueous solution. The free energy profiles yield the corresponding free energies of activation ($\Delta G_X^{\neq}$) and free energies of reaction (ΔG_X), where X specifies either *DDC* or *aq*, denoting the reaction in the enzyme or in

aqueous solution, respectively, which are summarized in Table 4.1. For comparison, results from the polarizable continuum model (*4.51–4.60*) (PCM) along with available experimental data are also provided for comparison and validation on the performance of the present QM/MM molecular dynamics simulations.
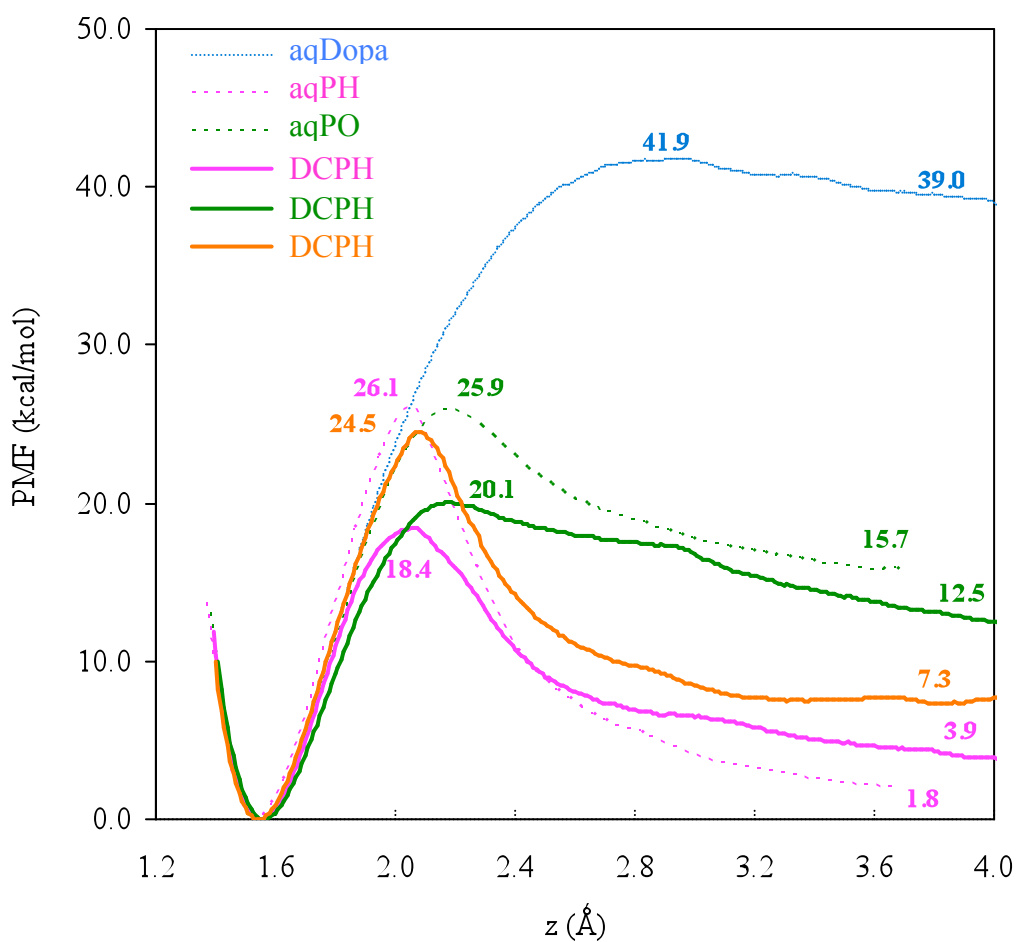
Figure 4.3   Computed potentials of mean force for the decarboxylation reactions of L-dopa in aqueous solution and in dopa decarboxylase

Table 4.1 Calculated Free Energies of Activation for the Decarboxylation Reaction of L-dopa and PLP−L-dopa in Aqueous Solution and DDC at 298.15 K [a]

	$\Delta G^{\neq}_{aq,PCM}$ [b]	$\Delta G^{\neq}_{aq,PMF}$	$C_\alpha - C^{CO_2^-}$ [c] (Å)	expt.
in aqueous solution				
aqDOPA	50.3	41.9	2.93	40.6 [d]
aqPH	24.2	26.1	2.06	--
aqPO	27.6	25.9	2.17	--
in DDC				
DCPH		18.4	2.07	
DCPO		20.1	2.17	16.1–16.5 [e-h]
DOPH		24.5	2.08	

[a] All energies are reported to one decimal place in kcal/mol. [b] Solution free energies of activation ($\Delta G^{\neq}_{aq,PCM}$) were calculated at MP2/6-31+G(d) using the geometries fully optimized at HF/6-31+G(d) in the gas phase. [c] Bond distance between C_α atom and the carboxylate carbon atom ($C^{CO_2^-}$) of L-dopa at the transition state of the decarboxylation reaction. [d] Activation energy (E_a) of phenylalanine (Phe) at 320 °C from Ref. (*4.10*). [e] Corresponding to k_{cat} value of 8 s^{-1} from Ref. (*4.7*). [f] Corresponding to k_{cat} value of 4.3 s^{-1} from Ref. (*4.9*). [g] Corresponding to k_{cat} value of 9.1 s^{-1} from Ref. (*4.5*). [h] Corresponding to k_{cat} value of 5.0 s^{-1} from Ref. (*4.6*)

For the uncatalyzed decarboxylation of L-dopa in aqueous solution, we obtained a free energy barrier of 41.9 kcal/mol, which may be compared with the experimental estimate of 40.6 kcal/mol for a closely related amino acid, phenylalanine at 320 $^{\circ}$C (*4.10*). The rates of spontaneous decarboxylation of amino acids are known to be exceedingly slow in the absence of catalysts and measurements of their rate constants are typically carried out in sealed tubes at high temperatures. The gas-phase reaction profile (not shown) exhibits a monotonically increasing free energy, reaching the plateau of fully separated CO_2 and the decarboxylation product. On the other hand, Figure 4.3 shows a small recombination barrier between the decarboxylation product of L-dopa and carbon dioxide, due to solvent reorganization. The recombination barrier is a characteristic feature that has been found in other decarboxylation reactions, in which the carbanion resulting from loss of CO_2 is not stabilized through π-conjugation. The computed free energy of reaction for the L-dopa decarboxylation in water is about 39 kcal/mol, which is 4 kcal/mol higher than that in the gas phase from MP2 calculations. Overall, solvent effects increase the decarboxylation "barrier" by 8.5 kcal/mol. At the transition state in water, the $C_\alpha - C^{CO_2^-}$ bond separation is 2.93 Å, corresponding to a Pauling bond order of 0.01. Thus, the carbon-carbon bond cleavage in water is fully completed at the transition state, consistent with the interpretation that the barrier is due to solvent reorganization effects for the recombination process. The QM/MM simulation results for the aqueous reactions may be compared with those estimated using the PCM continuum solvation model at the MP2/6-31+G(d) level (*4.51–4.60*), and the agreement is good between the two approaches.

The computed free energy of activation for the PLP-dependent decarboxylation reactions, *i.e.*, the decarboxylation of the external PLP–L-dopa aldimine complex without the enzyme, in aqueous solution is estimated to be 26.1 kcal/mol for the hydroxyimino tautomer (aqPH) and 25.9 kcal/mol for the oxoenamino configuration (aqPH) (in Table 4.1). Thus, the participation of the PLP cofactor dramatically reduces the decarboxylation barrier by as large as 16 kcal/mol, which corresponds to an intrinsic rate enhancement by 10^{11}-folds relative to the uncatalyzed reaction. The barrier reduction can be attributed to charge delocalization through π-electron delocalization, which is further enhanced by the pyridine moiety of PLP, acting as an electron sink to stabilize the carbanion transition state. The long-range ion-pair interactions between the pyridinium pyridine ring of PLP and the anion carbanion carbon of the substrate also provide transition state stabilization.

In the active site of dopa decarboxylase, the enzyme reduces the free energy barrier to 18.4 kcal/mol for the hydroxyimino aldimine (DCPH) and to 20.1 kcal/mol for the oxoenamino isomer (DCPH), which represent a further reduction in free energy barrier by 7.7 and 5.8 kcal/mol for the two cofactor tautomers, respectively. This translates to a rate acceleration by 10^6-folds over the PLP-assisted reactions in aqueous solution. Overall, the enzyme coupled with the remarkable action of the PLP cofactor stabilizes the transition state by more than 23 kcal/mol, resulting in a net rate acceleration of about 18 orders of magnitude relative to the uncatalyzed spontaneous process in water. In addition, the present QM/MM simulations reproduce the effects of aqueous solution as well as the effects of the enzyme on the decarboxylation reaction in comparison with the corresponding experimental data, validating the

87

computational procedure as well as the performance of the combined QM/MM potential surface used to model the decarboxylation reaction.

The participation of PLP in the decarboxylation of L-dopa also provides stabilization to the product intermediate (quinonoid intermediate) of the reaction both in the aqueous solution and in dopa decarboxylase. In particular, the free energy of reaction is reduced from the uncatalyzed reaction (aqDOPA) of 39.0 kcal/mol to about 23–37 kcal/mol upon formation of the external aldimine PLP complex in aqueous solution (Figure 4.3). The enhancement of PLP on the stabilization of the quinonoid intermediate formed after the decarboxylation step is due to the resonance of the pyridine-imine π-conjugation to the anion quinonoid intermediate. This has also been demonstrated in previous studies of the alanine racemization reaction in alanine racemase (*4.61, 4.62*).

To provide insight into transition state stabilization in the active site of DDC enzyme, a set of noncovalent interactions between the external PLP aldimine and key active-site residues have been analyzed. To estimate the contribution of individual residues (*4.I*) to catalysis, an electrostatic decomposition analysis was performed to account for electrostatic interactions between the reactive QM system with the MM enzyme environment (Figure 4.4). Such an interaction energy decomposition method has been widely applied to enzymes (*4.61*, *4.79–4.86*); although the computed interaction energies are not directly related to the computed PMF in Figure 4.3, the their change provides a reasonable probe to the electrostatic factors from each residue along the reaction pathway. Figure 4.4 displays the change in electrostatic interaction between the external aldimine substrate (QM subsystem) and a given residue in going

from the reactant state (RS) (Michaelis complex) to the distorted configuration of the transition state, thereby, a positive value of $\Delta\Delta E_{elec}(I)$ indicates that residue I has a greater stabilization of the reactant state than that of the transition state, and a negative value represents a net stabilizing contribution to the transition state (TS). The abscissa of Figure 4.4 is arranged in increasing distance between the center of mass of each residue and the C_α atom of the substrate.

In examining the changes of the intermolecular hydrogen-bonding interactions along the reaction path, we also analyzed the change in hydrogen-bond distances between different residues in the enzyme and the aldimine adduct (Table 4.2). Based on the free energy profiles in Figure 4.3, we used the structures saved in the PMF calculations, corresponding to the reactant state and transition state within a range of 0.1 Å of the reaction coordinate stationary points for these analyses. By this criterion, there are 400, 522, and 486 RS structures, and 575, 626, and 1124 TS structures, and 577, 508, and 479 PS structures for reactions DCPH, DCPO, and DOPH, respectively. All quantities herein have been averaged over these structures. Figure 4.5 shows the structural changes of the hydrogen-bonding networks in the binding pockets of DCPH, DCPO, and DOPH, which superimposes the reactant state and the transition state by overlapping the heavy atoms of the PLP pyridine ring.

The most striking feature of Figure 4.3 is that electrostatic interactions between the external PLP−L-dopa aldimine and residues surrounding the active site are dominantly in favor of the reactant state than the transition state. There is no single residue that can be identified in Figure 4.3 as a key contributor to provide transition state stabilization through electrostatic interactions, whereas several

residues are seen to stabilize the reactant state more than that of the transition state, including His192, Tyr332′, Asp271, and Arg447. Although this is surprising, it is in fact a very difficult task to catalyze decarboxylation reaction since the enzyme needs to stabilize the newly developed carbanion at the transition state only about 1.5 Å away from the carboxylate anion in the reaction state. The results in Figure 4.3 show that the dominant factor in barrier reduction in the dopa decarboxylation reaction is provided by the cofactor PLP by formation of the external aldimine.

His192, a highly conserved residue among group II PLP-dependent decarboxylases (*4.21*), has been proposed involving in the decarboxylation of dopa decarboxylase as the specific acid to reprotonate the quinonoid carbanion (*4.21, 4.63–4.66*). PLP-dependent decarboxylases are highly stereospecific; for example, Gani et al. observed that the decarboxylation reaction of Glu decarboxylase (group II) occurs with retention of configuration at the substrate C_α atom, revealing that a specific residue acts as a proton donor from the same face of the carboxylate group (*4.65*). Kinetic isotope effects measurements on the decarboxylation of L-methionine in L-methionine decarboxylase verify the notion that a monoprotic acid such as an imidazolium side chain of a His residue protonates the quinonoid intermediate (*4.66*). His192 is located directly above the pyridine ring in close proximity to the C_α atom of the substrate, which could fulfill multiple roles during the overall enzymatic transformation, which first acts as a proton acceptor during transaldimine process (*4.63*), and then serves as a counterion to position the substrate carboxylate group in a proper orientation for decarboxylation, and finally becomes a conjugated acid to protonate the C_α atom of quinonoid intermediate after the decarboxylation step.

90

Indeed, mutation of H192A incurs the loss of the enzymatic activation (*4.21*). Throughout the simulations, the imidazolium ring of His192 donates the Hε2 atom to make hydrogen-bonding interactions with the leaving carboxylate group at an average distance of about 1.8–2.0 Å in the reactant states, which is elongated to 1.9–2.1 Å in the transition states. Note that the ion-pair interaction with the substrate carboxylate group in the Michaelis complex also provides a key mechanism for differentiation of various reaction pathways in PLP-dependent enzymes (*4.87*). Although the electrostatic decomposition analysis reveals that His192 contributes more to the RS stabilization than to the TS stabilization, it is consistent with its key role to fix the substrate conformation such than the dissociating C_α–C bond is perpendicular to the plane of the Schiff base imino group. The average $PLP - C_\alpha{}^{\delta-} \cdots {}^+ H\varepsilon 2 - His192$ distance is less than 3.8 Å in the product states of the decarboxylation reaction, implying the proposed role as a proton donor for the protonation of the C_α atom in quinonoid.

Thr246 forms hydrogen-bonding interaction with the phenolate/phenolic group of PLP. The mutation of T246A has shown a decrease in k_{cat} and K_m by a factor of 143 and 1.54, respectively (*4.9, 4.21*). Energy decomposition analysis herein shows that Thr246 stabilizes both the reactant state and transition state, and it stabilizes the former over the latter by 1–3 kcal/mol. Altogether, both experimental and theoretical results suggest the residue enhances the catalytic efficiency of the enzyme and the binding affinity of the PLP aldimine in the active site by stabilizing the PLP aldimine throughout the decarboxylation reaction via hydrogen-bonding interactions.

91

The carboxylate group of Asp271 forms a salt bridge with the protonated pyridine nitrogen (N1) at an average distance of about 2 Å along the entire reaction pathway. The mutation of D271E decreases the catalytic activity by 10^3-folds, and the mutation of D271A abolishes the catalytic ability (*4.21*). Energy decomposition analyses show that Asp271 contributes to both the RS and TS stabilizations through $PLP-N1^+\cdots^-O\delta1-Asp271$ electrostatic interactions, and it favors the RS stabilization over the TS stabilization by 12−18 kcal/mol.

The reduction in interaction energy between the protonated pyridine and Asp271 at the transition state can be attributed to a decreased positive charge character in the pyridine ring since it has traditionally been proposed as an electron sink that attracts the C_α carbanion through π-electron delocalization. Although Asp271 shows significant stabilizing interactions with the external PLP aldimine at the reactant state, the intrinsic charge delocalization effects is more important in stabilizing the decarboxylation carbanion at the α-carbon of the substrate. Both experimental observations and theoretical results show that Asp271 is essential in catalysis of the decarboxylation of aromatic amino acid in DDC.

Another noticeable contributions to the differential interactions at the reactant and transition state is Arg447, which resides at an average distance of more than 6.8 Å (Table 4.2), and provides non-specific electrostatic stabilizations with the substrate.
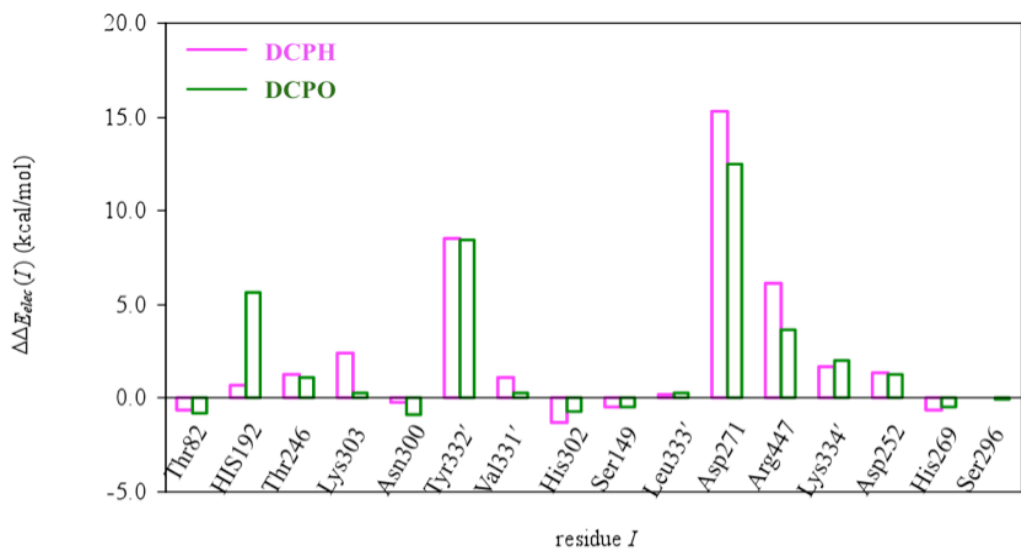
Figure 4.4   Residue contributions of dopa decarboxylase to transition state stabilization or destabilization for the decarboxylation reaction of L-dopa
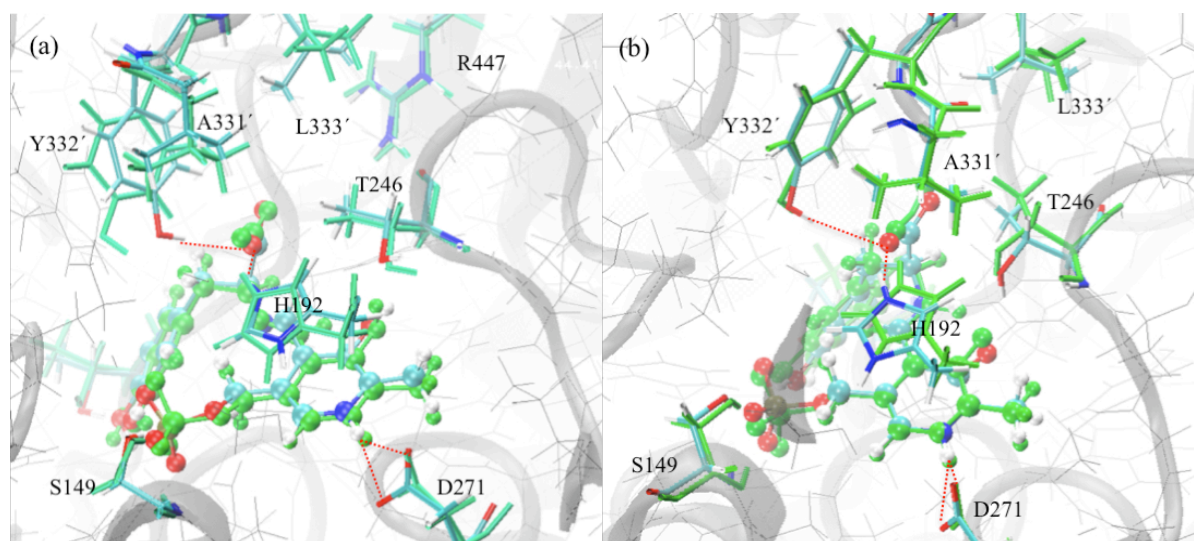
Figure 4.5 Superimpose of DDC active site and PLP external aldimine at reactant state and transition state (displayed in color of green) in reaction (1) DCPH, (2) DCPO, and (3) DOPH, respectively. PLP external aldimine is displayed in ball-and-stick model. Key residues are shown in thick-stick model. Structures of the reactant and transition states are snap shots of the umbrella sampling simulations.

Table 4.2 Selected Average Interaction Distances Between PLP−aldimine Cofactor and Active-site Residues in the Reactant (RS), Transition (TS), and Product (PS) States of Enzymatic Reactions DCPH, DCPO, and DOPH [a]

PLP−L-dopa [b]	Residue	DCPH			DCPO			DOPH		
		RS	TS	PS	RS	TS	PS	RS	TS	PS
O3	Hγ1, Thr246	3.50	3.46	3.57	2.49	2.09	2.27	3.51	3.28	3.60
H	Oγ1, Thr246	2.37	2.40	2.60	3.05	2.87	2.71	1.92	2.24	2.11
O3	Oγ1, Thr246	3.08	3.11	3.25	3.01	2.85	2.92	2.82	2.91	2.99
H1	Oδ1, Asp271	1.97	1.96	2.12	1.95	2.01	2.12	1.83	1.90	2.10
H1	Oδ2, Asp271	2.30	2.54	2.46	2.31	2.30	2.40	2.56	2.43	2.19
C$_\alpha$	Hε2, His192	3.80	3.63	3.81	3.60	3.73	3.63	3.41	3.61	3.95
O1	Hε2, His192	1.96	2.10	5.81	1.84	1.93	5.54	3.61	3.89	4.18
C4A	Hε2, His192	3.14	3.56	3.51	3.32	3.27	3.27	3.83	3.62	3.50
C$_\alpha$	Hη, Tyr332B	3.58	4.68	4.67	3.79	4. 46	4.15	3.91	4.94	6.73
O1	Hη, Tyr332B	2.08	3.51	6.28	2.07	2.89	6.27	3.46	3.84	3.47
C4A	Hη, Tyr332B	4.46	5.92	5.83	4.52	4.96	3.95	5.81	6.21	7.20
C$^{CO_2^-}$	Cζ, Arg447	7.17	6.83	4.79	8.29	8.15	6.50	10.18	10.00	11.11

[a] All bonds are reported to two decimal place in Å. [b] Numbering given in Figure 4.2.

Catalytic Effects of Loop Configuration: Closed- vs. Open-form Loop. It has been recognized that loop closure induced by substrate binding is an essential step for catalysis in numerous enzymes (*4.6*, *4.19*, *4.67–4.70*, *4.89–4.93*) which functions as a lid to occlude active center from the solvent or to provide critical stabilization of the transition state. Many PLP-dependent enzymes also consist of flexible loops at the binding site; however, it seems that some PLP-dependent enzymes do not require a loop conformation to enclose the active site based on analyses of X-ray crystallographic structures. These enzymes include dialkylglycine decarboxylase (*4.71*), ornithine aminotransferase (*4.72*), and D-amino acid aminotransferase (*4.73*). These observations raise an important question on the role of the active-site loop in DDC catalysis. To assess the loop function in the catalytic mechanism of DDC, we have determined the potential of mean force for the decarboxylation reaction of L-dopa in DDC at the hydroxyimino tautomeric configuration of the PLP cofactor, which has been found to be the preferred conformation (*4.17*).

In dopa decarboxylase, an 11-residue segment ranging from residue Met328′ through residue Asp339′, where primed numbers represents the residues in the second subunit, is located at the dimer interface in close proximity to the active site; however, its three dimensional coordinates are disordered in all available X-ray structures of DDC (*4.3*), presumably due to its dynamic flexibility under the crystallization conditions. Experimental studies demonstrated that the presence of this flexible loop is essential to catalysis since DDC completely loses its enzymatic activity when the loop residues are cleaved (*4.6, 4.19*). Furthermore, site-directed mutagenesis experiments show that the Tyr332′Ala mutation also completely

abolishes DDC activity, implicating that specific interactions between loop residues and the substrate complex is necessary in the enzyme catalyzed chemical process. Since a large number of hydrophobic residues are found in the loop sequence, it has been proposed that the flexible loop in DDC could act as a lid to enclose the active site and exclude the exposure of the substrate from the solvent upon ligand binding (*4.3, 4.6, 4.19*). To explore the function of the active-site loop in DDC catalysis, we have carried out free energy simulations along the decarboxylation coordinate both using a closed loop (DCPH) and an open loop configuration (DOPH) that were constructed as described previously (*4.17*). In the DCPH configuration, the loop is closed (Figure 4.6) over the active site, resulting in a solvent-occluded active site, whereas the loop in the DOPH structure is open and the active site is solvent-exposed. It is noteworthy that the enzyme with a closed loop at the active site (DCPH) has a lower free energy of activation for the L-dopa decarboxylation than that with an open loop (DOPH) by 6.1 kcal/mol (Table 4.1 and Figure 4.3), revealing a preference for a solvent-excluded active site in the catalysis by dopa decarboxylase.
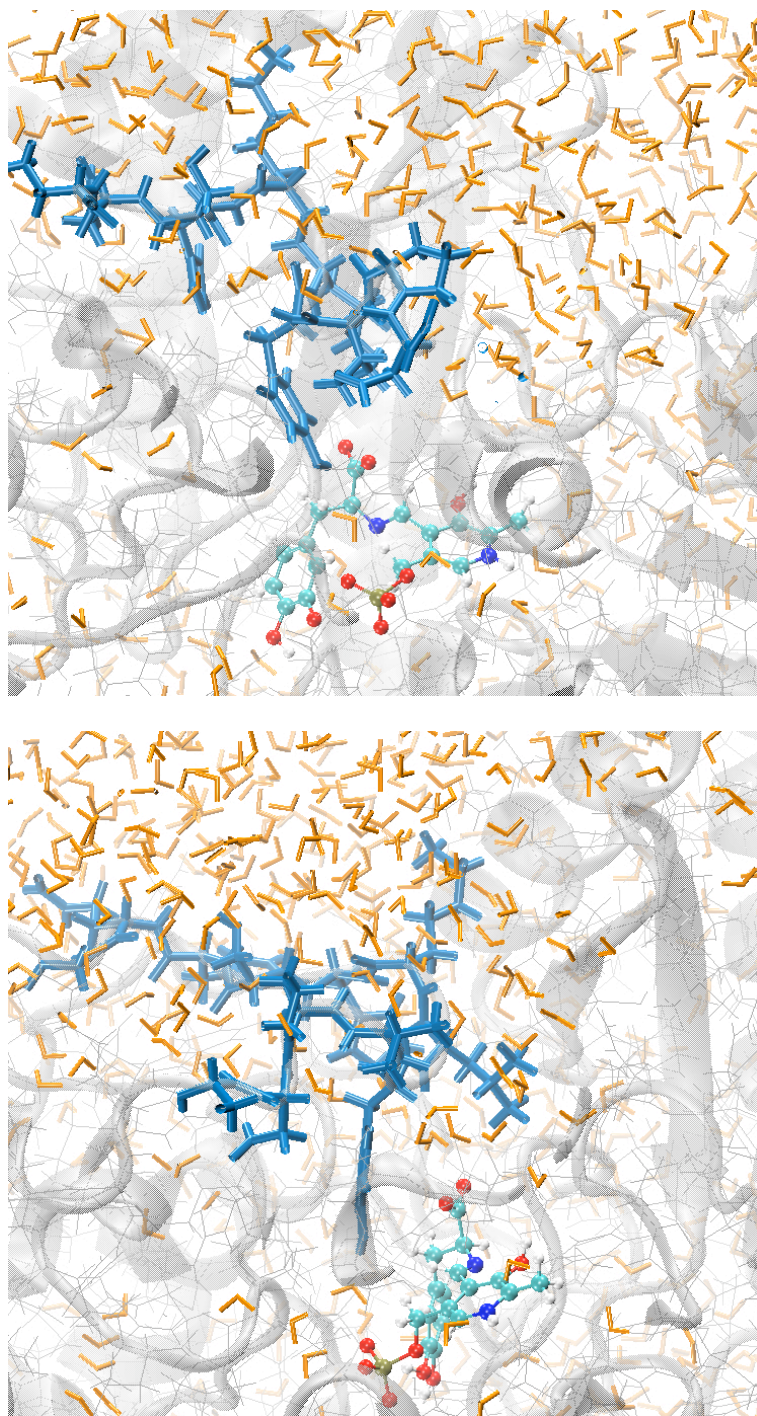
Figure 4.6 Molecular Surface and stereo views of the dynamic loop (residues 328–339) in (a) a closed or (b) an open conformation in dopa decarboxylase complex with PLP–L-dopa external aldimine. The PLP external aldimine is displayed in a ball-and-stick model, in which cyan, red, blue, and tan colors show carbon, oxygen, nitrogen, and phosphorus atoms, respectively. The loop is shown in blue molecular surface model or stick representation. Water molecules are shown in orange stick. The other residues are shown in molecular surface model. The structures are taken from the last configuration of the 250-ps equilibration of QM/MM MD simulations.

To elucidate the origin of the preference for a solvent-excluded active site in the DDC catalyzed decarboxylation reaction, we analyzed the correlation between the electrostatic interaction energies of the PLP aldimine and bulk water ($\Delta\Delta E_{elec}$) and the number of water molecules (N) within 4 Å of the carboxylate carbon atom of the leaving group along the reaction coordinate, z (Table 4.3 and Figure 4.7). We computed $\Delta\Delta E_{elec}$ in the following procedure. First, we saved the dynamics trajectory in every 50 fs (50 MD integration steps) during the umbrella sampling free energy simulations. For each configuration, we zeroed the partial atomic charges of all bulk water solvent molecules, and then calculated the QM/MM interaction energies. Afterwards, we determined the net electrostatic interaction energy at the reactant state (ΔE_{elec}^{RS}) and at the transition state (ΔE_{elec}^{TS}) between bulk water and the QM subsystem (i.e., the PLP–L-dopa external aldimine substrate) as the difference between the total QM energies that included and excluded bulk water (note that the former is already determined during the dynamics simulation). Finally, we obtain the differential stabilization effects relative to that in the reactant state at a given value of the reaction coordinate z as follows,

$$\Delta\Delta E_{elec}(z) = \Delta E_{elec}(z) - \Delta E_{elec}^{RS}(z_R). \tag{4.1}$$

where z_R is the reaction coordinate in the reactant state. The definition in eq. (4.1) indicates that a negative value of $\Delta\Delta E_{elec}(z)$ denotes that bulk water has a greater TS stabilization than the RS stabilization.

Figure 4.7 depicts the change in $\Delta\Delta E_{elec}(z)$ as a function of the reaction coordinate and the average number of water molecules within 4 Å of the carboxylate group, both in the DCPH and in the DOPH loop configurations. On average, there is

essentially no water molecule in close proximity of the carboxylate group in the closed loop configuration, DCPH, which is accompanied by increased stabilization (more negative value in $\Delta\Delta E_{elec}(z)$) along the reaction path, reaching a plateau of about -15 kcal/mol. In contrast, about four water molecules are found to be within 4 Å of the carboxylate group along most part of the reaction path, but reaching a maximum number of about 8 water molecules at $z = 3.5$ Å. Significantly, the electrostatic stabilization of the distorted substrate relative to that in the reactant state is entirely positive, revealing that the active site solvent molecules strongly stabilize the carboxylate group in the reactant state. As the decarboxylation reaction takes place, the negative charge of the carboxylate group retreats to the C_α carbon, weakening the electrostatic interactions with the water molecules. At the transition state, bulk water provides a net electrostatic stabilization of the transition state in DCPH by -5.3 kcal/mol over that in the reactant state. However, in DOPH, the solvent has an opposite effect, destabilizing the transition state by 13.8 kcal/mol. Although the interaction energies obtained here cannot be used to rationalize the free energy difference from the computed PMFs, the qualitative trends demonstrate that the exclusion of the bulk solvent from the active site by loop closure is essential in DDC catalysis.
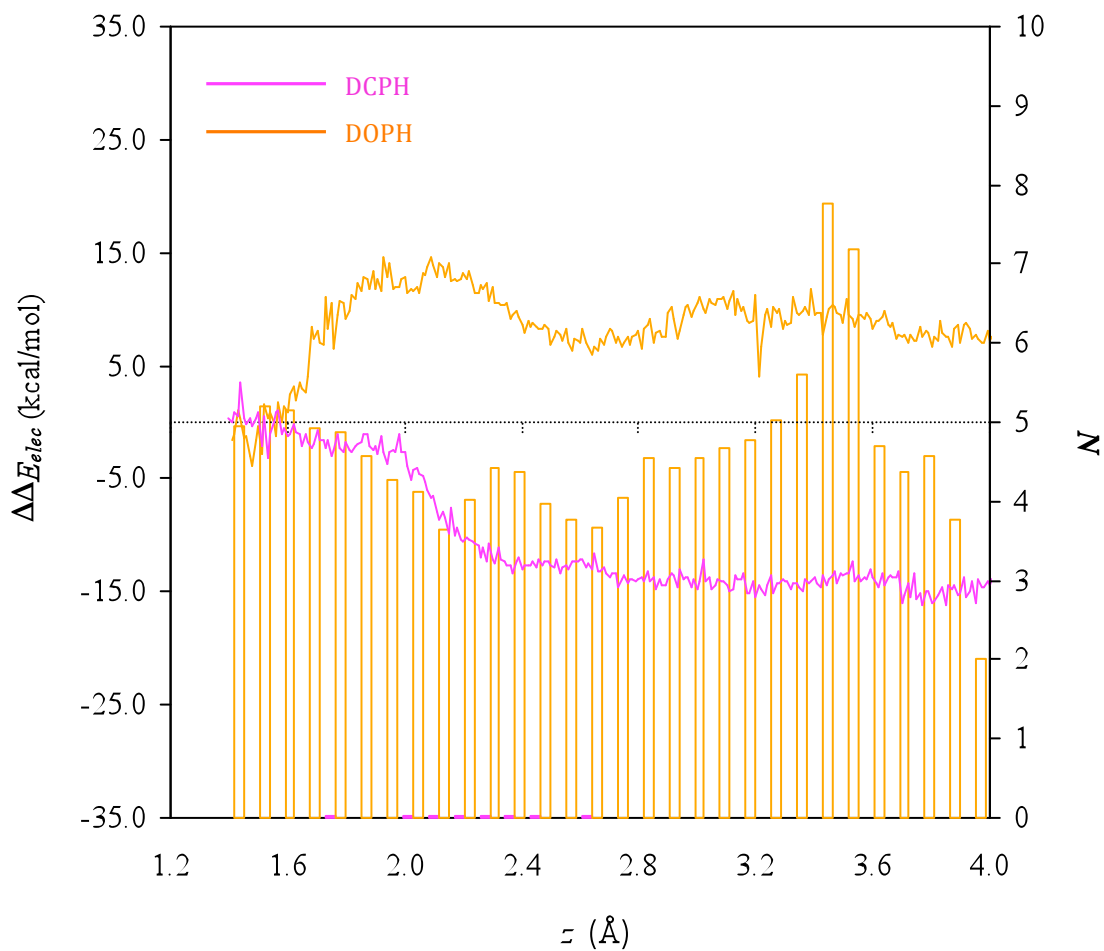
Figure 4.7   Correlation between external aldimine-bulk water electrostatic interaction energies ($\Delta\Delta E_{elec}$) and number of water molecules (N) within 4 Å of the carboxylate carbon atom of L-dopa along the reaction coordinate, z.

Table 4.3   Relative Electrostatic Interaction Energies of PLP–aldimine Complex and Bulk Water ($\Delta\Delta E_{elec}$) As Well As Number of Water Molecules Within 4 Å of the Carboxylate Carbon Atom of L-dopa (*N*) at the Reactant States (RS) and Transition States (TS) of Reactions DCPH and DOPH

	DCPH		DOPH	
	$\Delta\Delta E_{elec}$ (kcal/mol)	*N*	$\Delta\Delta E_{elec}$ (kcal/mol)	*N*
RS	0.0	0.003	0.0	5.178
TS	−5.3	0.013	13.8	4.199

Multiple sequence alignment of group II decarboxylases (*4.21*) has shown that Tyr332′ and Leu333′ are highly conservative among the loop residues. The mutations of Y332F and Y332A show increased K_m, but the decarboxylase activity is suppressed (*4.9, 4.21*), suggesting that the hydroxyl group of Tyr332′ is required in the catalytic process. The energy decomposition analyses in the QM/MM simulations of DCPH system (as well as DCPO) show that Tyr332′ provide stronger RS stabilization energy by ~ 8 kcal/mol than the stabilization of the TS. Tyr332′ forms a hydrogen bond with the carboxylate group of the substrate, which provides a key driving force for the loop closure, enabling the expulsion of the solvent molecules from the active site as the hydrophobic side chains of other residue encapsulate the active site cavity. However, this favorable interaction is reduced in the transition state due to the reduction of the negative charges on the carboxylate oxygen atoms in TS.

Comparison of the DCPH and DOPH simulations, Leu333′ in DCPH system has hydrophobic contacts with the aliphatic side chains of Val331′, Thr246, and Val436 at the carboxylate binding site during the entire reaction path of the catalysis. This hydrophobic surface appears to be responsible for the exclusion of solvent molecules from the active site. Indeed, the PMF calculations show that dopa decarboxylase with a solvent-occluded active site (DCPH) has a lower free energy barrier for the decarboxylation reaction than that with a solvent-exposed loop (DOPH). This phenomenon can be attributed to the hydrophobic clusters of Val331′, Leu333′, Thr246, and Val436 that provide a hedge around the carboxylate group of substrate to preclude the unfavorable solvent effects on the reaction, and also form more favorable interactions with the more neutral transition state structure than with the negatively charged reactant state structure to facilitate the decarboxylation relative to the reaction in aqueous solution. Leu333′ could mediate the loop closure over the active site and stabilize the transition state during catalysis.

Effects of Tautomeric Conformations: Hydroxyimine vs. Oxoenamine. To elucidate the catalytic power of dopa decarboxylase, it is essential to have an understanding of the tautomeric equilibrium and the intriguing protonation state of the PLP Schiff base. For the decarboxylations of the hydroxyimino (aqPH) and the oxoenamino (aqPO) external aldimine in water, we obtained similar free energies of activation of about 26 kcal/mol, but the former is predicted to be faster than the latter in the active site of dopa decarboxylase; the PMFs in Figure 4.3 exhibit a lower free energy barrier for the reaction using the enol (hydroxyimino) external aldimine tautomer by 1.7 kcal/mol than the keto (oxoenamino) form. Interestingly, it was

found in the study presented in Chapter 3 that the hydroxyimino aldimine of the Michaelis complex is more stable than the oxoenamino tautomer in the active site of dopa decarboxylase (*4.17*). Thus, the enol, hydroxyimino isomer of the PLP cofactor is favored both in the Michaelis complex and in the computed free energy of activation at the transition state. In addition, comparison of the free energies of the reaction for both conformations, the hydroxyimino aldimine is less endergonic than the oxoenamino isomer by about 9 kcal/mol in the enzyme, consisting with the spectroscopic finding that the quinonoid adduct in dopa decarboxylase is hydroxyimino conformation. Taking together the computational results and available experimental data, we conclude that the hydroxyimino PLP−L-dopa Schiff base is preferred for the decarboxylation of L-dopa in dopa decarboxylase.

The hydroxyimino (DCPH) and oxoenamino (DCPO) external PLP aldimine in dopa decarboxylase exhibit some differences in the intramolecular and intermolecular geometries on going from the reactant state to the transition state (Table 4.2 and Table 4.4). The bond lengths of $C_\alpha - C^{CO_2^-}$ in the reactions in DCPH and DCPO are 2.17 and 2.07 Å, respectively, indicating that the reaction for the hydroxyimino aldimine has an earlier transition state. A major difference between the DCPH and DCPO reactions is found in the conformational flexibility of the Schiff base unit, characterized by the dihedral angle of $N - C4A - C4 - C3$ (χ). In the DCPO configuration, the intramolecular $O3 \cdots H \cdots N$ hydrogen bond is maintained throughout the reaction path with an average torsion angle of $-1° \pm 2°$ to $2° \pm 3°$ in going from the reactant state to the transition state due to its strong electrostatic, ion-pair interactions. However, in the DCPH configuration, the imine group is rotated out

of the plane of the phenyl group of the PLP cofactor with an average dihedral angle χ of $-135° \pm 15°$ in the RS and $-133° \pm 4°$ in the TS. Note that there is also greater flexibility in the χ torsion for the Michaelis complex in the DCPH configuration than that in DCPO. The large average χ values in the reaction in DCPH reveals that the PLP aldimine favors the intermolecular hydrogen-bonding interactions with the active-site residue, including Thr246 and Lys303, (Table 4.2) as that in aqueous solution (*4.88*). Thr246 generates different interactions with the hydroxyimino and oxoenamino PLP−L-dopa aldimines (Table 4.2). In reaction DCPH, the residue accepts a hydrogen bond from the H atom of the hydroxyimino adduct with its side chain Oγ1 at a distance of 2.37, 2.40, and 2.60 in the RS, TS, and PS, respectively. However in reaction DCPO, Thr246 donates its side chain Hγ1 atom to form a stronger hydrogen bond with the O3 of the oxoenamino adduct at a distance of 2.49, 2.09, and 2.27 in the RS, TS, and PS, respectively.

Table 4.4   Selected Average Bond Distances, Angles, and Dihedral Angles Within PLP−Aldimine Cofactor at the Transition States of DCPH, DCPO, and DOPH in DDC [c]

	DCPH	DCPO	DOPH
Bonds (Å) [a,b]			
N1 – H1	1.03	1.03	1.03
C3 – O3	1.38	1.27	1.38
O3 – H	0.98	2.05	0.98
N – H	4.16	1.01	4.06
O3 – N	4.01	2.77	4.04
$C_\alpha - C^{CO_2^-}$	2.07 [1.55]	2.17 [1.57]	2.08 [1.54]
Angle (degree) [a,b]			
$O1 - C^{CO_2^-} - O2$	149.3 [122.8]	153.3 [123.3]	148.5 [120.4]
Dihedral Angle (degree) [a,b]			
N – C4A – C4 – C3	−135.2 ± 14.9 [−133.2 ± 3.7]	−0.9 ± 2.4 [2.1 ± 3.4]	−97.5 ± 65.9 [−134.9 ± 2.6]
$C - C_\alpha - C4 - C3$	−63.1 ± 2.4 [−65.7 ± 1.5]	−81.9 ± 1.7 [−79.6 ± 2.2]	−73.2 ± 6.9 [−53.7 ± 2.0]

[a] Numbering given in Figure 4.2. [b] Values in brackets give values for the reactant state. [c] Averaged over 400 configurations saved from the last 30-ps MD sampling trajectories.

Proposed Mechanism for Decarboxylation Reaction in DDC Enzyme.
Based on the mechanism proposed by Kagamiyama et al. (*4.4, 4.7*) and the results herein, we propose a new mechanism for the catalytic cycle of the decarboxylation reaction in DDC (Scheme 4.2). The overall enzymatic transformation involves three main processes: transaldimination, decarboxylation, and proton transfer.

In the resting state of the enzyme in the absence of substrate, dopa decarboxylase forms an internal Schiff base with PLP cofactor through the ε-amino group of Lys303 (**E–PLP**) in the active site. Internal **E–PLP** aldimine primarily shows two structures equilibrated via keto-enol tautomerization: oxoenamino (**I**) and hydroxyimino (**II**) tautomers (*4.4, 4.7, 4.31, 4.32*). After the enzyme is bound with a substrate (for instance, L-dopa in Scheme 4.2) in the active site, His192 abstracts a proton from the protonated α-amino group of the substrate. With the appearance of the unprotonated substrate at the active site, the Michaelis complex **[E–PLP]•L-dopa** is formed in which the internal **E–PLP** complex is driven to the oxoenamino state. Subsequently, the enzyme undergoes a sequence of reactions in the transaldimination process (**[E–PLP]•L-dopa** → **[E–PLP-L-dopa]** → **E•[PLP–L-dopa]**), in which the substrate displaces Lys303 to generate the external aldimine complex **E•[PLP–L-dopa]**, for which the enzyme environment slightly shifts the tautomerization equilibration to the hydroxyimino state (**IV**) stabilized by Thr246, Asp271, and Lys303. The external aldimine complex **E•[PLP–L-dopa]** represents the Michaelis complex configuration, which undergoes an irreversible decarboxylation reaction to produce the quinonoid intermediate **E•[PLP–Q]** and carbon dioxide. We propose that His192 protonates the quinonoid intermediate with retention of stereochemistry

at the C_α carbon. The decarboxylation step in the active site of DDC has been established to be the rate-limiting step of the enzymatic cycle (*4.7*). In the decarboxylation process, the anionic charge on the carboxylate carbon atom of the substrate is transmitted to the C_α carbanion of the quinonoid intermediate, which is further stabilized by the PLP cofactor. The negatively charged C_α atom of the quinonoid intermediate attracts the imidazolium ring of His192 to reach an optimal position within a hydrogen-bonding distance. His192 acts as a conjugated acid to protonate the C_α carbanion of the intermediate to form the enzyme–dopamine complex **E•[PLP–dopamine]**. Finally, a reverse transaldimination of the **E•[PLP–dopamine]** complex takes place, in which the neutral ε-amino side chain of Lys303 attacks the C4A of the PLP–dopamine external aldimine to recover the enzyme back to its unliganded form (**E–PLP**) and the product dopamine is released.

It is noteworthy that in this proposed mechanism His192 plays the key role as an acid/base catalytic agent in the reaction cycle. It has been suggested that, based on kinetic isotopic effects experiments of other PLP-dependent decarboxylases, the proton acceptor for the α-amino group of substrate in the formation of the Michaelis complex could be on the same face of PLP cofactor as the proton donor for the C_α of the quinonoid intermediate after the decarboxylation (*4.66, 4.74*). Furthermore, in the present simulations the average bond distance between His192 side chain and the quinonoid C_α is within a hydrogen-bonding distance, indicating that His192 has the potentiality to protonate the quinonoid intermediate at the C_α position in the protonation step. Thus, His192 could mediate the deprotonation of substrate and the protonation of quinonoid intermediate in the catalytic cycle.
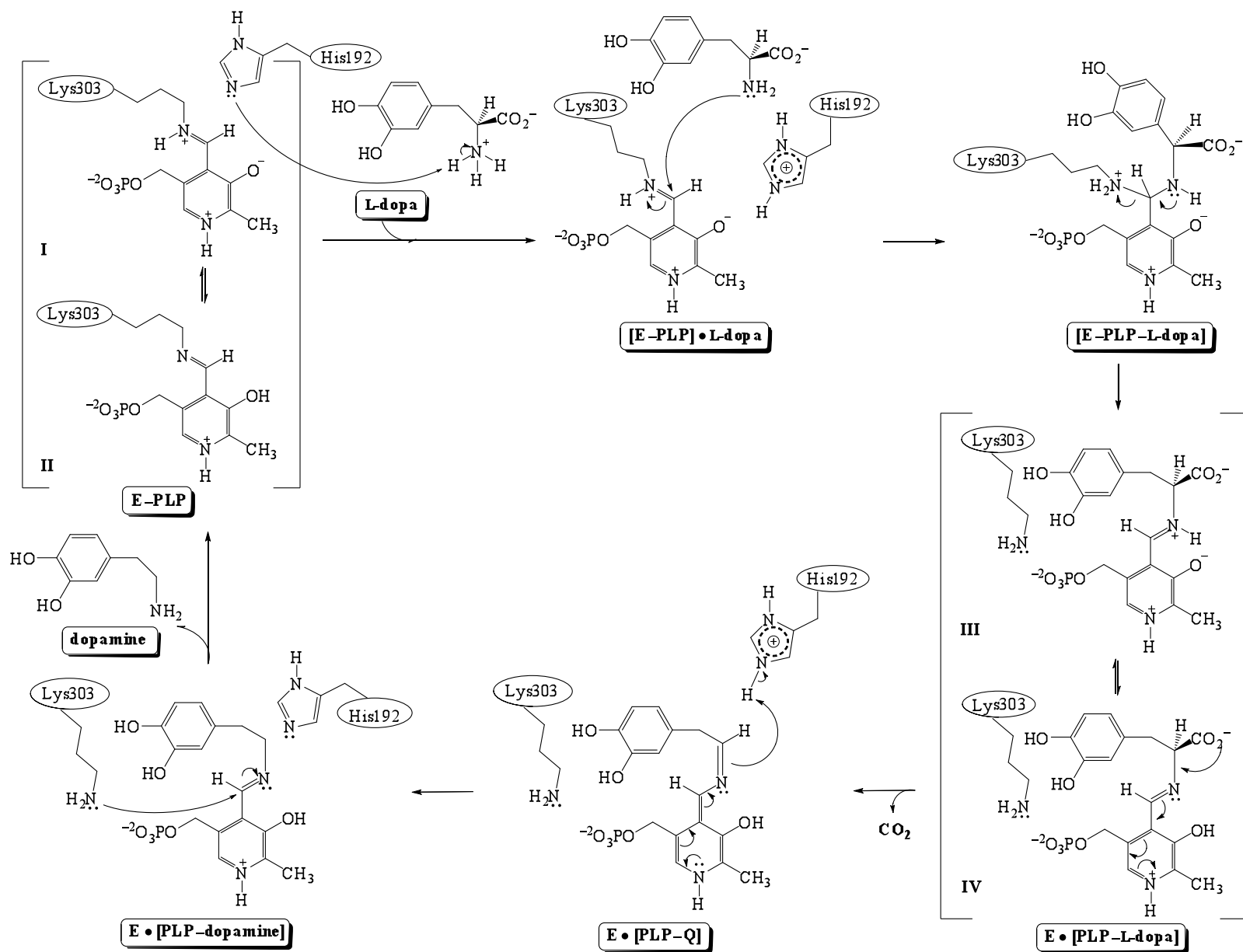
Scheme 4.2 Proposed Mechanism for the Decarboxylation Reaction of L-dopa by Dopa Decarboxylase

4.4   Conclusions

We have carried out combined QM/MM MD simulations of the decarboxylation reaction of L-dopa by the enzyme dopa decarboxylase. Free energy profiles for the decarboxylations of L-dopa and the external PLP aldimine complexes in water as well as in dopa decarboxylase are determined to gain an understanding of the origin of rate enhancement of the enzyme and the contributions of enzyme stabilization to the chemical reaction. Our computational results are agreeable to the experimental data. A main finding in this study is the origin of the catalytic proficiency of the enzyme in controlling the chemical reactivity. We have identified that the pyridine-protonated PLP cofactor contributes considerably to the enzymatic decarboxylation reaction that it dramatically enhances the reaction rate by about 11 orders of magnitude in aqueous solution compared with the uncatalyzed reaction. The cofactor plays an important role in the enzymatic reaction, and it stabilizes the transition state as well as the product intermediate in the reaction by reducing the negative charge on the C_α carbanion through its charge delocalization ability. In addition, the enzyme further lowers the free energy barrier of the decarboxylation reaction. It was found that the closed loop configuration, which expels water molecules from the active site, plays a key role in the enzymatic process. Since the solvent water molecules strongly stabilize the carboxylate group in the reactant state than that in the charge depleted transition state, the open-loop configuration, in which an average of four water molecules are present in close proximity to the carboxylate group, inhibits the decarboxylation reaction. Thus, a main contributing factor to barrier reduction in the enzyme is due to reduced reactant state stabilization by water

molecules due to the exclusion of water molecules in the active site in the closed loop configuration (*19*). An additional finding in this study is the tautomeric effects of PLP moieties on the decarboxylation reaction. For the enzymatic decarboxylation reaction of L-dopa, the external hydroxyimino aldimine is predicted to be preferred over the oxoenamino isomer by having a lower free energy barrier and a more stable carbanion intermediate. Thus, the small difference in structure of the PLP aldimines leads to distinct catalytic effects.

Chapter 5

Kinetic Isotope Effects in Dopa Decarboxylase

5.1   Introduction

Although of the dominant factor in enzyme catalysis is the lowering of the quasiclassical free energy of activation in comparison between catalyzed and the corresponding uncatalyzed reaction in aqueous solution (*5.1, 5.2*), nuclear quantum effects (NQEs) have been shown to be important in determining the rate of enzymatic reactions, which can make small but noticeable contributions to rate enhancement. Consequently, it is important to incorporate NQEs in computation of the rate constant (*5.1, 5.3*). Furthermore, it is of interest to determine kinetic isotope effects (KIEs) for chemical reactions in enzymes and in solution using computational methods because it is useful to use theory to rationalize experimental observations, and the ratio of the reaction rates between light and heavy isotopes provides direct experimental probe to

112

the nature of the transition state (*5.4*). Hence, KIEs have been widely used for inferring transition state structure in chemical and enzymatic reactions as well as been used for elucidating the reaction mechanisms by evaluating relative values of rate constants in certain specific chemical steps (*5.5, 5.6*). For certain types of enzyme-catalyzed reactions, heavy-atom isotope effects can provide useful information about the rates of various steps in the mechanism. Carbon isotope effects on decarboxylations are among the most studied heavy-atom isotope effects, both in nonenzymatic reactions and enzymatic reactions because the decarboxylation product, carbon dioxide, is especially amenable to convenient isotopic analysis (*5.6, 5.14*). To address the question of NQEs in enzymatic decarboxylation reactions, the present study was designed to make predictions on the KIEs both at the primary and at the secondary sites associated with the decarboxylation of L-dopa in solution and in the enzyme dopa decarboxylase (DDC). The findings provide further insights into the nature of the transition state structure of the decarboxylation of PLP−L-dopa aldimine in the enzymatic decarboxylation step.

Dopa decarboxylase catalyzes the conversion of L-dopa into dopamine in the treatment of Parkinson disease by means of a bound pyridoxal 5′-phosphate (PLP)-dependent cofactor, as shown in Scheme 5.1. In the absence of substrate, the cofactor is bound to the enzyme by means of a Schiff base linkage between the aldehyde carbon of the cofactor and an ε-amino group of a lysine residue of the enzyme. When substrate is bound in the active site of the enzyme, the internal aldimine undergoes a Schiff base interchange to form the so-called external aldimine Schiff base between L-dopa and the cofactor PLP. The decarboxylation reaction of the substrate L-dopa is

113

the rate limiting (*5.7*) step in the DDC-catalyzed reaction cycle, which is accelerated by a factor of ~10^{18} over the uncatalyzed, spontaneous decarboxylation reaction of aromatic amino acids in aqueous solution. Although KIEs for the present dopa decarboxylase have not yet been determined, carboxyl ^{13}C-KIE is widely used to probe the mechanism of PLP-dependent decarboxylation reactions (*5.8–5.13*). The observed KIEs are typically in the range of 1.01–1.06 (*5.6, 5.12, 5.14*), keeping in mind that these are not necessarily the intrinsic effects of the decarboxylation step. It is generally accepted that the intrinsic KIE for the carboxyl carbon in PLP-dependent decarboxylation reactions is about 1.05. In the closely related PLP-dependent L-ornithine decarboxylase, Swanson et al. found that the physiological substrate has an observed KIE of 1.0325, and it is increased to 1.0633 for the unphysiological substrate L-Lys (*5.12*). The latter values were considered to be close to the intrinsic KIE for the decarboxylation step as it becomes rate-limiting of the enzyme process. The L-ornithine decarboxylase reaction and the carbon kinetic isotope effects have been studied computationally, and a reasonable agreement with experimental results has been obtained (*5.60*). Thus, by means of theoretical methods within the framework of quantum transition state theory, computational studies can shed light on the extent of NQEs on the chemical reaction, including both zero-point energies and tunneling contributions.

In the present study, we employ a coupled free-energy perturbation and umbrella-sampling simulation technique in centroid path integral calculations (PI-FEP/UM) (*5.15*) to predict the ^{13}C primary (1°) and ^{2}H secondary (2°) KIEs on the decarboxylation of L-dopa catalyzed by the PLP-dependent enzyme, dopa
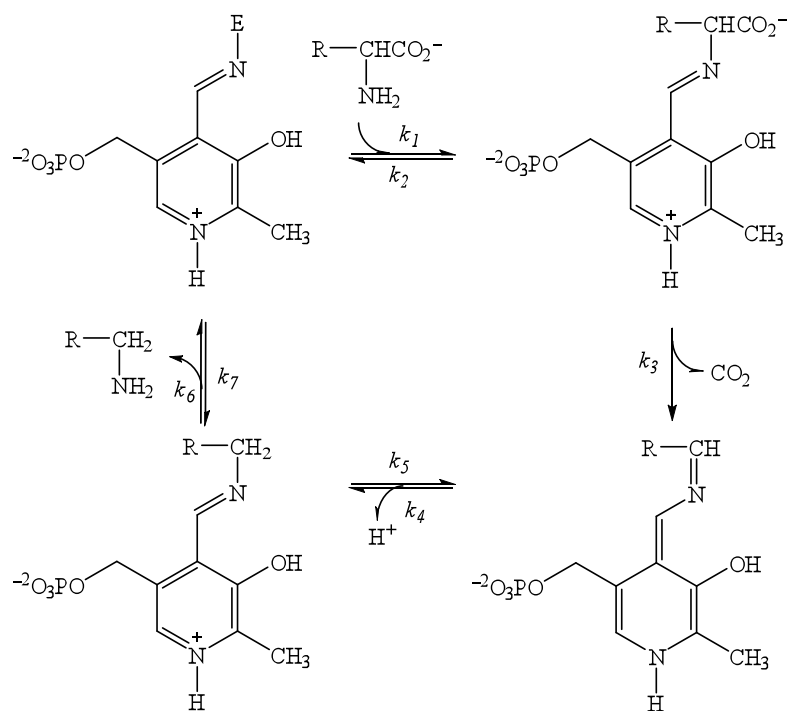
Scheme 5.1   Mechanism of the enzymatic decarboxylation of L-dopa by the pyridoxal 5′-phosphate-dependent dopa decarboxylase.

decarboxylase, with the cofactor Schiff base both in the oxoenamino (keto) and hydroxyimino (enol) conformation. In the centroid path integral simulations, we adopt a combined quantum mechanical and molecular mechanical (QM/MM) method to represent the potential energy surface (*5.16, 5.17*). Thus, both the electronic structure of the reacting system and the nuclear dynamics are treated quantum mechanically. In the PI-FEP/UM approach (*5.15*), we follow a two-step procedure (*5.18*) in which we first carry out molecular dynamics (MD) simulations to obtain the

classical potential of mean force (PMF) along the reaction coordinate for the L-dopa decarboxylation reaction using umbrella sampling (*5.19*). Then, the nuclear coordinates of atoms associated with the chemical reaction are quantized (Figure 5.1) and the configurations sampled in MD simulations are used in path integral simulations by constraining the centroid position of each quantized nucleus coinciding with its corresponding classical coordinates. In PI-FEP/UM, an important feature that enable accurate determination of the small free energy difference is that the ratio of the quantum partition functions for different isotopes, which yields the KIEs, is determined within a single centroid path-integral simulation by free-energy perturbation from a light isotope mass into a heavier one (*5.15*). This method is unique in that it yields accurate and converged results on computed KIEs, including secondary KIEs that are not possible by other path-integral simulation methods (*5.15, 5.20*). The computational technique has been demonstrated in a serious of studies of chemical reactions in solution and in enzymes (*5.20–5.24*).
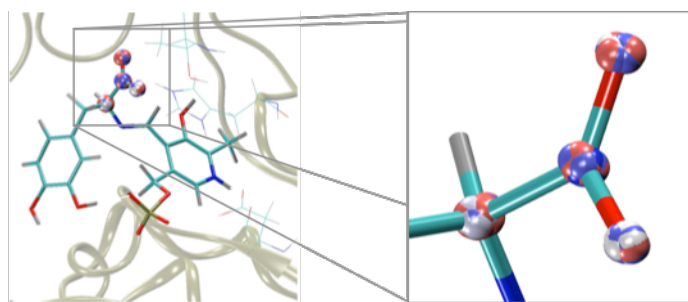
Figure 5.1   Selected quantized atoms (shown in multicolored ball-and-stick representation) in the PI-FEP/UM calculations of the decarboxylation reaction of L-dopa in dopa decarboxylase

116

5.2   Theoretical Background

The theoretical framework of the PI-FEP/UM method is centroid path integral quantum transition state theory (PI-QTST). In PI-QTST, each quantized nucleus is described by a ring of P quasiparticles, and the imaginary time-average position, or called centroid position, of the quantized nucleus, $\bar{r}$, is the geometric centre of the quasiparticles (*5.25*):

$$\bar{r} = \frac{1}{P}\sum_{i=1}^{P} r_i \tag{5.1}$$

where the discrete paths are closed with $r_{P+1} = r_1$. The centroid position, $\bar{r}$, in the centroid path integration is used as a principle variable (*5.26–5.32*), and the canonical quantum mechanical partition function of a hybrid quantum and classical system is written as follows (*5.25*),

$$Q_P^{qm} = \int d\mathbf{S} \int d\mathbf{s} \left(\frac{P}{2\pi\lambda^2}\right)^{3P/2} \int d\mathbf{R}\, e^{-\beta V^{qm}(r,S)} \tag{5.2}$$

where $\int d\mathbf{R} = \int d\mathbf{r}_1 \ldots \int d\mathbf{r}_P\, \delta(\bar{r} = s)$, $\mathbf{S}$ represents the coordinates of all classical particles of the system, $\mathbf{s}$ corresponds to the classical position vector of the quantized particle and it is exchangeable with the centroid position of the particle, $\bar{r}$, defined by eq. (5.1), and the effective quantum mechanical potential $V^{qm}(r,S)$ is given by (*5.25*)

$$V^{qm}(r,S) = \frac{P}{2\beta\lambda^2}\sum_i^P (r_i - r_{i+1})^2 + \frac{1}{P}\sum_i^P U(r_i,S) \tag{5.3}$$

In eq (5.3), λ is the de Broglie thermal wavelength of the particle of mass M:

$$\lambda = \sqrt{\frac{\beta\hbar}{M}} \tag{5.4}$$

117

where $\hbar$ is Planck's constant divided by 2π, and $\beta = 1/k_B T$ with k_B being Boltzmann's constant and T the temperature. Equation (5.3) corresponds to a classical system in which each bead is connected to its two neighbors via a harmonic spring and contributes a fraction of the full classical potential energy of the system, $U(r_i, S)$. Moreover, in this approach the quantum mechanical equilibration properties as well as the free energy are obtained from the classical averages for a system governed by V^{qm}.

In the following discussion, it is convenient to introduce a reference classical partition function defined as follows,

$$Q_P^{cm} = \int dS \int ds \, e^{-\beta U(s,S)} \left(\frac{P}{2\pi\lambda^2}\right)^{3P/2} \int dR \, e^{-\beta\left[-\frac{P}{2\lambda^2}\sum_i^P (r_i - r_{i+1})^2\right]} \tag{5.5}$$

In centroid path integral method, eq. (5.2) can be rewritten related to the reference classical partition function as follows,

$$Q_P^{qm} = Q_P^{cm} \times \frac{\int dS \int ds \, e^{-\beta U(s,S)} \int dR \, e^{-\frac{P}{2\lambda^2}\sum_i^P (r_i - r_{i+1})^2} e^{-\beta \bar{U}(\bar{r},S)}}{\int dS \int ds \, e^{-\beta U(s,S)} \int dR \, e^{-\frac{P}{2\lambda^2}\sum_i^P (r_i - r_{i+1})^2}}$$

$$= e^{-\beta\left[F^{cm} + F_{FP}^{\circ}\right]} \left\langle e^{-\beta\left[F(\bar{r}=s,S) - F_{FP}^{\circ}\right]} \right\rangle_U \tag{5.6}$$

where $e^{-\beta F^{cm}}$ and $e^{-\beta F_{FP}^{\circ}}$ are the centroid path integral free energies of the classical system and the free particle, respectively, the ensemble average $\langle ... \rangle_U$ is a purely classical ensemble average obtained according to the classical potential of the hybrid system, $U(\bar{r}, S)$, and the average differential potential $\Delta U(\bar{r}, S)$ is given by

$$\Delta U(\bar{r}, S) = \frac{1}{P} \sum_{i}^{P} \left[U(r_i, S) - U(\bar{r}, S) \right] \tag{5.7}$$

Importantly, eq. (5.6) can be rewritten exactly as a double average shown by eq. (5.8) (*5.15, 5.18, 5.33, 5.34*):

$$Q_P^{qm} = Q_P^{cm} \left\langle \left\langle e^{-\beta \Delta \bar{U}(\bar{r}, S)} \right\rangle_{FP, \bar{r}} \right\rangle_U \tag{5.8}$$

The inner average $\langle \ldots \rangle_{FP, \bar{r}}$ describes a path-integral free-particle (FP) sampling that is carried out without the external potential $U(\bar{r}, S)$:

$$\langle \ldots \rangle_{FP, \bar{r}} = \frac{\int dR \{\ldots\} \, \delta(\bar{r}) e^{-\frac{P}{2\lambda^2} \sum_i^P (r_i - r_{i+1})^2}}{\int dR \, \delta(\bar{r}) e^{-\frac{P}{2\lambda^2} \sum_i^P (r_i - r_{i+1})^2}} \tag{5.9}$$

where the integration of beads is constrained at the centroid position of the quantized particle, $\bar{r}$ (*5.30, 5.33–5.35*). Thus, the free energy sampling of eq. (5.9) yields the quantum free-energy difference relative to the reference free particle for a fixed classical configuration.

The double average formulation shown in eq. (5.8) is the theoretical basis in the simulation approach of PI-FEP/UM method (*5.15, 5.17, 5.22*) and it has been explored in the simulation approaches of the hybrid classical and path integral by Sprik et al. (*5.18*) as well as that of the quantized classical path (QCP) method by Hwang and Warshel (*5.33, 5.34*). Warshel et al. have further employed the idea of eq. (5.8) in the QCP approach to estimate the NQE for chemical reactions by separating the classical simulations and quantum corrections (*5.33, 5.34, 5.36, 5.37*). Therefore, this double-average procedure in eq. (5.8) is particularly useful because the outer ensemble average $\langle \ldots \rangle_U$ can be obtained by performing classical molecular dynamics

or Monte Carlo (MC) simulations according to the classical distribution $e^{-\beta U(s,S)}$, and then the nuclear quantum effects can be determined through path-integral free-particle samplings based on the free-particle distribution, $e^{-\frac{P}{2\lambda^2}\sum_i^P (r_i - r_{i+1})^2}$, in the absence of external potential. In other words, one can sample the free-particle distribution separately at each classical configuration and then average over all classical configurations obtained from MD or MC simulations.

The centroid path integral potential of mean force (PI PMF), defined as a function of the centroid reaction coordinate, $\bar{z}$, then can be expressed on the basis of eq. (5.8):

$$e^{-\beta w(\bar{z})} = e^{-\beta w_{cm}(\bar{z})} \left\langle \delta(z = \bar{z}) \left\langle e^{-\beta \Delta \bar{U}(\bar{z}[\bar{r}],S)} \right\rangle_{FP,\bar{z}} \right\rangle_U \tag{5.10}$$

where $w(\bar{z})$ and $w_{cm}(\bar{z})$ are the PI and classical PMFs as a function of the centroid reaction coordinate $\bar{z} = z[\bar{r}]$, respectively, and the average difference potential energy $\Delta U(\bar{z}[r],S)$ is given by eq. (5.7).

Another key feature of PI-QTST is that it is expressed analogously to classical TST, as the theory was designed to have this similarity, including a quantum activation term and a dynamical correction factor (*5.26, 5.32, 5.38*). Thus, the PI-QTST reaction rate is given by,

$$k_{QTST} = \frac{1}{2} \left\langle |\dot{z}| \right\rangle_{z^{\neq}} e^{-\beta w(z^{\neq})} \bigg/ \int_{-\infty}^{z^{\neq}} dz\, e^{-\beta w(z)} \tag{5.11}$$

where $z^{\neq}$ is the value of $z[\bar{r}]$ at the maximum of the PMF, $\left\langle |\dot{z}| \right\rangle_{z^{\neq}} = \left(2/\pi\beta M_{eff} \right)^{1/2}$ is a dynamical frequency factor approximated by the velocity for a free particle of effective mass M_{eff} along the reaction coordinate $z[\bar{r}]$ direction, and $w(z)$ is the PMF

as a function of the centroid reaction coordinate $z[\bar{r}]$. As in classical TST, the PI PMF, $w(z)$, can be computed from the equilibrium ensemble average, and it is defined by

$$e^{-\beta[w(z)-w(z_R)]} = e^{-\beta\Delta F(z)} = \frac{\langle\delta(z[\bar{r}]-z)\rangle}{\langle\delta(z[\bar{r}]-z_R)\rangle} \tag{5.12}$$

where z_R is the lowest point at the reaction state in the PMF and the ensemble average $\langle...\rangle$ is determined by the effective potential of eq. (5.3). Equation (5.12) also defines the centroid path integral free energy $\Delta F(z)$ at z relative to that at the reactant state minimum. From eq. (5.12), eq. (5.11) can be rewritten as follows

$$k_{QTST} = \frac{1}{\beta h}e^{\Delta F^{\neq}_{CPI}} \tag{5.13}$$

where the centroid path integral free energy of activation, $\Delta F^{\neq}_{CPI}$, is expressed by

$$\Delta F^{\neq}_{CPI} = \Delta F(z^{\neq}) - F^R_{CPI} \tag{5.14}$$

and

$$F^R_{CPI} = -\frac{1}{\beta}\ln\frac{1}{\lambda_{eff}}\int_{-\infty}^{z^{\neq}}dz\,e^{-\beta\Delta F(z)} \tag{5.15}$$

represents the centroid path integral free energy of the system relative to the lowest point of the reactant state z_R, and it depends on the effective masses associated with the centroid reaction coordinate at the transition state. λ_{eff} is the de Broglie thermal wavelength of the centroid reaction coordinate with an effective mass M_{eff} at the dividing surface, which is determined in the centroid path transition state ensemble.

On the basis of eq. (5.13), the kinetic isotope effects between a light isotope L and a heavy isotope H can be determined by

121

$$KIE = \frac{k_L}{k_H} = e^{-\beta\left[\Delta F_L(\bar{z}_L^{\neq})-\Delta F_H(\bar{z}_H^{\neq})\right]} e^{-\beta\left[F_{CPI,L}^R(\bar{z}_L^R)-F_{CPI,H}^R(\bar{z}_H^R)\right]} \qquad (5.16)$$

where $\bar{z}_L^{\neq}$, $\bar{z}_H^{\neq}$, $\bar{z}_L^R$, and $\bar{z}_H^R$ are the values of the centroid reaction coordinate at the transition state and the reactant state minimum for the light and heavy isotopes, respectively, and $F_{CPI,L}^R(\bar{z}_L^R)$ and $F_{CPI,H}^R(\bar{z}_H^R)$ are the free energies defined by eq. (5.15) for the two isotopes. Equation (5.16) shows that KIEs may be obtained by computing the PMF separately for the light and heavy isotopes (*5.15, 5.22, 5.33, 5.34, 5.39*), and this can be achieved by first performing umbrella sampling using classical MD to obtained classical PMF and then by performing the centroid path integral free-particle samplings in which the centroid position of the quantized particle is constrained to its classical coordinates sampled in the classical trajectory.

In the PI-FEP/UM method, eq. (5.16) is alternatively expressed as the ratios of the partial quantum partition functions for two different isotopes shown as follows (*5.15*):

$$KIE = \frac{k_L}{k_H} = \left[\frac{Q_L^{qm}(\bar{z}_L^{\neq})}{Q_H^{qm}(\bar{z}_H^{\neq})}\right]\left[\frac{Q_H^{qm}(\bar{z}_H^R)}{Q_L^{qm}(\bar{z}_L^R)}\right] e^{-\beta\left[F_{CPI,L}^R(\bar{z}_L^R)-F_{CPI,H}^R(\bar{z}_H^R)\right]} \qquad (5.17)$$

And the ratio of the two isotopic partial quantum partition functions at a given centroid reaction coordinate value $\bar{z}$ is obtained directly through FEP method over the mass from light to heavy isotopes in one simulation:

$$\frac{Q_H^{qm}(\bar{z})}{Q_L^{qm}(\bar{z})} = \frac{\left\langle \delta(z-\bar{z})\left\langle e^{-\frac{\beta}{P}\sum_i \Delta U_i^{L\to H}} e^{-\beta\Delta\bar{U}_L}\right\rangle_{FP,L}\right\rangle_U}{\left\langle \delta(z-\bar{z})^{-\beta\left[F_L(\bar{z},S)-F_{FP}^{\circ}\right]}\right\rangle} \qquad (5.18)$$

where $\Delta \overline{U}_L$ is defined by eq. (5.7). F_{FP}° is the free energy of the free particle reference state for the quantized particles (*5.15*). $\Delta U_i^{L \to H} = U(\mathbf{r}_{i,H}) - U(\mathbf{r}_{i,L})$ corresponds to the difference in classical potential energy at the light and heavy bead positions, $\mathbf{r}_{i,L}$ and $\mathbf{r}_{i,H}$.

Considering an atom transfer reaction where the light atom of mass M_L is replaced by a heavier isotope of mass M_H, a bisection sampling method coupled with the QCP approach (*5.33, 5.40*), called BQCP (*5.17, 5.22*), is employed to compute the light and heavy bead positions by using exactly the same sequence of random numbers:

$$\frac{\mathbf{r}_{i,L}}{\mathbf{r}_{i,H}} = \frac{\lambda_{M_L} \theta_i}{\lambda_{M_H} \theta_i} = \sqrt{\frac{M_H}{M_L}} \qquad i = 1, 2, \dots, P \qquad (5.19)$$

where the position vector θ_i is properly scaled and generated randomly according to the free-particle distribution. Consequently, the positions of all beads are translated such that the geometric center of the polymer ring $\bar{r} = \lambda_M \left(1/P \sum_i^P \theta_i \right)$ is made to coincide with the coordinate of the classical particle. Note that the resulting coordinates of these two bead distributions are dependent on the ratio of the corresponding particle masses via λ_M. The BQCP method samples the free-particle distribution based on the combination of the multilevel sampling (*5.41*) and Lévy Brownian bridge construction (*5.42*). The bisection method takes advantage that the exact free particle distribution is known and path integral configurations can be sampled independently, yielding rapid convergence (*5.17, 5.22*) which is a major problem for application to enzymes. Taken altogether, the free energy (inner average)

123

difference between the heavy and light atoms is obtained by eq. (5.18) by carrying out the bisection path integral sampling with light isotope and then perturbing the heavy isotope position according to eq. (5.19). Then, the free energy difference between the light and heavy isotope ensembles is weighted by a Boltzmann factor for each quantized configuration (outer average).

5.3   Computational Details

Potential Energy Function. We used a combined QM/MM potential in MD simulations (*5.43–5.46*) to study the decarboxylation step in dopa decarboxylase catalysis. The QM part was described by the semiempirical AM1 Hamiltonian (*5.47*), and the MM region was modeled by the CHARMM22 all-atom force field (*5.48–5.50*) for biopolymers and by TIP3P (*5.51*) model for water. The preparations of the system setup, the details of the dynamic methodology and MD umbrella sampling simulations have been described in Chapter 4, and thus only a brief description with emphasis on aspects specific to the present calculations is presented here.

We employed PI-FEP/UM approach (*5.15, 5.17, 5.22*) to evaluate the correction for the nuclear quantum effects on the decarboxylation reaction of L-dopa that is not included in the PMF obtained from the classical MD simulations. Two decarboxylation reactions catalyzed by dopa decarboxylase (DCP<u>H</u> and DCP<u>O</u>, see Computational Methods in Chapter 4 for details) in addition to their corresponding reactions in aqueous solution (aqPH and aqPO) are studied in the present work. The former enzymatic reaction represents the decarboxylation of the hydroxyimine PLP aldimine and that of the oxoenamine tautomer for the latter. In essence, the centroid

path integral simulations were carried out for the light isotopic reactions, and the ratio of the partition function between two isotopic reactions were determined by free energy perturbation scheme from the light mass into the heavier one. Consequently, in the first step, the classical PMF profiles for the decarboxylation reactions of PLP−L-dopa aldimine in dopa decarboxylase and in aqueous solution had been determined in Chapter 4, by combining the QM/MM force field, periodic boundary condition (PBC) dynamics scheme, and MD umbrella sampling simulations at 25°C and 1 atm. The reaction coordinate was defined as the distance of the cleaving $C_\alpha - C^{CO_2^-}$ bond. In the next step, the PMFs were quantized. In this step, path-integral free-particle samplings coupled with the bisection sampling scheme (BQCP) and a mass perturbing scheme were utilized for the classical configurations sampled along the classical PMF reaction coordinate to predict the KIEs for the decarboxylation reactions of L-dopa.

To make the analysis tractable, we focus on two pivotal regions along the reaction pathway for the decarboxylation step, that is the reactant state (RS) and the transition state (TS). Based on the free energy profiles in Figure 5.2, we made the following definitions for the KIE analysis: the RS structures are defined and collected by the reaction coordinates in the range of RC = 1.56 ± 0.3 Å, 1.55 ± 0.3 Å, 1.54 ± 0.3 Å, and 1.54 ± 0.3 Å, and the TS structures by RC = 2.17 ± 0.3 Å, 2.07 ± 0.3 Å, 2.17 ± 0.3 Å, and 2.06 ± 0.3 Å, for the ketoimine or enolamine PLP aldimine in dopa decarboxylase or in aqueous solution, respectively. By these criteria, BQCP employed 174 and 124 classical RS configurations, and 187 and 177 TS configurations, respectively, for the KIEs for the enzymatic decarboxylation of

ketoimine or enolamine PLP aldimine, combined with 200 path-integral steps, plus 100 PI integration steps for equilibration (although this is unnecessary in the PI-FEP/UM algorithm), per classical step. For the reactions in aqueous solution, 499 and 552 classical RS configurations, and 392 and 222 TS coordinates were used for each isotope for the reaction of ketoimine or enolamine PLP aldimine, combined with 200 path-integral steps (and 100 steps of equilibration) per classical step. The PI-FEP/UM results of the configurations sorted in the RS and TS bins are averaged, respectively.

Three KIEs for the decarboxylation of L-dopa were studied in the present work. In the first case (case 1 in Scheme 5.2), the primary ^{13}C KIEs of the carboxylate carbon atom $C^{CO_2^-}$ for the decarboxylation reaction are investigated, where the $C^{CO_2^-}$ atom is labeled as isotope and is quantized in addition to its three neighboring atoms (C_α atom and the oxygen atoms of the carboxylate group) in the PI simulations. Analogously, case 2 corresponds to the primary ^{13}C KIEs of the C_α atom for the decarboxylation reaction, where the C_α atom is isotopic labeled and C_α, $C^{CO_2^-}$, H_α, and the amide N atoms are quantized. We have investigated the effect of including C_β atom in the quantized particle list, and found that it does not make major contributions to alter the calculated KIE for the C_α atom (see Table 5.3). Case 3 illustrates the secondary deuterium KIEs of the H_α atom on the decarboxylation reaction, thus the secondary hydrogen H_α is quantized in addition to its three neighboring atoms, those are C_α, $C^{CO_2^-}$, and the amide N atoms. Each quantized atom was described by 32 beads (in blue dashed circle in Scheme 5.2). Previous studies

have shown that this treatment yields good convergence in the overall quantum corrections for modeling chemical reactions in aqueous solution (*5.22*). To estimate the standard errors in the computed KIEs, the entire path integral simulations were divided into 10 separated blocks, each treated independently. The standard uncertainties were determined from these 10 block averages.

All simulations were carried out using the CHARMM program (*5.49, 5.50*), in which the methods described here have been implemented, and all path-integral simulations used a parallel version that efficiently distributes integral calculations for the quantized beads (*5.22*).
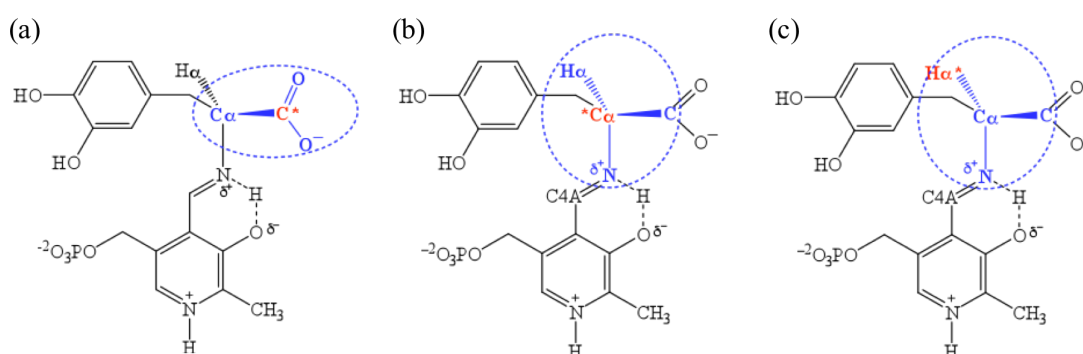
Scheme 5.2   Isotope-labeled atom (in red) and quantized atoms (in blue) used for the KIE calculations in (a) case 1, (b) case 2, and (c) case 3 of the decarboxylation reaction of PLP−L-dopa aldimine.

5.4   Results and Discussion

We have studied the NQEs on the decarboxylation reaction of L-dopa in dopa decarboxylase using PI-FEP/UM method, in which both the potential energy surface

(PES) for the chemical process as well as the NQEs are treated quantum-mechanically. To represent the PES, we used a combined QM/MM potential, in which the external aldimine complex of PLP and L-dopa is treated by the AM1 method. As it turns out, the AM1 model performs exceptionally well for decarboxylation reactions, including the present L-dopa decarboxylation (*5.15, 5.52, 5.53*) (see Chapter 4), and it is not necessary to further adjust the original QM model. To describe the NQEs, we employed a centroid path-integral approach that couples umbrella sampling for the determination of the potential of mean force and free energy perturbation between heavy and light isotopes (*5.15, 5.17, 5.20, 5.22, 5.24*). The two simulation steps corresponding to configuration sampling by molecular dynamics followed by centroid PI simulations to determine NQEs yields the QM-PMF along the centroid reaction coordinate (*5.15, 5.17, 5.20, 5.22, 5.24*). In the following, we first determine the quasiclassical PMFs both for the DDC-catalyzed and the corresponding uncatalyzed process in aqueous solution with PLP either in the oxoenamine (keto) or hydroxyimine (enol) tautomeric structures. The NQEs on the decarboxylation reaction profiles is examined by comparing the two 1° ^{13}C and the 2° ^{2}H$_\alpha$ KIEs with the transition- and reactant-state structures.

Potential of Mean Force. The computed PMFs for the DDC-catalyzed and the corresponding uncatalyzed PLP−L-dopa external aldimine in aqueous solution are depicted in Figure 5.2, and the classical free energies of activation ($\Delta G_{cm}^{\neq}$), quantum-mechanical free energies of activation ($\Delta G_{qm}^{\neq}$), and the phenomenological free energy that lowers the classical barrier due to the NQEs are listed in Table 5.1. The key geometric results at the reactant and transition states are reported in Table 5.2.

The enzyme lowers the free energy barrier by about 8 (6) kcal/mol with the PLP cofactor in the hydroxyimine (oxoenamine) configuration (Figure 5.2). A distinctive feature in Figure 5.2 is that the transition state is earlier for the hydroxyimine PLP aldimine than that for the oxoenamine isomer, both in the enzymatic and in the uncatalyzed reactions. This is reflected by the $C_\alpha - C^{CO_2^-}$ distance in the transition state and the concomitant change in the $O - C^{CO_2^-} - O$ angel of the carboxylate group (Table 5.2). With the inclusion of NQEs in the QM-PMF simulations, the computed free energies of activation for the ^{12}C decarboxylation reaction of the hydroxyimine (oxoenamine) PLP aldimines in the enzyme is 0.64 (1.18) kcal/mol lower than the corresponding "classical" barrier (Table 5.1), resulting in a computed barrier height of 17.8 (18.9) kcal/mol that is in agreement with the experimental value of 16.0 kcal/mol (*5.7, 5.54–5.56*). The computed NQEs for the oxoenamine isomer is significantly greater than that for the hydroxyimine case; the origin of this difference is not yet clear at this time. The decarboxylation of the external aldimine in aqueous solution in the absence of an enzyme, is known to be slow (*5.57*), which takes place at temperature above 100 °C. Note that the barrier for the decarboxylation of the amino acid L-dopa itself is exceedingly high in water, consistent with Wolfenden's experimental data (ca. 40 kcal/mol) on similar systems (*5.58*) (see reaction aqDOPA in Chapter 4). We found the NQEs in the enzymatic reaction are slightly enhanced over the unenzymatic reactions in aqueous solution that reduce the classical barrier by 0.61 (0.60) kcal/mol for the ^{12}C decarboxylation reaction in the hydroxyimino (oxoenamino) tautomer configuration. The results emphasize the importance of including NQEs in enzyme kinetics modeling to obtain

accurate rate constants (*5.3, 5.59*), although the dominant factor for the enzymatic rate enhancement is due to lowering the "classical" free energy barrier in the active site of the enzyme. In Chapter 4, it was shown that the exclusion of solvent molecules in the active site by loop closure is critical to the enzymatic rate enhancement, which may be attributed to reduced stabilization of the carboxylate group due to desolvation in comparison with the loop-open structure where solvent stabilization of the carboxylate group in the reactant state is significant as in bulk solution (Figure 5.2). In this Chapter, we focus the study on NQEs in catalysis, which has not been quantified previously.

Table 5.1   Computed classical mechanical ($\Delta G_{cm}^{\neq}$) and quantum mechanical ($\Delta G_{qm}^{\neq}$) free energies of activation and total nuclear quantum effects ($\Delta\Delta G_{qm}^{\neq}$) for the ^{12}C and ^{13}C decarboxylation reactions of PLP−L-dopa aldimine complex in aqueous solution and in dopa decarboxylase [a]

reaction	$\Delta G_{cm}^{\neq}$	$\Delta G_{qm}^{\neq}(^{12}C)$	$\Delta G_{qm}^{\neq}(^{13}C)$	$\Delta\Delta G_{qm}^{\neq}(^{12}C)$	$\Delta\Delta G_{qm}^{\neq}(^{13}C)$
aqPH	26.1	25.49	25.52	−0.61	−0.58
aqPO	25.9	25.34	25.33	−0.60	−0.57
DCPH	18.4	17.79	17.82	−0.64	−0.61
DCPO	20.1	18.88	18.91	−1.18	−1.15
expt.		16.1−16.5 [b-e]			

[a] Free energies are given in kcal/mol. Two decimal digits are reported with NQE contributions in view of kinetic isotope effect calculations. [b] Corresponding to k_{cat} value of 8 s^{-1} from Ref. (*5.7*). [c] Corresponding to k_{cat} value of 4.3 s^{-1} from Ref. (*5.56*). [d] Corresponding to k_{cat} value of 9.1 s^{-1} from Ref. (*5.54*). [e] Corresponding to k_{cat} value of 5.0 s^{-1} from Ref. (*5.55*).
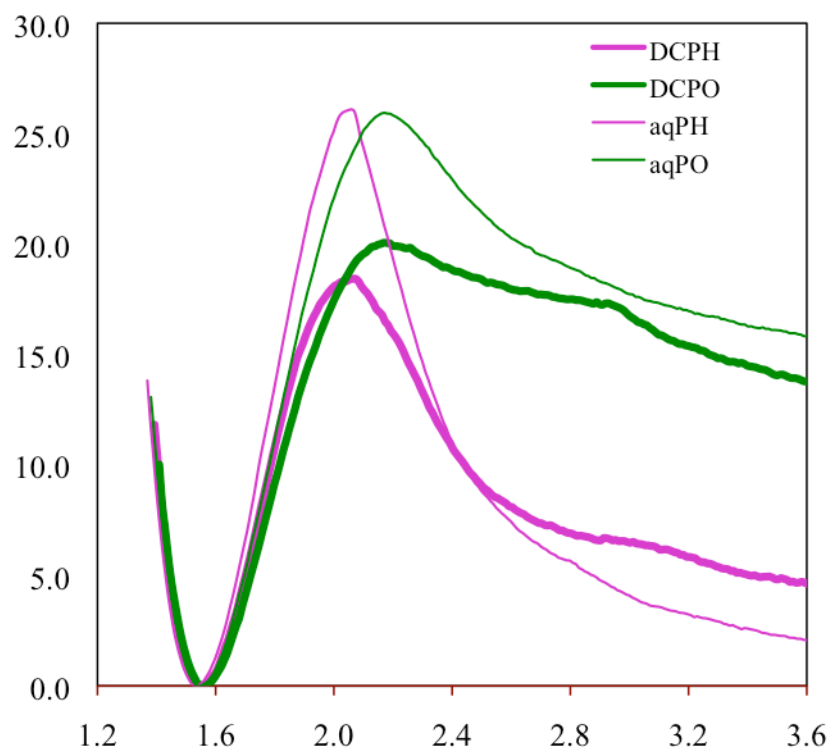
Figure 5.2 Computed free energy profiles for the decarboxylation reactions of PLP−L-dopa in aqueous solution (dashed lines) or catalyzed by dopa decarboxylase (solid lines). The PLP aldimine complex with L-dopa is either in enolamine tautomeric form (*i.e.* hydroxyimine, displayed in color of magenta) or ketoimine isomeric structure (*i.e.* oxoenamine, displayed in color of green).

Table 5.2 Selected Average Bond Distances, Angles, and Dihedral Angles Within PLP–Aldimine Cofactor at the Reactant

States (RS) and the Transition States (TS) of aqPH, aqPO, DCPH, and DCPO

reaction	aqPH		aqPO		DCPH		DCPO	
	RS	TS	RS	TS	RS	TS	RS	TS
Bonds (Å) [a]								
$C_\alpha - C^{CO_2^-}$	1.54	2.06	1.54	2.17	1.55	2.07	1.56	2.17
$C_\alpha - N$	1.45	1.34	1.46	1.39	1.45	1.36	1.46	1.39
$N - C4A$	1.28	1.34	1.43	1.32	1.29	1.32	1.32	1.32
Angle (degree) [a]								
$O - C^{CO_2^-} - O$	119.7	150.3	120.6	154.2	122.9	148.6	122.9	153.8
$H_\alpha - C_\alpha - N$	105.2	111.8	105.3	112.3	106.3	118.9	105.1	112.2
$H_\alpha - C_\alpha - C_\beta$	107.9	114.0	109.7	116.0	107.8	114.3	109.3	115.6
$C_\beta - C_\alpha - N$	110.9	126.3	114.0	122.6	113.5	118.2	118.1	123.8
Dihedral Angle (degree) [a]								
$H_\alpha - C_\alpha - N - C_\beta$	−116.5	−147.8	−120.4	−146.6	−118.4	−146.7	−112.2	−147.4

[a] Numbering given in Figure 4.2.

Kinetic Isotope Effects. The computed 1° and 2° KIEs for the decarboxylation reaction of L-dopa in water and in the enzyme are listed in Table 5.3, while the computed NQEs and KIEs from the individual blocks are documented in Table 5.4. The standard errors in ±1s deviation are estimated from these individual block averages with respect to the full average (note that the total computed average KIE is not a simple mean of the individual block-averages because of the configuration integral is a sum of the exponential function of energy difference over configurations, see Appendix C). From path-integral and free-energy perturbation simulations, we found that, the hydroxyimine PLP complex has somewhat greater NQEs on the enzymatic reaction than the aqueous-phase reaction, consistent with the free energy results shown in Table 5.1. Contrary to the hydroxyimine complex, the oxoenamine isomer reflects slightly smaller NQEs on the enzymatic reaction relative to the corresponding reaction in water. Overall, by comparing the enzymatic and the model systems in water solution that mimics exactly the same enzymatic reaction and mechanism, the calculated primary C13-KIEs in the DDC-catalyzed decarboxylation have approximately the same magnitude as those observed in the non-enzymatic decarboxylations considering statistical uncertainties. Evidently, the enzyme does not operate to significantly alter the size of the KIEs, and thus, NQEs. In the following, three KIE cases (see Computational Details) of the decarboxylation reaction are considered, including (*i*) the primary $^{12}C/^{13}C$ KIE of the carboxylate $C^{CO_2^-}$ atom, (*ii*) the primary $^{12}C/^{13}C$ KIE of the C_α atom, and (*iii*) the secondary H/D KIE of the H_α atom.

Table 5.3 Computed 1° and 2° KIEs for the decarboxylation reactions of L-dopa in PLP-dependent dopa decarboxylase and in PLP-catalyzed aqueous solution

case: quantized atom [a]	KIE	aqPH	aqPO	DCPH	DCPO	expt. [b]
case 1: $C^{CO_2^-}$	$^{12}k/^{13}k$	1.0529 ± 0.0026	1.0498 ± 0.0024	1.0554 ± 0.0028	1.0461 ± 0.0049	1.05–1.06
case 2: C_α	$^{12}k/^{13}k$	1.0298 ± 0.0031	1.0340 ± 0.0026	1.0390 ± 0.0056	1.0283 ± 0.0064	
				1.0401 [c] ± 0.0054	1.0282 [c] ± 0.0059	
case 3: H_α	$^{1}k/^{2}k$	0.9746 ± 0.0509	1.1368 ± 0.0513	1.0622 ± 0.0366	1.1266 ± 0.0528	

[a] Numbering given in Figure 4.2. [b] From Ref. (*5.6*). [c] Including the data averaged with the C_β atom included.

The finding of larger $^{13}C^{CO_2^-}$-KIEs for the decarboxylation reaction (case 1 in Table 5.3) in the hydroxyimine PLP–L-dopa aldimine than that in the oxoenamine PLP aldimine suggests that there are differential NQEs between these two tautomeric isomers. A diagnostic analysis of the concomitant change in the $O-C^{CO_2^-}-O$ angel of the carboxylate group from the reactant state to the transition state sheds light on the KIE results on the reactions in the enzyme as well as in aqueous solution. The difference in the calculated carboxyl $^{13}C^{CO_2^-}$-KIEs in Table 5.3 is closely related to the small, but significant, variations in the $O-C^{CO_2^-}-O$ bond angle. The average bond angles are close to 150°, about mid-way from the reactant to the product states,

consistent with the estimated Pauling bond order of about 0.4 for the $C_\alpha - C^{CO_2^-}$ bond

at the transition state (computed using c = 0.3). The transition state of the

hydroxyimine tautomer in the active site has smallest $O - C^{CO_2^-} - O$ bond angle of the

four decarboxylation reactions, including an early transition state with the least gain

in bond order of the CO_2 leaving group. On the other hand, the oxoenamine isomer

has the largest $O - C^{CO_2^-} - O$ angle, indicative of enhanced force constant of the CO_2

stretch and a late transition state. Thus, for the oxoenamine aldimine, the reduction in

the difference of zero-point energy (ZPE) (between ^{12}C and ^{13}C) associated with the

$C_\alpha - C^{CO_2^-}$ bond stretch is accompanied by a gain in the $O - C^{CO_2^-} - O$ stretching

frequency, orthogonal to the reaction coordinate, in going from the reactant state to

the transition state. The two compensating factors yield a relatively small computed

KIE in the oxoenamine isomer of the cofactor. However, the change of the

$O - C^{CO_2^-} - O$ angle is smaller for the hydroxyimine configuration, resulting in a

greater KIE at the carboxyl site. In addition, these $^{13}C^{CO_2^-}$-KIE results fit nicely into

the current interpretation of PLP-dependent enzymatic decarboxylations, for which

the intrinsic ^{13}C KIEs are believed to be in the range of 1.05–1.06 (*5.6*). Values in this

range are also found for nonenzymatic decarboxylations (*5.60*). The agreement

between the experimentally observed values and the predicted intrinsic ^{13}C KIEs for

the enzymatic reactions suggests that the computational approach in the present study

is reasonable. It would be interesting to test the present theoretical prediction of the

intrinsic ^{13}C KIE for the decarboxylation reaction catalyzed by dopa decarboxylase,

and to elucidate whether or not the decarboxylation step is solely rate-determining such that the intrinsic value is fully expressed in the experimentally observed KIEs.

The rehybridization at the C_α position from an sp^3 to an sp^2 center along the decarboxylation path is reflected by the primary $^{13}C_\alpha$- and secondary ^{2}H-KIEs in aqueous-phase reactions. In water, the $^{12}k/^{13}k$ effects of C_α are most sensitively associated with the development of the $C_\alpha = N$ double bond character as indicated by the $C_\alpha - N$ bond distance in Table 5.2. Consequently, a shorter $C_\alpha = N$ distance corresponds to greater sp^2 character and a larger force constant of bond stretching, giving rise to increased ZPE difference in the transition state between $^{13}C_\alpha$ and $^{12}C_\alpha$ isotopic reactions due to the $C_\alpha = N$ stretching mode, which compensates for the loss in the ZPE difference from the reaction coordinate motion ($C_\alpha - C$ stretch). The combined results yield a smaller $^{13}C_\alpha$ KIE, and the trend is mirrored in Table 5.3. On the other hand, the deuteron $^1k/^2k$ (H_α) KIE for the decarboxylations in water solution is directly coupled to the out-of-plane bending mode about the $C_\alpha = N$ π-plane that is being developed at the transition state, which corresponding to a change from about 1400 cm^{-1} for an sp^3 C–H deformation mode in the reactant state to about 800 cm^{-1} for an sp^2 out-of plane bending mode in the product. Thus, the smaller deformation angles (out-of-plane torsion, e.g. the improper dihedral angle $\phi_{H_\alpha - C_\alpha - N - C_\beta}$ in Table 5.2) in the reaction of the oxoenamine aldimine isomer in aqueous solution leads to relatively larger 2° KIE than the reaction of the hydroxyimine isomer in aqueous solution. In the enzyme, however, both $^{13}C_\alpha$- and $^2H_\alpha$-KIEs are not fully explained by the trends of the $C_\alpha - N$ bond distance and the $H_\alpha - C_\alpha - N - C_\beta$ improper dihedral

136

angle at the transition states, respectively. It could be conceivable that the enzyme might provide some limited steric compression, leading to a change in bending vibrations in the quinonoid-like transition state aldimine complexes. Further calculations are required to analyze the intrinsic origins of the difference in the KIEs for the enzymatic reaction.

5.5   Discussion

We have predicted the primary carbon and secondary deuterium kinetic isotope effects for the PLP-dependent decarboxylation reaction of L-dopa in dopa decarboxylase as well as in water using am integrated path integral and free-energy perturbation-umbrella sampling method. Calculations show that there is a small contribution from nuclear quantum effects to the enzymatic rate enhancement in the PLP-dependent L-dopa decarboxylation, although the effects is negligible in comparison with the effect of classical barrier reduction. The calculated primary ^{13}C KIE of the carboxyl carbon atom in the enzymatic reaction is in agreement with the experimental results from other PLP-dependent decarboxylation reactions. Furthermore, the difference in the tautomeric structure of PLP−L-dopa aldimine leads to distinctive KIE results. Particularly, there is a distinguishing difference in the secondary $^{2}H_{\alpha}$ KIEs between the hydroxyimine and oxoenamine tautomers in water, where the former shows an inverse secondary KIE, whereas the latter exhibits a normal secondary KIE. Overall, the carboxylate carbon atom shows somewhat larger KIEs than the C_{α} atom, and the hydroxyimine tautomer of the PLP cofactor has greater primary carbon KIEs in both positions than the oxoenamine isomer in the

enzyme. The computed ^{13}C carboxyl KIE for the hydroxyimine substrate is within the range of the intrinsic effects expected for similar reactions, although the carboxyl ^{13}C-KIE for the oxoenamine tautomer is lightly below the normal range (but within statistical errors). It would be interesting to see if experimental measurements of all three (two primary carbon-13 and one secondary deuterium) KIEs can be used along with the computational analyses to distinguish the active form of tautomerism of the cofactor.

Chapter 6

Dynamics of Histone Lysine Demethylase LSD1 in Complex With CoREST and the H3 Tail Peptide

6.1   Introduction

Lysine-specific demethylase (LSD1), a flavin adenine dinucleotide (FAD)-dependent amine oxidase, is associated with gene expression through its capability of protein lysine modifications (*6.1−6.6*). LSD1 shows both coactivator and corepressor functions modulated by its interacting factors (*6.7−6.9*). It is relevant to the gene activation of the androgen receptor (AR) by interacting with AR to demethylate Lys9 on histone H3 peptide of nucleosomes (H3K9) (*6.10, 6.11*). It is also linked to the neuronal-specific gene repression in nonneuronal cells with the participation of CoREST, a SANT-domain-containing transcriptional corepressor, to target histone

H3 demethylation at Lys4 (H3K4) within nucleosomes (*6.7, 6.12−6.18*). Surprisingly, it has been recently discovered that LSD1 represses p53-mediated apoptosis by demethylating methylated mark on p53 at Lys370 to inhibit the interaction of p53 with DNA (*6.19*). These findings demonstrate that gene regulatory activities of LSD1 associated with its substrate specificity can be fine tuned through different binding factors (*6.7*).

Crystal structures of LSD1, either in the native state or in complex with CoREST, have been determined (*6.20−6.24*). LSD1 comprises three distinct structural domains: an N-terminal SWIRM domain, an C-terminal AOL (amine oxidase-like) domain, and a protruding Tower domain. The SWIRM domain (residues 172−270), consisting of a helix-turn-helix compact fold, has been found in a number of chromatin-associated proteins and was implicated to have a function of DNA binding (*6.25−6.27*). In addition, it interacts with the AOL domain through a hydrophobic interface, maintaining structural stability of the AOL domain as well as the catalytic activity of LSD1 toward substrate binding (*6.21, 6.22*). The AOL domain, adopting an amine oxidase fold, consists of a FAD-binding subdomain (AOL1: residues 271−417) and a substrate-binding subdomain (AOL2: residues 523−833), separated by the insert sequence comprising the Tower domain. The negatively-charged active-site pocket of LSD1 is located at the interface of the two (FAD-binding and substrate-binding) subdomains and goes deep into the protein interior where the flavin ring of FAD is located. The binding cavity is capacious and opening-wide relative to other FAD-dependent enzymes, indicating the capacity to accommodate diverse substrates (*6.21, 6.22*). In addition, extensive mutagenesis

studies on the active-site residues of LSD1 showed that the LSD1 demethylation activity is either reduced or abolished, indicating the importance of the AOL domain in substrate recognition and in catalysis (*6.20, 6.21, 6.28*).

The Tower domain (residues 418–522) contains two elongated antiparallel-orientated a helices that connect the two subdomains of the AOL lobe. It presents a surface for the binding of CoREST that is critical to mediate the demethylation of H3K4 through association with nucleosomal substrates (*6.7, 6.18, 6.22, 6.23, 6.29*). It also has been suggested that binding of the Tower domain to different proteins is critical for the capability of LSD1 to interact with a variety of targets.

Among LSD1 structures, two LSD1•CoREST complexes have been cocrystallized with a peptide containing the residues of the N-terminal H3 tail in the catalytic cavity of LSD1. One ternary complex structure contains a 21-amino acid H3 peptide in the active site, in which the peptide Lys4 is mutated to a methionine (*6.24*). The other LSD1•CoREST•H3 crystal structure contains a peptide comprising the N-terminal 21 residues of H3 tail with *N*-methylpropargy-Lys at position 4 (*6.23*). These structures reveal spacious active site cavity to accommodate several residues of the histone tail to properly position the substrate lysine residue. Both structures exhibit similar binding patterns, and the peptide substrate makes both intramolecular hydrogen-bonding interactions and intermolecular contacts with LSD1. In addition, mutagenesis studies on both LSD1 and the H3 peptide substrate have identified several crucial residues for the catalytic activity of LSD1 (*6.21, 6.24*).

The present Chapter presents a molecular dynamics (MD) simulation study of the LSD1•CoREST complex complexed with a substrate consisting of 16-amino-acid

residues of the N-terminal histone H3 tail (H3-p16) in which Lys4 is dimethylated. We focus the present investigation on the dynamics properties of Tower domain, including the azimuthal rotation angle and tilting angle of relative to the catalytic domain in order to gain an understanding of the flexibility of the LSD1 Tower domain associated with CoREST. In addition, upon release of crystal constraints, the conformational rearrangement of H3-p16 in the protein-peptide complex was observed in solution and compared with the crystal structure. Overall, by sampling the conformational space of the LSD1•CoREST complex bound to the target H3-p16 peptide, the results of MD simulations in water provide insights into the ternary complex's behavior in solution.

6.2   Computational Details

Initial Setup. The initial structure for the simulation of LSD1 in complex with CoREST and a substrate-like peptide inhibitor, ARTMQTARKSTGGAKAP in the active site of the catalytic domain was taken from the X-ray structure of PDB (*6.30*) entry 2V1D (*6.23*). His259 was protonated due to its side-chain nitrogen atoms were within 3.5 Å of a hydrogen acceptor in the crystal structure. The remaining histidine residues were treated as neutral by protonating at $N^{\delta 1}$ or $N^{\epsilon 2}$ according to their local environment. Thus, His185, 250, 394, 422, 430, 484, 532, 551, 564, 680, and 812 were protonated at $N^{\delta 1}$; His253, 459, and 802 were protonated at $N^{\epsilon 2}$. A bridging water molecule was added to form hydrogen bonds with the N1F atom of the FAD and the N^{ξ} atom of Lys661 (*6.20*). The active site residue Lys661, which plays a role as a general base in the hydride transfer reaction (Scheme 6.1) (*6.20, 6.22*), was

modeled as a neutral residue. Dimethylated Lys4 (methyl-pLys4) on the histone H3-p16 peptide was constructed by modifying the structure of pMet4 in the crystal structure. We treated the remaining titratable residues corresponding to their ionization states at pH 7. Hydrogen atoms were added to the crystal structure using the HBUILD module of the CHARMM program (*6.31, 6.32*). The complex was fully solvated in a box of explicit water molecules with dimensions of $128 \times 171 \times 128$ Å^3. Five Cl$^-$ ions were added to neutralize the system. The final system contains 269709 atoms, 12967 of which belong to the ternary protein complex (Figure 6.1).
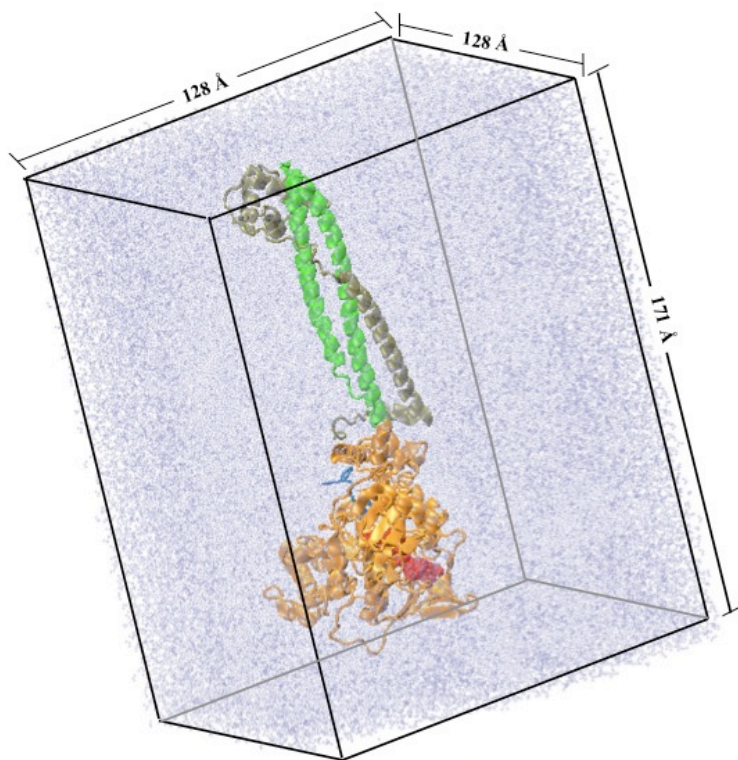
Figure 6.1
A model of LSD1•CoREST complex bound to 16-a.a. peptide of the N-terminal H3 tail (H3-p16) was solvated in a rectangular water box. Tower domain, SWIRM and AOL domains of LSD1, CoREST, and H3-p16 peptide are in a cartoon representation colored in green, orange, silver, and blue, respectively. FAD is in ball model colored in red.
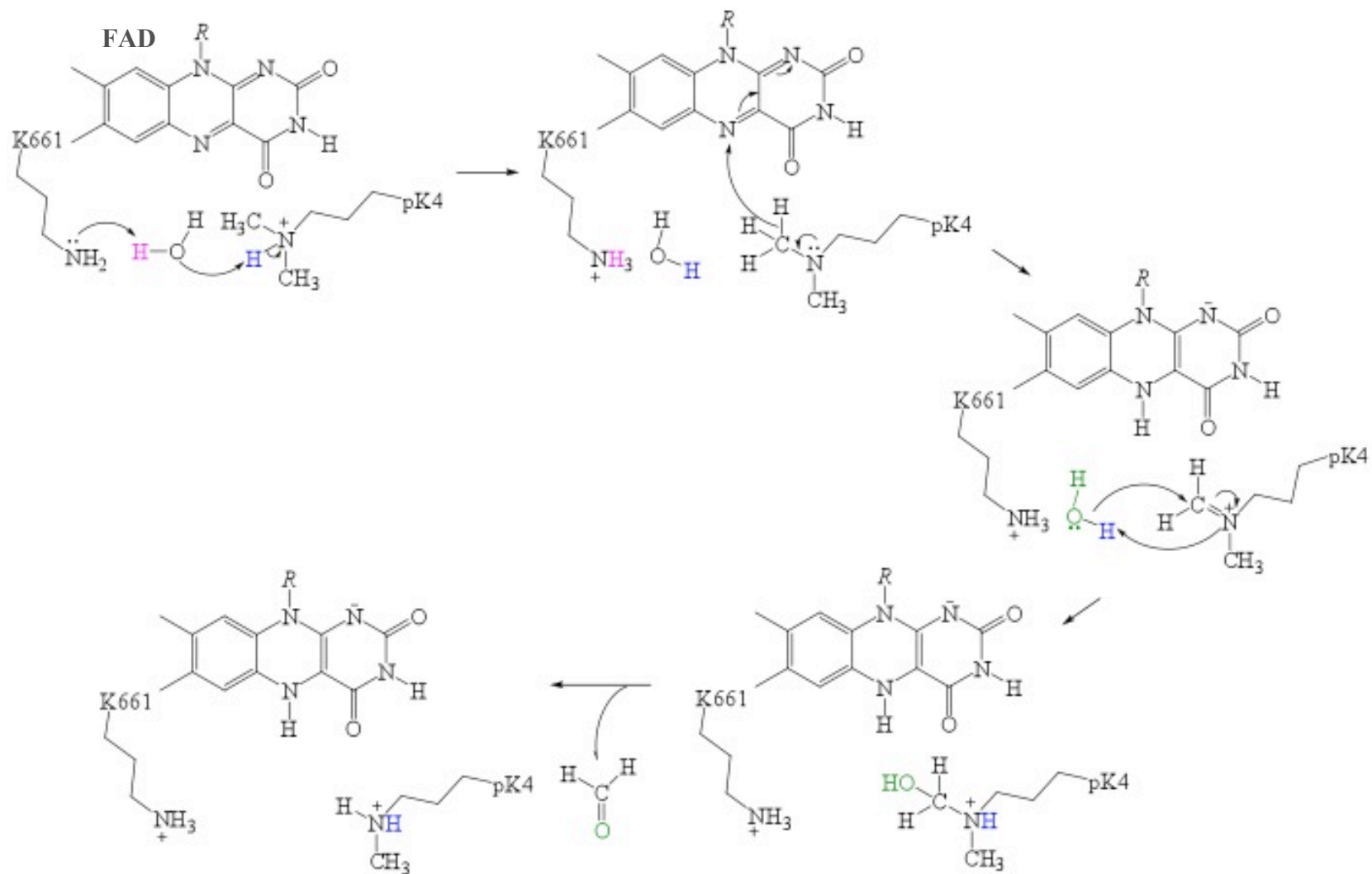
Scheme 6.1 Proposed catalytic mechanism for the demethylation by LSD1

Simulation Protocol. All MD simulations herein were performed using the program NAMD (*6.33*) with the CHARMM27 force field (*6.34*) for protein residues, FAD, and ions along with the TIP3P (*6.35*) model for water molecules. Periodic boundary condition (PBC) and the isothermal-isobaric (NPT) ensemble at 298.15 K and 1 atm were used throughout the MD simulations. The impulse-based Verlet integration scheme was used to integrate the equations of motion with a time step of 1 fs. The van der Waals energies were feathered to zero in the region between 11 Å and 13 Å with a smooth switching function, while long-rang electrostatic interactions were treated by means of the particle-mesh Ewald (PME) method. All bonds involving hydrogen atoms were constrained to their equilibrium distances using the SHAKE algorithm during the MD simulations.

Water molecules, ions, and the dimethylated pLys4 of the resulting structure was initially subjected to 4000 steps of energy minimization using the adopted-basis Newton–Raphson (*6.31*) (ABNR) method, whereas the rest of the system was held constrained. This relieved close contacts between the protein complex and solvent molecules without disrupting its overall conformation. After the minimization, the system was gradually heated from 0 K to 298.15 K within 80 ps of molecular dynamics simulations, which was subsequently equilibrated at 298.15 K for 19.97 ns under constant pressure and temperature conditions. The early part of the simulations ($0 < t < 10$ ns) were treated as the equilibration phase, and the last half of the simulations at $t \geq 10$ ns were employed in structural analysis. Trajectories of the simulation in equilibration were saved in every 5 ps, resulting in a total of 3989 snapshot coordinates.

6.3   Results and Discussion

Overview of MD Simulations. Figure 6.2 shows the time evolutions of the root-mean-square deviations (RMSDs) for the backbone heavy atoms of the full-length LSD1, the AOL and SWIRM domains, the Tower domain, and the substrate H3-p16 peptide, respectively, by best fit to superimpose the backbone heavy atoms of the AOL and SWIRM domains as well as the FAD with respect to the experimental starting structure. During the 20 ns of MD simulations, the RMSD value of the entire backbone heavy atoms of LSD1-CoREST-pH3-16 complex has large fluctuations in a range of 0.9 Å to 8.3 Å. Examination of the RMSD fluctuations of individual domains shows that the dominant fluctuation comes from the Tower domain with a conformational fluctuation over 20 ± 4 Å. On the other hand, the AOL and SWIRM domains exhibits dynamic fluctuations found in typical globular proteins over similar time evolution. The backbone root-mean-square fluctuation (RMSF) in Figure 6.3 for each residue averaged over the entire trajectory provides a different perspective of the local protein flexibility. The RMSF results show that dynamic fluctuations of the Tower domain dominate the overall protein flexibility of the LSD1 ternary complex in aqueous solution (Figure 6.3), whereas the RMS fluctuations of the AOL and SWIRM domains are more stable. Altogether, the time-dependent and time-averaged fluctuations reveal an important feature of the intrinsic flexibility of LSD1, which can be separated into two distinctive dynamic regions: (a) the globular dynamics region consisting of the catalytic AOL domain and the SWIRM domain of LSD1, and (b) the mobile region of the Tower domain that undergoes large amplitude motions.

146
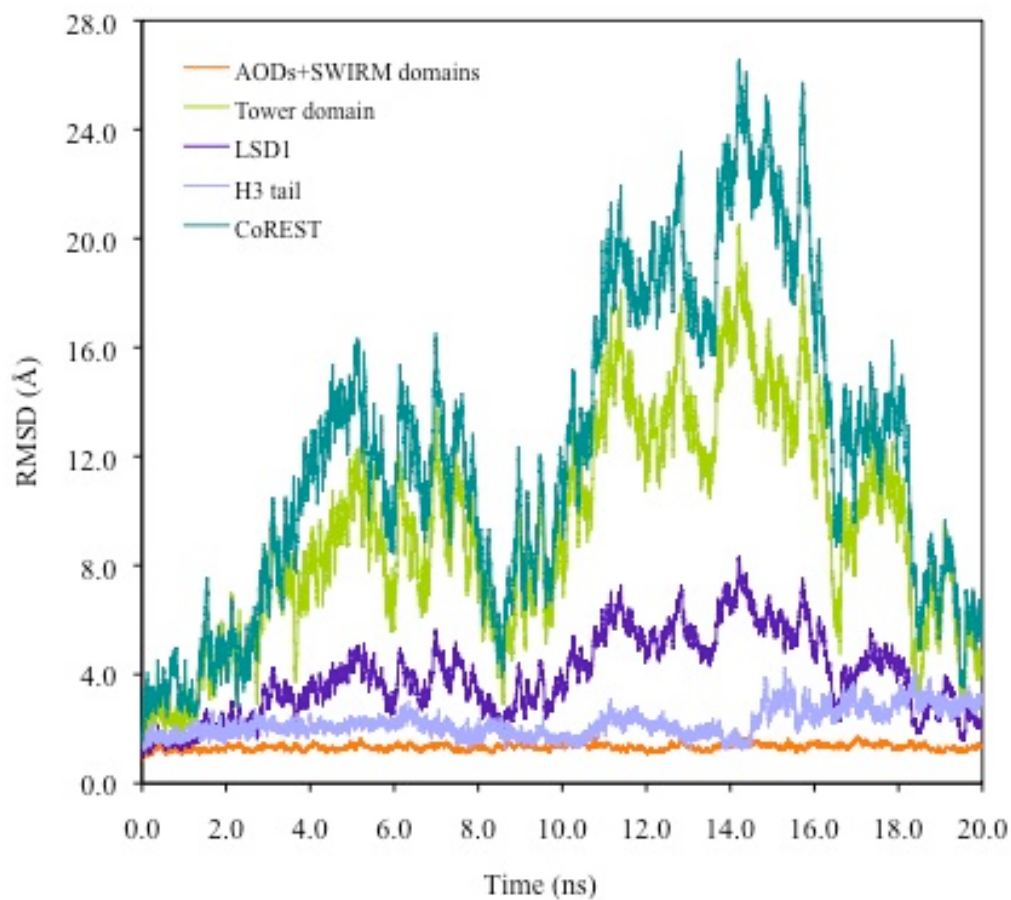
Figure 6.2   RMSD versus simulation time for the backbone atoms of LSD1 relative to the crystal structure. The RMSD of the full-length LSD1, AOL and SWIRM domains, Tower domain, and the N-terminal 16 residues of H3 tail (H3-p16) are colored in violet, orange, green, and light purple, respectively.
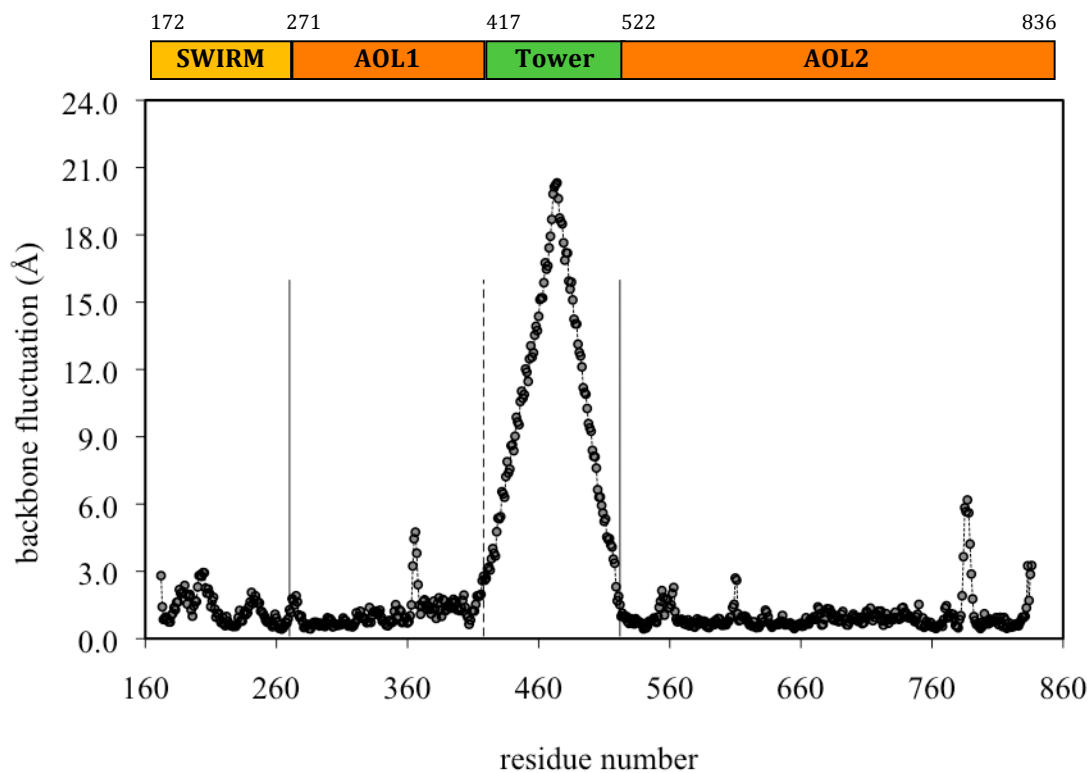
Figure 6.3 Average RMS fluctuation (RMSF) of LSD1 calculated from 19.97 ns of MD simulations

Dynamic Motions of the Tower Domain And CoREST. To characterize the dynamic motions of the Tower domain as well as CoREST, we have examined the fluctuations of the helices relative to the globular region of LSD1•CoREST complex. To this end, we have analyzed the azimuthal, θ, and latitude tilting (colatitudes) angle, φ, fluctuations of the helices vector relative to the fixed frame of the X-ray crystal structure. These angles are defined in Figure 6.4. To calculate the colatitudes and azimuthal angles, we first define the vector of the Tower domain as the average vector direction of the two antiparallel helices, and the helical directional vectors are calculated using the method described by Aqvist (*6.36*) using CHARMM. In the initial 10 ns of MD simulations, the colatitude angle of the Tower domain fluctuates between 0.03° and 13.1° with an average of 6.8 ± 2.6°, which appear to have reached equilibrium (Figure 6.4a). However, after 10 ns, even larger fluctuations took place in the colatitudes motions and the angle increases to 21° before it recovers to the initial values. Clearly, longer simulations are required to fully address the large amplitude motions, but the results presented in Figure 6.4a are sufficient to demonstrate that the Tower domain undergoes rather large colatitudes dynamic fluctuations relative to the globular catalytic domain with a range of 0–21°. Considering the length of the Tower domain helices (~70 Å), the tip of the mobile region traverses more than 8 Å as a result of the colatitude angular fluctuations within 5 ns time scale. The number distribution of the colatitudes angle fluctuations in the last 10 ns of MD simulations is shown in Figure 6.4b (computed with bin size of 1° thus the normalized distribution can be obtained by dividing the total number of configurations used in the analysis), and the average φ angle is 12.5 ± 3.9°.

149

The azimuthal angle of the Tower domain traverses the full range of 360°

during the 20 ns molecular dynamics simulations (Figures 6.4c), clearly illustrating

the dynamics flexibility is not restricted to colatitudes motions. The number

distribution determined from structures collected during the final 10 ns simulations

are displayed in Figure 6.4d, which already covers nearly the entire azimuthal range,

except the area from −20° to −100° that has not been sampled during this period. The

results presented in Figure 6.4 demonstrate that the Tower domain undergoes large

amplitude colatitudal and azimuthal motions in the time scale of 10 to 20 ns, keeping

in mind that the present simulations barely covered one period of these fluctuations. It

would be interesting to investigate the biological implications of these dynamic

fluctuations in the future, particularly on recognition and binding of the Tower

domain to the histone particle. The installation of the Blue Waters Systems at the

NCSA at the University of Illinois in 2011 provides an opportunity to extend the

length of the present dynamics simulations and the possibility to examine the

complex of LSD-CoREST-Histone structures.

It is not surprising that the CoREST protein also experiences large amplitude

motions since it noncovalently interacts with the Tower domain in the LSD-CoREST-

Histone complex. During the 20 ns of MD simulations, the RMSD value of the

CoREST, which follows the motion of the Tower domain, has fluctuations in a range

of 1.5 Å to 26.6 Å (Figure 6.2). Here, we defined the vector of the CoREST helix

(residues 330−362) as the average vector direction of the $C\alpha$ carbon atoms, and the

colatitudal vibration motion (φ) was defined as the angle of the vector of the CoREST

helix relative to the fixed frame of the Tower helix (residues 418−466) of the X-ray

crystal structure. The azimuthal rotation motion (θ) was defined as, setting the X-ray frame of the Tower helix along z axis, the angle of the projection of the CoREST helix vector to x axis (similar as shown in Figure 6.4). The analysis of motion of CoREST helix show that CoREST helix domain undergoes large amplitude colatitudal motion (φ) in the time scale of 10 to 20 ns within a range from 8° to 32° (Figure 6.5b), but has smaller amplitude azimuthal motion (θ) within a range from −40° to −120°. The latter is attributed to the manner that CoREST helix domain is not parallel with Tower helices but maintains an interhelical crossing angle (Ω) of 10°−25° with Tower helices during the entire simulations (Figure 6.7). In addition, the dynamic motion of CoREST SANT2 domain (residues 385−440) is analyzed to shed light on the ability of conformational mobility in liquid phase (Figure 6.8). The vibration motion (φ) and the azimuthal motion (θ) of CoREST SANT2 domain were described, similarly as shown in Figure 6.4, as the vector defined from Cα atom of residue 418 to the center of mass of Cα atoms of CoREST SANT2 domain (residues 385−440) relative to the vector of the Tower helix (residues 418−466) of the X-ray crystal structure. The analyses reveal that CoREST SANT2 domain experiences similar amplitude of colatitudal vibration motion (φ) with CoREST helix domain but occurs larger amplitude of azimuthal rotation motion (θ).
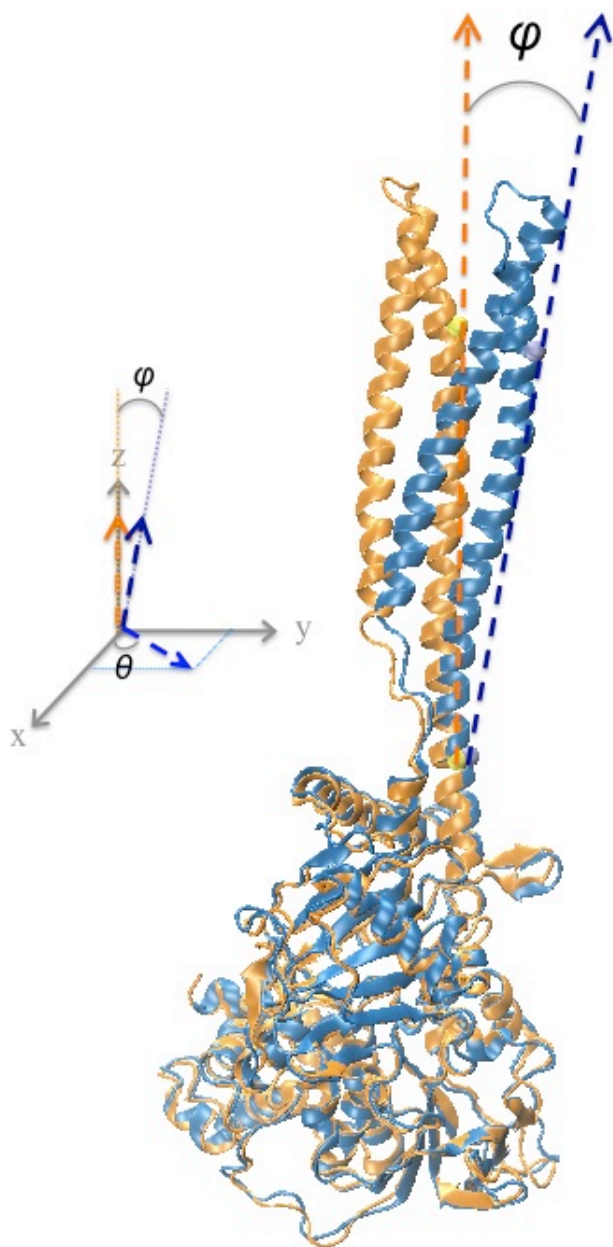
Figure 6.4

(a) Time evolution for the entire Tower domain during entire MD simulations and (b) distribution of the vibration angle (φ) in the last 10 ns simulations. (c) Time evolution for the entire Tower domain during entire MD simulations and (d) distribution of the azimuthal rotation angle (θ) in the last 10 ns simulations. Descriptions of vibration angle, φ, and rotation angle, θ, are shown in the inset. The experimental structure and a snapshot of the simulation are colored in orange and blue, respectively.
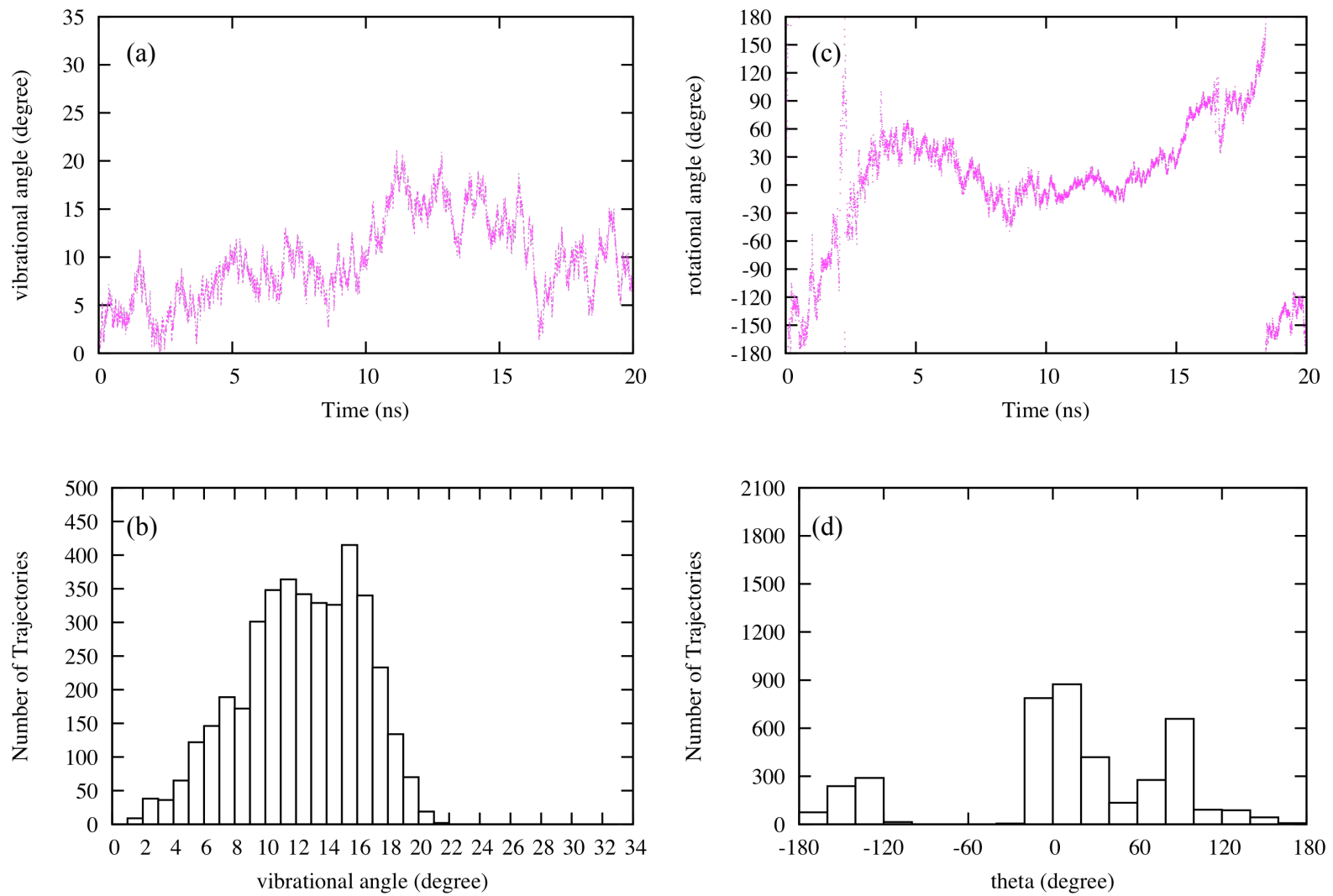
Figure 6.4 (cont.)
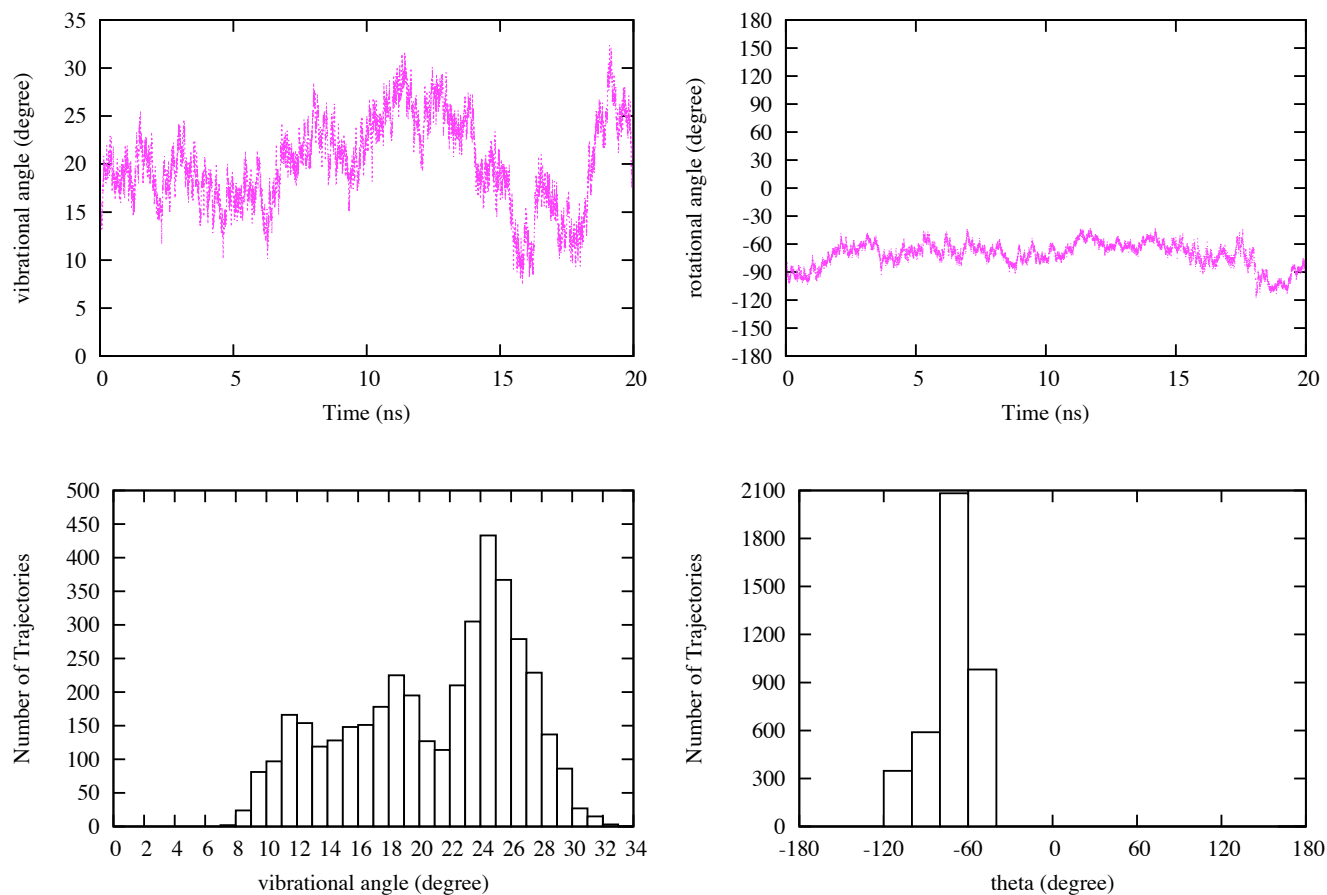
Figure 6.5 (a) Time evolution for the azimuthal vibration angle (φ) of CoREST helix during entire MD simulations and (b) distribution of the vibration angle (φ) in the last 10 ns simulations. (c) Time evolution for the rotation angle (θ) of CoREST helix during entire MD simulations and (d) distribution of the rotation angle (θ) in the last 10 ns simulations. The vibration (φ) and rotation (θ) angles are defined as the vector of the CoREST helix (residues 330–362) as the average vector direction relative to the fixed frame of the Tower helix (residues 418–466) of the X-ray crystal structure.
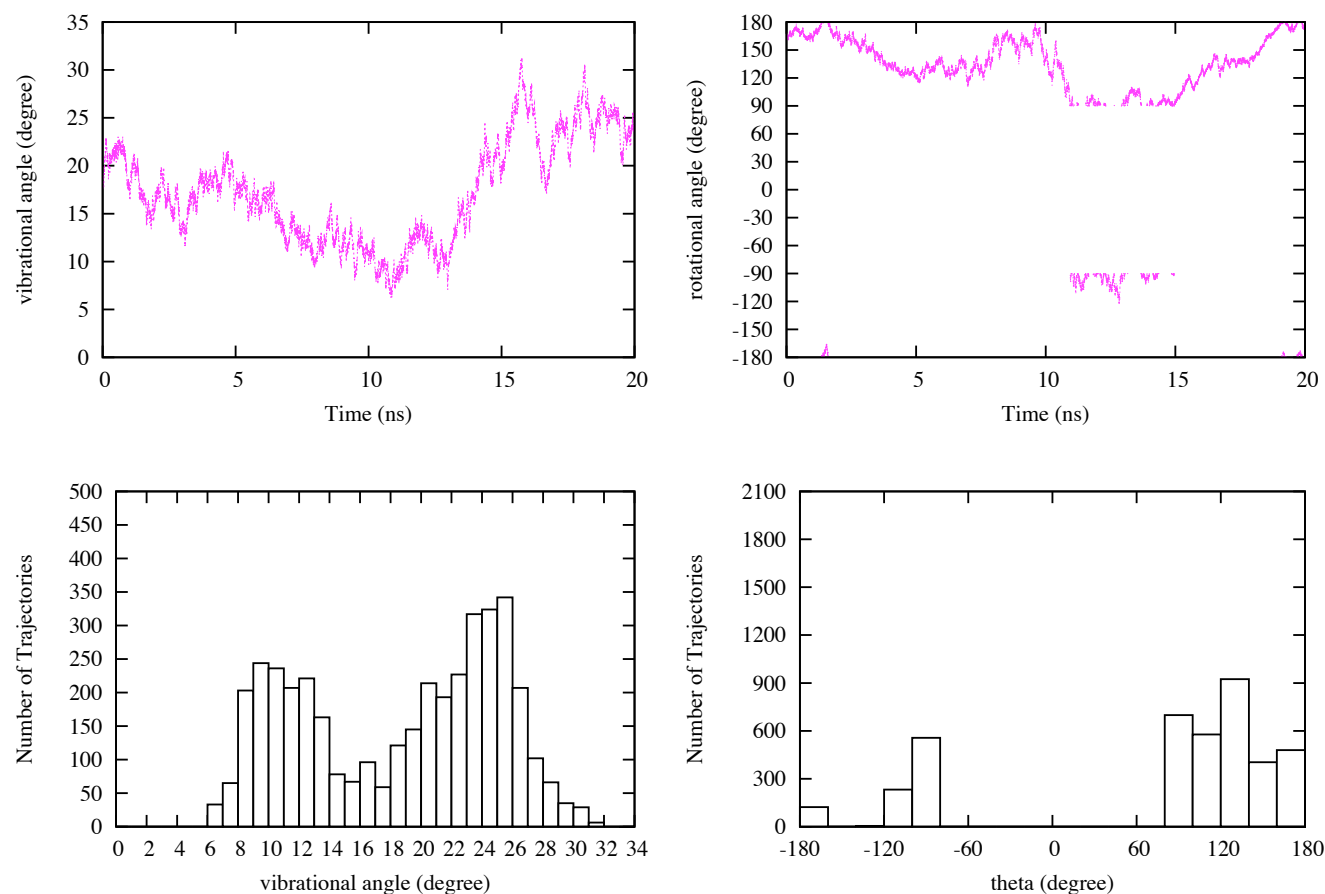
Figure 6.6   (a) Time evolution for the azimuthal vibration angle (φ) of CoREST SANT2 domain during entire MD simulations and (b) distribution of the vibration angle (φ) in the last 10 ns simulations. (c) Time evolution for the rotation angle (θ) of CoREST SANT2 domain during entire MD simulations and (d) distribution of the rotation angle (θ) in the last 10 ns simulations. The vibration (φ) and rotation (θ) angles are defined as the vector from Cα atom of residue 418 to the center of mass of Cα atoms of CoREST SANT2 domain (residues 385–440) relative to the fixed frame of the Tower helix (residues 418–466) of the X-ray crystal structure.
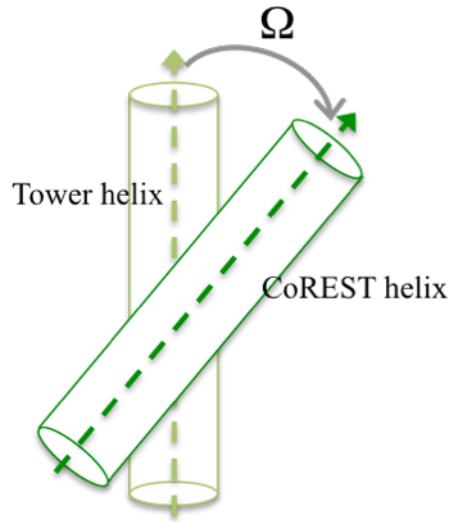
Figure 6.7
Tower helix-CoREST helix crossing angle (Ω). (a) Time evolution of Ω during the entire MD simulations and (b) distribution of Ω in the last 10-ns simulations.
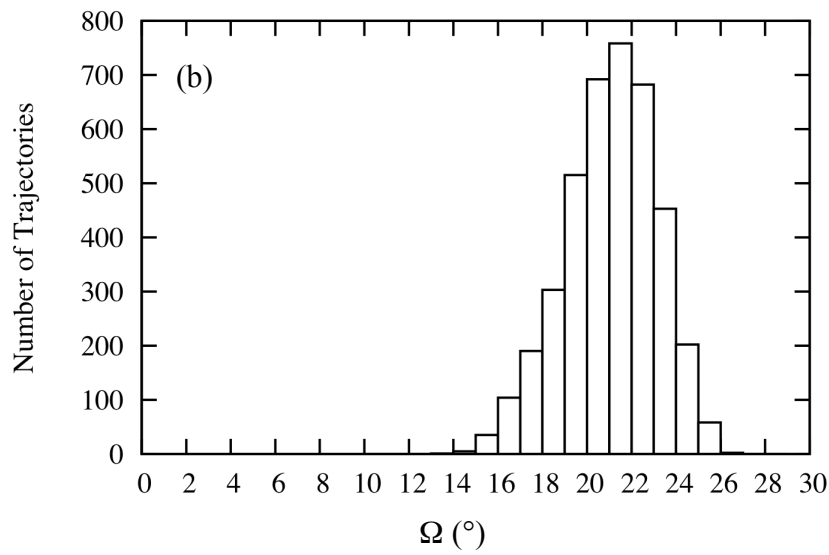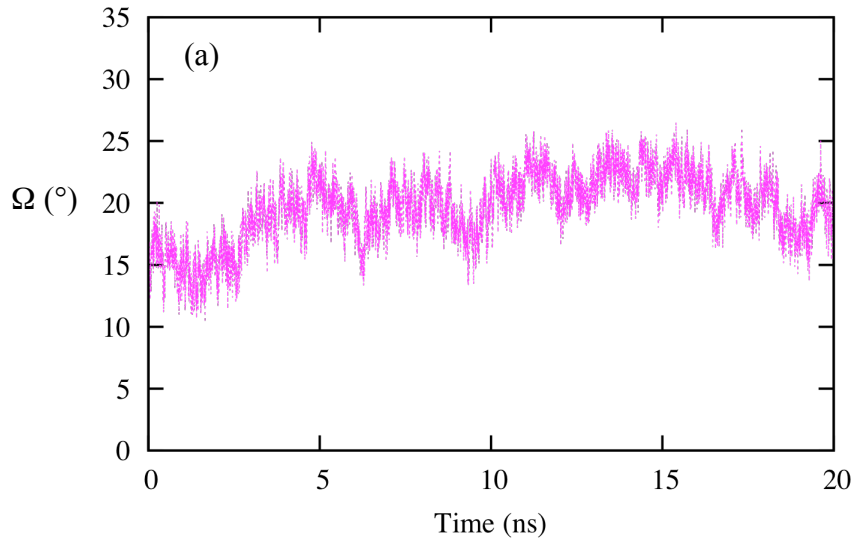
Simulations of H3-p16 in LSD1. The catalytic center of LSD1 consists of a remarkably large cavity that goes deep into the interior of the catalytic domain in which the enzyme cofactor FAD and the histone H3 tail peptide are accommodated. Biochemical studies have shown that the minimal number of residues required for LSD1 recognition is 21 (*6.28*), and the two crystal structures complexed with H3 tail peptides consisted 16 (*6.23*) or 7 (*6.24*) residues in the PDB files, respectively. In the active site, the folded conformations of the H3-p16 peptide have only local fluctuations throughout the 20-ns simulations, and the substrate is held in the binding pocked through a large network of hydrogen bonding interactions and hydrophobic surface contacts. The backbone RMSD of the H3 peptide is about 2.0 Å for the initial 15 ns, and settles to a value of 2.9 Å towards the end of the simulation. It should be noted that the fluctuation of the H3 peptide is primarily due to the flexible C-terminus of the peptide that is solvent-exposed, which undergoes significantly larger fluctuations than those buried in the binding site.

The binding interactions between LSD1 and the H3-p16 peptide from our simulations are analyzed and compared with the crystal structure. The H3 peptide is located in a remarkable large cavity that consists of a series of acidic residues, including D375, D379, E387, D553, D555, D556, and E559 that contribute to the electrostatic interactions with the peptide substrate. In the catalytic cavity, the peptide structure establishes a broad, sophisticated network of interactions by forming inter- and intramolecular hydrogen bonds as well as van der Waals contacts. In Table 6.1, hydrogen bonds between the residues of LSD1···H3-p16 and H3-p16···H3-p16 are

listed from the MD simulations, together with the occupancy of the bonding during the last 10 ns of simulations. Hydrogen bonds are defined as a distance ≤ 3.0 Å and are reported only if they exist for $\geq 5\%$ of the investigating time period. If more than one hydrogen bond is formed between two residues, only the largest occupancy value is reported. Occupancy of hydrogen-bonding interactions is calculated as the percentage of the number of occurrence of a specific hydrogen bond formed along the last 10-ns trajectory (configurations are saved every 5 ps). Figure 6.8 depicts the binding patterns (side chain, backbone amide hydrogen H^{bkb}, and backbone carbonyl oxygen O^{bkb}) of each H3-p16 residue in the catalytic cavity of LSD1 and the occupancy of the bindings in the last 10-ns simulation. Snapshots of a number of selected interactions in the substrate-binding cavity are depicted in Figure 6.9. For the following convenience of description, the one-letter code (ARTMQTARKSTGGAKAP) and the three-letter code (Ala^1-Arg^2-Thr^3-mLys^4-Gln^5-Thr^5-Ala^6-Arg^7-Lys^8-Ser^9-Thr^{10}-Gly^{11}-Gly^{12}-Ala^{13}-Lys^{14}-Ala^{15}-Pro^{16}) for the sequence of H3-p16 peptide are used in the following Tables, Figures, and discussions.

Table 6.1   Hydrogen bond formation between LSD1 and the N-terminal 16 residues of the histone H3 peptide (H3-p16) [a]

H3–p16 [d]	Occupancy (%) [b, c, e]							
pA1	––							
pA1–NH$_3^+$	D555	100	D556	99	W552bkb	97		
pA1–Obkb	N540	42	pK4bkb	20	pT3bkb	5		
pR2	pG12bkb	99	D556	97	pG13bkb	64		
pR2–Hbkb	D556	100						
pR2–Obkb	pQ5bkb	86	pK9	4				
pT3	D555	100	pK4bkb	83				
pT3–Hbkb	D555	60	D556	39	pA1bkb	5		
pT3–Obkb	pK9	100	T335	9	pQ5bkb	8		
pK4	water	31	K661*	4				
pK4–Hbkb	pT3	83	pA1bkb	20				
pK4–Obkb	pT6bkb	71	pT6	9				
pQ5	pS10bkb	89	pS10	44	pR8bkb	9	N535bkb	7
	pA7bkb	6						
pQ5–Hbkb	pR2bkb	86	pT3bkb	8				
pQ5–Obkb	pK9	3						
pT6	H564	50	N535	28	pA7bkb	10	pK4bkb	9
pT6–Hbkb	pK4bkb	71						
pT6–Obkb	N535	7						
pA7	––							
pA7–Hbkb	N535	23	pT6	10				
pA7–Obkb	N535	31	pQ5	6				
pR8	C360bkb	100	D375	100	H532	50	E379	19
pR8–Hbkb	––							
pR8–Obkb	pQ5	9	pS10bkb	5	N535	4		

pK9	H564	100	pT3^{bkb}	100	pR2^{bkb}	4	pQ5^{bkb}	3
pK9–H^{bkb}	––							
pK9–O^{bkb}	––							
pS10	N383	44	pQ5	44	pG12^{bkb}	46	pT11^{bkb}	13
pS10–H^{bkb}	pQ5	89	pR8^{bkb}	5				
pS10–O^{bkb}	pG12^{bkb}	4						
pT11	N383	3	pG12^{bkb}	3				
pT11–H^{bkb}	N383	18	pS10	13				
pT11–O^{bkb}	pA15^{bkb}	56	pK14^{bkb}	30	pA13^{bkb}	4		
pG12	––							
pG12–H^{bkb}	pS10	46	N383	22	pS10^{bkb}	4	pT11	3
pG12–O^{bkb}	pR2	99	pK14^{bkb}	7				
pG13	––							
pG13–H^{bkb}	E387	18	pT11^{bkb}	4				
pG13–O^{bkb}	pR2	64						
pK14	D556	100	E559	100				
pK14–H^{bkb}	pT11^{bkb}	30	pG12^{bkb}	7				
pK14–O^{bkb}	––							
pA15	––							
pA15–H^{bkb}	pT11^{bkb}	56						
pA15–O^{bkb}	––							

[a] Hydrogen bonds are defined as a distance ≤ 3.0 Å. Hydrogen bonds are reported only if they exist for ≥ 5% of the investigating time period. [b] Salt bridge is marked by asterisks (*). [c] Only one hydrogen bond is calculated if there are more than one hydrogen bonds are formed between two residues. [d] H^{bkb} represents the backbone amide hydrogen atom; O^{bkb} symbolizes the backbone carbonyl carbon atom. [e] bkb represents the backbone group of the residue.
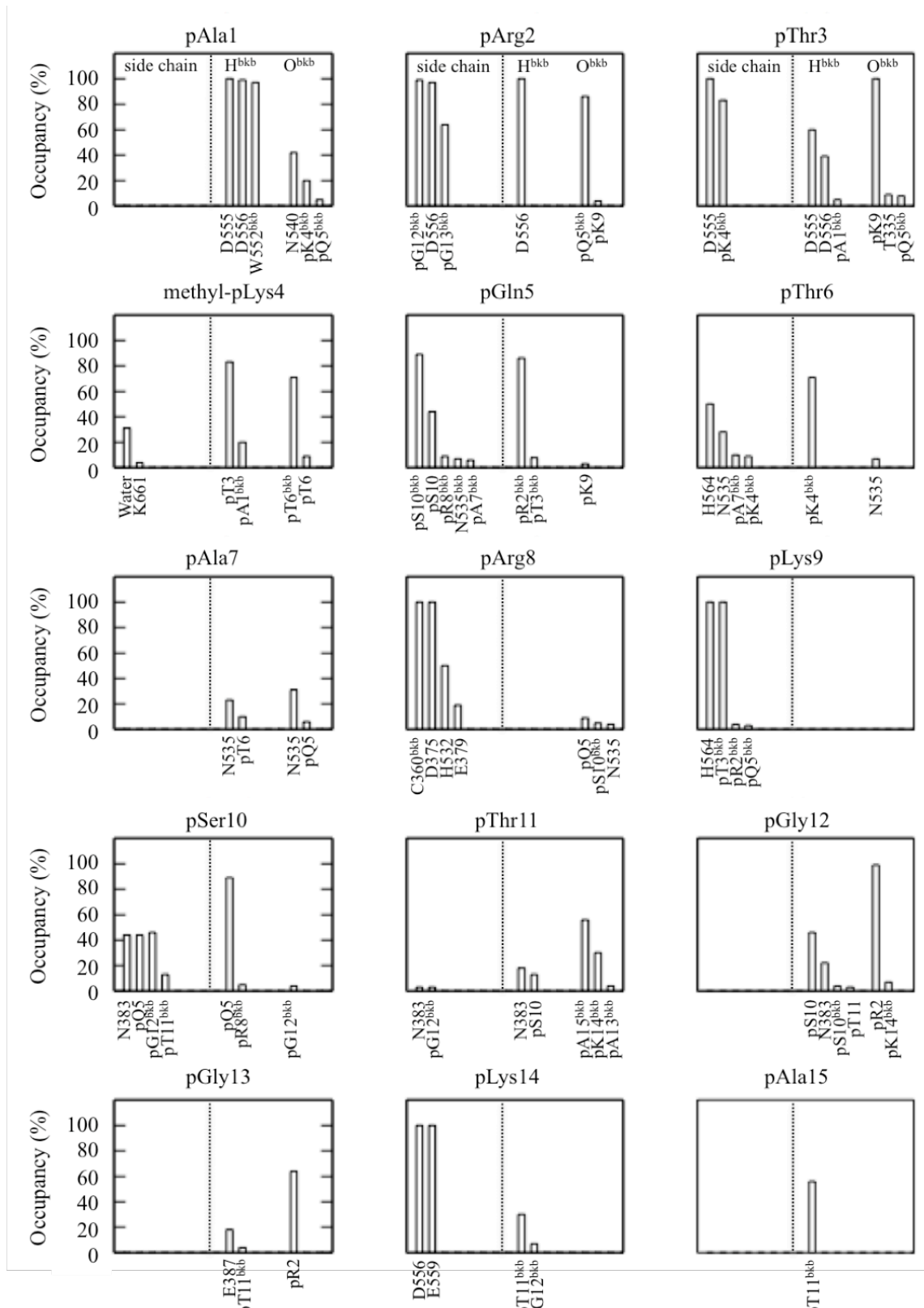
Figure 6.8 Contacts (< 3.0 Å) between residues of LSD1···H3-p16 and H3-p16···H3-p16 obtained from the last 10-ns MD simulations. The side chains, backbone amide hydrogen (H^{bkb}), and backbone carbonyl oxygen (O^{bkb}) of the H3-p16 residues that interact with the surrounding residues are listed, respectively.
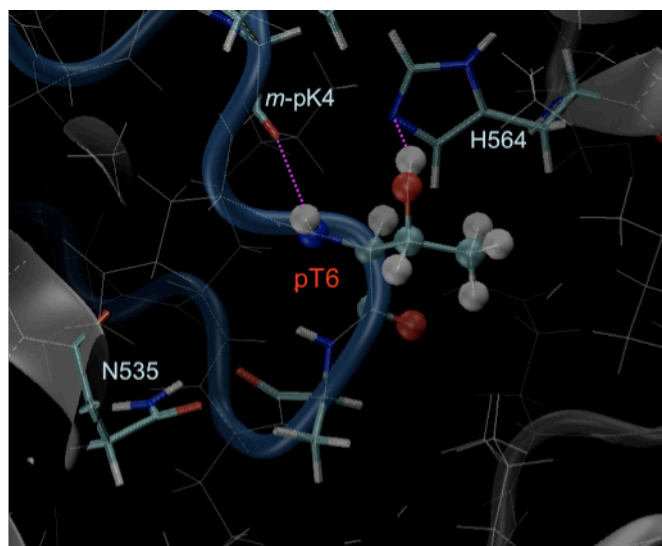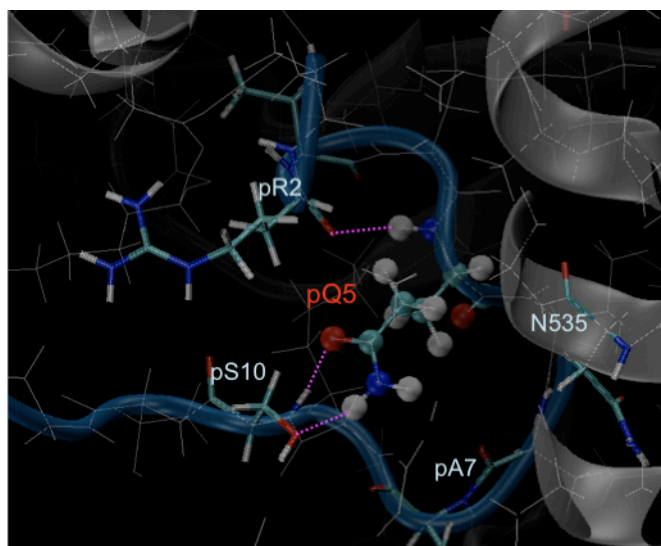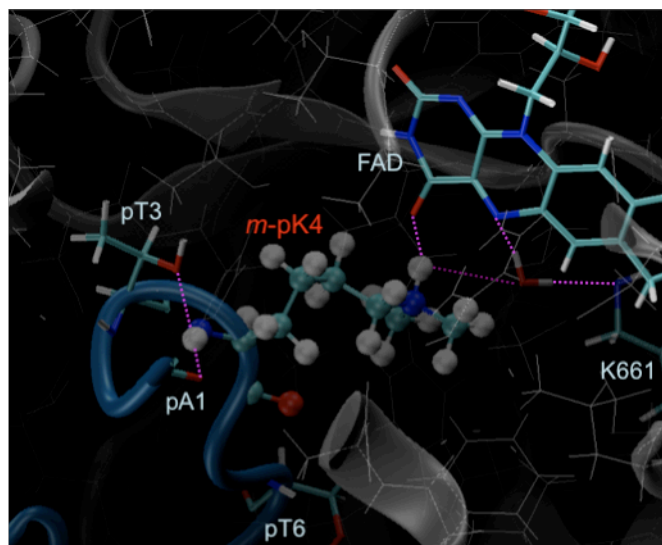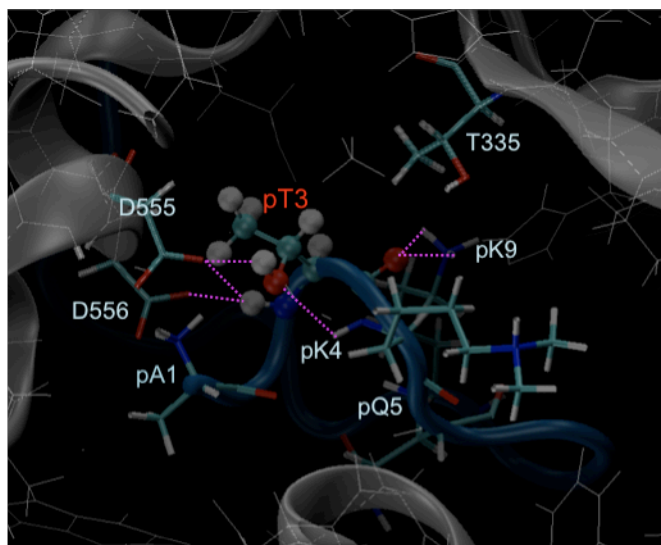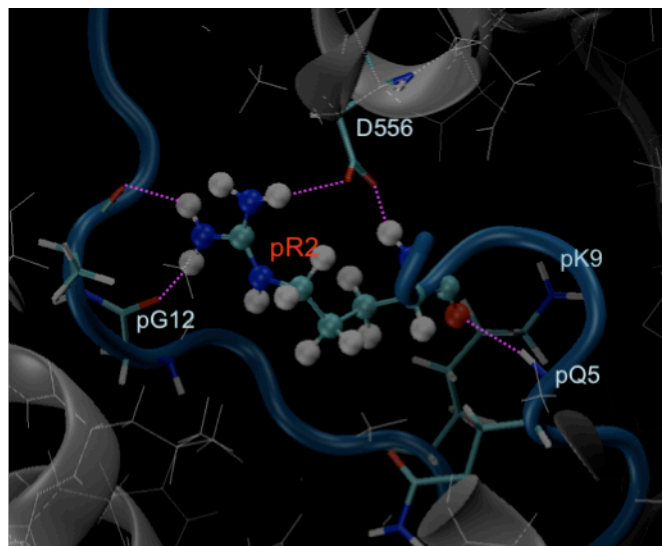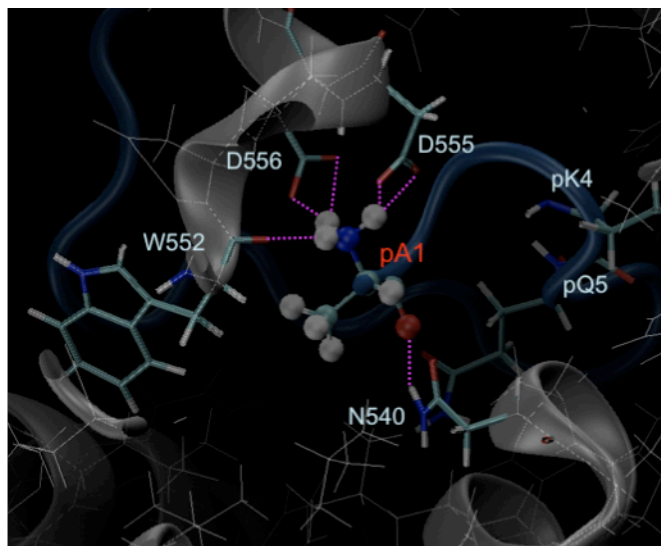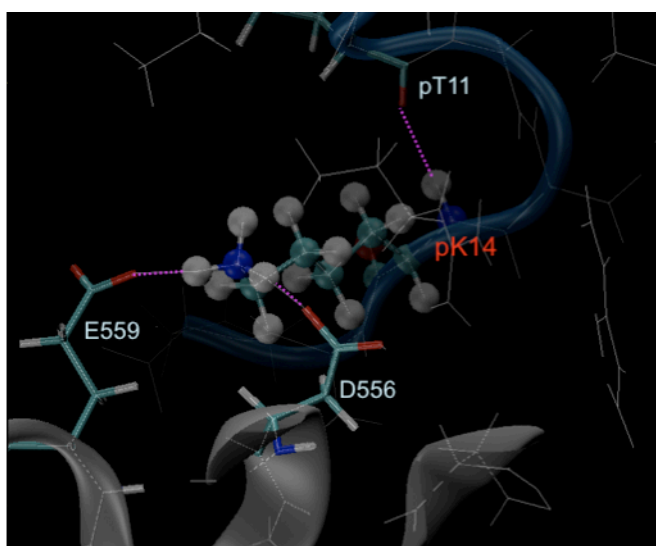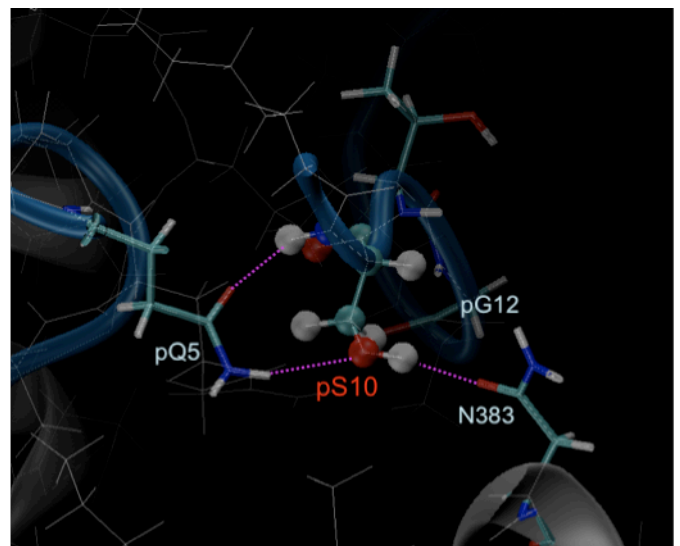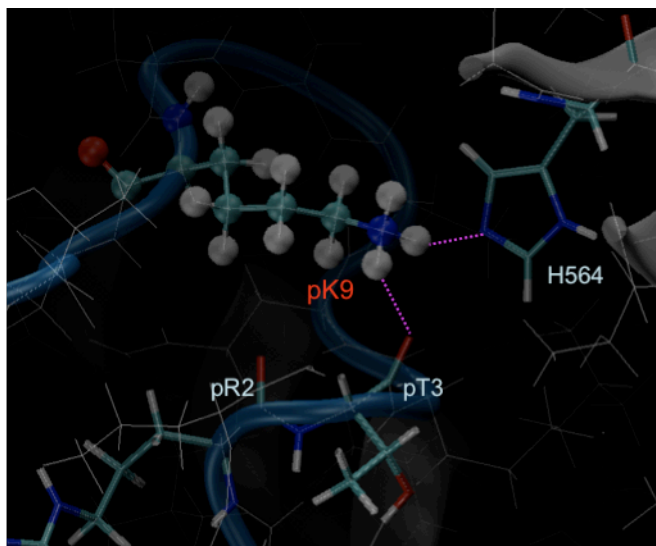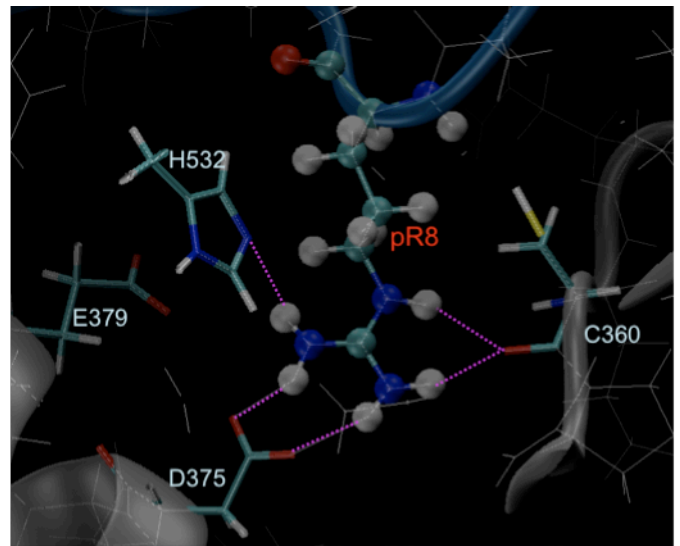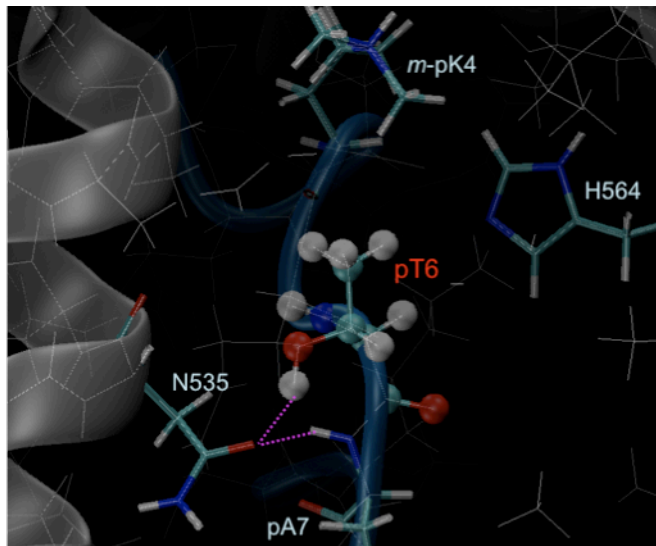
Figure 6.9

Snapshots of the binding mode of some key H3-p16 residues within the catalytic cavity of LSD1

It is of interest that the stabilization of the backbone of the peptide structure is mainly achieved through the protein-peptide interactions with the side chains of N383, N535, N540, D555 and D556 of LSD1. Importantly, the H3 peptide itself also forms a range of hydrogen bonding interactions, among which includes pR2, pT3, pQ5, pK9 and pS10 of H3-p16. In particular, the N-terminal amine of H3 pA1 inserts into an anionic pocket consisting of N540, W552, D555, and D556 to form favorable electrostatic interactions and hydrogen-bond contacts with LSD1. Different from the simulation results, the crystal structure showed that the N-terminal amine of pA1 forms hydrogen bonds with the backbone carbonyl group of A539 (A539 $O^{bkb}\cdots$pA1 N^{NH3+} = 2.37 Å) and with the side chain of N540 (N540 $O^{\delta 1}\cdots$ pA1 N^{NH3+} = 3.23 Å). Moreover in the simulations, D555 and D556 also form favorable interactions with the backbone groups of pR2 (D556 $O^{\delta}\cdots$pR2 H^{bkb} = 2.03 ± 0.16 Å), pT3 (D555 $O^{\delta}\cdots$pT3 H^{bkb} = 2.54 ± 0.20 Å and D556 $O^{\delta}\cdots$pT3 H^{bkb} = 2.54 ± 0.20 Å), and with the side chains of pR2 (D556 $O^{\delta}\cdots$pR2 H^{η} = 1.98 ± 0.28 Å) and pK14 (D556 $O^{\delta}\cdots$pK14 H^{ξ} = 1.77 ± 0.33 Å). Therefore, the anionic pocket composed of N540, W552, D555, and D556 is important for histone tail peptide binding and recognition.

The side chains of pR2, pT3, pK9 and pS10 provide additional inter- and intramolecular interactions that contribute to H3-p16 stabilization in the cavity of the protein complex. pR2 forms salt bridges with D553 O^{δ} and D556 O^{δ} in the crystal structure, but the former interaction is absent during the simulations because the carboxylate group of D553 directs to the solvent. It also donates hydrogen bonds to the pG12 O^{bkb} and pG13 O^{bkb} both in the crystal structure and during the dynamics simulation. Additionally, the pR2 $N^{\varepsilon}\cdots$pS10 O^{γ} contact in the crystal structure is

attenuated by an intervening water molecule, i.e. pR2 $N^\varepsilon \cdots$water$\cdots$pS10 O^γ. pT3 donates hydrogen bonds intermolecularly to the D555 O^δ (D555 $O^\delta \cdots$pT3 H^γ = 1.77 ± 0.12 Å) and intramolecularly to the methyl-pK4 H^{bkb} (pT3 $O^\gamma \cdots$pK4 H^{bkb} = 2.68 ± 0.19 Å), both of which are found in the crystal structure. pK9 makes hydrogen-bonding interactions with the H564 N^ε of LSD1 (H564 $N^\varepsilon \cdots$pK9 H^ζ = 2.09 ± 0.33 Å) and the pT3 O^{bkb} of H3-p16 (pT3 $O^{bkb} \cdots$pK9 H^ζ = 2.59 ± 0.26 Å), both of which, however, are missing in the crystal structure. pS10 favorably interacts with the side chains of N383 of LSD1 and pQ5 of H3-p16 as well as with the backbone amide group of pG12, in accord with the crystal structure. Indeed, pR2, pT3, pK9 and pS10 of the histone H3 tail are potential histone modification marks and the epigenetic modifications on these histone residues reduce the catalytic activity of LSD1 (*6.3, 6.28*). Thus, these sequence-specific interactions involving pA1, pR2, pT3, pQ5, pK9 and pS10 are expected to contribute to the substrate specificity of LSD1 toward H3K4 demethylation, and the additional modifications on these residues perturb the binding stability of histone H3 tail in the active site of LSD1.

The simulations show that the peptide residue in position 4 is in location position in the front side of the flavin ring of FAD, an orientation perfectly suited for the catalytic chemical transformation through an initial hydride transfer from the methyl group of dimethylated pLys4 to FAD. For comparison, a similar conformational feature is found in the three-dimensional structure of the complex. Although there was no interpretable electron density corresponding to water molecules in the crystal structure, our simulations reveal that the H^ζ of methyl-pK4 is 1.82 ± 0.16 Å to the FAD O4 atom and 3.28 ± 0.49 Å to a water molecule that is in

turn positioned by K661 of LSD1. This water molecule is ideally located to shuttle a proton from the reactive methyl-pK4 of the histone to K661 during the oxidation reaction (*6.20, 6.22*). Consistent with previous reports, another astonishing feature is that the backbone stabilization of the methyl-pK4 is mediated through an extensive chain of interactions, involving a number of residues of the H3 tail and LSD1 (*6.22*). Its amide H^{bkb} is intramolecularly hydrogen-bonded to the pT3 $O^{\gamma 1}$ and/or pA1 O^{bkb} that insert into an anionic pocket and form intermolecular interactions with N540, D555, and D556. Indeed, mutations of N540, D555 and D556 to alanine residues substantially diminish the catalytic activity of the enzyme on the dimethylated H3K4 peptide, indicating a crucial binding mode for LSD1 towards substrate H3 tail.

In addition, the preferred methyl-pK4 demethylation site is directly anchored to the substrate-binding site by hydrophobic interactions with several aliphatic and aromatic residues of LSD1 (V333, F530, A539, L659, Y761, A809, and T810) (*6.21*). This allows LSD1 to place the side chain of methyl-pK4 in the proximity to FAD. Altogether, residues N540, W552, D555, and D556 of LSD1 could modulate the backbone position of methyl-pK4 residue (Figure 6.8) and optimize the methylated head group of the residue through a hydrophobic surface of the pocket to place the H^{ξ} of the substrate residue in the catalytic competent position relative to the flavin ring of FAD and the catalytic water (*6.21, 6.24*). Studies of the LSD1 catalysis are planned, but they are beyond the scope of the current investigation. The structural analysis paved a way to use the truncated catalytic domain only in the enzyme simulations since the present dynamics simulations demonstrate that the globular features of the AOL and SWRIM domains are well preserved. It is also known

experimentally that catalysis does not require the presence of the CoREST protein, which interacts with the TOWER domain; its presence in experimental studies as well as crystallization is due to increased stability for analysis.

6.4   Conclusion

By means of molecular dynamics simulations, we have investigated the conformational dynamics of LSD1•CoREST complex in the presence of a N-terminal 16-a.a. H3-tail substrate in aqueous solution. The overall conformation of the ternary complex in water is in reasonable agreement with the X-ray crystal structure, except the Tower domain of LSD1 that tolerates a significantly conformational fluctuation with respect to the other domains of the protein. The present dynamics simulations provided important insights into the dynamic nature of LSD1 Tower domain in water by analyzing its azimuthal and colatitudal angle motions over the 20 ns of MD simulations. It turns out that LSD1 moderately swings its Tower domain in a rather large range of conformations. Together with the fact that LSD1 can interact with a variety of substrate proteins to mediate its multiple gene regulatory functions, the specificity of LSD1 for histone and non-histone substrates may require the fluctuating capability of its Tower domain.

The simulations revealed the dynamics of the histone H3 peptide in the catalytic cleft of LSD1 with the presence of explicit water solvent. No significant conformational changes in the histone H3 peptide structure have been observed in the catalytic cleft of LSD1 during the 20 ns molecular dynamics simulation. Notably, some intermolecular interactions between LSD1 and the H3 peptide and

167

intramolecular contacts within H3 peptide in the catalytic cavity were observed from the simulations, but they are absent in the X-ray crystal structure. A bridge water molecule between the methylated lysine substrate and K661 of LSD1 was observed in the simulations. In addition, the first three residues of the histone H3 tail are found to be importance to position the methylated Lys4 on histone H3 tail towards the catalytic center of LSD1, consistent with experimental proposals. Furthermore, pR2, pT3, pQ5, pK9, and pS10 are postulated to play a role in stabilizing the overall conformation of the histone H3 tail in the cavity based on structural analyses.

Bibliography

Chapter 1

(1.1)   Karplus, M. and McCammon, J. (2002) Molecular dynamics simulations of biomolecules. *Nature Struct. Biol. 9*, 646-653.

(1.2)   Warshel, A. (2002) Molecular dynamics simulations of biological reactions. *Acc. Chem. Res. 35*, 385-395.

(1.3)   Gao, J. (2003) Catalysis by enzyme conformational change as illustrated by orotidine 5'-monophosphate decarboxylase. *Curr. Opin. Struct. Biol. 13*, 184-192.

(1.4)   Benkovic, S. J. and Hammes-Schiffer, S. (2003) A perspective on enzyme catalysis. *Science 301*, 1196-1202.

(1.5)   Gao, J., Ma, S., Major, D. T., Nam, K., Pu, J., and Truhlar, D. G. (2006) Mechanisms and free energies of enzymatic reactions. *Chem. Rev. 106*, 3188-3209.

(1.6)   Warshel, A. (1991) *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. John Wiley and Sons, New York, 1991.

(1.7)   Becker, O. M., MacKerell, A. D. Jr., Roux, B., and Watanabe, M. (2001) *Computational Biochemistry and Biophysics*. Marcel Dekker, New York.

(1.8)   Garcia-Viloca, M., Gao, J., Karplus, M., and Truhlar, D. G. (2004) How enzymes work: Analysis by modern rate theory and computer simulations, *Science 303*, 186-195.

(1.9)   Warshel, A. (2003) Computer simulations of enzyme catalysis: methods, process, and insight. *Annu. Rev. Biophys. Struct. 32*, 425-443.

(1.10)   Kollman, P. A. (1993) Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev. 93*, 2395-2417.

(1.11)   Frenkel, D. and Smit, B. (2002) *Understanding Molecular Simulation.* Academic Press, 2edn.

(1.12)   Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., and Casero, R. A. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1, *Cell 119*, 941-953.

(1.13)   Forneris, F., Binda, C., Vanoni, M. A., Mattevi, A., and Battaglioli, E. (2005) Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process, *Febs Lett. 579*, 2203-2207.

(1.14)   Forneris, F., Binda, C., Dall'Aglio, A., Fraaije, M. W., Battaglioli, E., and Mattevi, A. (2006) A highly specific mechanism of histone H3-K4 recognition by histone demethylase LSD1, *J. Biol. Chem. 281*, 35289-35295.

(1.15)   Forneris, F., Binda, C., Battaglioli, E., and Mattevi, A. (2008) LSD1: oxidative chemistry for multifaceted functions in chromatin regulation, *Trend. Biochem. Sci. 33*, 181-189.

(1.16)   Shi, Y., and Whetstine, J. R. (2007) Dynamic regulation of histone lysine methylation by demethylases, *Mol. Cell 25*, 1-14.

(1.17)   Marmorstein, R., and Trievel, R. C. (2009) Histone modifying enzymes: structures, mechanisms, and specificities, *Biochim. Biophys. Acta. 1789*, 58-68.

(1.18)   Shi, Y. J., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005) Regulation of LSD1 histone demethylase activity by its associated factors, *Mol. Cell 19*, 857-864.

(1.19)   Rajamani, R. and Gao, J. (2002) Combined QM/MM study of the opsin shift in bacteriorhodopsin. *J. Comput. Chem. 23*, 96–105.

(1.20)   Houjou, H., Inoue, Y., and Sakurai, M. (2001) Study of the opsin shift of bacteriorhodopsin: insight from QM/MM calculations with electronic polarization effects of the protein environment. *J. Phys. Chem. B 105*, 867-879.

(1.21)   Houjou, H., Inoue, Y., and Sakurai, M. (1998) Physical origin of the opsin shift of bacteriorhodopsin. Comprehensive analysis based on medium effect theory of absorption spectral. *J. Am. Chem. Soc. 120*, 4459-4470.

Chapter 2

(2.1)   Bennett, G. E.; Johnston, K. P. *J. Phys. Chem.* **1994**, *98*, 441.

(2.2)   Bermejo, M. D.; Cocero, M. J. *AIChE Journal* **2006**, *52*, 3933.

(2.3)   Marrone, P. A.; Hodes, M.; Smith, K. A.; Tester, J. W. *J. Supercr. Fluids* **2004**, *29*, 289.

(2.4)   Marrone, P. A.; Cantwell, S. D.; Dalton, D. W. *Ind. Eng. Chem. Res.* **2005**, *44*, 9030.

(2.5)   Ryan, E. T.; Xiang, T.; Johnston, K. P.; Fox, M. A. *J. Phys. Chem. A* **1997**, *101*, 1827.

(2.6)   Williams, P. T.; Onwudili, J. A. *Envir. Tech.* **2006**, *27*, 823.

(2.7)   Shaw, J. P.; Petsko, G. A.; Ringe, D. *Biochemistry* **1997**, *36*, 1329.

(2.8)   Savage, P. E. *Chem. Rev.* **1999**, *99*, 603.

(2.9)   Minett, S.; Fenwick, K. *European Water Management* **2001**, *4*, 54.

(2.10)   Haar, L.; Gallagher, J. S.; Kell, G. S. *NBSATRC Steam Tables*; Hemisphere: Washington, D.C., 1984.

(2.11)   Archer, D. G.; Wang, P. *J. Phys. Chem. Reg. Data* **1990**, *19*, 371.

(2.12)   Paulaitis, M. E.; Krukonis, V. J.; Kurnik, R. T.; Reid, R. C. *Rev. Chem. Eng.* **1983**, 179.

(2.13)   Johnston, K. P.; Rossky, P. J. *NATO Sci. Ser., Ser. E* **2000**, *366*, 323.

(2.14)   Galkin, A. A.; Lunin, V. V. *Russ. Chem. Rev.* **2005**, *74*, 21.

(2.15)   Gao, J. *J. Am. Chem. Soc.* **1993**, *115*, 6893.

(2.16)   Reichardt, C. *Solvents and Solvent Effects in Organic Chemistry*, 2nd ed.; VCH: Weinheim, 1990.

(2.17)   Amos, A. T.; Hall, G. G. *Proc. R. Soc. London* **1961**, *Ser. A263*, 482.

(2.18)   Karelson, M. M.; Katritzky, A. R.; Zerner, M. C. *Int. J. Quant. Chem., Quant. Chem. Symp.* **1986**, *20*, 521.

(2.19)   Lerf, C.; Suppan, P. *J. Chem. Soc., Faraday Trans.* **1992**, *88*, 963.

(2.20)   Li, J.; Cramer, C. J.; Truhlar, D. G. *Int. J. Quant. Chem.* **2000**, *77*, 264.

(2.21)   Nugent, S.; Ladanyi, B. M. *J. Chem. Phys.* **2004**, *120*, 874.

(2.22)   Gao, J. *J. Am. Chem. Soc.* **1994**, *116*, 9324.

(2.23)   Gao, J.; Byun, K. *Theor. Chem. Acc.* **1997**, *96*, 151.

(2.24)   Rajamani, R.; Gao, J. *J. Comput. Chem.* **2002**, *23*, 96.

(2.25)   Poulsen, T. D.; Ogilby, P. R.; Mikkelsen, K. V. *J. Chem. Phys.* **2002**, *116*, 3730.

(2.26)   Buncel, E.; Rajagopal, S. *Acc. Chem. Res.* **1990**, *23*, 226.

(2.27)   Reichardt, C. *Chem. Rev.* **1994**, *94*, 2319.

(2.28)   Bayliss, N. S.; McRae, E. G. *J. Phys. Chem.* **1954**, *58*, 1006.

(2.29)   Canuto, S.; Coutinho, K.; Zerner, M. C. *J. Chem. Phys.* **2000**, *112*, 7293.

(2.30)   Roesch, N.; Zerner, M. C. *J. Phys. Chem.* **1994**, *98*, 5817.

(2.31)   Cossi, M.; Barone, V. *J. Chem. Phys.* **2000**, *112*, 2427.

(2.32)   Pappalardo, R. R.; Reguero, M.; Robb, M. A.; Frish, M. *Chem. Phys. Lett.* **1993**, *212*, 12.

(2.33)   Minezawa, N.; Kato, S. *J. Chem. Phys.* **2007**, *126*, 054511/1.

(2.34)   Mennucci, B.; Cammi, R.; Tomasi, J. *J. Chem. Phys.* **1998**, *109*, 2798.

(2.35)   Martin, M. E.; Sanchez, M. L.; Olivares del Valle, F. J.; Aguilar, M. A. *J. Chem. Phys.* **2000**, *113*, 6308.

(2.36)   Bernasconi, L.; Sprik, M.; Hutter, J. *J. Chem. Phys.* **2003**, *119*, 12417.

(2.37)   Fonseca, T. L.; Coutinho, K.; Canuto, S. *J. Chem. Phys.* **2007**, *126*, 034508/1.

(2.38)   Neugebauer, J.; Louwerse, M. J.; Baerends, E. J.; Wesolowski, T. A. *J. Chem. Phys.* **2005**, *122*, 094115/1.

(2.39)   Coutinho, K.; Canuto, S. *Theochem* **2003**, *632*, 235.

(2.40)   Martin, M. E.; Sanchez, M. L.; Olivares del Valle, F. J.; Aguilar, M. A. *J. Chem. Phys.* **2000**, *113*, 6308.

(2.41)   Luzhkov, V.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4491.

(2.42)   Blair, J. T.; Krogh-Jespersen, K.; Levy, R. M. *J. Am. Chem. Soc.* **1989**, *111*, 6948.

(2.43)   DeBolt, S. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 7515.

(2.44)   Gao, J.; Li, N.; Freindorf, M. *J. Am. Chem. Soc.* **1996**, *118*, 4912.

(2.45)   Gao, J.; Alhambra, C. *J. Am. Chem. Soc.* **1997**, *119*, 2962.

(2.46)   Gao, J. *J. Comput. Chem.* **1997**, *18*, 1062.

(2.47)   Thompson, M. A.; Schenter, G. K. *J. Phys. Chem.* **1995**, *99*, 6374.

(2.48)   Thompson, M. A. *J. Phys. Chem.* **1996**, *100*, 14492.

(2.49)   Heitz, M. P.; Bright, F. V. *J. Phys. Chem.* **1996**, *100*, 6889.

(2.50)   Wyatt, V. T.; Bush, D.; Lu, J.; Hallett, J. P.; Liotta, C. L.; Eckert, C. A. *J. Supercr. Fluids* **2005**, *36*, 16.

(2.51)   Takebayashi, Y.; Sugeta, S. Y., T.; Otake, K.; Nakahara, M. *J. Phys. Chem. B* **2003**, *107*, 9847.

(2.52)   Takebayashi, Y.; Yoda, S. S., T.; Otake, K.; Sako, T.; Nakahara, M. *J. Chem. Phys.* **2004**, *120*, 6100.

(2.53)   Balbuena, P. B.; Johnston, K. P.; Rossky, P. J. *J. Phys. Chem.* **1996**, *100*, 2716.

(2.54)   Gao, J. *J. Phys. Chem.* **1994**, *98*, 6049.

(2.55)   Kubo, M.; Levy, R. M.; Rossky, P. J.; Matubayasi, N.; Nakahara, M. *J. Phys. Chem. B* **2002**, *106*, 3979.

(2.56)   Dang, L. X. *J. Chem. Phys.* **1992**, *97*, 2659.

(2.57)   Gao, J. *J. Phys. Chem.* **1992**, *96*, 537.

(2.58)   Gao, J. MCQUB; 3.0 ed. Buffalo, N.Y., 1998.

(2.59)   Gao, J. MCQUM; 4.0 ed. Department of Chemistry, University of Minnesota, 2000.

(2.60)   Stewart, J. J. P. *J. Computer-aided Desgin* **1990**, *4*, 1.

(2.61)   Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(2.62)   Gao, J.; Xia, X. *Science* **1992**, *258*, 631.

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

(2.64)   Owicki, J. C.; Scheraga, H. A. *Chem. Phys. Lett.* **1977**, *47*, 600.

(2.65)   Westacott, R. E.; Johnston, K. P.; Rossky, P. J. *J. Am. Chem. Soc.* **2001**, *123*, 1006.

(2.66)   Westacott, R. E.; Rossky, P. J.; Johnston, K. P. *Abstr. Pap. - Am. Chem. Soc.* **2000**, *220th*, COMP.

(2.67)   Balbuena, P. B.; Johnston, K. P.; Rossky, P. J.; Hyun, J.-K. *J. Phys. Chem. B* **1998**, *102*, 3806.

(2.68)   Suppan, P. *J. Photochem. Photobiol.* **1990**, *A50*, 293.

(2.69)   De Vries, A. H.; Van Duijnen, P. T. *Int. J. Quant. Chem.* **1996**, *57*, 1067.

(2.70)   Fox, T.; Roesch, N. *Chem. Phys. Lett.* **1992**, *191*, 33.

(2.71)   Georg, H. C.; Coutinho, K.; Canuto, S. *Chem. Phys. Lett.* **2006**, *429*, 119.

(2.72)   Roehrig, U. F.; Frank, I.; Hutter, J.; Laio, A.; VandeVondele, J.; Rothlisberger, U. *ChemPhysChem* **2003**, *4*, 1177.

Chapter 3

(3.1)   Frey, P. A., and Hegeman, A. D. (2007) Enzymatic Reaction Mechanisms, Oxford University Press, New York.

(3.2)   Snell, E. E., and Di Mari, S. J. (1970) *The Enzymes: Kinetics and Mechanism,* 3rd ed., Boyer, P. D. Ed., Academic Press, New York. Vol. 2, pp. 335-362.

(3.3)   Sharif, S., Huot, M. C., Tolstoy, P. M., Toney, M. D., Jonsson, K. H. M., and Limbach, H.-H. (2007) [15]N Nuclear Magnetic Resonance Studies of Acid-Base Properties of Pyridoxal 5'-Phosphate Aldimines in Aqueous Solution. *J. Phys. Chem. B 111*, 3869-3876.

(3.4)   Burkhard, P., Dominici, P., Voltattorni, C. B., Jansonius, J. N., and Malashkevich, V. N. (2001) Structural Insight into Parkinson's Disease Treatment from Drug-Inhibited DOPA Decarboxylase. *Nat. Struct. Biol. 8*, 963-967.

(3.5)   Christen, P., and Metzler, D. E. (1985) *Transaminases,* John Wiley & Sons, New York.

(3.6)   Dolphin, D., Poulson, R., and Avramović, O. (1986) *Vitamin B_6: Pyridoxal Phosphate: Chemical, Biochemical, and Medical Aspects,* John Wiley & Sons, New York.

(3.7)   Poupon, A., Jebai, F., Labesse, G., Gros, F., Thibault, J., Mornon, J.-P., and Krieger, M. (1999) Structure Modelling and Site-Directed Mutagenesis of the Rat Aromatic L-Amino Acid Pyridoxal 5´-Phosphate-Dependent Decarboxylase: A Functional Study. *Proteins 37*, 191-203.

(3.8)   Bertoldi, M., Castellani, S., and Votattorni, C. B. (2001) Mutation of residues in the coenzyme binding pocket of Dopa decarboxylase. *Eur. J. Biochem. 268*, 2975-2981.

(3.9)   Hansen, P. E., Sitkowski, J., Stefaniak, L., Rozwadowski, Z., and Dziembowska, T. (1998) One-Bond Deuterium Isotope Effects on [15]N Chemical Shifts in Schiff Bases. *Berichte der Bunsen-Gesellschaft. 102*, 410-413.

(3.10)   Filarowski, A., Koll, A., Rospenk, M., Krol-Starzomska, I., and Hansen, P. E. (2005) Tautomerism of Sterically Hindered Schiff Bases. Deuterium Isotope Effects on [13]C Chemical Shifts. *J. Phys. Chem. A 109*, 4464-4473.

(3.11)   Sharif, S., Denisov, G. S., Toney, M. D., and Limbach, H.-H. (2006) NMR Studies of Solvent-Assisted Proton Transfer in a Biologically Relevant Schiff Base: Toward a Distinction of Geometric and Equilibrium H-Bond Isotope Effects. *J. Am. Chem. Soc. 128*, 3375-3387.

(3.12)   Hill, M. P., Carroll, E. C., Toney, M. D., and Larsen, D. S. (2008) Rapid Photodynamics of Vitamin B6 Coenzyme Pyridoxal 5'-Phosphate and its Schiff Bases in Solution. *J. Phys. Chem. B 112*, 5867-5873.

(3.13)   Claramunt, R. M., Lopez, C., Santa Maria, M. D., Sanz, D., and Elguero, J. (2006) The use of NMR Spectroscopy to Study Tautomerism. *Prog. Nucl. Magn. Res. Spectr. 49*, 169-206.

(3.14)   Tong, H., and Davis, L. (1995) *2*-Amino-*3*-Ketobutyrate-CoA Ligase from Beef Liver Mitochondria: An NMR Spectroscopic Study of Low-Barrier Hydrogen Bonds of a Pyridoxal 5'-Phosphate-Dependent Enzyme. *Biochemistry 34*, 3362-3367.

(3.15)   Alarcon, S. H., Olivieri, A. C., Cravero, R. M., Labadie, G., and Gonzalez-Sierra, M. (1995) Ground- and Excited-State Prototropic Tautomerism in Anils of Aromatic α-Hydroxyaldehydes Studied by Electronic Absorption, Fluorescence and ^{1}H and ^{13}C NMR Spectroscopies and Semi-Empirical Calculations. *J. Phys. Org. Chem. 8*, 713-720.

(3.16)   Hayashi, H., Mizuguchi, H., and Kagamiyama, H. (1993) Rat Liver Aromatic L-Amino Acid Decarboxylase: Spectroscopic and Kinetic Analysis of the Coenzyme and Reaction Intermediates. *Biochemistry 32*, 812-818.

(3.17)   Barboni, E., Voltattorni, C. B., D'Erme, M., Fiori, A., Minelli, A., and Rosei, M. A. (1982) Inhibitors Binding to L-Aromatic Amino Acid Decarboxylase. *Life Sciences 31*, 1519-1524.

(3.18)   Minelli, A., Charteris, A. T., Voltattorni, B. C., and John, R. A. (1979) Reactions of DOPA (*3,4*-Dihydroxyphenylalanine) Decarboxylase with DOPA. *Biochem. J. 183*, 361-368.

(3.19)   Dominici, P., Tancini, B., Barra, D., and Voltattorni, C. B. (1987) Purification and Characterization of Rat-Liver *3,4*-Dihydroxyphenylalanine Decarboxylase. *Eur. J. Biochem. 169*, 209-213.

(3.20)   Voltattorni, C. B., Minelli, A., Vecchini, P., Fiori, A., and Turano, C. (1979) Purification and Characterization of *3,4*-Dihydroxyphenylalanine Decarboxylase from Pig Kidney. *Eur. J. Biochem. 93*, 181-188.

176

(3.21)   Ando-Yamamoto, M., Hayashi, H., Sugiyama, T., Fukui, H., Watanabe, T., and Wada, H. (1987) Purification of L-DOPA Decarboxylase from Rat Liver and Production of Polyclonal and Monoclonal Antibodies Against it. *J. Biochem. (Tokyo, Japan) 101*, 405-414.

(3.22)   Nakazawa, H., Kumagai, H., and Yamada, H. (1974) Constitutive Aromatic L-Amino Acid Decarboxylase from Micrococcus Percitreus. *Biochem. Biophys. Res. Commun. 61*, 75-82.

(3.23)   Fiori, A., Turano, C., Voltattorni, C. B., Minelli, A., and Codini, M. (1975) Interaction of L-DOPA Decarboxylase with Substrates: A Spectrophotometric Study. *FEBS Letters 54*, 122-125.

(3.24)   Kirsch, J. F., Eichele, G., Ford, G. C., Vincent, M. G., Jansonius, J. N., Gehring, H., and Christen, P. (1984) Mechanism of Action of Aspartate Aminotransferase Proposed on the Basis of its Spatial Structure. *J. Mol. Biol. 174*, 497-525.

(3.25)   Dominiak, P. M., Grech, E., Barr, G., Teat, S., Mallinson, P., and Wozniak, K. (2003) Neutral and Ionic Hydrogen Bonding in Schiff Bases. *Chemistry: a European Journal 9*, 963-970.

(3.26)   Moustakali-Mavridis, I., Hadjoudis, E., and Mavridis, A. (1978) Crystal and Molecular Structure of some Thermochromic Schiff Bases. *Acta Crystall., Sec. B. B34*, 3709-3715.

(3.27)   Krygowski, T. M., Wozniak, K., Anulewicz, R., Pawlak, D., Kolodziejski, W., Grech, E., and Szady, A. (1997) Through-Resonance Assisted Ionic Hydrogen Bonding in *5*-Nitro-*N*-Salicylideneethylamine. *J. Phys. Chem. A 101*, 9399-9404.

(3.28)   Perona, A., Sanz, D., Claramunt, R. M., Pinilla, E., Torres, M. R., and Elguero, J. (2007) Acid Assisted Proton Transfer in *4*-[(*4-R*-Phenylimino)Methyl]Pyridin-*3*-Ols: NMR Spectroscopy in Solution and Solid State, X-Ray and UV Studies and DFT Calculations. *J. Phys. Org. Chem. 20*, 610-623.

(3.29)   Kolehmainen, E., Osmialowski, B., Krygowski, T. M., Kauppinen, R., Nissinen, M., and Gawinecki, R. (2000) Substituent and Temperature Controlled Tautomerism: Multinuclear Magnetic Resonance, X-Ray, and Theoretical Studies on *2*-Phenacylquinolines. *J. Chem. Soc., Perkin Trans. 2*, 1259-1266.

177

(3.30)   Sharif, S., Schagen, D., Toney, M. D., and Limbach, H.-H. (2007) Coupling of Functional Hydrogen Bonds in Pyridoxal-5'-Phosphate-Enzyme Model Systems Observed by Solid-State NMR Spectroscopy. *J. Am. Chem. Soc. 129*, 4440-4455.

(3.31)   Sharif, S., Denisov, G. S., Toney, M. D., and Limbach, H.-H. (2007) NMR Studies of Coupled Low- and High-Barrier Hydrogen Bonds in Pyridoxal-5'-Phosphate Model Systems in Polar Solution. *J. Am. Chem. Soc. 129,* 6313-6327.

(3.32)   Alarcon, S. H., Olivieri, A. C., Labadie, G. R., Cravero, R. M., and Gonzalez-Sierra, M. (1995) Tautomerism of Representative Aromatic α-Hydroxy Carbaldehyde Anils as Studied by Spectroscopic Methods and AM1 Calculations. Synthesis of *10*-Hydroxyphenanthrene-*9*-Carbaldehyde. *Tetrahedron 51*, 4619-4626.

(3.33)   Koll, A., Parasuk, V., Parasuk, W., Karpfen, A., and Wolschann, P. (2004) Theoretical Study on the Intramolecular Hydrogen Bond in Chloro-Substituted *N,N*-Dimethylaminomethylphenols. I. Structural Effects. *J. Mol. Struct. 690*, 165-174.

(3.34)   Koll, A. (2003) Specific Features of Intramolecular Proton Transfer Reaction in Schiff Bases. *Int. J. Mol. Sci. 4*, 434-444.

(3.35)   Enchev, V., Ugrinov, A., and Neykov, G. D. (2000) Intramolecular Proton Transfer Reactions in Internally Hydrogen-Bonded Schiff Bases: Ab Initio and Semiempirical Study. *J. Mol. Struct. (Theochem). 530*, 223-235.

(3.36)   Kabak, M., Elmali, A., and Elerman, Y. (1999) Keto-Enol Tautomerism, Conformations and Structure of *N*-(*2*-Hydroxy-*5*-Methylphenyl)-*2*-Hydroxybenzaldehydeimine. *J. Mol. Struct. 477*, 151-158.

(3.37)   Bach, R. D., Canepa, C., and Glukhovtsev, M. N. (1999) Influence of Electrostatic Effects on Activation Barriers in Enzymatic Reactions: Pyridoxal 5'-Phosphate-Dependent Decarboxylation of α-Amino Acids. *J. Am. Chem. Soc. 121*, 6542-6555.

(3.38)   Bach, R. D., and Canepa, C. (1997) Theoretical Model for Pyruvoyl-Dependent Enzymatic Decarboxylation of *a*-Amino Acids. *J. Am. Chem. Soc. 119*, 11725-11733.

(3.39)   Kiruba, G. S. M., and Wong, M. W. (2003) Tautomeric Equilibria of Pyridoxal-5'-Phosphate (Vitamin B$_6$) and 3-Hydroxypyridine Derivatives: A Theoretical Study of Solvation Effects. *J. Org. Chem. 68*, 2874-2881.

(3.40)   Zgierski, M. Z. (2001) Theoretical Study of Photochromism of *N*-Salicylidene-α-Methylbenzylamine. *J. Chem. Phys. 115*, 8351-8358.

(3.41)   Voltattorni, C. B., Minelli, A., and Turano, C. (1971) Spectral Properties of the Coenzyme Bound to DOPA Decarboxylase from Pig Kidney. *FEBS Letters 17*, 231-235.

(3.42)   Manousek, O., and Zuman, P. (1960) Polarographic Identification of the Reactive Form in the Hydrolysis of Pyridoxal 5'-Phosphate. *Biochim. Biophys. Acta. 44*, 393-394.

(3.43)   Metzler, D. E., and Snell, E. E. (1955) Spectra and Ionization Constants of the Vitamin B6 Group and Related *3*-Hydroxypyridine Derivatives. *J. Am. Chem. Soc. 77*, 2431-2437.

(3.44)   InsightII. Accelrys Molecular Simulations Inc., San Diego, CA.

(3.45)   Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem. 4*, 187-217.

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

(3.47)   Fiser, A., Do, R. K., and Sali, A. (2000) Modeling of Loops in Protein Structures. *Protein Sci. 9*, 1753-1773.

(3.48)   Xiang, Z. X., Soto, C. S., and Honig, B. (2002) Evaluating Confirmation Free Energies: The Colony Energy and its Application to the Problem of Loop Predictions. *Proc. Nat. Acad. Sci. 99*, 7432-7437.

(3.49)   DePristo, M. A., deBakker, P. I. W., Lovell, S. C., and Blundell, T. L. (2003) Ab Initio Construction of Polypeptide Fragments: Efficient Generation of Accurate, Representative Ensembles. *Proteins 51*, 41-55.

(3.50)   Rohl, C. A., Strauss, C. E. M., Chivian, D., and Baker, D. (2004) *Proteins 55*, 656-677.

(3.51)   Zhu, K., Shirts, M., and Friesner, R. (2007) Improved Methods for Side Chain and Loop Predictions Via the Protein Local Optimization Program: Variable Dielectric Model for Implicitly Improving the Treatment of Polarization Effects. *J. Chem. Theory Comput. 3*, 2108-2119.

(3.52)   Felts, A. K., Gallicchio, E., Chekmarev, D., Paris, K. A., Friesner, R. A., and Levy, R. M. (2008) Prediction of Protein Loop Conformations using the AGBNP Implicit Solvent Model and Torsion Angle Sampling. *J. Chem. Theory Comput. 4*, 855-868.

(3.53)   Gao, J. (1993) Potential of Mean Force for the Isomerization of DMF in Aqueous Solution: A Monte Carlo QM/MM Simulation Study. *J. Am. Chem. Soc. 115*, 2930-2935.

(3.54)   Gao, J. (1994) Origin of the Solvent Effects on the Barrier to Amide Isomerization from Combined QM/MM Monte Carlo Simulations. *Proceedings - Indian Academy of Sciences, Chemical Sciences 106*, 507-519.

(3.55)   Wong, K.-Y., and Gao, J. (2007) The Reaction Mechanism of Paraoxon Hydrolysis by Phosphotriesterase from Combined QM/MM Simulations. *Biochemistry 46*, 13352-13369.

(3.56)   Marti, S., Moliner, V., and Tunon, I. (2005) Improving the QM/MM Description of Chemical Processes: A Dual Level Strategy to Explore the Potential Energy Surface in very Large Systems. *J. Chem. Theory Comput. 1*, 1008-1016.

(3.57)   Ferrer, S., Ruiz-Pernia, J. J., Tunon, I., Moliner, V., Garcia-Viloca, M., Gonzalez-Lafont, A., and Lluch, J. M. (2005) A QM/MM Exploration of the Potential Energy Surface of Pyruvate to Lactate Transformation Catalyzed by LDH. Improving the Accuracy of Semiempirical Descriptions. *J. Chem. Theory Comput. 1*, 750-761.58. Gao, J., and Xia, X. (1992) A prior evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science* 258**,** 631-635.

(3.58)   Gao, J. (1992) Absolute Free Energy of Solvation from Monte Carlo Simulations using Combined Quantum and Molecular Mechanical Potentials. *J. Phys. Chem. 96*, 537-540.

(3.59)   Gao, J. (1996) Hybrid QM/MM Simulations: An Alternative Avenue to Solvent Effects in Organic Chemistry. *Acc. Chem. Res. 29*, 298-305.

(3.60)   Gao, J., Ma, S., Major, D. T., Nam, K., Pu, J., and Truhlar, D. G. (2006) Mechanisms and free energies of enzymatic reactions. *Chem. Rev. 106*, 3188-3209.

(3.61)   Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P. (1985) Development and use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc. 107*, 3902-3909.

(3.62)   Koll, A., Rospenk, M., Jagodzinska, E., and Dziembowska, T. (2000) Dipole Moments and Conformation of Schiff Bases with Intramolecular Hydrogen Bonds. *J. Mol. Struct. 552*, 193-204.

(3.63)   Amara, P., Field, M. J., Alhambra, C., and Gao, J. (2000) The Generalized Hybrid Orbital Method for Combined Quantum mechanical/molecular Mechanical Calculations: Formulation and Tests of the Analytical Derivatives. *Theor. Chem. Acc. 104*, 336-343.

(3.64)   Gao, J., Amara, P., Alhambra, C., and Field, M. J. (1998) A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations. *J. Phys. Chem. A. 102*, 4714-4721.

(3.65)   Xie, W., and Gao, J. (2007) Design of a Next Generation Force Field: The X-POL Potential. *J. Chem. Theory Comput. 3*, 1890-1900.

(3.66)   MacKerell Jr., A. D., Bashford, D., Bellott, M., L., D. R.,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., III, R. W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B 102*, 3586-3616.

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

(3.68)   Nam, K., Gao, J., and York, D. M. (2005) An Efficient Linear-Scaling Ewald Method for Long-Range Electrostatic Interactions in Combined QM/MM Calculations. *J. Chem. Theory Comput. 1*, 2-13.

(3.69)   Darden, T., York, D. M., and Pedersen, I. (1993) Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys. 98*, 10089-10092.

(3.70) Verlet, L. (1967) Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev. 159*, 98-103.

(3.71) Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys. 23*, 327-341.

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

(3.73) Davenport, R. C., Bash, P. A., Seaton, B. A., Karplus, M., Petsko, G. A., and Ringe, D. (1991) Structure of the Triosephosphate Isomerase-Phosphoglycolohydroxamate Complex: An Analog of the Intermediate on the Reaction Pathway. *Biochemistry 30*, 5821-5826.

(3.74) Cunningham, M. A., Ho, L. L., Nguyen, D. T., Gillilan, R. E., and Bash, P. A. (1997) Simulation of the Enzyme Reaction Mechanism of Malate Dehydrogenase. *Biochemistry 36*, 4800-4816.

(3.75) Chatfield, D. C., Eurenius, K. P., and Brooks, B. R. (1998) HIV-1 Protease Cleavage Mechanism: A Theoretical Investigation Based on Classical MD Simulation and Reaction Path Calculations using a Hybrid QM/MM Potential. *J. Mol. Struct. (Theochem). 423*, 79-92.

(3.76) Dinner, A. R., Blackburn, G. M., and Karplus, M. (2001) Uracil-DNA Glycosylase Acts by Substrate Autocatalysis. *Nature 413*, 752-755.

(3.77) Devi-Kesavan, L. S., Garcia-Viloca, M., and Gao, J. (2003) Semiempirical QM/MM Potential with Simple Valence Bond (SVB) for Enzyme Reactions. Application to the Nucleophilic Addition Reaction in Haloalkane Dehalogenase. *Theor. Chem. Acc. 109*, 133-139.

(3.78) Hensen, C., Hermann, J. C., Nam, K., Ma, S., Gao, J., and Hoeltje, H.-D. (2004) A Combined QM/MM Approach to Protein-Ligand Interactions: Polarization Effects of the HIV-1 Protease on Selected High Affinity Inhibitors. *J. Med. Chem. 47*, 6673-6680.

(3.79) Garcia-Viloca, M., Truhlar, D. G., and Gao, J. (2003) Reaction-Path Energetics and Kinetics of the Hydride Transfer Reaction Catalyzed by Dihydrofolate Reductase. *Biochemistry 42*, 13558-13575.

(3.80)   Major, D. T., and Gao, J. (2006) A Combined Quantum Mechanical and
           Molecular Mechanical Study of the Reaction Mechanism and α-Amino
           Acidity in Alanine Racemase. *J. Am. Chem. Soc. 128*, 16345-16357.

(3.81)   Frisch, M. J. et al. (2003) *Gaussian 03.* Gaussian, Inc., Pittsburgh, PA.

(3.82)   Stephens, P. J., Devlin, F. J., Chabalowski, C. F., and Frisch, M. J. (1994)
           Ab initio calculation of vibrational absorption and circular dichroism
           spectra using density funcational force fields. *J. Phys. Chem.*, *98*, 11623-
           11627.

(3.83)   Tomasi, J., Mennucci, B., and Cammi, R. (2005) Quantum mechanical
           continuum solvation models. *Chem. Rev. 105*, 2999-3094.

(3.84)   Barone, V., Cossi, M., and Tomasi, J. (1998) Geometry Optimization of
           Molecular Structures in Solution by the Polarizable Continuum Model. *J.
           Comput. Chem. 19*, 404-417.

(3.85)   Shaw, J., Petsko, G. A., and Ringe, D. (1997) Determination of the
           structure of alanine racemase from bacillus stearothermophilus at 1.9-Å
           resolution. *Biochemistry 36*, 1329-1342.

(3.86)   Sun, S., Toney, M. D. (1999) Evidence for a two-base mechanism
           involving tyrosine-265 from arginine-219 mutants of alanine racemase.
           Biochemistry 38, 4058-4065.

(3.87)   Hayashi, H., Tsukiyama, F., Ishii, S., Mizuguchi, H., and Kagamiyama,
           H. (1999) Acid-Base Chemistry of the Reaction of Aromatic L-Amino
           Acid Decarboxylase and Dopa Analyzed by Transient and Steady-State
           Kinetics: Preferential Binding of the Substrate with its Amino Group
           Unprotonated. *Biochemistry 38*, 15615-15622.

Chapter 4

(4.1)   Jansonius, J. N. (1998) Structure, evolution and action of vitamin B_6-
           dependent enzymes, *Curr. Opin. Struct. Biol. 8*, 759-769.

(4.2)   Christen, P., and Metzler, D. E. (1985) *Transaminases* John Wiley &
           Sons, New York

(4.3)   Burkhard, P., Dominici, P., Borri-Voltattorni, C., Jansonius, J. N., and Malashkevich, V. N. (2001) Structural insight into Parkinson's disease treatment from drug-inhibited DOPA decarboxylase, *Nat. Struct. Biol. 8*, 963-967.

(4.4)   Hayashi, H., Mizuguchi, H., and Kagamiyama, H. (1993) Rat liver aromatic L-amino acid decarboxylase: Spectroscopic and kinetic analysis of the coenzyme and reaction intermediates, *Biochemistry 32*, 812-818.

(4.5)   Dominici, P., Moore, P. S., Castellani, S., Bertoldi, M., and Borri Voltattorni, C. (1997) Mutation of cysteine 111 in dopa decarboxylase leads to active site perturbation, *Protein Sci. 6*, 2007-2015.

(4.6)   Ishii, S., Hayashi, H., Okamoto, A., and Kagamiyama, H. (1998) Aromatic L-amino acid decarboxylase: conformational change in the flexible region around Arg334 is required during the transaldimination process, *Protein Sci. 7*, 1802-1810.

(4.7)   Hayashi, H., Tsukiyama, F., Ishii, S., Mizuguchi, H., and Kagamiyama, H. (1999) Acid-base chemistry of the reaction of aromatic L-amino acid decarboxylase and dopa analyzed by transient and steady-state kinetics: preferential binding of the substrate with its amino group unprotonated, *Biochemistry 38*, 15615-15622.

(4.8)   Snider, M. J., and Wolfenden, R. (2000) The rate of spontaneous decarboxylation of amino acids, *J. Am. Chem. Soc. 122*, 11507-11508.

(4.9)   Bertoldi, M., Gonsalvi, M., Contestabile, R., and Borri Voltattorni, C. (2002) Mutation of tyrosine 332 to phenylalanine converts dopa decarboxylase into a decarboxylation-dependent oxidative deaminase, *J. Biol. Chem. 277*, 36357-36362.

(4.10)   Li, J., and Brill, T. B. (2003) Decarboxylation mechanism of amino acids by density functional theory, *J. Phys. Chem. A 107*, 5993-5997.

(4.11)   Walsh, C. (1979) *Enzymatic Reaction in Mechanism*, W. H. Freeman Co., New York.

(4.12)   Eliot, A. C., and Kirsch, J. F. (2004) Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations, *Annu. Rev. Biochem. 73*, 383-415.

(4.13)   Toney, M. D. (2005) Reaction specificity in pyridoxal phosphate enzymes, *Arch. Biochem. Biophys. 433*, 279-287.

184

(4.14)   Bach, R. D., Canepa, C., and Glukhovtsev, M. N. (1999) Influence of electrostatic effects on activation barriers in enzymatic reactions: pyridoxal 5'-phosphate-dependent decarboxylation of α-amino acids, *J. Am. Chem. Soc. 121*, 6542–6555.

(4.15)   Rios, A., Amyes, T. L., and Richard, J. P. (2000) Formation and stability of organic zwitterions in aqueous solution: enolates of the amino acid glycine and its derivatives, *J. Am. Chem. Soc. 122*, 9373–9385.

(4.16)   Rios, A., Crugeiras, J., Amyes, T. L., and Richard, J. P. (2001) Glycine enolates: The large effect of iminium ion formation on alpha-amino carbon acidity, *J. Am. Chem. Soc. 123*, 7949-7950.

(4.17)   Lin, Y. L., and Gao, J. (2010) Internal proton transfer in the external pyridoxal 5'-phosphate Schiff base in dopa decarboxylase, *Biochemistry 49*, 84-94.

(4.18)   Sandmeier, E., Hale, T. I., and Christen, P. (1994) Multiple evolutionary origin of pyridoxal-5'-phosphate-dependent amino acid decarboxylases, *Eur. J. Biochem. 221*, 997-1002.

(4.19)   Matsuda, N., Hayashi, H., Miyatake, S., Kuroiwa, T., and Kagamiyama, H. (2004) Instability of the apo form of aromatic L-amino acid decarboxylase in vivo and in vitro: implications for the involvement of the flexible loop that covers the active site, *J. Biochem. (Tokyo, Japan) 135*, 33-42.

(4.20)   Tancini, B., Dominici, P., Simmaco, M., Schinina, M. E., Barra, D., and Borri Voltattorni, C. (1988) Limited tryptic proteolysis of pig kidney 3,4-dihydroxyphenylalanine decarboxylase, *Arch. Biochem. Biophys. 260*, 569-576.

(4.21)   Ishii, S., Mizuguchi, H., Nishino, J., Hayashi, H., and Kagamiyama, H. (1996) Functionally important residues of aromatic L-amino acid decarboxylase probed by sequence alignment and site-directed mutagenesis, *J. Biochem. (Tokyo, Japan) 120*, 369-376.

(4.22)   Bertoldi, M., Frigeri, P., Paci, M., and Borri Voltattorni, C. (1999) Reaction specificity of native and nicked 3,4-dihydroxyphenylalanine decarboxylase, *J. Biol. Chem. 274*, 5514-5521.

(4.23)   Tong, H., and Davis, L. (1995) 2-Amino-3-ketobutyrate-CoA ligase from beef liver mitochondria: an NMR spectroscopic study of low-barrier hydrogen bonds of a pyridoxal 5'-phosphate-dependent enzyme, *Biochemistry 34*, 3362-3367.

(4.24)   Alarcon, S. H., Olivieri, A. C., Labadie, G. R., Cravero, R. M., and Gonzalez-Sierra, M. (1995) Tautomerism of representative aromatic α-hydroxy carbaldehyde anils as studied by spectroscopic methods and AM1 calculations. Synthesis of 10-hydroxyphenanthrene-9-carbaldehyde, *Tetrahedron 51*, 4619-4626.

(4.25)   Hansen, P. E., Sitkowski, J., Stefaniak, L., Rozwadowski, Z., and Dziembowska, T. (1998) One-bond deuterium isotope effects on ^{15}N chemical shifts in Schiff bases, *Ber. Bunsen-Ges. 102*, 410-413.

(4.26)   Filarowski, A., Koll, A., Rospenk, M., Krol-Starzomska, I., and Hansen, P. E. (2005) Tautomerism of sterically hindered Schiff bases. Deuterium isotope effects on ^{13}C chemical shifts, *J. Phys. Chem. A 109*, 4464-4473.

(4.27)   Claramunt, R. M., Lopez, C., Santa Maria, M. D., Sanz, D., and Elguero, J. (2006) The use of NMR spectroscopy to study tautomerism, *Prog. Nucl. Magn. Reson. Spectrosc. 49*, 169-206.

(4.28)   Sharif, S., Denisov, G. S., Toney, M. D., and Limbach, H.-H. (2006) NMR Studies of solvent-assisted proton transfer in a biologically relevant Schiff base: Toward a distinction of geometric and equilibrium H-bond isotope effects, *J. Am. Chem. Soc. 128*, 3375-3387.

(4.29)   Sharif, S., Fogle, E., Toney, M. D., Denisov, G. S., Shenderovich, I. G., Buntkowsky, G., Tolstoy, P. M., Huot, M. C., and Limbach, H.-H. (2007) NMR localization of protons in critical enzyme hydrogen bonds, *J. Am. Chem. Soc. 129*, 9558-9559.

(4.30)   Hill, M. P., Carroll, E. C., Toney, M. D., and Larsen, D. S. (2008) Rapid Photodynamics of Vitamin B_6 Coenzyme Pyridoxal 5'-Phosphate and Its Schiff Bases in Solution, *J. Phys. Chem. B 112*, 5867-5873.

(4.31)   Shaltiel, S., and Cortijo, M. (1970) The mode of binding of pyridoxal 5'-phosphate in glycogen phosphorylase, *Biochem. Biophys. Res. Commun. 41*, 594-600.

(4.32)   Johnson, G. F., Tu, J. I., Bartlett, M. L., and Graves, D. J. (1970) Physical-chemical studies on the pyridoxal phosphate binding site in sodium borohydride-reduced and native phosphorylase, *J. Biol. Chem. 245*, 5560-5568.

(4.33)   Barboni, E., Borri Voltattorni, C., D'Erme, M., Fiori, A., Minelli, A., and Rosei, M. A. (1982) Inhibitors binding to L-aromatic amino acid decarboxylase, *Life Sciences 31*, 1519-1524.

186

(4.34)   Borri Voltattorni, C., Minelli, A., and Turano, C. (1971) Spectral properties of the coenzyme bound to DOPA decarboxylase from pig kidney, *Febs Lett 17*, 231-235.

(4.35)   Olmo, M., Sanchez-Jimenez, F., Medina, M., and Hayashi, H. (2002) Spectroscopic analysis of recombinant rat histidine decarboxylase, *J. Biochem. (Tokyo, Japan) 132*, 433-439.

(4.36)   Fleming, J. V., Sanchez-Jimenez, F., Moya-Garcia, A. A., Langlois, M. R., and Wang, T. C. (2004) Mapping of catalytically important residues in the rat L-histidine decarboxylase enzyme using bioinformatic and site-directed mutagenesis approaches, *Biochem. J. 379*, 253-261.

(4.37)   Minelli, A., Charteris, A. T., Voltattorni, C. B., and John, R. A. (1979) Reactions of DOPA (3,4-dihydroxyphenylalanine) decarboxylase with DOPA, *Biochem. J. 183*, 361-368.

(4.38)   Kirsch, J. F., Eichele, G., Ford, G. C., Vincent, M. G., Jansonius, J. N., Gehring, H., and Christen, P. (1984) Mechanism of action of aspartate aminotransferase proposed on the basis of its spatial structure, *J. Mol. Biol. 174*, 497-525.

(4.39)   Toney, M. D. (2001) Computational studies on nonenzymatic and enzymatic pyridoxal phosphate catalyzed decarboxylations of 2-aminoisobutyrate, *Biochemistry 40*, 1378-1384.

(4.40)   Sicinska, D., Truhlar, D. G., and Paneth, P. (2005) Dependence of transition state structure on substrate: The intrinsic C-13 kinetic isotope effect is different for physiological and slow substrates of the ornithine decarboxylase reaction because of different hydrogen bonding structures, *J. Am. Chem. Soc. 127*, 5414-5422.

(4.41)   Warshel, A., and Levitt, M. (1976) *J. Mol. Biol. 103*, 227.

(4.42)   Field, M. J., Bash, P. A., and Karplus, M. (1990) *J. Comput. Chem. 11*, 700.

(4.43)   Gao, J., and Xia, X. (1992) A prior evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations, *Science 258*, 631-635.

(4.44)   Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P. (1985) *J. Am. Chem. Soc. 107*, 3902-3909.

(4.45)   Amara, P., Field, M. J., Alhambra, C., and Gao, J. (2000) The generalized hybrid orbital method for combined quantum mechanical/molecular mechanical calculations: formulation and tests of the analytical derivatives, *Theor. Chem. Acc. 104*, 336-343.

(4.46)   Gao, J., Amara, P., Alhambra, C., and Field, M. J. (1998) A Generalized Hybrid Orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations, *J. Phys. Chem. A 102*, 4714-4721.

(4.47)   MacKerell, A. D., Jr., Brooks, B., Brooks, C. L., III, Nilsson, L., Roux, B., Won, Y., and Karplus, M. (1998) *The Encyclopedia of Computational Chemistry*, Vol. 1, John Wiley & Sons, Chichester.

(4.48)   Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *Journal of Chemical Physics 79*, 926-935.

(4.49)   Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and Rosenberg, J. M. (1992) The weighted histogram analysis method for free-fnergy calculations on biomolecules .1. The method, *J. Comput. Chem. 13*, 1011-1021.

(4.50)   Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1995) Multidimensional free-energy calculations using the weighted histogram analysis method, *J. Comput. Chem. 16*, 1339-1350.

(4.51)   Miertus, S., and Tomasi, J. (1982) *Chem. Phys. 65*, 239

(4.52)   Cossi, M., Barone, V., Cammi, R., and Tomasi, J. (1996) *Chem. Phys. Lett. 255*, 327.

(4.53)   Cancès, M. T., Mennucci, B., and Tomasi, J. (1997) *J. Chem. Phys. 107*, 3032

(4.54)   Barone, V., Cossi, M., and Tomasi, J. (1997) *J. Chem. Phys. 107*, 3210

(4.55)   Cossi, M., Barone, V., Mennucci, B., and Tomasi, J. (1998) *Chem. Phys. Lett. 286*, 253.

(4.56)   Barone, V., and Cossi, M. (1998) *J. Phys. Chem. A 102*, 1995

(4.57)   Barone, V., Cossi, M., and Tomasi, J. (1998) *J. Comp. Chem. 19*, 404.

(4.58)   Mennucci, B., and Tomasi, J. (1997) *J. Chem. Phys. 106*, 5151.

(4.59)   Tomasi, J., Mennucci, B., and Cancès, E. (1999) *J. Mol. Struct. (Theochem) 464*, 211.

(4.60)   Silla, E., Tunon, I., and Pascual-Ahuir, J. L. (1994) *J. Comp. Chem. 15*, 1127.

(4.61)   Major, D. T., and Gao, J. (2006) A combined quantum mechanical and molecular mechanical study of the reaction mechanism and a-amino acidity in alanine aacemase, *J. Am. Chem. Soc. 128*, 16345-16357.

(4.62)   Major, D. T., Nam, K., and Gao, J. (2006) Transition state stabilization and α-amino carbon acidity in alanine racemase, *J. Am. Chem. Soc. 128*, 8114-8115.

(4.63)   Dominici, P., Tancini, B., and Voltattorni, C. B. (1985) Chemical modification of pig-kidney 3,4-dihydroxyphenylalanine decarboxylase with diethyl pyrocarbonate - evidence for an essential histidyl residue, *J. Biol. Chem. 260*, 583-589.

(4.64)   Bertoldi, M., Castellani, S., and Bori Voltattorni, C. (2001) Mutation of residues in the coenzyme binding pocket of Dopa decarboxylase. Effects on catalytic properties, *Eur. J. Biochem. 268*, 2975-2981.

(4.65)   Tilley, K., Akhtar, M., and Gani, D. (1994) The stereochemical course of decarboxylation, transamination and elimination-reactions catalyzed by escherichia-coli glutamic-acid decarboxylase, *J. Chem. Soc. Perkin Trans. 1*, 3079-3087.

(4.66)   Akhtar, M., Stevenson, D. E., and Gani, D. (1990) Fern L-methionine decarboxylase - kinetics and mechanism of decarboxylation and abortive transamination, *Biochemistry 29*, 7648-7660.

(4.67)   Joseph, D., Petsko, G. A., and Karplus, M. (1990) Anatomy of a conformational change - hinged lid motion of the triosephosphate isomerase loop, *Science 249*, 1425-1428.

(4.68)   First, E. A., and Fersht, A. R. (1993) Mutational and kinetic-analysis of a mobile loop in tyrosyl-transfer RNA-synthetase, *Biochemistry 32*, 13658-13663.

(4.69)   Kato, H., Tanaka, T., Yamaguchi, H., Hara, T., Nishioka, T., Katsube, Y., and Oda, J. (1994) Flexible loop that is novel catalytic machinery in a ligase - atomic-structure and function of the loopless glutathione synthetase, *Biochemistry 33*, 4995-4999.

(4.70)   Larson, E. M., Larimer, F. W., and Hartman, F. C. (1995) Mechanistic insights provided by deletion of a flexible loop at the active-site of ribulose-1,5-bisphosphate carboxylase oxygenase, *Biochemistry 34*, 4531-4537.

(4.71)   Toney, M. D., Hohenester, E., Keller, J. W., and Jansonius, J. N. (1995) Structural and mechanistic analysis of two refined crystal structures of the pyridoxal phosphate-dependent enzyme dialkylglycine decarboxylase, *J. Mol. Biol. 245*, 151-179.

(4.72)   Shen, B. W., Hennig, M., Hohenester, E., Jansonius, J. N., and Schirmer, T. (1998) Crystal structure of human recombinant ornithine aminotransferase, *J. Mol. Biol. 277*, 81-102.

(4.73)   Peisach, D., Chipman, D. M., Van Ophem, P. W., Manning, J. M., and Ringe, D. (1998) D-cycloserine inactivation of D-amino acid aminotransferase leads to a stable noncovalent protein complex with an aromatic cycloserine-PLP derivative, *J. Am. Chem. Soc. 120*, 2268-2274.

(4.74)   Yamada, H., and O'Leary, M. H. (1978) Stereochemistry of reactions catalyzed by glutamate decarboxylase, *Biochemistry 17*, 669-672.

(4.75)   Alhambra, C., Corchado, J., Sanchez, M. L., Garcia-Viloca, M., Gao, J., and Truhlar, D. G., (2001) Canonical variational theory for enzyme kinetics with the potential of mean force and multidimensional quantum mechanical tunneling dynamics. Theory and application to liver alcohol dehydrogenase. *J. Phys. Chem. B*, *105*, 11326-11340.

(4.76)   Gao, J. (1995) An automated procedure for simulating chemical reactions in solution. Application to the decarboxylation of 3-carboxybenzisoxazole in water, *J. Am. Chem. Soc. 117*, 8600-8607.

(4.77)   Wu, N., Mo, Y., Gao, J., and Pai, E. F. (2000) Electrostatic stress in catalysis: structure and mechanism of the enzyme orotidine monophosphate decarboxylase, *Proc. Natl. Acad. Sci. 97*, 2017-2022.

(4.78)   Major, D. T., and Gao, J. (2007) An integrated path integral and free-energy perturbation-umbrella sampling method for computing kinetic isotope effects of chemical reactions in solution and in enzymes, *J. Chem. Theor. Comput. 3*, 949-960.

(4.79)   Wong, K.-Y., and Gao, J. (2007) The reaction mechanism of paraoxon hydrolysis by phosphotriesterase from combined QM/MM simulations. *Biochemistry 46*, 13352-13369.

(4.80)   Davenport, R. C., Bash, P. A., Seaton, B. A., Karplus, M., Petsko, G. A., and Ringe, D. (1991) Structure of the triosephosphate isomerase-phosphoglycolohydroxamate complex: an analog of the intermediate on the reaction pathway. *Biochemistry 30*, 5821-5826.

(4.81)   Cunningham, M. A., Ho, L. L., Nguyen, D. T., Gillilan, R. E., and Bash, P. A. (1997) Simulation of the enzyme reaction mechanism of malate dehydrogenase. *Biochemistry 36*, 4800-4816.

(4.82)   Chatfield, D. C., Eurenius, K. P., and Brooks, B. R. (1998) HIV-1 protease cleavage mechanism: a theoretical investigation based on classical MD simulation and reaction path calculations using a hybrid QM/MM potential. *J. Mol. Struct. (Theochem). 423*, 79-92.

(4.83)   Dinner, A. R., Blackburn, G. M., and Karplus, M. (2001) Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature 413*, 752-755.

(4.84)   Devi-Kesavan, L. S., Garcia-Viloca, M., and Gao, J. (2003) Semiempirical QM/MM potential with simple valence bond (SVB) for enzyme reactions. Application to the nucleophilic addition reaction in haloalkane dehalogenase. *Theor. Chem. Acc. 109*, 133-139.

(4.85)   Hensen, C., Hermann, J. C., Nam, K., Ma, S., Gao, J., and Hoeltje, H.-D. (2004) A combined QM/MM approach to protein-ligand interactions: polarization effects of the HIV-1 protease on selected high affinity inhibitors. *J. Med. Chem. 47*, 6673-6680.

(4.86)   Garcia-Viloca, M., Truhlar, D. G., and Gao, J. (2003) Reaction-path energetics and kinetics of the hydride transfer reaction catalyzed by dihydrofolate reductase. *Biochemistry 42*, 13558-13575.

(4.87)   Frey, P. A., and Hegeman, A. D. (2007) Enzymatic Reaction Mechanisms, Oxford University Press, New York.

(4.88)   Sharif, S., Huot, M. C., Tolstoy, P. M., Toney, M. D., Jonsson, K. H. M., and Limbach, H.-H. (2007) ^{15}N nuclear magnetic resonance studies of acid-base properties of pyridoxal 5'-phosphate aldimines in aqueous solution. *J. Phys. Chem. B 111*, 3869-3876.

(4.89)   Banner, D. W., Bloomer, A. C.; Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D., and Waley, S. G. (1976) Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 Å resolution: using amino acid sequence data. *Nature, 255*, 609-614.

(4.90)   Alber, T., Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, David, Rivers, P. S., and Wilson, I. A. (1981) On the three-dimensional structure and catalytic mechanism of triose phosphate isomerase. *Philos. Trans. Roy. Soc. B Biol. Sci. 293*, 159-171.

(4.91)   Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K., and Jansonius, J. N. (1991) Structural conservation in parallel β/α-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry, 30*, 9161-9169.

(4.92)   Schubert, H. L., Fauman, E. B., Stuckey, J. A., Dixon, J. E., and Saper, M. A. (1995) A ligand-induced conformational change in the Yersinia protein tyrosine phosphatase. *Protein Sci.*, *4*, 1904-1913.

(4.93)   Wierenga, R. K., Borcher, T. V., and Noble, M. E. M. (1992) Crystallographic binding studies with triosephosphate isomerases: conformational changes induced by substrate and substrate-analogues. *FEBS Lett.*, *307*, 34-39.

Chapter 5

(5.1)   Garcia-Viloca, M., Gao, J., Karplus, M., and Truhlar, D. G. (2004) How enzymes work: Analysis by modern rate theory and computer simulations, *Science 303*, 186-195.

(5.2)   Nagel, Z. D., and Klinman, J. P. (2006) Tunneling and dynamics in enzymatic hydride transfer, *Chem. Rev. 106*, 3095-3118.

(5.3)   Pu, J., Gao, J., and Truhlar, D. G. (2006) Multidimensional tunneling, recrossing, and the transmission coefficient for enzymatic reactions, *Chem. Rev. 106*, 3140-3169.

(5.4)   Kohen, A., and Limbach, H.-H. (2005) *Isotope Effects in Chemistry and Biology*, Taylor and Francis Group, CRC press, New York.

(5.5)   O'Leary, M. H. (1982) *Stable Isotopes*, Elsevier, Amsterdam.

(5.6)   O'Leary, M. H. (1988) Transition-state structures in enzyme-catalyzed decarboxylations, *Acc. Chem. Res. 21*, 450-455.

(5.7)   Hayashi, H., Tsukiyama, F., Ishii, S., Mizuguchi, H., and Kagamiyama, H. (1999) Acid-base chemistry of the reaction of aromatic L-amino acid decarboxylase and dopa analyzed by transient and steady-state kinetics: preferential binding of the substrate with its amino group unprotonated, *Biochemistry 38*, 15615-15622.

(5.8)   O'Leary, M. H., Richards, D. T., and Hendrickson, D. W. (1970) Carbon isotope effects on the enzymatic decarboxylation of glutamic acid, *J. Am. Chem. Soc. 92*, 4435-4440.

(5.9)   O'Leary, M. H., and Piazza, G. J. (1981) Medium effects in enzyme-catalyzed decarboxylations, *Biochemistry 20*, 2743-2748.

(5.10)   O'Leary, M. H., Yamada, H., and Yapp, C. J. (1981) Multiple isotope effect probes of glutamate decarboxylase, *Biochemistry 20*, 1476-1481.

(5.11)   Abell, L. M., and Oleary, M. H. (1988) Isotope effect studies of the pyridoxal 5'-phosphate dependent histidine-decarboxylase from *Morganella Morganii*, *Biochemistry 27*, 5927-5933.

(5.12)   Swanson, T., Brooks, H. B., Osterman, A. L., O'Leary, M. H., and Phillips, M. A. (1998) Carbon-13 isotope effect studies of *Trypanosoma brucei* ornithine decarboxylase, *Biochemistry 37*, 14943-14947.

(5.13)   Zhou, X., Jin, X., Medhekar, R., Chen, X., Dieckmann, T., and Toney, M. D. (2001) Rapid kinetic and isotopic studies on dialkylglycine decarboxylase, *Biochemistry 40*, 1367-1377.

(5.14)   O'Leary, M. H. (1976) *Studies of Enzyme Reaction Mechanisms by Means of Heavy-atom Isotope Effects*, University Park Press, Baltimore.

(5.15)   Major, D. T., and Gao, J. (2007) An integrated path integral and free-energy perturbation-umbrella sampling method for computing kinetic isotope effects of chemical reactions in solution and in enzymes, *J. Chem. Theor. Comput. 3*, 949-960.

(5.16)   Gao, J., and Xia, X. (1992) A prior evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations, *Science 258*, 631-635.

(5.17)   Major, D. T., and Gao, J. (2005) Implementation of the bisection sampling method in path integral simulations, *J. Mol. Graph. Modell. 24*, 121-127.

(5.18)   Sprik, M., Klein, M. L., and Chandler, D. (1985) Staging - a sampling technique for the Monte-Carlo evaluation of path-integrals, *Phys. Rev. B 31*, 4234-4244.

(5.19)   Valleau, J. P., and Torrie, G. M. (1977) *In Morden Theoretical Chemistry*, Vol. 5, Plenum, New York.

(5.20)   Wong, K. Y., and Gao, J. (2008) Systematic approach for computing zero-point energy, quantum partition function, and tunneling effect based on Kleinert's variational perturbation, *J. Chem. Theor. Comput. 4*, 1409-1422.

(5.21)   Major, D. T., York, D. M., and Gao, J. (2005) Solvent polarization and kinetic isotope effects in nitroethane deprotonation and implications to the nitroalkane oxidase reaction, *J. Am. Chem. Soc. 127*, 16374-16375.

(5.22)   Major, D. T., Garcia-Viloca, M., and Gao, J. (2006) Path integral simulations of proton transfer reactions in aqueous solution using combined QM/MM potentials, *J. Chem. Theor. Comput. 2*, 236-245.

(5.23)   Major, D. T., Heroux, A., Orville, A. M., Valley, M. P., Fitzpatrick, P. F., and Gao, J. (2009) Differential quantum tunneling contributions in nitroalkane oxidase catalyzed and the uncatalyzed proton transfer reaction, *Proc. Natl. Acad. Sci. 106*, 20734-20739.

(5.24)   Rubinstein, A., and Major, D. T. (2009) Catalyzing racemizations in the absence of a cofactor: the reaction mechanism in proline racemase, *J. Am. Chem. Soc. 131*, 8513-8521.

(5.25)   Feynman, R. P., and Hibbs, A. R. (1965) *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York.

(5.26)   Voth, G. A., Chandler, D., and Miller, W. H. (1989) Rigorous formulation of quantum transition-state theory and its dynamical corrections, *J. Chem. Phys. 91*, 7749-7760.

(5.27)   Gillan, M. J. (1988) The quantum simulation of hydrogen in metals, *Philos. Mag. A 58*, 257-283.

(5.28)   Voth, G. A. (1996) Path-integral centroid methods in quantum statistical mechanics and dynamics, *Adv. Chem. Phys. 93*, 135-218.

(5.29)   Cao, J. S., and Voth, G. A. (1994) The formulation of quantum-statistical mechanics based on the feynman path centroid density .5. Quantum

instantaneous normal-mode theory of liquids, *J. Chem. Phys. 101*, 6184-6192.

(5.30) Gillan, M. J. (1987) Quantum simulation of hydrogen in metals, *Phys. Rev. Lett. 58*, 563-566.

(5.31) Gehlen, J. N., Chandler, D., Kim, H. J., and Hynes, J. T. (1992) Free-energies of electron transfer, *J. Phys. Chem. 96*, 1748-1753.

(5.32) Messina, M., Schenter, G. K., and Garrett, B. C. (1993) Centroid-density quantum rate theory - Variational optimization of the dividing surface, *J. Chem. Phys. 98*, 8525-8536.

(5.33) Hwang, J. K., and Warshel, A. (1993) A quantized classical path approach for calculations of quantum-mechanical rate constants, *J. Phys. Chem. 97*, 10053-10058.

(5.34) Hwang, J. K., and Warshel, A. (1996) How important are quantum mechanical nuclear motions in enzyme catalysis?, *J. Am. Chem. Soc. 118*, 11745-11751.

(5.35) Hwang, J. K., Chu, Z. T., Yadav, A., and Warshel, A. (1991) Simulations of quantum mechanical corrections for rate constants of hydride-transfer reactions in enzymes and solutions, *J. Phys. Chem. 95*, 8445–8448.

(5.36) Feierberg, I., Luzhkov, V., and Aqvist, J. (2000) Computer simulation of primary kinetic isotope effects in the proposed rate-limiting step of the glyoxalase I catalyzed reaction, *J. Biol. Chem. 275*, 22657-22662.

(5.37) Villa, J., and Warshel, A. (2001) Energetics and dynamics of enzymatic reactions, *J. Phys. Chem. B 105*, 7887-7907.

(5.38) Voth, G. A. (1993) Feynman path-integral formulation of quantum-mechanical transition-state theory, *J. Phys. Chem. 97*, 8365-8377.

(5.39) Olsson, M. H. M., Siegbahn, P. E. M., and Warshel, A. (2004) Simulations of the large kinetic isotope effect and the temperature dependence of the hydrogen atom transfer in lipoxygenase, *J. Am. Chem. Soc. 126*, 2820-2828.

(5.40) Ceperley, D. M. (1995) Path-integrals in the theory of condensed helium, *Rev. Mod. Phys. 67*, 279-355.

(5.41)　　Ceperley, D. M., and Pollock, E. L. (1986) Path-integral computation of the low-temperature properties of liquid-He-4, *Phys. Rev. Lett. 56*, 351-354.

(5.42)　　Lévy, P. (1939) *Compos. Math. 7*, 283.

(5.43)　　Field, M. J., Bash, P. A., and Karplus, M. (1990) *J. Comput. Chem. 11*, 700.

(5.44)　　Gao, J. (1995) Methods and applications of combined quantum mechanical and molecular mechanical potentials, *Rev. Comput. Chem. 7*, 119-185.

(5.45)　　Amara, P., Field, M. J., Alhambra, C., and Gao, J. (2000) The generalized hybrid orbital method for combined quantum mechanical/molecular mechanical calculations: formulation and tests of the analytical derivatives, *Theor. Chem. Acc. 104*, 336-343.

(5.46)　　Gao, J., Amara, P., Alhambra, C., and Field, M. J. (1998) A Generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations, *J. Phys. Chem. A 102*, 4714-4721.

(5.47)　　Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P. (1985) *J. Am. Chem. Soc. 107*, 3902-3909.

(5.48)　　MacKerell, A. D., Jr., Brooks, B., Brooks, C. L., III, Nilsson, L., Roux, B., Won, Y., and Karplus, M. (1998) *The Encyclopedia of Computational Chemistry*, Vol. 1, John Wiley & Sons, Chichester.

(5.49)　　Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (2009) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem. 4*, 187-217.

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

(5.51)　　Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys. 79*, 926-935.

(5.52) Gao, J. (1995) An automated procedure for simulating chemical reactions in solution. Application to the decarboxylation of 3-carboxybenzisoxazole in water, *J. Am. Chem. Soc. 117*, 8600-8607.

(5.53) Wu, N., Mo, Y., Gao, J., and Pai, E. F. (2000) Electrostatic stress in catalysis: structure and mechanism of the enzyme orotidine monophosphate decarboxylase, *Proc. Natl. Acad. Sci. 97*, 2017-2022.

(5.54) Dominici, P., Moore, P. S., Castellani, S., Bertoldi, M., and Borri Voltattorni, C. (1997) Mutation of cysteine 111 in dopa decarboxylase leads to active site perturbation, *Protein Sci. 6*, 2007-2015.

(5.55) Ishii, S., Hayashi, H., Okamoto, A., and Kagamiyama, H. (1998) Aromatic L-amino acid decarboxylase: conformational change in the flexible region around Arg334 is required during the transaldimination process, *Protein Sci. 7*, 1802-1810.

(5.56) Bertoldi, M., Gonsalvi, M., Contestabile, R., and Borri Voltattorni, C. (2002) Mutation of tyrosine 332 to phenylalanine converts dopa decarboxylase into a decarboxylation-dependent oxidative deaminase, *J. Biol. Chem. 277*, 36357-36362.

(5.57) Kalyankar, G. D., and Snell, E. E. (1962) Pyridoxal-catalyzed decarboxylation of amino acids, *Biochemistry 1*, 594-600.

(5.58) Snider, M. J., and Wolfenden, R. (2000) The rate of spontaneous decarboxylation of amino acids, *J. Am. Chem. Soc. 122*, 11507-11508.

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

(5.60) Sicinska, D., Truhlar, D. G., Paneth, P. (2001) Solvent-dependent ttransition states for decarboxylations. *J. Am. Chem. Soc. 123*, 7683-7686.

Chapter 6

(6.1) Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., and Casero, R. A. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell 119*, 941-953.

(6.2)   Forneris, F., Binda, C., Vanoni, M. A., Mattevi, A., and Battaglioli, E. (2005) Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. *FEBS Lett. 579*, 2203-2207.

(6.3)   Forneris, F., Binda, C., Dall'Aglio, A., Fraaije, M. W., Battaglioli, E., and Mattevi, A. (2006) A highly specific mechanism of histone H3-K4 recognition by histone demethylase LSD1. *J. Biol. Chem. 281*, 35289-35295.

(6.4)   Forneris, F., Binda, C., Battaglioli, E., and Mattevi, A. (2008) LSD1: oxidative chemistry for multifaceted functions in chromatin regulation. *Trend. Biochem. Sci. 33*, 181-189.

(6.5)   Shi, Y., and Whetstine, J. R. (2007) Dynamic regulation of histone lysine methylation by demethylases. *Mol. Cell 25*, 1-14.

(6.6)   Marmorstein, R., and Trievel, R. C. (2009) Histone modifying enzymes: structures, mechanisms, and specificities. *Biochim. Biophys. Acta. 1789*, 58-68.

(6.7)   Shi, Y. J., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005) Regulation of LSD1 histone demethylase activity by its associated factors. *Mol. Cell 19*, 857-864.

(6.8)   Smith, B. C., and Denu, J. M. (2009) Chemical mechanisms of histone lysine and arginine modifications. *Biochim. Biophys. Acta 1789*, 45-57.

(6.9)   Allis, C. D., Berger, S. L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhattar, R., Shilatifard, A., Workman, J., and Zhang, Y. (2007) New nomenclature for chromatin-modifying enzymes. *Cell 131*, 633-636.

(6.10)   Wissmann, M., Yin, N., Muller, J. M., Greschik, H., Fodor, B. D., Jenuwein, T., Vogler, C., Schneider, R., Gunther, T., Buettner, R., Metzger, E., and Schule, R. (2007) Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nat. Cell Biol. 9*, 347-353.

(6.11)   Garcia-Bassets, I., Kwon, Y. S., Telese, F., Prefontaine, G. G., Hutt, K. R., Cheng, C. S., Ju, B. G., Ohgi, K. A., Wang, J., Escoubet-Lozach, L., Rose, D. W., Glass, C. K., Fu, X. D., and Rosenfeld, M. G. (2007) Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. *Cell 128*, 505-518.

(6.12)   Ballas, N., Battaglioli, E., Atouf, F., Andres, M. E., Chenoweth, J., Anderson, M. E., Burger, C., Moniwa, M., Davie, J. R., Bowers, W. J., Federoff, H. J., Rose, D. W., Rosenfeld, M. G., Brehm, P., and Mandel, G. (2001) Regulation of neuronal traits by a novel transcriptional complex. *Neuron 31*, 353-365.

(6.13)   Humphrey, G. W., Wang, Y., Russanova, V. R., Hirai, T., Qin, J., Nakatani, Y., and Howard, B. H. (2001) Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1. *J. Biol. Chem. 276*, 6817-6824.

(6.14)   You, A., Tong, J. K., Grozinger, C. M., and Schreiber, S. L. (2001) CoREST is an integral component of the CoREST-human histone deacetylase complex. *PNAS 98*, 1454-1458.

(6.15)   Hakimi, M. A., Bochar, D. A., Chenoweth, J., Lane, W. S., Mandel, G., and Shiekhattar, R. (2002) A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *PNAS 99*, 7420-7425.

(6.16)   Hakimi, M. A., Dong, Y., Lane, W. S., Speicher, D. W., and Shiekhattar, R. (2003) A candidate X-linked mental retardation gene is a component of a new family of histone deacetylase-containing complexes. *J. Biol. Chem. 278*, 7234-7239.

(6.17)   Shi, Y., Sawada, J., Sui, G., Affar el, B., Whetstine, J. R., Lan, F., Ogawa, H., Luke, M. P., and Nakatani, Y. (2003) Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature 422*, 735-738.

(6.18)   Lee, M. G., Wynder, C., Cooch, N., and Shiekhattar, R. (2005) An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature 437*, 432-435.

(6.19)   Huang, J., Sengupta, R., Espejo, A. B., Lee, M. G., Dorsey, J. A., Richter, M., Opravil, S., Shiekhattar, R., Bedford, M. T., Jenuwein, T., and Berger, S. L. (2007) p53 is regulated by the lysine demethylase LSD1. *Nature 449*, 105-108.

(6.20)   Chen, Y., Yang, Y., Wang, F., Wan, K., Yamane, K., Zhang, Y., and Lei, M. (2006) Crystal structure of human histone lysine-specific demethylase 1 (LSD1). *PNAS 103*, 13956-13961.

(6.21) Stavropoulos, P., Blobel, G., and Hoelz, A. (2006) Crystal structure and mechanism of human lysine-specific demethylase-1. *Nat. Struct. Mol. Biol. 13*, 626-632.

(6.22) Yang, M., Gocke, C. B., Luo, X., Borek, D., Tomchick, D. R., Machius, M., Otwinowski, Z., and Yu, H. (2006) Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. *Mol. Cell 23*, 377-387.

(6.23) Forneris, F., Binda, C., Adamo, A., Battaglioli, E., and Mattevi, A. (2007) Structural basis of LSD1-CoREST selectivity in histone H3 recognition. *J. Biol. Chem. 282*, 20070-20074.

(6.24) Yang, M., Culhane, J. C., Szewczuk, L. M., Gocke, C. B., Brautigam, C. A., Tomchick, D. R., Machius, M., Cole, P. A., and Yu, H. (2007) Structural basis of histone demethylation by LSD1 revealed by suicide inactivation. *Nat. Struct. Mol. Biol. 14*, 535-539.

(6.25) Qian, C. M., Zhang, Q., Li, S. D., Zeng, L., Walsh, M. J., and Zhou, M. M. (2005) Structure and chromosomal DNA binding of the SWIRM domain. *Nat. Struct. Mol. Biol. 12*, 1078-1085.

(6.26) Da, G. P., Lenkart, J., Zhao, K. H., Shiekhattar, R., Cairns, B. R., and Marmorstein, R. (2006) Structure and function of the SWIRM domain, a conserved protein module found in chromatin regulatory complexes. *Proc. Natl. Acad. Sci. USA 103*, 2057-2062.

(6.27) Tochio, N., Umehara, T., Koshiba, S., Inoue, M., Yabuki, T., Aoki, M., Seki, E., Watanabe, S., Tomo, Y., Hanada, M., Ikari, M., Sato, M., Terada, T., Nagase, T., Ohara, O., Shirouzu, M., Tanaka, A., Kigawa, T., and Yokoyama, S. (2006) Solution structure of the SWIRM domain of human histone demethylase LSD1. *Structure 14*, 457-468.

(6.28) Forneris, F., Binda, C., Vanoni, M. A., Battaglioli, E., and Mattevi, A. (2005) Human histone demethylase LSD1 reads the histone code. *J. Biol. Chem. 280*, 41360-41365.

(6.29) Wysocka, J., Milne, T. A., and Allis, C. D. (2005) Taking LSD 1 to a new high. *Cell 122*, 654-658.

(6.30) Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. J., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. J. (1977) *J. Mol. Biol. 112*, 535-542.

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

(6.32)   Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (2009) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem. 4*, 187-217.

(6.33)   Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem. 26*, 1781-1802.

(6.34)   MacKerell, A. D., Jr., Brooks, B., Brooks, C. L., III, Nilsson, L., Roux, B., Won, Y., and Karplus, M. (1998) *The Encyclopedia of Computational Chemistry*, Vol. 1, John Wiley & Sons, Chichester.

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

(6.36)   Aqvist, J. (1986) A simple way to calculate the axis of an α-helix. *Comput. Chem., 10*, 97-99.

Appendix A

Supplementary Material for Chapter 2:

Supercritical Solvation with Explicit Polarization: Towards an Understanding of Solvent Effects and the Solvatochromic Shifts of Acetone from Steam Vapor to Ambient Aqueous Solution

Table A1. Computed solvatochromic shift (Δv) and the corresponding reduced density (ρ_r) in the $n \rightarrow \pi^*$ excitation of acetone in water fluid at various temperatures

Temperature (°C)	ρ_r (g/cm^3)	Δv (cm^{-1})
25	3.11	1244.87
50	3.07	1023.75
50	3.10	1053.52
100	3.00	991.50
100	3.02	1003.71
100	3.08	910.30
100	3.10	1041.48
200	2.74	706.58
200	2.88	807.85
200	3.09	955.59
300	2.53	663.80
300	2.68	702.55
300	2.81	754.53
400	1.90	416.84
400	2.02	501.19
400	2.15	555.80

400	2.36	603.43
450	1.11	289.93
450	1.49	380.18
450	1.75	366.71
450	1.91	399.46
500	0.05	48.36
500	0.09	75.23
500	0.21	111.35
500	0.28	141.84
500	0.45	192.32
500	0.70	210.45
500	1.05	280.93
500	1.42	317.69
500	1.64	387.04

Supplementary Material for Chapter 3:

Internal Proton Transfer in the External Pyridoxal 5´-Phosphate

Schiff Base in Dopa Decarboxylase

Figure B1   Gas-phase structures of oxoenamine and hydroxyimine Schiff bases fully optimized at B3LYP/6-311+G(d,p) for reactions **a1**, **a2**, **b1**, and **c1**; HF/6-311+G(d,p) for reactions **b2**, **c2**, and **d2**; HF/6-31+G(d) for reaction **d1**

oxoenamine tautomer   hydroxyimine tautomer

oxoenamine tautomer   hydroxyimine tautomer

oxoenamine tautomer   hydroxyimine tautomer

c1

c2
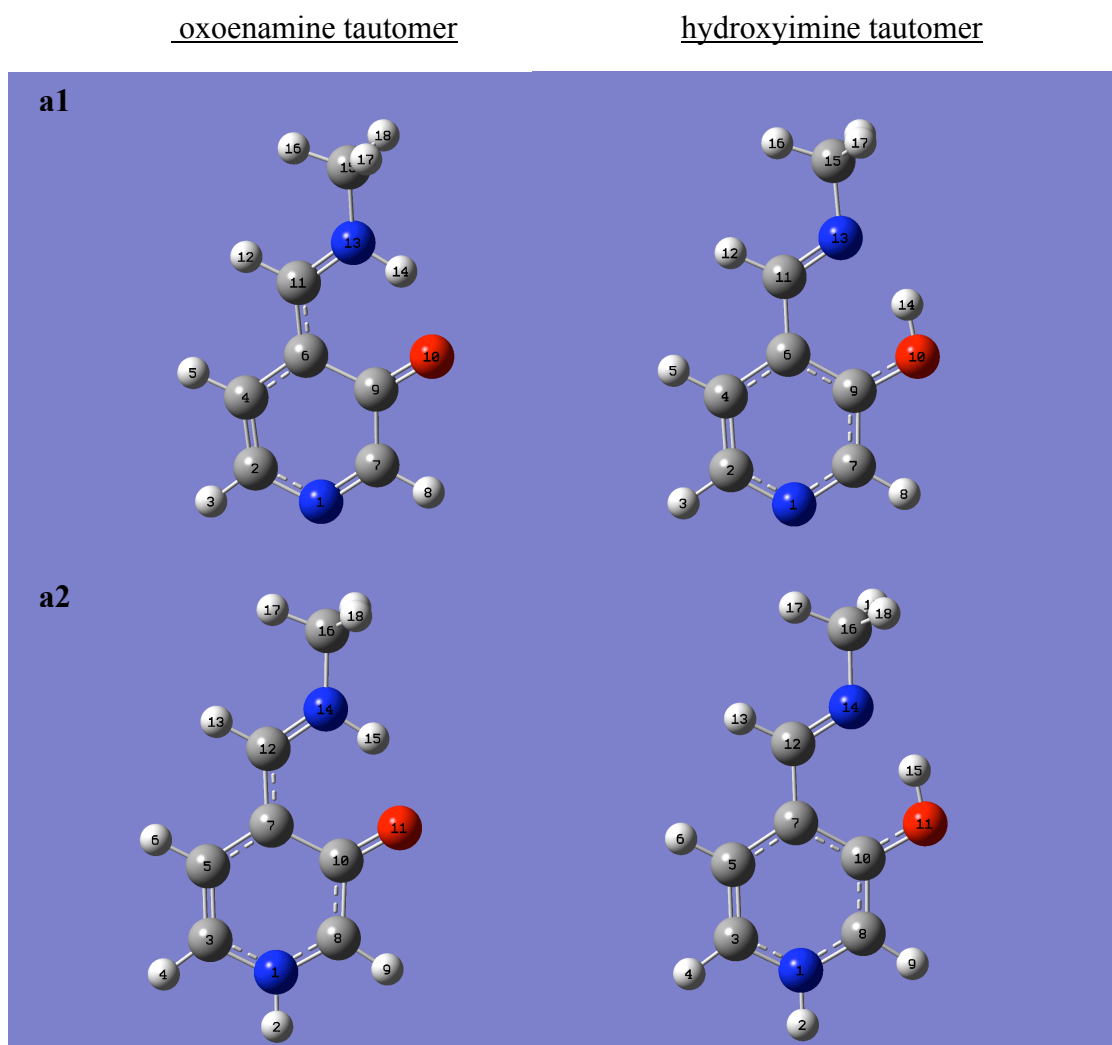
Figure B1 (continued)

oxoenamine tautomer   hydroxyimine tautomer

d1
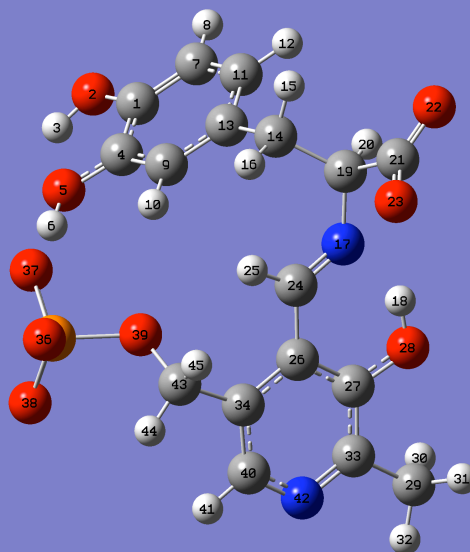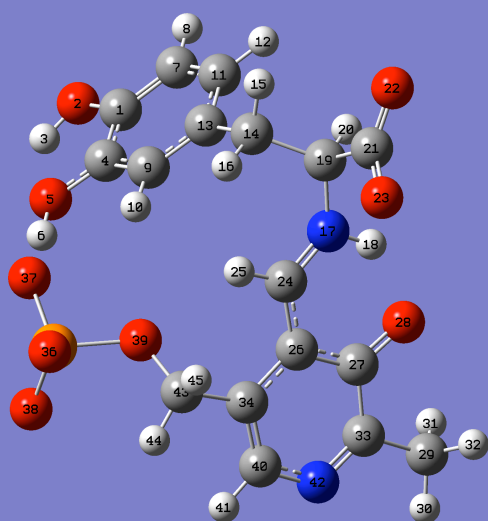
d2
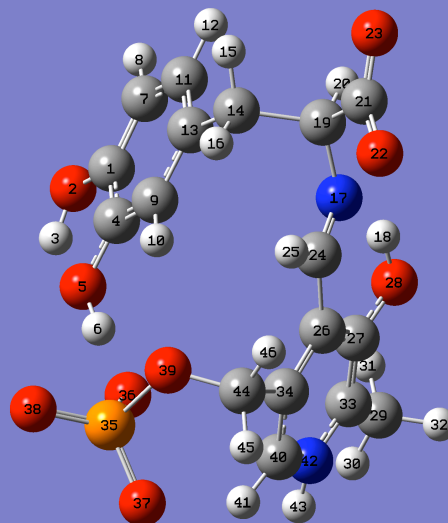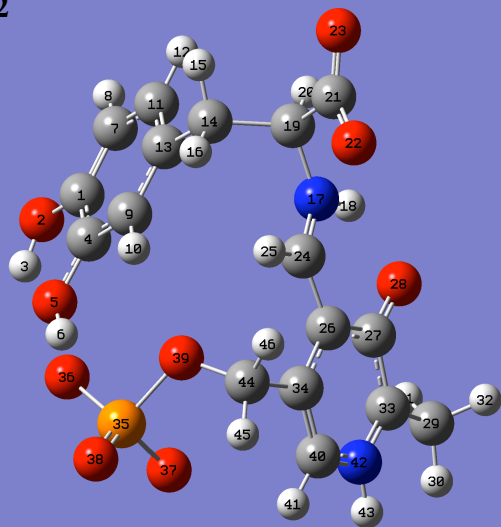
Table B1　Atomic Coordinates of the Fully Optimized

Oxoenamine/Hydroxyimine Tautomers in Reactions **a–d** in the Gas Phase

reaction	tautomeric state	
a1 [a]	oxoenamine	N,0,2.83444974,0.1944606003,0.000078857
		C,0,2.4381635778,-1.1178602128,-0.0004995841
		H,0,3.2360934076,-1.8511212499,-0.0008998659
		C,0,1.1246099014,-1.4920277806,-0.0006082279
		H,0,0.8529799133,-2.5438524461,-0.0011071375
		C,0,0.1063067337,-0.4971155533,-0.0000726827
		C,0,1.9301355379,1.134302632,0.0005748102
		H,0,2.2615339149,2.170467091,0.001006973
		C,0,0.4936175178,0.9029291695,0.0006519341
		O,0,-0.3382357345,1.8500034752,0.0010189154
		C,0,-1.2547169008,-0.8390424614,-0.0001977601
		H,0,-1.5538620442,-1.886166933,-0.0007118454
		N,0,-2.2160217261,0.0626528305,0.0002981078
		H,0,-1.8457015376,1.0369386114,0.000740986
		C,0,-3.6372746593,-0.2300666349,0.0001138512
		H,0,-3.7947361008,-1.3098942168,0.0001070357
		H,0,-4.1127558516,0.1961805299,0.8870679486
		H,0,-4.1125701339,0.1961715044,-0.8869514282
a1 [a]	hydroxyimine	N,0,-2.838330711,0.1306793385,-0.0001167351
		C,0,-2.413249445,-1.1430761336,0.0000610482
		H,0,-3.1820500412,-1.9085129412,0.0001026225
		C,0,-1.0701783808,-1.4831166806,0.000170936
		H,0,-0.7684016109,-2.5255867581,0.0002839756
		C,0,-0.0970058563,-0.4729702485,0.0001006054
		C,0,-1.9311636937,1.09896027,-0.0001726359
		H,0,-2.2859176981,2.1254647514,-0.0002821559
		C,0,-0.5451084822,0.86539563,-0.0000443195
		O,0,0.2908999999,1.9132190841,-0.000002479
		C,0,1.3257212135,-0.7982469432,0.0001319946
		H,0,1.5874102731,-1.8655256504,0.0002236818
		N,0,2.2284386279,0.1065729486,-0.0000920551
		H,0,1.2151308252,1.5529750669,0.0002927613
		C,0,3.6263833128,-0.2799023437,-0.0001376851
		H,0,3.7681624118,-1.3686931814,0.0000928359
		H,0,4.1183694612,0.1438599556,-0.8802077007
		H,0,4.1185386024,0.1442558475,0.8796473459

[a] Calculated using the B3LYP/6-311+G(d,p) method

Table B1 (continued)

reaction	tautomeric state	
a2 [a]	oxoenamine	N,0,3.9397036378,2.0000379275,-0.0956814617
		H,0,3.4120517699,2.0661401404,0.7709895908
		C,0,5.3034072026,1.9924500897,-0.0024674223
		H,0,5.7272262822,2.0573224608,0.9873468227
		C,0,6.0298463798,1.9035410088,-1.1641864407
		H,0,7.1117299259,1.8966826757,-1.1007508302
		C,0,5.3880382921,1.8228390683,-2.4119090175
		C,0,3.2526016278,1.927909521,-1.2284428792
		H,0,2.1719201155,1.9424686924,-1.1802865126
		C,0,3.9338999617,1.8320747009,-2.4995152709
		O,0,3.2976521325,1.7635036045,-3.5694487928
		C,0,6.1559225578,1.7303739071,-3.6162485541
		H,0,7.2424195534,1.7227045458,-3.5651032957
		N,0,5.6019619977,1.6537941252,-4.7885964066
		H,0,4.5611134699,1.6663775089,-4.7602901134
		C,0,6.3119556079,1.5571683338,-6.0609317685
		H,0,7.3877737604,1.5580685485,-5.8899315468
		H,0,6.0192536505,0.6350424797,-6.5661358534
		H,0,6.0375220744,2.405500661,-6.6904102471
a2 [a]	hydroxyimine	N,0,3.6997322949,1.6709321232,-0.2926507177
		H,0,3.0902371452,1.7000570576,0.5192900415
		C,0,5.0354646069,1.6655530762,-0.0913148064
		H,0,5.3814038531,1.6927959252,0.9315637232
		C,0,5.8665002679,1.6260179194,-1.1923547969
		H,0,6.9387251343,1.6214039738,-1.0398831541
		C,0,5.3382160771,1.5923048751,-2.4865082454
		C,0,3.1229213739,1.6400112259,-1.5066086747
		H,0,2.0432142074,1.647540378,-1.5624901537
		C,0,3.9233989791,1.5994380415,-2.6522809222
		O,0,3.3313725912,1.5699106792,-3.8267263798
		C,0,6.2181011746,1.5500867337,-3.6616073244
		H,0,7.2990930994,1.5461758041,-3.4860555906
		N,0,5.7111096618,1.519588293,-4.8312251501
		H,0,4.057598618,1.5432687137,-4.5267952523
		C,0,6.5643412539,1.4775883633,-6.0043549185
		H,0,7.6302476256,1.476082143,-5.7560967175
		H,0,6.3173369727,0.5832774386,-6.5823307873
		H,0,6.331985063,2.3399672354,-6.6345701732

Table B1 (continued)

reaction	tautomeric state	
b1 [a]	oxoenamine	N,0,3.4290763435,0.6098044989,2.639438662 C,0,2.892984463,1.8248487743,2.9418745763 H,0,3.2074569621,2.2668117472,3.8816194518 C,0,2.0015086838,2.4636743234,2.1128249916 H,0,1.5909957187,3.4306390862,2.3908554083 C,0,1.6040333638,1.8656519493,0.8952816311 C,0,3.0770555268,0.0281005125,1.5161632119 H,0,3.5115078986,-0.9437719112,1.2877836741 C,0,2.1473734736,0.570049615,0.5441916111 O,0,1.8519086084,-0.0505216877,-0.5200228917 C,0,0.6948104371,2.5352482863,0.028089262 H,0,0.2827445957,3.5019494784,0.2948208347 N,0,0.3082821753,2.0192201214,-1.1092939638 C,0,-0.637467147,2.6360479511,-2.0298997835 H,0,-0.1631532365,3.4867038721,-2.5300443875 H,0,-0.8885737647,1.9010277692,-2.7925261415 C,0,-1.9836885204,3.2048525698,-1.3922635408 O,0,-2.8443965998,3.4345112178,-2.259454265 O,0,-1.9785775296,3.3861805751,-0.1594818234 H,0,0.6997476187,1.0772789534,-1.2792452288
b1	hydroxyimine	N,0,3.3696641582,0.5759102043,2.6675834708 C,0,2.9238266864,1.8151749956,2.9232030083 H,0,3.2644161951,2.2687692978,3.8497047196 C,0,2.0712230202,2.5026724885,2.0695190396 H,0,1.7307378564,3.5020938347,2.3214591829 C,0,1.6370488126,1.9055154622,0.8799289439 C,0,2.9622327017,-0.0082973844,1.541018776 H,0,3.322147388,-1.0134408233,1.3374216352 C,0,2.101360445,0.5955484514,0.6120376313 O,0,1.745610508,-0.0687696253,-0.4959888734 C,0,0.750149759,2.5984979406,-0.0527058745 H,0,0.4193379091,3.6045858594,0.2121661325 N,0,0.3760698834,2.0247492652,-1.1392735015 C,0,-0.5553996393,2.6939100358,-2.0078229972 H,0,-0.1868954359,3.6895778243,-2.2890394294 H,0,-0.6951353607,2.1103005797,-2.9158823166 C,0,-2.0069855295,2.9913382396,-1.3637316429 O,0,-2.9031723289,3.0555149782,-2.2237214938 O,0,-2.0209793006,3.1745330358,-0.1338156739 H,0,1.1246723912,0.5555085741,-1.000893591

Table B1 (continued)

reaction	tautomeric state	
b2 [b]	oxoenamine	N,0,3.3664942173,0.6217487109,2.5427039893 H,0,3.9948860463,0.1402719424,3.1534581224 C,0,3.475224152,1.9699902987,2.4706313785 H,0,4.2297505945,2.4334994774,3.0692410034 C,0,2.6197982085,2.6215651447,1.6308640373 H,0,2.6940018701,3.6898114833,1.5431322212 C,0,1.6719208467,1.9294583469,0.8785807416 C,0,2.5094724976,-0.0993743127,1.8728501096 H,0,2.5221174369,-1.1632975463,2.009441439 C,0,1.5654296756,0.5033919323,0.9549014059 O,0,0.7761105682,-0.1886362767,0.3302485644 C,0,0.8221582604,2.6900327411,-0.0190700702 H,0,0.9215273699,3.7569385978,-0.0394283394 N,0,-0.0257414139,2.1660562827,-0.8077663674 H,0,-0.087820826,1.163511001,-0.7975749523 C,0,-0.764216148,2.9350572141,-1.8176124208 H,0,-1.2297449091,2.2276527744,-2.4858777499 H,0,-1.5437272125,3.5079925561,-1.32999853 C,0,0.1397735923,3.9333898237,-2.6369508591 O,0,-0.446775038,4.3927846892,-3.5918028854 O,0,1.2645387812,4.124421879,-2.1823175983
b2	hydroxyimine	N,0,3.3784924295,0.6105182438,2.4817657698 H,0,4.0221067071,0.0991158757,3.0472529678 C,0,3.4281854457,1.9407387348,2.4956604381 H,0,4.1674565897,2.3971597226,3.1230259326 C,0,2.5522142321,2.6452632441,1.71560515 H,0,2.5955731883,3.717308513,1.7078783278 C,0,1.6228145538,1.9836489244,0.9184084693 C,0,2.5114445685,-0.0820350741,1.7322206241 H,0,2.5535431263,-1.1518634237,1.7791958406 C,0,1.6094679997,0.5707676328,0.9276500944 O,0,0.7840724078,-0.1538974674,0.2083966562 C,0,0.719149358,2.7606906381,0.0533077006 H,0,0.8084937247,3.8338789132,0.0964099002 N,0,-0.0430464878,2.1580114598,-0.7365081467 C,0,-0.7606460445,2.9371127245,-1.7170401844 H,0,-1.4191010513,2.2931915749,-2.2772911805 H,0,-1.3541772051,3.7079697091,-1.2307546569 C,0,0.2135854706,3.7083843982,-2.713150228 O,0,-0.3188420796,3.9697751256,-3.7716039731 O,0,1.3235478542,3.9677516932,-2.2610457903 H,0,0.2676925814,0.4373704765,-0.3576620761

[b] Calculated using the HF/6-311+G(d,p) method

Table B1 (continued)

reaction	tautomeric state	
c1 [a]	oxoenamine	C,0,-5.0656338407,-0.321110353,-0.2811825559
		O,0,-6.3668893501,-0.4833563427,-0.7001549312
		H,0,-6.808559806,0.369812227,-0.6449371768
		C,0,-4.2198284075,-1.4343916286,-0.3134096772
		O,0,-4.7404344363,-2.6267414299,-0.7605336611
		H,0,-4.0404736797,-3.2871649247,-0.7446183372
		C,0,-4.559916999,0.8947644823,0.1716804971
		H,0,-5.2175439446,1.7609526744,0.1940463785
		C,0,-2.897960133,-1.3012952907,0.1057027835
		H,0,-2.2542680653,-2.1784081632,0.0777200061
		C,0,-3.2367131217,1.0226184568,0.5900768854
		H,0,-2.8445797475,1.9761747626,0.9237129154
		C,0,-2.3813121842,-0.0827148884,0.5617970058
		C,0,-0.9348819836,0.0179406289,0.999167902
		H,0,-0.871613114,0.5656389651,1.9425178654
		H,0,-0.5557677148,-0.9962591527,1.1676719616
		N,0,1.3375628072,0.202333521,0.1141371585
		H,0,1.5424425059,-0.5836798277,0.7580389739
		C,0,-0.0285472661,0.7384399778,-0.0154682377
		H,0,-0.3689309106,0.5259789164,-1.0340598052
		C,0,-0.0214113787,2.3362910662,0.1117060102
		O,0,-0.7035815073,2.8135948362,1.0353329504
		O,0,0.6734004835,2.8962417448,-0.7630435036
		C,0,2.3937201078,0.6116210315,-0.5408558065
		H,0,2.2439067571,1.4539594264,-1.2114907114
		C,0,3.6683434127,0.0037737227,-0.3712631548
		C,0,3.8412377727,-1.1226671102,0.5224846877
		O,0,2.9023543619,-1.6195159191,1.2127482642
		C,0,5.1901436064,-1.6531388352,0.5652553069
		H,0,5.3609120646,-2.5038926426,1.2227861154
		C,0,4.7889390744,0.4799334017,-1.0915671562
		H,0,4.6718894227,1.3289104468,-1.7593874727
		C,0,6.0149541534,-0.1225436757,-0.9429826582
		H,0,6.8870919642,0.2299122847,-1.483567888
		N,0,6.2107116278,-1.1881475859,-0.1163405505

Table B1 (continued)

reaction	tautomeric state	
c1 [a]	hydroxyimine	C,0,-5.0796772793,-0.3001180738,-0.2804106651
		O,0,-6.3801910776,-0.4589935886,-0.7082786485
		H,0,-6.7939552437,0.4094013599,-0.7322475481
		C,0,-4.2675083493,-1.4346644773,-0.2002673684
		O,0,-4.8147479189,-2.6486163265,-0.5530178776
		H,0,-4.1294647276,-3.3197226863,-0.4746409898
		C,0,-4.5437676054,0.9356212378,0.0742926213
		H,0,-5.1751863923,1.8192001782,0.0080836075
		C,0,-2.9491296572,-1.3031910555,0.2313648392
		H,0,-2.3316749024,-2.1977770585,0.2910433751
		C,0,-3.2243425972,1.0614176834,0.5044298488
		H,0,-2.8075018406,2.0285896008,0.7598749753
		C,0,-2.4007517347,-0.0660662339,0.5907026692
		C,0,-0.959105487,0.0306651088,1.0425527781
		H,0,-0.8929288639,0.657341867,1.9354434818
		H,0,-0.6057169576,-0.9723472961,1.3033565907
		N,0,1.3367236047,0.1425966048,0.2093712872
		H,0,2.2884998997,-0.4590802559,1.4678479349
		C,0,-0.005690265,0.6257772424,-0.0117093652
		H,0,-0.3422138846,0.3542616723,-1.0205467922
		C,0,0.0227531026,2.2564054917,-0.0306531681
		O,0,-0.662509727,2.8139554572,0.845402125
		O,0,0.7226664097,2.7167491241,-0.9482116911
		C,0,2.2023098181,0.2145916805,-0.736415358
		H,0,1.9388963799,0.6016132763,-1.7232270127
		C,0,3.5850986341,-0.2016572274,-0.5140934837
		C,0,4.0126780213,-0.6991119136,0.7393478107
		O,0,3.1796124139,-0.807123858,1.7846715098
		C,0,5.3518591806,-1.0933517102,0.8779455223
		H,0,5.6827240878,-1.4765109612,1.8396201862
		C,0,4.5450501009,-0.1345685477,-1.5318973337
		H,0,4.2644650043,0.2498193173,-2.5073991607
		C,0,5.8474200117,-0.5492266869,-1.2888611569
		H,0,6.5999751466,-0.4986859293,-2.0705041895
		N,0,6.2522218504,-1.0258558858,-0.101882293

213

Table B1 (continued)

reaction	tautomeric state	
c2 [b]	oxoenamine	C,0,-2.6956818756,-4.1950866282,-4.546769881
		O,0,-3.8912924709,-3.6180174031,-4.3038801269
		H,0,-3.8818906144,-2.7264168657,-4.6100718435
		C,0,-1.6719324614,-3.4929782644,-5.1614046849
		O,0,-1.9440107969,-2.2040714061,-5.5053888967
		H,0,-1.1510043777,-1.7199627623,-5.6559908132
		C,0,-2.4797784332,-5.509171926,-4.1797007841
		H,0,-3.2812778339,-6.0500712379,-3.7104076719
		C,0,-0.4573221501,-4.101315133,-5.4102858062
		H,0,0.3105065836,-3.5338559231,-5.9103565055
		C,0,-1.254223671,-6.1105509658,-4.411672777
		H,0,-1.1079498902,-7.1351390666,-4.1173762871
		C,0,-0.2202632822,-5.4150059815,-5.024939104
		C,0,1.1358542491,-6.0549455917,-5.240782069
		H,0,1.0315868791,-7.0918192211,-5.5236895063
		H,0,1.6556250593,-5.5614945139,-6.057939485
		N,0,2.3007206142,-4.6524873589,-3.5961977964
		H,0,1.8712301556,-3.8976336405,-4.1005309705
		C,0,2.0379387303,-6.0464911775,-4.0028665032
		H,0,1.5257295897,-6.5322945403,-3.1797630256
		C,0,3.3848359333,-6.8482835464,-4.2343953072
		O,0,4.4129513942,-6.2782767367,-3.8795515194
		O,0,3.1897135141,-7.9487574584,-4.7054982193
		C,0,3.1549723742,-4.3260132473,-2.7166315144
		H,0,3.6290859592,-5.1121571924,-2.1645901999
		C,0,3.5368099122,-2.9530464206,-2.4271571546
		C,0,2.9428892507,-1.8565345856,-3.1318473629
		O,0,2.0744998389,-1.94236725,-3.9900828126
		C,0,3.4668717894,-0.5671438479,-2.7408943145
		H,0,3.0962045959,0.3271453787,-3.2027790895
		C,0,4.530824348,-2.7408163335,-1.4744953588
		H,0,4.976666505,-3.5797124986,-0.9728515627
		C,0,4.9690868469,-1.484230024,-1.1698374337
		H,0,5.7302108344,-1.249610524,-0.4568762889
		N,0,4.3998188091,-0.4526803752,-1.8332282752
		H,0,4.7199940912,0.4692942694,-1.6162690487

Table B1 (continued)

reaction	tautomeric state	
c2 [b]	hydroxyimine	C,0,-2.6716649163,-4.1426987323,-4.3560398265
		O,0,-3.8599055837,-3.5779071707,-4.0429666875
		H,0,-3.9938257867,-2.8091000911,-4.5714551961
		C,0,-1.8120140611,-3.5509084518,-5.2646996448
		O,0,-2.2334030608,-2.3754955625,-5.8139946955
		H,0,-1.7203232538,-2.150091803,-6.5691101647
		C,0,-2.3006501567,-5.3361623731,-3.7680350463
		H,0,-2.9786920691,-5.7994598179,-3.0742519522
		C,0,-0.5988952657,-4.1395293632,-5.5708108352
		H,0,0.0487018279,-3.6572666992,-6.2853724894
		C,0,-1.0874881515,-5.9257833372,-4.0800512809
		H,0,-0.8288599749,-6.8626863997,-3.6184402187
		C,0,-0.2110108554,-5.3374647379,-4.9845137948
		C,0,1.1250014457,-5.9714982227,-5.3167521887
		H,0,0.9930252857,-7.0081249006,-5.5939571029
		H,0,1.5612519282,-5.4820082227,-6.1812532893
		N,0,2.408347991,-4.5737604493,-3.7703800152
		C,0,2.1325839609,-5.9341254013,-4.1703683742
		H,0,1.740011597,-6.4831248085,-3.3159125836
		C,0,3.5145139513,-6.6768263279,-4.5109087047
		O,0,4.4498124134,-6.3465751138,-3.7913652522
		O,0,3.4241242973,-7.4990919616,-5.3988895228
		C,0,2.99817249,-4.3772518915,-2.6850615654
		H,0,3.2433164666,-5.161896398,-1.9875890334
		C,0,3.4648075247,-3.0213055674,-2.3503650637
		C,0,3.1961866793,-1.9150063734,-3.1871959444
		O,0,2.4995893737,-1.9942604341,-4.2974221279
		C,0,3.6916107974,-0.6875852491,-2.8202089859
		H,0,3.5232503604,0.188676031,-3.4137366194
		C,0,4.223800208,-2.8173746974,-1.2017929764
		H,0,4.4568339797,-3.6459444031,-0.5612212771
		C,0,4.6938591119,-1.5710162293,-0.8891423139
		H,0,5.28667971,-1.3578951422,-0.0222104123
		N,0,4.4128763215,-0.5534229754,-1.6999568305
		H,0,4.7596384926,0.3538237587,-1.4714621567
		H,0,2.2667412724,-2.9204858293,-4.4494674575

Table B1 (continued)

reaction	tautomeric state	
d1 [c]	oxoenamine	C,0,3.6469094129,1.0276854343,1.9170703877
		O,0,4.4673830565,0.5522287872,2.9036811723
		H,0,4.5080127981,-0.3902802458,2.7676201237
		C,0,3.1331366634,0.1189055978,0.9875785488
		O,0,3.4857945678,-1.1502789157,1.1446528635
		H,0,3.0830830263,-1.7729801784,0.4141103129
		C,0,3.3328206602,2.3666630067,1.8383565503
		H,0,3.7432268237,3.0451349401,2.5686777457
		C,0,2.3010357495,0.591776522,-0.0220612188
		H,0,1.9223385064,-0.12854312,-0.7222183127
		C,0,2.4924117706,2.8270365891,0.8266669872
		H,0,2.2517116395,3.8762781383,0.7728842596
		C,0,1.9650148824,1.9404303865,-0.1072382042
		C,0,1.0549392602,2.4365937125,-1.2197206145
		H,0,1.5211977532,3.279658223,-1.7164374715
		H,0,0.9402842735,1.6654204316,-1.9735824097
		N,0,-1.205165994,1.9195033707,-0.2161648337
		H,0,-1.9518877298,2.2434152953,0.3758800123
		C,0,-0.3235488473,2.9501560745,-0.7721309386
		H,0,-0.171190649,3.650131767,0.0383820587
		C,0,-0.9961510646,3.7567368369,-1.9332511489
		O,0,-0.5569096382,4.9118006949,-2.0393699037
		O,0,-1.8392460453,3.1761840848,-2.6118930988
		C,0,-1.2283960005,0.6493154292,-0.4291564427
		H,0,-0.4762128155,0.225834008,-1.0637903209
		C,0,-2.1882126462,-0.2404880299,0.15699243
		C,0,-3.1885251262,0.2667779087,1.0382937554
		O,0,-3.3122877009,1.4667015368,1.3733019624
		C,0,-5.2087195067,-0.236072128,2.4883033353
		H,0,-5.8244671049,-1.0757012568,2.7942530527
		H,0,-4.7867438575,0.2423963034,3.3686642826
		H,0,-5.8303210587,0.5112897494,2.001203198
		C,0,-4.1227243432,-0.7245280072,1.5650278896
		C,0,-2.1381741007,-1.6407989774,-0.1266087781
		P,0,1.3324252482,-3.2123328553,-1.178767107
		O,0,1.179420491,-3.2874460606,-2.6718666056
		O,0,2.6164061511,-2.51913389,-0.7188379667
		O,0,0.9710528159,-4.4509921204,-0.410086526
		O,0,0.1746404622,-2.0419387643,-0.695207009
		C,0,-3.0783737468,-2.4404197866,0.454153261
		H,0,-3.0723549511,-3.5018243185,0.2908641764
		N,0,-4.0548100054,-1.9785687663,1.2810624281
		C,0,-1.1344299628,-2.275024438,-1.0816110696
		H,0,-1.3419863799,-3.3375908889,-1.1164160586
		H,0,-1.3067453351,-1.884583729,-2.0855504136

[c] Calculated using the HF/6-31+G(d) method

216

Table B1 (continued)

reaction	tautomeric state	
d1 [c]	hydroxyimine	C,0,3.9079854051,0.6662247725,1.717853096
		O,0,4.747819366,0.0958549779,2.6368964037
		H,0,4.6572828473,-0.8447285646,2.5121868637
		C,0,3.1921493998,-0.1760590922,0.8616543099
		O,0,3.3937297938,-1.479221137,1.0282331032
		H,0,2.8923994805,-2.0762275236,0.353547807
		C,0,3.769589185,2.033369308,1.634234056
		H,0,4.3340356737,2.6602923107,2.3057336305
		C,0,2.3380335732,0.3892130567,-0.0790752757
		H,0,1.7953410484,-0.2811950239,-0.7189629204
		C,0,2.9072060217,2.588199615,0.6903463472
		H,0,2.8019130852,3.6593004144,0.6323531537
		C,0,2.1795924095,1.7725177864,-0.1707503126
		C,0,1.2437309895,2.3918202476,-1.1985996971
		H,0,1.7616414804,3.2105621104,-1.6896877532
		H,0,1.01415045,1.6724452676,-1.9755188701
		N,0,-1.0381980871,2.0770813105,-0.0827157825
		H,0,-2.4313520277,2.23062781,1.0117549907
		C,0,-0.0544306897,3.0039243338,-0.6249792322
		H,0,0.2240524437,3.6411576629,0.2033933321
		C,0,-0.6934142854,3.9487601833,-1.6991747048
		O,0,-0.3310356338,5.1338752538,-1.6157400875
		O,0,-1.4394268911,3.4293704276,-2.5299164446
		C,0,-1.1437456715,0.8765720536,-0.4375421381
		H,0,-0.49496432,0.4116780845,-1.157883128
		C,0,-2.1928871473,-0.0079330544,0.1435301112
		C,0,-3.1374108622,0.4896896687,1.0449648656
		O,0,-3.1648754553,1.7707558187,1.4490374772
		C,0,-5.1430691729,0.181611189,2.5300071056
		H,0,-4.6717458696,0.5858233219,3.4223171399
		H,0,-5.7183051739,0.990471786,2.086344686
		H,0,-5.8159342277,-0.6191143735,2.8160865858
		C,0,-4.1176683718,-0.3581686746,1.5648837615
		C,0,-2.2667349759,-1.3683405042,-0.1949351342
		P,0,0.938277887,-3.4076791609,-1.1527651775
		O,0,0.7855370038,-3.519324645,-2.6446807229
		O,0,2.3070113872,-2.911775391,-0.6930299295
		O,0,0.3657049909,-4.5411906508,-0.3475777962
		O,0,-0.0251591935,-2.0505659048,-0.7333594159
		C,0,-3.2734668337,-2.1171696932,0.3951553585
		H,0,-3.3436780829,-3.1681051457,0.1794833965
		N,0,-4.1743867767,-1.6333002515,1.2404823739
		C,0,-1.330664113,-2.0547498938,-1.1830026456
		H,0,-1.6914350245,-3.0665481245,-1.3266350675
		H,0,-1.4028174523,-1.5563633262,-2.1499361134

Table B1 (continued)

reaction	tautomeric state	
d2 [b]	oxoenamine	C,0,-2.8883351436,2.4338486827,1.5896969964
		O,0,-2.8912064442,3.4851698843,2.4531120135
		H,0,-2.1306333847,4.0052282359,2.2371953844
		C,0,-1.8691117165,2.3602335741,0.632338793
		O,0,-0.9799133998,3.33806169,0.6563869668
		H,0,-0.224555997,3.2510205723,-0.0325695829
		C,0,-3.86283556,1.4676167046,1.6519859437
		H,0,-4.6364421685,1.5538095167,2.3961601217
		C,0,-1.8628199977,1.2917521787,-0.2540042973
		H,0,-1.0756154352,1.2568392422,-0.9857907704
		C,0,-3.8423879063,0.3967106587,0.7616139671
		H,0,-4.6126026137,-0.3546775065,0.8177303895
		C,0,-2.8417561652,0.2997425619,-0.1925957586
		C,0,-2.7983681099,-0.8750665741,-1.1525288629
		H,0,-3.787993392,-1.0814252664,-1.543406961
		H,0,-2.1770626437,-0.6252331784,-2.0059001731
		N,0,-0.9912954961,-2.0773236896,0.0883456041
		H,0,-0.8535710154,-2.4202519558,1.0245228583
		C,0,-2.3145746074,-2.2028218882,-0.5365982405
		H,0,-2.9819602207,-2.4828554804,0.2651184283
		C,0,-2.3364410086,-3.3384388659,-1.61771435
		O,0,-1.2720793838,-3.5571911555,-2.1852494667
		O,0,-3.4455804822,-3.8366347913,-1.7749983372
		C,0,0.049027601,-1.6288026488,-0.4833375209
		H,0,-0.0463411929,-1.268933061,-1.4856256648
		C,0,1.3457018732,-1.524405161,0.1775102986
		C,0,1.4935647447,-2.0139000058,1.5134607446
		O,0,0.6204599054,-2.6308789717,2.1525580399
		C,0,3.0112125362,-2.1169835945,3.5445961093
		H,0,4.0014307696,-1.8063622305,3.8682937936
		H,0,2.269288604,-1.6691753558,4.1969272369
		H,0,2.922368086,-3.1928093018,3.6544541514
		C,0,2.7480985895,-1.7189104455,2.1277188045
		C,0,2.3688460778,-0.8362945864,-0.4813223351
		P,0,2.0280752061,2.3974901974,-1.3939486853
		O,0,0.7773622676,3.177245098,-1.0601795805
		O,0,2.9103423958,2.0596622461,-0.2220335218
		O,0,2.7384864996,2.7953798611,-2.6395077629
		O,0,1.3268484836,0.8611386661,-1.7742372238
		C,0,3.5414520576,-0.6065347366,0.1929243779
		H,0,4.3080493438,0.036691079,-0.1786812718
		N,0,3.6612133431,-1.047276124,1.4481554111
		H,0,4.4944859861,-0.7897543362,1.9319183843
		C,0,2.1943867434,-0.1988809044,-1.8560502398
		H,0,3.1704571058,0.1015805587,-2.2252671848
		H,0,1.8052039981,-0.9396035034,-2.5522590278

Table B1 (continued)

reaction	tautomeric state	
d2 [b]	hydroxyimine	C,0,-1.1898532853,3.5709495045,0.8049016684
		O,0,-0.6141733317,4.6047772779,1.4783178482
		H,0,0.3163232661,4.5346513497,1.3225205451
		C,0,-0.3646806677,2.7002154261,0.0839267883
		O,0,0.936093876,2.9682519904,0.1302549807
		H,0,1.5472789373,2.2171026023,-0.1587029864
		C,0,-2.5498475974,3.3817237157,0.8313742647
		H,0,-3.1618307982,4.0685741056,1.3913408517
		C,0,-0.9467909541,1.6545832011,-0.6146063638
		H,0,-0.2993495074,1.0140126762,-1.1859363111
		C,0,-3.1190538788,2.3088980132,0.148128142
		H,0,-4.1854391869,2.161447892,0.1876861907
		C,0,-2.3260548567,1.4315036992,-0.572875267
		C,0,-2.9376581258,0.2275646449,-1.2706300807
		H,0,-3.8929439375,0.493225681,-1.7086776104
		H,0,-2.2932788064,-0.0834291089,-2.0862431198
		N,0,-2.0134903847,-1.3278814791,0.4256939262
		H,0,-1.5681016253,-1.4332078845,2.0993870186
		C,0,-3.2058453548,-0.982636864,-0.3386317956
		H,0,-3.9514503925,-0.6852552686,0.3859735904
		C,0,-3.8200389043,-2.1660130712,-1.1651869723
		O,0,-3.0448497341,-3.0380118404,-1.5464778462
		O,0,-5.0289400985,-2.048902967,-1.3549732052
		C,0,-0.9521886005,-1.6351580369,-0.159344444
		H,0,-0.880333105,-1.7353237823,-1.2255403382
		C,0,0.3141328646,-1.7344557394,0.6153921767
		C,0,0.288564107,-1.5930517776,2.0318153821
		O,0,-0.8453420833,-1.5657392188,2.7309246011
		C,0,1.5137311632,-1.2243207816,4.2025816556
		H,0,2.482801512,-0.8329254068,4.4976088862
		H,0,0.7496679821,-0.5174234073,4.4978320299
		H,0,1.3355537528,-2.1500835993,4.743207535
		C,0,1.4555498065,-1.4445717711,2.7206643662
		C,0,1.530765699,-1.7972819477,-0.0391098389
		P,0,3.0560928195,0.4558962155,-1.6809383546
		O,0,2.4881205195,1.0675924997,-0.3978635795
		O,0,4.1872812228,-0.4945142539,-1.365747008
		O,0,3.193196568,1.3710219882,-2.8312158946
		O,0,1.7902911858,-0.6091666931,-2.1269546379
		C,0,2.6864649558,-1.6300375516,0.7144804658
		H,0,3.6410358533,-1.5123230968,0.2244111493
		N,0,2.6109687919,-1.4724140439,2.0196401398
		H,0,3.4565172255,-1.2578007962,2.5000779798
		C,0,1.7282767191,-1.8585002931,-1.5569591405
		H,0,2.6450139067,-2.4128397437,-1.7364380492
		H,0,0.9144400997,-2.4098208204,-2.0127589508

219

Table B2   Calculated Energies for the Tautomerization Reaction of Reactions

a–d in the Gas Phase at 298.15 K [a,b]

			oxoenamine Schiff base → hydroxyimine Schiff base				
reaction	ΔE_{elec} [b]	ΔE_T	$T\Delta S$	ΔG_{gas} [h]	$\Delta G_s(oxo)$ [h]	$\Delta G_s(hyd)$ [i]	ΔG_{PCM} [j]
a1 [c]	−4.9	0.1 [f]	−0.4	−4.4	−15.6	−9.5	1.7
a2 [c]	−0.6	0.7 [f]	−0.3	0.4	−62.3	−59.7	3.0
b1 [c]	2.7	−0.6 [f]	−0.2	2.3	−65.2	−65.5	2.0
b2 [d]	7.7	−0.5 [g]	−0.2	7.4	−52.0	−58.2	1.2
c1 [c]	0.9	−0.4 [f]	−0.4	0.9	−75.2	−76.0	0.1
c2 [d]	7.9	−0.7 [g]	0.2	7.0	−59.2	−67.5	−1.3
d1 [e]	5.4	−0.1 [g]	0.3	5.0	−317.2	−318.6	3.6
d2 [d]	7.0	0.1 [g]	0.0	7.1	−188.4	−184.8	10.7

[a] All energies are reported to 1 decimal place in kcal/mol. [b] Single-point B3LYP/6-311+G(d,p) energy. [c] Geometries optimized at the B3LYP/6-311+G(d,p) level. [d] Geometries optimized at the HF/6-311+G(d,p) level. [e] Geometries optimized at the HF/6-31+G(d) level. [f] The frequencies were scaled by an empirical factor of 0.9613 (*1*). [g] The frequencies were scaled by an empirical factor of 0.8929 (*1*). [h] Gas-phase free energy of tautomerization reaction, ΔG_{gas}, was computed according to $\Delta G_{gas} = \Delta E_{elec} + \Delta E_T − T\Delta S$, where ΔE_{elec}, ΔE_T, and ΔS are the differences in the electronic energy, thermal energy, and entropy between the hydroxyimine conformation and oxoenamine isomer, respectively. Note that the thermal energy includes the zero-

point energy (*2*). [h] Solvation free energies of the oxoenamine-form Schiff bases in aqueous solution were calculated at the HF/6-311+G(d,p) level. [i] Solvation free energies of the hydroxyimine-form Schiff bases in aqueous solution were calculated at the HF/6-311+G(d,p) level. [j] Solution free energies were computed according to $\Delta G_{PCM} = \Delta G_{gas} + \Delta G_s(\text{hyd}) - \Delta G_s(\text{oxo})$.

References

1.   Hehre, W. J., Radom, L., Schleyer, P. v. R., and Pople, J. A. (1986) *Ab Initio Molecular Orbital Theory.* John Wiley and Sons, New York.

2.   Foresman, J. B., and Frisch, E. (1996) *Exploring Chemistry with Electronic Structure Methods.* 2nd ed., Gaussian, Inc., Pittsburgh, PA.

Appendix C

Supplementary Material for Chapter 4:

Catalyzing Decarboxylation of L-dopa in Dopa Decarboxylase

Figure C1   Computed potentials of mean force for the decarboxylation reaction of L-dopa in dopa decarboxylase with different loop conformations
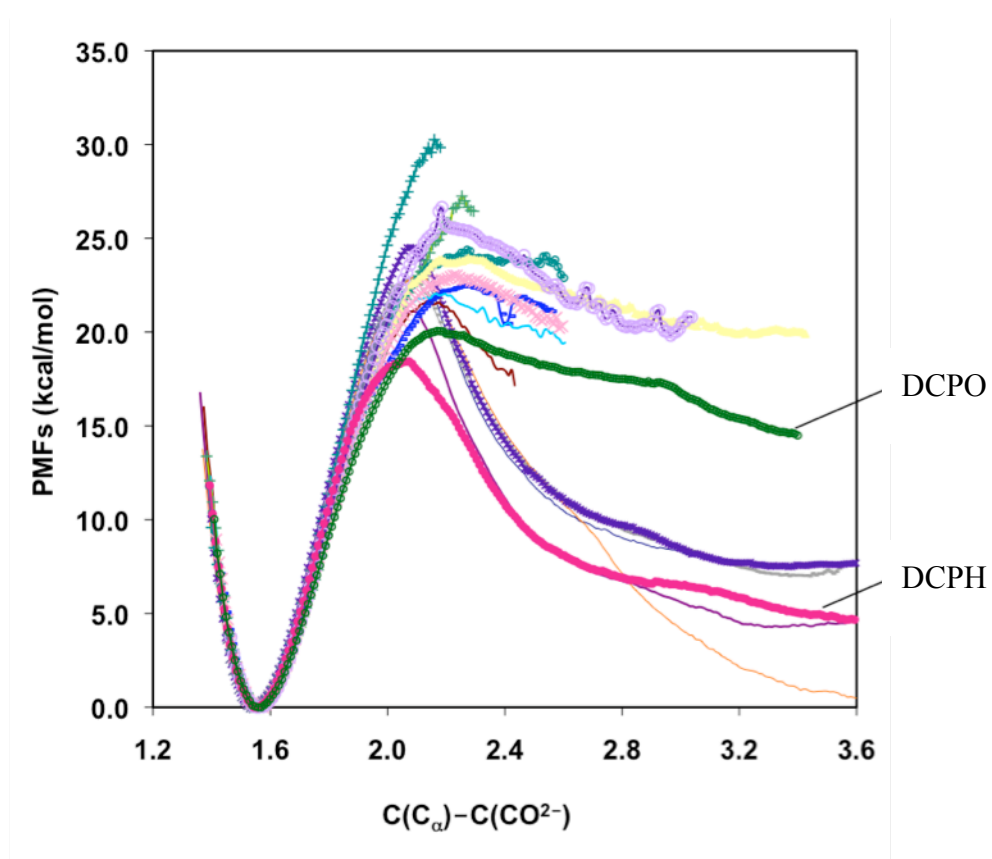
Appendix D

Supplementary Material for Chapter 5:

Kinetic Isotope Effects in Dopa Decarboxylase

Table D1   Computed primary ^{13}C and secondary ^{2}H KIEs of each individual block average (10 blocks) for the decarboxylation reactions of PLP–L-dopa *in dopa decarboxylase*

	hydroxyimine PLP aldimie			oxoenamine PLP aldimie		
block	case 1: $^{13}C^{CO_2^-}$ KIE	case 2: $^{13}C_\alpha$ KIE	case 3: $^{2}H_\alpha$ KIE	case 1: $^{13}C^{CO_2^-}$ KIE	case 2: $^{13}C_\alpha$ KIE	case 3: $^{2}H_\alpha$ KIE
1	1.0584	1.0351	1.0527	1.0445	1.0234	1.0046
2	1.0541	1.0279	1.0415	1.0403	1.0147	1.1404
3	1.0512	1.0398	1.0408	1.0446	1.0296	1.1076
4	1.0561	1.0362	1.1063	1.0497	1.0293	1.1537
5	1.0554	1.0421	1.0443	1.0424	1.0307	1.0827
6	1.0517	1.0421	1.0404	1.0476	1.0319	1.1004
7	1.0586	1.0383	1.0203	1.0538	1.0249	1.1516
8	1.0565	1.0487	1.0443	1.0491	1.0322	1.1869
9	1.0593	1.0430	1.1086	1.0389	1.0298	1.1484
10	1.0553	1.0361	1.1264	1.0515	1.0390	1.1662
average of blocks: [a]						
	1.0554 ± 0.0028	1.0390 ± 0.0056	1.0622 ± 0.0366	1.0461 ± 0.0049	1.0283 ± 0.0064	1.1266 ± 0.0528

[a] Standard deviation is calculated as $\sqrt{\sum_{i=1}^{N}(\langle A\rangle - A_i)^2 / (N-1)}$, where $\langle A\rangle$ is the average of the blocks, A_i is the block average of individual block, and N and the number of blocks, which is 10 in this study.

Table D2   Computed primary ^{13}C and secondary ^{2}H KIEs of each individual block average (10 blocks) for the decarboxylation reactions of PLP−L-dopa *in aqueous solution*

block	hydroxyimine PLP aldimie			oxoenamine PLP aldimie		
	case 1: $^{13}C^{CO_2^-}$ KIE	case 2: $^{13}C_\alpha$ KIE	case 3: $^{2}H_\alpha$ KIE	case 1: $^{13}C^{CO_2^-}$ KIE	case 2: $^{13}C_\alpha$ KIE	case 3: $^{2}H_\alpha$ KIE
1	1.0529	1.0281	0.9795	1.0465	1.0323	1.1102
2	1.0559	1.0350	1.0389	1.0515	1.0391	1.0933
3	1.0511	1.0295	1.0083	1.0468	1.0313	1.1952
4	1.0539	1.0328	1.0036	1.0495	1.0348	1.1776
5	1.0486	1.0310	0.9387	1.0462	1.0344	1.1105
6	1.0548	1.0291	0.8620	1.0514	1.0303	1.0963
7	1.0486	1.0248	1.0092	1.0487	1.0359	1.0658
8	1.0551	1.0318	0.9905	1.0493	1.0322	1.1903
9	1.0536	1.0254	0.9352	1.0524	1.0347	1.1344
10	1.0545	1.0293	0.9904	1.0523	1.0356	1.2126
average of blocks: [a]						
	1.0529 ± 0.0026	1.0298 ± 0.0031	0.9746 ± 0.0509	1.0495 ± 0.0024	1.0340 ± 0.0026	1.1368 ± 0.0513

[a] See footnote of Table D1.

Appendix E

The following pages include my other publications that I have contributed as a part of

collaboration during doctoral studies.

JCTC Journal of Chemical Theory and Computation

A Non-Orthogonal Block-Localized Effective Hamiltonian Approach for Chemical and Enzymatic Reactions

Alessandro Cembran, Apirak Payaka, Yen-lin Lin, Wangshen Xie, Yirong Mo,*,[†]
Lingchun Song,* and Jiali Gao*

*Department of Chemistry, Digital Technology Center and Supercomputing Institute,
University of Minnesota, Minneapolis, Minnesota 55455, and Department of
Chemistry, Western Michigan University, Kalamazoo, Michigan 49008*

Abstract: The effective Hamiltonian−molecular orbital and valence bond (EH-MOVB) method based on nonorthogonal block-localized fragment orbitals has been implemented in the program CHARMM for molecular dynamics simulations of chemical and enzymatic reactions, making use of semiempirical quantum mechanical models. Building upon *ab initio* MOVB theory, we make use of two parameters in the EH-MOVB method to fit the barrier height and the relative energy between the reactant and product state for a given chemical reaction to be in agreement with experimental or high-level *ab initio* or density functional results. Consequently, the EH-MOVB method provides a highly accurate and computationally efficient QM/MM model for dynamics simulation of chemical reactions in solution. The EH-MOVB method is illustrated by examination of the potential energy surface of the hydride transfer reaction from trimethylamine to a flavin cofactor model in the gas phase. In the present study, we employed the semiempirical AM1 model, which yields a reaction barrier that is more than 5 kcal/mol too high. We use a parameter calibration procedure for the EH-MOVB method similar to that employed to adjust the results of semiempirical and empirical models. Thus, the relative energy of these two diabatic states can be shifted to reproduce the experimental energy of the reaction, and the barrier height is optimized to reproduce the desired (accurate) value by adding a constant to the off-diagonal matrix element. The present EH-MOVB method offers a viable approach to characterizing solvent and protein-reorganization effects in the realm of combined QM/MM simulations.

1. Introduction

Combined quantum mechanical and molecular mechanical (QM/MM) methods offer an excellent opportunity for studying chemical and electron transfer reactions in solution and in biological systems.[1−3] In principle, the accuracy of combined QM/MM potentials can be systematically improved; however, it is still time-demanding to carry out QM/MM simulations using *ab initio* wave function theory (WFT) or density functional theory (DFT) for subsystems consisting of more than 100 atoms in the QM region. Consequently, it

is useful to develop efficient QM/MM techniques that can be made accurate for specific chemical and biomolecular applications, yet sufficiently fast for extensive conformational sampling. Aside from the brute force approach by increasing the level of theory and the size of basis set, there are two other ways to achieve this goal. The first is to parametrize purely empirical energy functions to model a specific process,[4,5] and the second is to parametrize quantum mechanical models against experimental data with specific reaction parameters (SRP) for a given class of reactions.[6−9] In this article, we describe an effective Hamiltonian approach based on the molecular orbital-valence bond (MOVB) theory developed in our laboratories for the treatment of reactive potential surfaces of reactions.[10−12] In particular, we illustrate

* To whom correspondence should be addressed. E-mail: yirong.mo@wmich.edu (Y.M.), lcsong2007@gmail.com (L.S.), gao@jialigao.org (J.G.).
† Western Michigan University.

this novel QM/MM technique in the hydride transfer reaction from trimethyl ammonium ion to flavin cofactor.

The mixed molecular orbital and valence bond (MOVB) theory,[10,11] initially developed at the *ab initio* level and recently extended to multistate density functional theory (called MSDFT or, equivalently, VBDFT),[13] is designed to treat reactive potential energy surfaces for chemical reactions and electron transfer processes. In this approach, the whole process is described with two or more resonating configurations, including the reactant and product states. In each specific state, molecular orbitals (MOs) are strictly localized within individual fragments of a molecular system.[14−20] However, the block-localized molecular orbitals (BLMOs) are still delocalized within each orbital block, making the MOVB method extremely efficient in comparison with *ab initio* valence bond theory.[21−25] Key features of the MOVB and MSDFT theories include (1) that the BLMOs (or block-localized Kohn−Sham orbitals)[13] within each fragment are orthogonal, which makes it computationally efficient, and (2) that the BLMOs between different fragments are nonorthogonal,[11,13,20] which retains important characters of valence bond (VB) theory. In the limiting case in which there is one fragment, MOVB and MSDFT reduce exactly to the standard Hartree−Fock theory and Kohn−Sham DFT, respectively.

Recently, we introduced an effective Hamiltonian MOVB approach,[26] in which the *ab initio* electronic matrix elements are adjusted to yield accurate barrier height and reaction enthalpy. This approach has an apparent similarity in the "calibration" process used to adjust the barrier height and the energy of reaction in semiempirical or empirical valence bond models,[27−32] although the theory and algorithm of MOVB are based on *ab initio* WFT and DFT approaches to define VB electronic configuration states. Effective Hamiltonian approaches are widely used in many different areas.[33−40] A major advantage of the EH-MOVB approach is that all VB matrix elements, including off-diagonal terms, are determined by an electronic structure method, which depends explicitly on all degrees of freedom in the system. In the empirical and semiempirical valence bond approaches, typically a simple function, depending on one degree of freedom, or a constant is used to treat the off-diagonal matrix elements in a VB-like Hamiltonian.[27,30,31,41] Note that empirical multiconfigurational models have been described to fit the energy, gradient, and Hessian of *ab initio* potential surfaces[40,42−44] using Gaussian and polynomial functions[45,46] or Shepard interpolation.[47−49]

In this paper, we show that the EH-MOVB method can be constructed using semiempirical QM models such as the Austin model 1 (AM1),[50] parametrization model 3 (PM3),[51] or Recife model 1 (RM1)[52] to yield the barrier height for a chemical reaction in agreement with experiments or with *ab initio* results. In the following, we first present the theoretical background, followed by computational details. Results and discussions are presented next. Finally, the paper concludes with a summary of the major findings of this study and future perspectives.

Scheme 1. Schematic Representation of the Block-Localization of Molecular Orbitals within Individual Molecular Fragments for the Reactant Diabatic State (left) and the Product Diabatic State (right) for the Hydride Transfer Reaction between Trimethylamine (**TMA-H**) and a Model for the Flavin Cofactor (**Nf**$^+$)^a
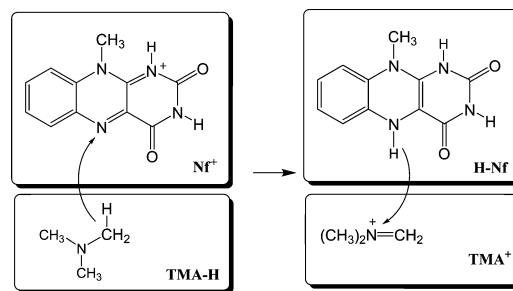
^{a} Atoms and charges in each rectangle specify the molecular block defined by the corresponding Lewis structure within which molecular orbitals are localized. The antisymetric wave function constructed from the two blocks on the left-hand side of the arrow, **TMA-H** and **Nf**$^+$, defines the reactant diabatic state, whereas that for the blocks on the right-hand side, **TMA**$^+$ and **H-Nf**, define the product diabatic state.

2. Method

A. The Mixed Molecular Orbital and Valence Bond (MOVB) Theory. In MOVB,[10−12,21] we use one Slater determinant wave function constructed using nonorthogonal block-localized molecular orbitals (BLMO) to define the reactant and product configurations. These electronic configurations are called diabatic states. The use of localized orbitals within molecular fragments has been explored by many groups in different applications such as reducing basis set superposition errors in weakly bound complexes,[17−19,53,54] and it has been used in other contexts.[14−16,55−61] For the hydride transfer reaction between trimethylamine, $(CH_3)_3N$ (**TMA-H**), and a flavin cofactor (**Nf**$^+$) model (hereafter simply called flavin), the wave function of the reactant diabatic state, $\Psi_r(\mathbf{R})$ (see Scheme 1), is defined by a single Slater determinant wave function in which molecular orbitals are block-localized on the two subsystems:

$$\Psi_r(\mathbf{R}) = \hat{A}\{\chi_r^{\mathbf{TMA-H}}\chi_r^{\mathbf{Nf}^+}\} \quad (1)$$

where $\mathbf{R}$ specifies all Cartesian atomic coordinates of the system and $\hat{A}$ is an antisymmetrization operator. The notations $\chi_r^{\mathbf{TMA-H}}$ and $\chi_r^{\mathbf{Nf}^+}$ in eq 1 specify the products of occupied BLMOs that are defined as linear combinations of atomic orbitals located on atoms in fragments **TMA-H** and **Nf**$^+$, respectively (Scheme 1). Similarly, the wave function of the product state (Scheme 1), $\Psi_p(\mathbf{R})$, is expressed as

$$\Psi_p(\mathbf{R}) = \hat{A}\{\chi_p^{\mathbf{TMA}^+}\chi_p^{\mathbf{H-Nf}}\} \quad (2)$$

where $\chi_p^{\mathbf{TMA}^+}$ and $\chi_p^{\mathbf{H-Nf}}$ denote the products of occupied BLMOs expanded over basis orbitals on atoms in fragments **TMA**$^+$ and **H-Nf**, respectively (Scheme 1).

It is important to note that the MOs within each fragment for each state are constrained to be orthogonal, but they are

nonorthogonal between different fragments.[11] Consequently, the MOVB model retains key characteristic features of valence bond theory in the use of nonorthogonal orbitals. The structure of the transformation matrix for the reactant and product states is block-diagonal.

$$\mathbf{C}_r = \begin{pmatrix} \mathbf{C}_r^{\mathbf{TMA-H}} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_r^{\mathbf{Nf^+}} \end{pmatrix} \text{ and } \mathbf{C}_p = \begin{pmatrix} \mathbf{C}_p^{\mathbf{TMA^+}} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_p^{\mathbf{H-Nf}} \end{pmatrix} \quad (3)$$

where $\mathbf{C}_r$ and $\mathbf{C}_p$ are the matrices of molecular orbital coefficients for the reactant and product states, respectively.[11] Note that the dimensions of the matrix elements in eq 3 are different as the hydride atom is grouped in different blocks. The total number of electrons within each fragment of each diabatic state is also fixed according to the corresponding Lewis structure (Scheme 1), and there is no chemical bond between the two fragments in each state.

In the present EH-MOVB model employing a semiempirical method, a special situation must be considered because of the neglect of diatomic differential overlap (NDDO) approximation.[62] The two-center, one-electron integral between two atoms that are located in different blocks is not included in the Fock matrix of either block, and it must be treated specifically. Note that these integrals are formally ignored in the NDDO approximation, but they are also treated as an exception in standard semiempirical methods because of the importance of these integrals in describing chemical bonding.[50] In MOVB, when the two bonding atoms involved in bond formation are in different molecular blocks (fragments), the two-center, one-electron integral is not treated by the standard semiempirical model, and the exclusion of this type of resonance integral affects the bonding properties as the chemical bonds are partially formed and broken across different blocks (fragments) at the transition state. Consequently, we need to include these resonance integrals for the corresponding bonds in the diabatic energy term as follows. For the reactant state in the present hydride transfer reaction, the reactant state diabatic energy is given as follows:

$$H_{rr}(\mathbf{R}) = \langle \Psi_r(\mathbf{R})|H|\Psi_r(\mathbf{R})\rangle + \alpha_{HN}S_{HN}\frac{1}{2}[\beta_s(H) + \beta_{sp}(N)] \quad (4)$$

and the energy for the product state is

$$H_{pp}(\mathbf{R}) = \langle \Psi_p(\mathbf{R})|H|\Psi_p(\mathbf{R})\rangle + \alpha_{CH}S_{CH}\frac{1}{2}[\beta_s(H) + \beta_{sp}(C)] \quad (5)$$

where S_{HN} and S_{CH} are the s-type overlap integrals between the acceptor nitrogen atom and the transferring hydrogen atom, and between the donor carbon atom and the migrating hydrogen atom, specified by the subscripts respectively, and $\beta_{sp}(X) = [\beta_s(X) + 3\beta_p(X)]/4$ at X = N or C and the β's being the standard semiempirical parameters for these atoms.[50,51] The use of s-type overlap integrals in eqs 4 and 5 is to preserve rotation invariance. In eqs 4 and 5, we treat α_{HN} and α_{CH} as semiempirical parameters, adjusted to yield the corresponding bond distances in agreement with DFT energies at the transition state. These two parameters associated with bonding interactions may also be considered

as EH-MOVB parameters, in addition to the two parameters to adjust diabatic coupling results.

The MOVB wave function for the reactive system is written as a linear combination of the diabatic states.

$$\Phi_g(\mathbf{R}) = a_r\Psi_r(\mathbf{R}) + a_p\Psi_p(\mathbf{R}) \quad (6)$$

where a_r and a_p are the configurational coefficients for the reactant and product diabatic states, respectively.[15,16,20,26] The potential energy of the adiabatic ground state, $V_g(\mathbf{R})$, is the lower energy root of the secular equation.

$$\begin{vmatrix} H_{rr}(\mathbf{R}) - V(\mathbf{R}) & H_{rp}(\mathbf{R}) - S_{rp}(\mathbf{R})V(\mathbf{R}) \\ H_{pr}(\mathbf{R}) - S_{pr}(\mathbf{R})V(\mathbf{R}) & H_{pp}(\mathbf{R}) - V(\mathbf{R}) \end{vmatrix} = 0 \quad (7)$$

where $V(\mathbf{R})$ is the adiabatic potential energy, $H_{rr}(\mathbf{R})$ and $H_{pp}(\mathbf{R})$ are the Hamiltonian matrix elements for the reactant and product diabatic states, respectively, $H_{rp}(\mathbf{R}) = H_{pr}(\mathbf{R})$ is the exchange integral (off-diagonal matrix element), and $S_{rp}(\mathbf{R}) = S_{pr}(\mathbf{R})$ is the overlap integral between the two diabatic states.

The Hamiltonian matrix elements in eq 7 are given as follows:[11,13]

$$H_{ab} = S_{ab}\Big\{Tr[(\mathbf{D}_{ab})^T\mathbf{h}] + \frac{1}{2}Tr[(\mathbf{D}_{ab})^T\mathbf{J}\mathbf{D}_{ab}] \\ - \frac{1}{4}Tr[(\mathbf{D}_{ab})^T\mathbf{K}\mathbf{D}_{ab}] + E_{nuc}\Big\} \quad (8)$$

where the subscripts a and b specify either the reactant (r) or the product (p) state or both; E_{nuc} is the nuclear Coulomb energy; S_{ab} and $\mathbf{D}_{ab}$ are the overlap integral and density matrix over nonorthogonal determinant wave functions; and $\mathbf{h}$, $\mathbf{J}$, and $\mathbf{K}$ are the standard one-electron, Coulomb, and exchange matrices. It is important to note that eq 8 is a general formula that is valid for *ab initio* and semiempirical WFT as well as for standard Kohn−Sham DFT.[13] In the latter case, the exchange integral $\mathbf{K}$ is replaced by the exchange-correlation potential.[13]

In reference,[20] we described two special situations to optimize the wave function of eq 6. In the first case, which is called the consistent diabatic configurations (CDC) MOVB, both the orbital coefficients (eq 3) and configurational coefficients are optimized as in the multiconfiguration self-consistent field method. An alternative approach is to variationally optimize the reactant and product state separately, followed by optimizing the configuraitonal coefficient in eq 6 with the orbital coefficients kept fixed. The latter configuration interaction procedure is called the variational diabatic configuration (VDC) MOVB to emphasize that the diabatic states are individually optimized. Both CDC and VDC states are useful in condensed phase simulations, although their applications will be addressed in future publications.

B. Effective Hamiltonian MOVB. We aim to develop an efficient (e.g., capable of carrying out nanosecond to microsecond dynamics simulations using the current computer architecture) and accurate (within 1 to 2 kcal/mol of experimental barrier height) QM/MM method for simulation of enzymatic reactions and chemical processes in solution using MOVB. Although *ab initio* MOVB and multistate

VBDFT provide a natural choice, and the former has indeed been applied to a number of condensed phase reactions,[10−12,59,63] it is still very time-demanding to carry out routine free energy simulations, in which a large number of atoms are treated quantum-mechanically. To this end, we have implemented the MOVB method into the CHARMM package,[64] based on the NDDO approximations.[62,65] The present implementation represents a significant advance in combined QM/MM methodology because (a) semiempirical methods are computationally efficient, allowing for statistical mechanical sampling in molecular dynamics simulations, and (b) the computational accuracy can be conveniently achieved using the nonorthogonal block-localized orbital approach described here.[26]

Experience shows that the qualitative features of the potential surface for chemical processes can be adequately represented by semiempirical models, such as AM1,[50] PM3,[51] or the self-consistent charge tight-bonding density functional algorithm (SCC-DFTB).[66] Consequently, we define and describe the reactant and product diabatic states using a semiempirical Hamiltonian. The quantitative errors in the computed barrier height and the energy of reaction inherited in the semiempirical method are eliminated by adjusting the EH-MOVB matrix elements[26] in a similar way to that in empirical or semiempirical VB models.[27,30,31,41] It should be realized that all combined QM/MM methods are semiempirical models in that one has to employ empirical potential functions such as the Lennard-Jones terms to approximate the quantum mechanical exchange repulsion and dispersion interactions between the QM and MM regions. Thus, the adjustment of the EH-MOVB matrix elements is no stranger to combined QM/MM methodologies.

Specifically, we introduce a parameter in the off-diagonal Hamiltonian matrix element H_{rp}, which is optimized in order to reproduce the barrier height for a given chemical reaction:

$$H_{rp}^{EH} = H_{rp} + \gamma_{rp} \tag{9}$$

In eq 9, H_{rp} is the MOVB off-diagonal matrix element that is determined directly (eq 8) using a given semiempirical model, γ_{rp} is a parameter that affects dominantly the computed barrier height, and H_{rp}^{EH} is the total effective Hamiltonian (EH) resonance (exchange) integral. Another formalism that we have explored is to scale the off-diagonal matrix element as follows:[26]

$$H_{rp}^{EH} = \zeta_{rp} H_{rp} \tag{10}$$

Both options can be useful, depending on the performance of the semiempirical model and the specific reaction considered, and both are available options in our implementation in CHARMM. In eq 9, the resonance integral is shifted by a constant value, whereas the scaling procedure in eq 10 affects the dependence of the resonance integral on the overlap between the reactant and product diabatic states. For the hydride transfer reaction between trimethylamine and flavin, we found that eq 9 yields the best results, and it is employed in the present study.

The second parameter that we introduce in the EH-MOVB model is the adjustment of the relative energy between the reactant and product diabatic states. Thus, if necessary, the diagonal MOVB matrix element for the product state, H_{pp}, is shifted by an amount of $\Delta\varepsilon$ to yield the desired energy of reaction for the process of interest:

$$H_{pp}^{EH} = H_{pp} + \Delta\varepsilon \tag{11}$$

The value of the parameter $\Delta\varepsilon$ is readily estimated as follows:

$$\Delta\varepsilon = \Delta E_{expt} - \Delta E_{MOVB} \tag{12}$$

where $\Delta E_{MOVB} = H_{pp}(\mathbf{R}_p) - H_{rr}(\mathbf{R}_r)$, which is the relative energy of the unshifted reactant and product diabatic state at their corresponding equilibrium geometries $\mathbf{R}_r$ and $\mathbf{R}_p$, and ΔE_{expt} is the experimental energy of reaction.

The procedure outlined above (eqs 9−12) is identical to that used in the parameter "calibration" of empirical valence bond models, such as that in refs 32 and 41, or more generally, of the semiempirical valence bond,[27−31] which allows the energies (barrier height and reaction energy) to be readily fitted to their targets exactly. In general, however, it is much more challenging to "calibrate" the variation of molecular structure along the entire reaction path, especially the precise geometry of the transition state. The sophistication of the mathematical algorithm used by Schlegel and Sonnenberg is a remarkable reflection of the difficulty in constructing an accurate potential energy surface employing empirical valence bond models.[45,46] The changes of the structural properties, including bond order and force constant, are critically important if one is interested in computing kinetic isotope effects, particularly the error-sensitive secondary effects (2° KIEs), for enzymatic reactions. Inaccuracy can easily be hidden in the large primary KIEs because they typically involve a significant loss of zero-point effects. Thus, agreement with the experiment in primary KIEs, which could be simply due to the loss of the reactant state stretching mode, is not necessarily an indication of good geometry at the transition state. In fact, it is essential to examine both the optimized structure and energy at the transition state to validate the quality of a two-state (or multistate) model against high-level electronic structural data.[20,26,45,46]

To this end, the off-diagonal matrix element in EH-MOVB (eq 9) is an explicit function of all degrees of freedom of the system, i.e., $H_{rp}(\mathbf{R}) = \langle\Psi_r(\mathbf{R})|H|\Psi_p(\mathbf{R})\rangle$.[10−13,20,26] Consequently, the full-dimensional potential surface can be adequately represented as accurately as the accuracy of the level of the electronic structure method permits, and the transition structure for a reaction can be obtained in accord with that optimized from WFT or DFT calculations. Note that the approach outlined in eqs 9−12 is in principle analogous to that used in effective Hamiltonian valence bond methods to parameterically model the *ab initio* matrix elements to reproduce the exact high-level results.[33−40,42−49]

3. Computational Details

All calculations are carried out using CHARMM c34a2,[67] modified with the implementation of the present EH-MOVB. The current QM/MM module in CHARMM at the semiempirical level, called SQUANTUM,[68] was implemented in our

group by Nam and Walker in 2004, based on a Fortran90 code.[69] SQUANTUM has been incorporated into the standard distribution and has become the default QM/MM module of CHARMM since version c33a1. The EH-MOVB method was implemented by Song in collaboration with Xie, and it has become a part of the SQUANTUM module with additional options to define the number of states and the number of blocks in each state as well as the associated options. The EH-MOVB method can provide a rigorous valence bond-like model for studying chemical reactions such that the users can conveniently calibrate the model to yield a potential energy surface with the desired barrier height and reaction energy as well as optimized geometry at the transition state. It should be noted that the present EH-MOVB is not a simple quantum mechanical representation of the ideas of empirical valence bond or semiempirical valence bond models such as the London-Eyring-Polanyi-Sato formalism. EH-MOVB is deeply rooted in the traditional approach of Heitler-London-Slater-Pauling function of valence bond theory.

The EH-MOVB module at the semiempirical level is computationally fast; for large systems, the computational bottleneck using our QM/MM potential is in the treatment of the classical long-range electrostatic effects with particle-mesh Ewald (PME) rather than the QM calculation itself. In addition, two options are available for determining the diabatic and adiabatic ground state energies: (1) the consistent diabatic state (CDC) method and (2) the variational diabatic state (VDC) model.[20] For those who are interested in using the energy gap between the product and reactant diabatic state as the reaction coordinate,[70] the VDC diabatic states should be used, since the variational diabatic state is of interest in this case.[10−12] The VDC determinants also provides the basis states in configuration interaction calculations to give the adiabatic ground state potential energy surface. On the other hand, if geometrical parameters are used to define the reaction coordinate on the adiabatic ground state potential surface, the CDC model is appropriate since this method yields the optimal adiabatic ground-state energy, and analytical gradients can be computed. Note that the CDC method is analogous to multiconfiguration self-consistent field (MCSCF) theory,[20,26] whereas the VDC approach is akin to a configuration interaction (CI) method.[10−12]

DFT calculations are carried out using Gaussian 03[71] modified to include the M06-2X functionals.[72,73] The 6-31+G(d) basis set is used throughout for all calculations. Geometries for the hydride transfer reaction between trimethylamine and flavin cofactor along the reaction coordinate defined below are optimized using the 6-31+G(d) basis set at each level of theory. The recently developed M06-2X functional, which produces similar energies in comparison with MP2 single point calculations, is used to calibrate the EH-MOVB model.

To describe the change in energy and wave function of the two Lewis bond states as the reaction takes place, we define the reaction coordinate here as the difference between the bond lengths of the central hydrogen atom, which is transferred, to the donor atom (C) of **TMA-H** and to the acceptor atom (N) of **Nf**$^+$:
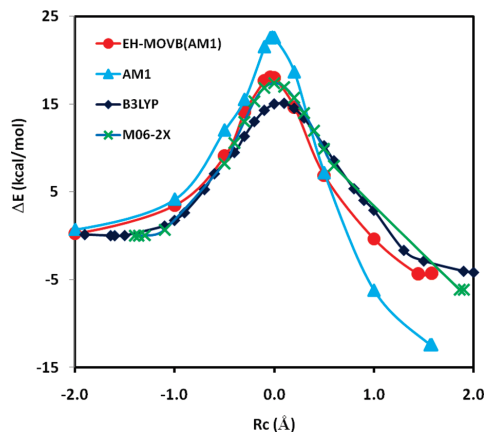
$$Rc = R(C - H) - R(H - N) \qquad (13)$$

Of course, one can use other definitions to monitor the progress of the reaction, including the difference between the corresponding bond orders or energies of the two Lewis bond states. The geometrical variable, corresponding to the asymmetric bond stretch coordinate, is a good choice and chemically intuitive.

4. Results and Discussion

The main goal of this study is to develop an effective Hamiltonian within MOVB theory to study chemical reactions in solution and in enzymes using CHARMM as a combined QM/MM potential. We hope to illustrate that the procedure can be conveniently used by biochemists as a research tool to help interpret experimental findings, with a straightforward calibration of the EH-MOVB model. We use the hydride transfer reaction from trimethylamine to a flavin cofactor model. The discussion of dynamics simulations is beyond the scope of this report and will be reported separately. We first carry out *ab initio* electronic structural calculations using DFT to yield the structures and energies along the hydride transfer reaction pathway. Then, we optimize the EH-MOVB Hamiltonian to reproduce the "high-level" data. The qualitative features and quantitative results of the diabatic configurations and the adiabatic potential surface will be discussed.

The adiabatic ground state potential energy surfaces determined using DFT with the B3LYP and M06-2X functionals are compared with the standard semiempirical AM1 model and the EH-MOVB method in Figure 1 as a function of the reaction coordinate Rc (eq 13) for the hydride transfer reaction between trimethylamine and a flavin cofactor. Optimized structures at the reactant state and product state complex and the transition state are illustrated in Figure 2 along with key structural parameters. The M06-2X density functional calculations yield an estimated barrier height of 17.4 kcal/mol and a relative energy of −6.1 kcal/mol between the product and reactant states. The popular hybrid B3LYP method underestimates the hydride transfer barrier at 15.1 kcal/mol. The semiempirical AM1 energy profile is qualitatively correct, but it contains two main problems; the computed energy of activation is 22.6 kcal/mol, about 5 kcal/mol too high, compared with the M06-2X value, and the predicted energy of reaction is too endothermic by 6.4 kcal/mol. The latter error is completely transferred into the MOVB relative energies of the reactant and product diabatic states, which can be easily corrected by shifting the product state up by an equal amount, which has no effect on gradient evaluations (Table 1). With an increase in the strength of diabatic coupling between the reactant and product states at the transition state, the barrier height can be lowered, and using the parameters listed in Table 1, we obtained an activation energy of 18.1 kcal/mol for the hydride transfer between trimethylamine and flavin and an energy of reaction of −4.4 kcal/mol. We note that the AM1 model finds another configuration in which the donor N−C−H unit is roughly coplanar with the flavin ring, and it is slightly lower in energy (by about 2 kcal/mol) than the configuration in which **TMA**

Figure 1. Computed potential energy profile along the minimum energy path (Rc = R[C−H] − R[H−N]) for the hydride transfer reaction between trimethylamine and the flavin model using EH-MOVB(AM1) (in red), AM1 (in light blue), B3LYP/6-31G(d) (in navy blue), and M06-2X/6-31G(d) (in green).
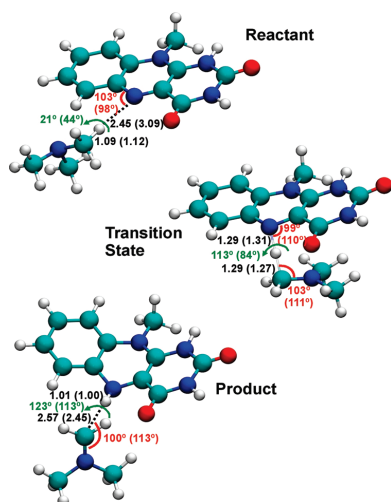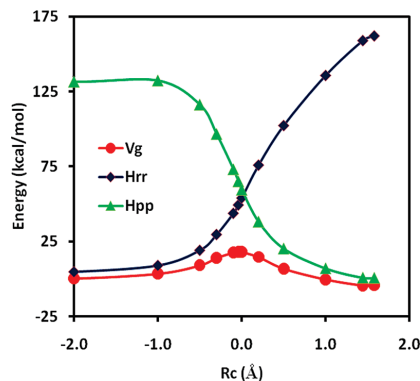

Figure 2. Optimized geometries for the reactant and product complexes and the transition state for the hydride transfer reaction depicted in Scheme 1. MOVB results are listed first, followed by DFT values in parentheses. Distances are given in angstroms and angles in degrees.

Table 1. EH-MOVB Parameters Used in This Study[a]

α_{CH}	α_{HN}	$\Delta\varepsilon$ (kcal/mol)	γ_{rp} (eV)
0.9	1.0	8.0	1.5

[a] The AM1 model is used to define the diabatic reactant and product states for the hydride transfer reaction between trimehylamine and a model flavin cofactor.

is under the plane of the flavin ring. The latter configuration is more closely aligned with the structure found in the active site in the human histone lysine-specific demethylase (LSD1) structure,[74] which is most relevant to the hydride transfer reaction pathway.
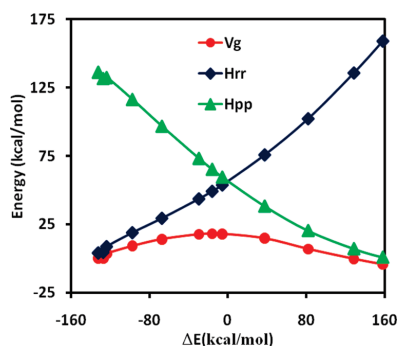

Figure 3. Computed potential energy surfaces for the diabatic reactant state (blue), the diabatic product state (green), and the adiabatic ground state (red) along the minimum energy path.

The optimized geometries at the reactant, product, and transition states from EH-MOVB(AM1) calculations are in accord with those obtained using M06-2X density functional theory. In particular, the donor (C−H) and acceptor (H−N) distances from the hydride atom transferred are 1.29 and 1.29 Å, respectively, which may be compared with the DFT (B3LYP) values of 1.27 and 1.31 Å. The potential energy surface about bond angles and torsional angles is relatively flat, and the accord between EH-MOVB(AM1) and M06-2X is reasonable (Figure 2).

The minimum energy path (MEP) for the hydride transfer from trimethylamine to flavin has been optimized as a function of the reaction coordinate defined by eq 13. In the present study, we have constrained the hydride migration to be collinear with the donor (C) and acceptor (N) atoms, whereas all other degrees of freedom are fully minimized using the ABNR algorithm in CHARMM.[64] The potential energy curves for the reactant and product diabatic states are shown in Figure 3 along with that for the adiabatic ground state. The reactant state potential shows a steady increasing as the reaction coordinate changes from the reactant to the product side. On the other hand, the potential energy surface is somewhat leveled off for the product state when the molecular geometry is in the reactant state configuration. The trend of the two diabatic potential energy curves is consistent with heterolytic bond cleavages of the reactant (C−H) and the product (N−H) species. At the diabatic state crossing point, which corresponds roughly to the location of the transition state of the hydride transfer reaction, the diabatic state is ca. 40 kcal/mol in energy above the adiabatic ground state, suggesting that there is significant electronic coupling between the reactant and product states. The coupling energy is similar to values determined for proton transfer and nucleophilic substitution reactions using *ab initio* WFT and DFT.[10−13,20,26]

Figure 4 exhibits the same potential curves shown in Figure 3, but they are plotted against the diabatic energy difference, or the energy-gap reaction coordinate.

$$\Delta E = H_{rr} - H_{pp} \qquad (14)$$

Figure 4. Computed potential energy surfaces in Figure 3 for the diabatic reactant state (blue), the diabatic product state (green), and the adiabatic ground state (red) represented as a function of the energy difference between the reactant and product diabatic states (i.e., the energy-gap reaction coordinate).

Figure 5. Computed potential energy profiles for the reactant (blue) and product (green) diabatic states along with the adiabatic ground state (red) and the reference potential as a function of the coupling parameter linearly connecting the reactant and product potentials. This reaction path is called the reference minimum energy path, which has a different meaning from that of Figure 1.
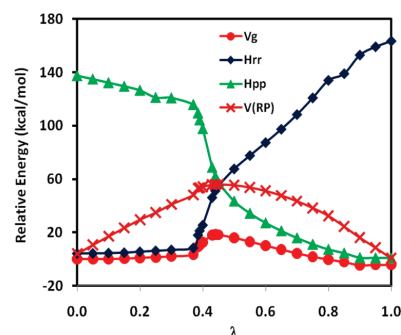
Figure 4 shows that the minimum energy potential surface for the adiabatic ground state and those for the diabatic states can be fully represented with the use of either a geometrical or an energy-gap reaction coordinate when the reaction profile is determined by optimizing the geometrical reaction coordinate.

For reactions in solutions or in enzymes, it is of interest to consider the effects of solvent or protein reorganization, and this is often presented using the energy-gap reaction coordinate (eq 14). Although this is easily modeled using an empirical force field to represent the diabatic states, it is far from straightforward if a combined QM/MM potential is employed. The MOVB theory is the first and only QM/MM approach at this time to provide well-defined diabatic states for condensed phase simulations, and *ab initio* MOVB-QM/MM methods have been utilized in the study of solvent effects and reorganization energies for several reactions in solution.[10−13] Of course, empirical potential functions have been used extensively to describe the energy-gap coordinate.[41,75,76] The present EH-MOVB approach in the context of a QM electronic structure theory can be conveniently calibrated to yield accurate results and applied to enzymatic catalysis using the program CHARMM. The free energy reaction profile as a function of the energy-gap reaction coordinate is typically obtained through a coupled free energy perturbation simulation,[10,11,41] which drives the solvent and protein configurations from the reactant state to the product state using a reference potential (which is also called a mapping potential),[41] $V_{RP}(\mathbf{R})$, and umbrella sampling that transforms the biased simulations with $V_{RP}(\mathbf{R})$ into the true adiabatic ground-state potential surface, $V_g(\mathbf{R})$.

The reference potential is typically expressed as a mixture of the diabatic reactant and product energy through a coupling parameter λ:

$$V_{RP}(\lambda) = (1 - \lambda)H_{rr}(\mathbf{R}) + \lambda H_{pp}(\mathbf{R}) \qquad (15)$$

where λ is a parameter that varies from 0 (reactant) to 1 (product), and $\mathbf{R}$ specifies the instantaneous geometry of the system. In the present study of the model hydride transfer

from **TMA-H** to flavin (**Nf**$^+$), we optimized the reference minimum-energy path (RMEP) defined by eq 15. Then, using the geometries along this reference minimum-energy path, we carried out single-point energy calculations to determine the adiabatic ground state energy. Note that this "RMEP" is not the true adiabatic ground-state MEP (Figures 3 and 4) determined using the EH-MOVB potential, $V_g(\mathbf{R})$, because the structures are optimized using different potential energy surfaces.

Figure 5 depicts the diabatic potential energies and the adiabatic ground state energy, along with the reference potential (eq 15), as a function of the coupling parameter. Since the reference potential is dominantly determined by the reactant diabatic state when λ is less than 0.5, there is a rapid geometry change in the hydride transfer coordinate, which is not explicitly specified by the coupling parameter λ and cannot be effectively restrained to yield a smooth variation. Consequently, there is a sudden change in the molecular geometry as the hydride is fully transferred to the carbon atom, corresponding to a geometrical description of $Rc = -0.7$ Å to $Rc = -1.8$ Å. The ground-state potential is shown as a function of the geometrical reaction coordinate in Figure 6. This is accompanied by a rather steep increase in the reactant diabatic state and the adiabatic ground state potential in the region of $\lambda = 0.4$ and 0.5 (Figure 5). Interestingly, the overall reference potential shows smoother variations (curve in maroon) due to the compensating contributions from the product diabatic state. The computed barrier height is 18.3 kcal/mol along the RMEP, similar to that of the MEP for the hydride transfer.

Figure 7 recasts the data illustrated in Figure 5, but the adiabatic ground state potential energy surface $V_g(\mathbf{R})$ is plotted against the energy-gap reaction coordinate ΔE. In contrast to Figures 5 and 6, the potential $V_g(\mathbf{R})$ appears to be surprisingly smooth, despite the fact that part of the geometrical variations along the reaction path in fact is discontinuous in Figure 5. Figure 7 shows that a nonsmooth geometrical transition that gives rise to an abrupt energy change can be hidden behind the seeming smooth energy
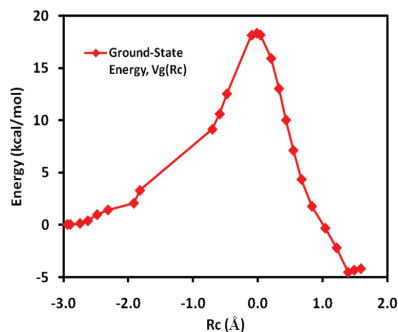
Figure 6. Potential energy profile for the hydride transfer reaction between methylamine and the model flavin cofactor plotted against the geometrical reaction coordinate (eq 14) following the reference minimum energy path in Figure 5.
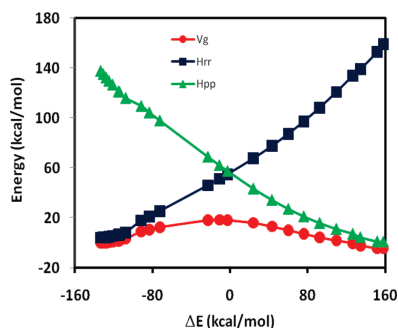
Figure 7. Computed potential energy profiles for the reactant (blue) and product (green) diabatic states and the adiabatic ground state (red) as a function of the energy gap-reaction coordinate for structures obtained along the reference minimum energy path in Figure 5.

curve when the adiabatic ground-state potential is given as a function of a geometry-implicit coordinate such as the energy-gap representation. This observation suggests that it is critically important to report and show both energy results and the corresponding geometries of the reactive molecule in calculations that employ the reference potential of eq 15.

Before we leave this section, we consider the procedure used in condensed phase and enzyme calculations.[10,11,41] In this case, the reference potential of eq 15 will be used in a series of discrete free energy perturbation simulations with fixed values of λ_i to yield the free energy differences as λ_i changes from 0 to 1. Thus, the free energy at λ relative to the reactant state ($\lambda_0 = 0$) is determined as follows:

$$\Delta G_{RP}(\lambda) = -RT \sum_{i=0}^{\lambda} \ln \langle e^{-[V_{RP}(\lambda_{i+1}) - V_{RP}(\lambda_i)]/RT} \rangle_i \qquad (16)$$

where $\langle ... \rangle_i$ specifies an ensemble average over the potential $V_{RP}(\lambda_i)$; the summation runs to a value $\lambda = \lambda_{i+1}$. Here, the use of the arbitrary reference potential is purely for the purpose of moving the system to go from molecular configurations corresponding to the reactant state ensemble into the product state. To obtain the free energy of the true

ground state potential surface, governed by the distribution $e^{-V_g(\Delta E)/RT}$, an umbrella sampling-like procedure is applied to the configurations sampled on the basis of the distribution of $e^{-V_{RP}(\lambda_i)/RT}$. Thus,

$$\Delta G(\Delta E) = \Delta G_{RP}(\lambda_i) - RT \ln\{\rho_{RP}^i(\Delta E)\langle e^{-[V_g - V_{RP}(\lambda_i)]/RT}\rangle_i\} \qquad (17)$$

where the quantity $\rho_{RP}^i(\Delta E)$ is the normalized distribution of configurations that have a value of ΔE in the ensemble sampled by the reference potential $V_{RP}(\lambda_i)$.

An important distinction that should be made is that the procedure outlined in eqs 16 and 17 yields the *free energy* profile, or the potential of mean force, as a function of an ensemble of configurations, all having the energy gap ΔE. Obviously, it is not and should not be compared with the *potential energy* surface. Furthermore, the "reaction path" mapped by eq 16 is not the minimum energy path of the adiabatic ground state, nor the reference minimum energy path. Thus, the energy computed, either by averaging over all configurations sampled on the basis of eq 16 or by selecting a single structure of its ensemble, is not directly comparable to results rigorously defined by the MEP. Obviously, it can be deceptive when potential energies free energies obtained along the minimum energy path and or single-point energy calculations on selected geometries from a statistical ensemble are mixed together and compared without rigorously specifying their origins.

5. Conclusions

The effective Hamiltonian−molecular orbital and valence bond (EH-MOVB) method based on nonorthogonal block-localized molecular orbitals has been implemented into the program CHARMM for molecular dynamics simulations of chemical and enzymatic reactions, making use of semiempirical quantum mechanical methods. Building upon previous results using *ab initio* MOVB theory, we introduce two parameters in the EH-MOVB method, along with the addition of the two-center, one-electron integrals across different molecular blocks which may be considered as parameters, such that the barrier height and the relative energy between the reactant and product state for a given chemical reaction can be fitted in good agreement with experimental or high-level *ab initio* and DFT results. The EH-MOVB method provides a highly accurate and computationally efficient QM/MM model for dynamics simulation of chemical reactions in solution. The MOVB theory is the first and currently the only QM/MM method that allows the potential of mean force to be determined as a function of the energy-gap reaction coordinate for characterization of solvent reorganization effects.

The EH-MOVB method is illustrated by examination of the potential energy surface of the hydride transfer reaction from trimethylamine to a flavin cofactor model in the gas phase. In the present study, we employ the semiempirical AM1 model, which yields a qualitatively correct energy profile along the minimum energy path (Figure 1). However, as in most practical applications using semiempirical Hamiltonians, the quantitative results are not satisfactory. Tradi-

tionally, there is no systematic way of improving the semiempirical model, even though the qualitative features of structure and energy are reasonable. In EH-MOVB, the barrier height is optimized to reproduce the desired (accurate) value in the gas phase (i.e., the intrinsic performance of the effective Hamiltonian) either by scaling or by adding a constant to the off-diagonal matrix element. The present EH-MOVB method offers an alternative approach to characterization of solvent and protein-reorganization effects in the realm of truly combined QM/MM simulations.

Supporting Information Available: Structures optimized using the standard AM1 method, the EH-MOVB method along the minimum energy path and along the reference minimum energy path, and the B3LYP/6-31+G(d) method along the minimum energy path. All semiempirical calculations were performed with the SQUANTM module of CHARMM. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198.

(2) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.

(3) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 3188.

(4) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 154.

(5) Gao, J. *J. Am. Chem. Soc.* **1991**, *113*, 7796.

(6) Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *233*, 231.

(7) Marti, S.; Moliner, V.; Tunon, I. *J. Chem. Theory Comput.* **2005**, *1*, 1008.

(8) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *Biochemistry* **2003**, *42*, 13558.

(9) Nam, K.; Cui, Q.; Gao, J.; York, D. M. *J. Chem. Theory Comput.* **2007**, *3*, 486.

(10) Mo, Y.; Gao, J. *J. Comput. Chem.* **2000**, *21*, 1458.

(11) Mo, Y.; Gao, J. *J. Phys. Chem. A* **2000**, *104*, 3012.

(12) Gao, J.; Garcia-Viloca, M.; Poulsen, T. D.; Mo, Y. *Adv. Phys. Org. Chem.* **2003**, *38*, 161.

(13) Cembran, A.; Song, L.; Mo, Y.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 2702.

(14) Mo, Y.; Peyerimhoff, S. D. *J. Chem. Phys.* **1998**, *109*, 1687.

(15) Mo, Y.; Zhang, Y.; Gao, J. *J. Am. Chem. Soc.* **1999**, *121*, 5737.

(16) Mo, Y.; Gao, J.; Peyerimhoff, S. D. *J. Chem. Phys.* **2000**, *112*, 5530.

(17) Stoll, H.; Preuss, H. *Theor. Chem. Acc.* **1977**, *46*, 12.

(18) Gianinetti, E.; Raimondi, M.; Tornaghi, E. *Int. J. Quantum Chem.* **1996**, *60*, 157.

(19) Gianinetti, E.; Vandoni, I.; Famulari, A.; Raimondi, M. *Adv. Quantum Chem.* **1998**, *31*, 251.

(20) Song, L.; Gao, J. *J. Phys. Chem. A* **2008**, *112*, 12925.

(21) Gao, J.; Mo, Y. *Prog. Theor. Chem. Phys.* **2000**, *5*, 247.

(22) Cooper, D. L.; Gerratt, J.; Raimondi, M. *Adv. Chem. Phys.* **1987**, *69*, 319.

(23) Hiberty, P. C.; Flament, J. P.; Noizet, E. *Chem. Phys. Lett.* **1992**, *189*, 259.

(24) Wu, W.; Song, L.; Cao, Z.; Zhang, Q.; Shaik, S. *J. Phys. Chem. A* **2002**, *106*, 2721.

(25) Song, L.; Mo, Y.; Zhang, Q.; Wu, W. *J. Comput. Chem.* **2005**, *26*, 514.

(26) Song, L.; Mo, Y.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 174.

(27) Sato, S. *J. Chem. Phys.* **1955**, *23*, 592.

(28) Kuntz, P. J.; Nemeth, E. M.; Polanyi, J. C.; Rosner, S. D.; Young, C. E. *J. Chem. Phys.* **1966**, *44*, 1168.

(29) Raff, L. M.; Stivers, L.; Porter, R. N.; Thompson, D. L.; Sims, L. H. *J. Chem. Phys.* **1970**, *52*, 3449.

(30) Raff, L. M. *J. Chem. Phys.* **1974**, *60*, 2220.

(31) Silver, D. M.; Brown, N. J. *J. Chem. Phys.* **1980**, *72*, 3859.

(32) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218.

(33) Sheppard, M. G.; Freed, K. F. *J. Chem. Phys.* **1981**, *75*, 4507.

(34) Hurtubise, V.; Freed, K. F. *Adv. Chem. Phys.* **1993**, *83*, 465.

(35) Martin, C. H.; Graham, R. L.; Freed, K. F. *J. Phys. Chem.* **1994**, *98*, 3467.

(36) Bernardi, F.; Olivucci, M.; Robb, M. A. *J. Am. Chem. Soc.* **1992**, *114*, 1606.

(37) Bearpark, M. J.; Robb, M. A.; Bernardi, F.; Olivucci, M. *Chem. Phys. Lett.* **1994**, *217*, 513.

(38) Bearpark, M. J.; Bernardi, F.; Olivucci, M.; Robb, M. A. *J. Phys. Chem. A* **1997**, *101*, 8395.

(39) Bearpark, M. J.; Smith, B. R.; Bernardi, F.; Olivucci, M.; Robb, M. A. *ACS Symp. Ser.* **1998**, *712*, 148.

(40) Chang, Y. T.; Miller, W. H. *J. Phys. Chem.* **1990**, *94*, 5884.

(41) Aqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523.

(42) Schmitt, U. W.; Voth, G. A. *J. Phys. Chem. B* **1998**, *102*, 5547.

(43) Day, T. J. F.; Soudackov, A. V.; Cuma, M.; Schmitt, U. W.; Voth, G. A. *J. Chem. Phys.* **2002**, *117*, 5839.

(44) Maupin, C. M.; Wong, K. F.; Soudackov, A. V.; Kim, S.; Voth, G. A. *J. Phys. Chem. A* **2006**, *110*, 631.

(45) Schlegel, H. B.; Sonnenberg, J. L. *J. Chem. Theory Comput.* **2006**, *2*, 905.

(46) Sonnenberg, J. L.; Schlegel, H. B. *Mol. Phys.* **2007**, *105*, 2719.

(47) Kim, Y.; Corchado, J. C.; Villa, J.; Xing, J.; Truhlar, D. G. *J. Chem. Phys.* **2000**, *112*, 2718.

(48) Tishchenko, O.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 13530.

(49) Lin, H.; Zhao, Y.; Tishchenko, O.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 1237.

(50) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(51) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.

235

(52) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. *J. Comput. Chem.* **2006**, *27*, 1101.

(53) Stoll, H.; Wagenblast, G.; Preuss, H. *Theor. Chim. Acta* **1980**, *57*, 169.

(54) Raimondi, M.; Famulari, A.; Specchio, R.; Sironi, M.; Moroni, F.; Gianinetti, E. *THEOCHEM* **2001**, *573*, 25.

(55) Mo, Y.; Gao, J. *J. Phys. Chem. A* **2001**, *105*, 6530.

(56) Mo, Y.; Subramanian, G.; Gao, J.; Ferguson, D. M. *J. Am. Chem. Soc.* **2002**, *124*, 4832.

(57) Mo, Y.; Schleyer, P. v. R.; Wu, W.; Lin, M.; Zhang, Q.; Gao, J. *J. Phys. Chem. A* **2003**, *107*, 10011.

(58) Mo, Y.; Wu, W.; Song, L.; Lin, M.; Zhang, Q.; Gao, J. *Angew. Chem., Int. Ed.* **2004**, *43*, 1986.

(59) Mo, Y.; Gao, J. *J. Phys. Chem. B* **2006**, *110*, 2976.

(60) Khaliullin, R. Z.; Head-Gordon, M.; Bell, A. T. *J. Chem. Phys.* **2006**, *124*, 204105/1.

(61) Mo, Y.; Gao, J. *Acc. Chem. Res.* **2007**, *40*, 113.

(62) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129.

(63) Mo, Y.-r.; Alhambra, C.; Gao, J.-l. *Huaxue Xuebao* **2000**, *58*, 1504.

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

(65) Pople, J. A.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S136.

(66) Elstner, M.; Porezag, D.; Juugnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Sukai, S.; Seifect, G. *Phys. Rev. B* **1998**, *58*, 7260.

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

(68) Nam, K.; Prat-Resina, X.; Garcia-Viloca, M.; Devi-Kesavan, L. S.; Gao, J. *J. Am. Chem. Soc.* **2004**, *126*, 1369.

(69) Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comput. Chem.* **2008**, *29*, 1019.

(70) Marcus, R. A. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1111.

(71) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision D.01; Gaussian, Inc.: Pittsburgh, PA, 2004.

(72) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 1009.

(73) Zheng, J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2007**, *111*, 4632.

(74) Chen, Y.; Yang, Y.; Wang, F.; Wan, K.; Yamane, K.; Zhang, Y.; Lei, M. *Proc. Natl. Acad. Sc. U.S.A.* **2006**, *103*, 13956.

(75) Billeter, S. R.; Webb, S. P.; Agarwal, P. K.; Iordanov, T.; Hammes-Schiffer, S. *J. Am. Chem. Soc.* **2001**, *123*, 11262.

(76) Hatcher, E.; Soudackov, A. V.; Hammes-Schiffer, S. *J. Am. Chem. Soc.* **2007**, *129*, 187.

CT1001686

Explicit Polarization (X-Pol) Potential Using ab Initio Molecular Orbital Theory and Density Functional Theory[†]

Lingchun Song,* Jaebeom Han, Yen-lin Lin, Wangshen Xie, and Jiali Gao*

Department of Chemistry, Digital Technology Center and Supercomputing Institute University of Minnesota, Minneapolis, Minnesota 55455-0431

The explicit polarization (X-Pol) method has been examined using ab initio molecular orbital theory and density functional theory. The X-Pol potential was designed to provide a novel theoretical framework for developing next-generation force fields for biomolecular simulations. Importantly, the X-Pol potential is a general method, which can be employed with any level of electronic structure theory. The present study illustrates the implementation of the X-Pol method using ab initio Hartree−Fock theory and hybrid density functional theory. The computational results are illustrated by considering a set of bimolecular complexes of small organic molecules and ions with water. The computed interaction energies and hydrogen bond geometries are in good accord with CCSD(T) calculations and B3LYP/aug-cc-pVDZ optimizations.

1. Introduction

Previously, we introduced an explicit polarization (X-Pol) method for condensed-phase and macromolecular simulations.[1−6] The X-Pol potential is based on electronic structure theory as a framework for developing next-generation force fields,[1−3] that is, to go beyond the traditional Lifson-type empirical potential (also known as molecular mechanics (MM))[7,8] used in essentially all current atomistic simulations of proteins and nucleic acids. The X-Pol potential is designed to make the fundamental paradigm change in the functional form of MM force field and in the representation of biomolecular systems. In the X-Pol method, a molecular system is partitioned into fragments, such as an individual solvent molecule or a peptide unit or a group of such entities. The electronic interaction within each fragment is treated using electronic structure theory, and the interactions between two fragments are described by Hartree product of the antisymmetric wave functions of individual fragments. Short-range exchange repulsion and long-range dispersion-like attraction between fragments are included by pairwise functions.[1−3] The variational X-Pol theory allows analytical gradients of the energy to be efficiently evaluated in dynamics simulations.[5] The X-Pol method is a general theory that can be implemented using any electronic structural methods including wave function theory (WFT) and density functional theory (DFT). Recently, we reported a molecular dynamics simulation of a fully solvated BPTI protein, consisting of 14 281 atoms and 29 026 basis functions,[6] employing the quantal X-Pol potential based on an approximate MO theory. It is possible to run about 3.5 ps (1 fs integration step) per day on a single 2.66 GHz processor.[6] That work demonstrated the feasibility of an entirely new concept in force field development for large-scale simulations. In this article, we illustrate the X-Pol method by making use of ab initio WFT MO theory and DFT, and present computational results on bimolecular interactions.

Before we begin, it is useful to briefly consider the past and current developments of molecular mechanical force fields,

which date back to the pioneering studies of steric effects, independently by Hill[9] and by Westheimer,[10] and the subsequent developments of algorithms and applications to organic compounds.[11,12] The force field for biomolecular simulations was established by Lifson in the 1960s,[7,13] which led to the first molecular dynamics simulation of a protein by McCammon, Gelin, and Karplus.[14] Significant progress has been made in the accuracy of conventional force fields thanks to the tremendous efforts of parametrization by numerous groups in the past forty years. In fact, a major current push is to incorporate explicit polarization terms to treat electrostatic interactions.[15] Nevertheless, it is a sobering fact to note that the fundamental representation of biomolecular systems and the basic functional forms, including polarization terms,[16] in these force fields have hardly changed in the past forty years.[7,14,15,17,18] Undoubtedly, classical force fields will continue to be widely used and the accuracy will be further improved. However, despite the success of molecular mechanics in biomacromolecular modeling, there are also shortcomings, such as the arbitrariness in the choice of energy terms and the associated degrees of freedom, a lack of systematic approach to treat anharmonicity and cross coupling of energy terms, and the difficulty in describing electronic polarization and charge transfer. A fundamental change in force field development is warranted to increase the predictability of quantitative computational biology.

In the X-Pol potential, the system is treated explicitly by electronic structure theory and the wave function (or electron density) for each fragment is optimized by self-consistent field (SCF) method in the presence of the electric field due to all other fragments until the convergence is achieved for the entire system.[1−6] The internal energy terms and electrostatic potentials used in the classical force field are replaced and described explicitly by electronic structure theory. Consequently, they are obtained directly from quantum mechanical calculations, and electronic polarization and charge transfer are represented naturally in the theory. Furthermore, such a method can be directly used to model chemical reactions, electron transfer, and electronically excited states. The X-Pol potential, implemented using the semiempirical Austin Model 1 (AM1) method,[19] has been tested and applied to the simulation of liquid water[2] and

X-Pol Potential

J. Phys. Chem. A, Vol. 113, No. 43, 2009 **11657**

liquid hydrogen fluoride,[20] and has been extended to dynamics simulations of a fully solvated protein in aqueous solution.[6]

The X-Pol potential and its associated linear scaling method represent an entirely different approach from divide-and-conquer-type algorithms.[21−24] The divide-and-conquer (D&C)[21−24] and localized orbital methods[25] are linear scaling approaches to efficiently obtain a solution of the Hartree−Fock or Kohn−Sham equations for large molecular systems. In contrast, the X-Pol method is a quantum-mechanical force field (QMFF), whose energy is not the Hartree−Fock or Kohn−Sham DFT energy of the entire system.[1−6] Furthermore, the X-Pol potential is variational and can be efficiently used to carry out molecular dynamics simulations of solvated proteins,[5,6] whereas the D&C remains too time-consuming to carry out millions of electronic structure calculations for condensed-phase systems. The X-Pol potential also differs from combined quantum mechanical and molecular mechanical (QM/MM) approaches that employ a polarizable force field in the MM region.[26,27] The difficulties and uncertainties of treating molecular polarization in a classical force field are still present in coupled QM/MM-pol models. The mutual polarization of the entire system is treated consistently and equally in the X-Pol method.[1−3] Following our initial work,[1,2] Kitaura et al. described an approach, called fragment molecular orbital,[28,29] a procedure similar to the nonvariational, double self-consistent-field (DSCF) X-Pol method described in ref 1, and it has been successfully applied to numerous applications.[30] Recently, Gascon et al. published a self-consistent space-domain decomposition implementation of the X-Pol method for computing electrostatic potentials of proteins.[31] These authors used Morokuma's ONIOM scheme to carry out the DSCF optimization. Field also described a similar implementation making use of both the AM1 and HF/STO-3G method for water.[32] A closely related approach is the effective fragment potential (EFP) model developed by Gordon and co-workers.[33,34] The EFP method represents the electronic energies of interacting fragments by a set of analytical potentials optimized to fit the electronic structural data.

In what follows, we first present the X-Pol method with emphasis on the implementation using density functional theory. Then, in section 3, we describe the algorithm and computational details. Section 4 highlights the optimization strategies and results on bimolecular complexes. Finally, we summarize the major findings of this study and future perspectives on multilevel approaches for biomolecular modeling and simulations.

2. Theoretical Background

The X-Pol potential is based on a hierarchy of three levels of approximations.[1−5] At each level, if the approximation is not made, the X-Pol method reduces to the standard electronic structure theory at that stage. By applying these approximations, we achieve computational efficiency. Furthermore, it allows the introduction of one set of justifiable parameters associated with each approximation to achieve computational accuracy. For convenience of discussion, we consider a system of N closed-shell molecules, called fragments, that are not covalently connected. The generalization for treating covalently connected fragments has been described previously,[3] making use of the generalized hybrid orbital (GHO) scheme developed for combined QM/MM simulations.[35−39]

2.1. Level-1 (L1) Approximation and the Energy Expression of the X-Pol Potential. The first approximation in the X-Pol method is on the construction of the total molecular wave function, Φ (and the electron density), of the system, which is assumed to be a Hartree product of the antisymmetric wave functions of the individual fragments ($\{\Psi^A; A = 1,..., N\}$):[1]

$$\Phi = \prod_{A=1}^{N} \Psi^A \tag{1}$$

Here, the wave function Ψ^A may be approximated by a single determinant or by multiconfigurational methods such as the complete active space self-consistent field (CASCF) model or valence bond (VB) theory. In the rest of this article, the individual molecular wave function for fragment A is written as a single Slater determinant of m^A doubly occupied molecular orbitals, $\{\psi_i^A; i = 1,..., m^A\}$, which are linear combinations of atomic orbitals located on atoms of fragment A, subject to the orthonormal constraint:

$$\sum_{\mu\nu} c_{i\mu}^A c_{j\nu}^A S_{\mu\nu}^A - \delta_{ij} = 0 \tag{2}$$

where $c_{i\mu}^A$ and $c_{j\nu}^A$ are orbital coefficients and $S_{\mu\nu}^A$ is the overlap integral between atomic orbitals χ_μ^A and χ_ν^A in fragment A.

The approximation of eq 1 is equivalent to neglect of the exchange repulsion between electrons in different fragments, and the partition of the system into fragments ignores electron transfer and dispersion interactions.[1−3] However, this approximation is quite reasonable in the spirit of force field development and significantly simplifies computation. Nevertheless, it is essential to account for the energies due to exchange repulsion and dispersion attraction to prevent collapse of electrostatic interactions and to include van der Waals forces between different fragments, respectively, both of which are, in principle, dependent on the instantaneous molecular wave functions (or electron densities). An empirical or semiempirical approach based on perturbation theory is required to determine these energy terms,[1−3] for which one set of empirical parameters is introduced.

$$E_{\text{int},ed} = \frac{1}{2} \sum_{A \neq B}^{N} E_{\text{int},ed}^{AB}[\rho^A, \rho^B, \{\zeta^{AB}\}] \tag{3}$$

where ρ^A and ρ^B are the electron densities of fragments A and B, $E_{\text{int},ed}^{AB}[\rho^A, \rho^B, \{\zeta^{AB}\}]$ specifies the exchange-repulsion and dispersion (ed) interaction functional of the two charge fragments, and $\{\zeta^{AB}\}$ is a set of atomistic parameters to correct the errors due to the Hartree-product approximation in eq 1.

The total energy of the system is given as follows:[1]

$$E_{\text{tot}} = \sum_{A=1}^{N} \left(E^A + \frac{1}{2} E_{\text{int}}^A \right) + E_{\text{int,ed}} \tag{4}$$

where E^A is the energy of fragment A with the wave function Ψ^A, E_{int}^A is the Coulomb interaction energy between fragment A and all other fragments, and $E_{\text{int,ed}}$ is the exchange-repulsion and dispersion energy among all fragments (see below).

In Hartree−Fock (HF) theory, the energy for fragment A, which is denoted by the subscript HF, is written as

$$E^A = E_{\text{HF}}^A = \sum_i 2H_i^A + 2\sum_{i,j}\left(J_{ij}^A - \frac{1}{2}K_{ij}^A\right) + E_{\text{nuc}}^A \tag{5}$$

238

where the subscripts i and j specify doubly occupied MOs in fragment A and the terms in eq 5 are respectively the one-electron integrals (H_i^A) that include the electronic kinetic energy and the electron−nucleus attraction, the Coulomb integrals (J_{ij}^A), the exchange integrals (K_{ij}^A), and the nuclear repulsion energy (E_{nuc}^A).

In DFT, the molecular orbitals, subject to the constraint of eq 2, are the Kohn−Sham (KS) orbitals,[40] from which the electron density for fragment A is obtained (this can readily be generalized to spin-polarized systems):

$$\rho^A(\mathbf{r}) = 2 \sum_i^{m^A} |\psi_i^A(\mathbf{r})|^2 \qquad (6)$$

The energy for fragment A is then

$$E^A = E_{DFT}^A[\rho^A(\mathbf{r})] = \sum_i 2H_i^A + \sum_{i,j} 2J_{ij}^A + E_{xc}^A[\rho^A(\mathbf{r})] + E_{nuc}^A \quad (7)$$

where the first two terms have expressions identical to those in HF theory (eq 5), and the third term, $E_{xc}^A[\rho^A(\mathbf{r})]$, is the energy component accounting for all the effects of exchange and correlation of electrons in fragment A.

2.2. Level-2 (L2) Approximation and Coulomb Interactions between QM Fragments. The Coulomb interaction energy between fragment A and all other fragments arises from summations over electrons and nuclei from fragment A, and it has identical expressions for both DFT and HF theory:

$$E_{int}^A = \sum_i 2I_i^A + \sum_a L_a^A \qquad (8)$$

where a specifies an atom in fragment A. The two terms of eq 8, I_i^A and L_a^A, represent the interaction energies of orbital i and the nuclear charge of atom a of fragment A with the total external electrostatic potential (ESP) due to all other fragments in the system.[1,3] The external ESP on fragment A is given as follows:

$$V^A(\mathbf{r}) = \sum_{B \neq A} \left[-\int \frac{\rho^B(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \sum_b \frac{Z_b^B}{|\mathbf{r} - \mathbf{R}_b^B|} \right] \quad (9)$$

where the summation is over all fragments except A, $\rho^B(\mathbf{r}')$ is the electron density of fragment B and $\mathbf{R}_b^B$ and Z_b^B are the nucleus position and charge of atom b in fragment B. Thus,

$$I_i^A = \langle \psi_i^A | V^A(\mathbf{r}) | \psi_i^A \rangle \qquad (10)$$

$$L_a^A = Z_a^A V^A(\mathbf{R}_a^A) \qquad (11)$$

The second, L2, approximation in the X-Pol method is concerned with the numerical calculation of the two-electron four-index integrals of eq 10. In principle, they can be enumerated analytically in exactly the same way as in solving the HF or KS equations,[4] which requires no approximation at this level of the X-Pol hierarchy. However, the enormous

amount of such integrals between fragment pairs for a solvated protein present a serious limitation in computational efficiency, a crucial feature required in a force field. Fortunately, extensive investigations of the electrostatic potential of eq 9 using electronic structure methods show that it can be accurately and efficiently determined by a variety of approximate approaches.[41−43] Thus, eq 10 can be determined by a multipole expansion of the two-electron integrals,[44] or transformed into one-electron integrals using an effective ESP that best reproduces the exact results from the electronic structure theory. Such an effective ESP can be expressed in terms multipoles located either on a single center of the fragment or on individual atoms, or in terms of atom-centered partial charges (monopoles).[33,45] Thus, depending on the user's flavor, a series of X-Pol potentials can be developed on the basis of the specific choice of representing the ESP in eq 9.

We choose to use atom-centered monopoles, i.e., partial charges, in the present study to represent the ESP in eq 9.[1−6] Still, there are further considerations to be made. The partial atomic charges can be derived directly from population analysis of the molecular wave function for each fragment, such as the Mulliken population[46] and Lowdin population method. The advantage of using population analysis is that analytical derivatives can be easily computed. The natural orbital technique and the class IV charge models (CM4) are alternative selections.[47] A popular approach, used in classical force field development, which in principle yields the closest agreement with the original ESP, is the ESP-fitted partial charges.[41,42,48] However, it poses difficulty to derive expression for analytical gradient (note that we do not use ESP-fitted charges to compute the Coulomb energy and forces as a substitute for electronic structure calculations as is done in some QM/MM calculations). In this and previous work, we have used the Mulliken charges to approximate the potential defined in eq 9. Recently, Raghavachari and co-workers developed a QM/QM electronic embedding approach using Mulliken popularion charges within the ONIOM framework.[49,50]

The Mulliken population charges for atoms in fragment A are given as follows:

$$q_a^A(\mathbf{R}_a^A) = Z_a^A - \sum_{\mu \in a} (\mathbf{PS})_{\mu\mu} \qquad (12)$$

where $\mathbf{P}$ and $\mathbf{S}$ are standard density and overlap matrices. To best reproduce the ESP in eq 9, we introduce a parameter,[1,2] for neutral fragments only, associated with the second approximation in the X-Pol method to scale the Mulliken charges such that it minimizes the difference in ESP between the electronic structure theory and the monopole representation:

$$\min\left\{ \left| V_{qm}^A(\mathbf{r}) - \lambda \sum_a \frac{q_a^A}{|\mathbf{r} - \mathbf{R}_a^A|} \right| \right\} \qquad (13)$$

This is akin to the ESP-fitting procedure, although there is only one adjustable variable for a given theoretical level and basis set for all neutral molecules here. An alternative interpretation of the parameter λ is that it is optimized to yield the best agreement with experiment for the interaction between two fragments. A more sophisticated scaling scheme may be important when diffuse functions are used to specifically account for charge penetration effects, and this issue has been thoroughly examine in other context.[51,52]

239

X-Pol Potential

J. Phys. Chem. A, Vol. 113, No. 43, 2009 **11659**

Having defined the partial atomic charges, we write the electrostatic potential at fragment *A* as

$$V^A(\mathbf{r}) = \sum_{B \neq A} \lambda \left[\sum_b \frac{q_b^B}{|\mathbf{r} - \mathbf{R}_b^B|} \right] \quad (14)$$

Note that $V_{qm}^A(\mathbf{r})$ in eq 13 is the ESP of molecule *A*, whereas $V^A(\mathbf{r})$ is the ESP on *A* due to other fragments. It is interesting to point out that inclusion of eq 14 in the Fock matrix to optimize the wave function and determine the energy of the system is identical to carrying out a total of *N* combined QM/MM calculations,[53] one for each fragment.

In the present study, we found that it is not necessary to scale the Mulliken population charge at the ab initio Hartree–Fock and DFT level of theory using the 6-31G(d) basis set. It is known that HF/6-31G(d) tends to yield slightly overpolarized atomic charges, almost perfectly mimicking the average polarization effects in aqueous solution.[42] Thus, it is not surprising that for the present set of simple bimolecular complexes, a value of unity is adequate for λ. However, in general, especially when semiempirical quantum models are used, λ ought to be optimized to best reproduce the target experimental data.

2.3. Level-3 (L3) Approximation and Parameterization of the X-Pol Force Field. At this point the X-Pol potential is a fully quantum chemical model such that each molecule or a group of molecules (fragments) is explicitly represented and treated by electronic structure theory, and its interactions with the rest of the system consist of Coulomb (Eint *A*) and exchange-repulsion and dispersion terms ($E_{int,ed}^A$). Clearly, it would be ideal to use the most accurate electronic structure method along with a large basis set to describe each QM fragment; unfortunately, this would not be feasible, nor is it practical for molecular dynamics simulations of biomolecular systems. Thus, to develop the X-Pol potential into a force field, one must consider two important factors: (1) computational efficiency and (2) the capability for empirical parametrization.[1,2] The first factor is obvious, but the latter may not be so obvious, which may even be counterintuitive because an ab initio theory, albeit with the approximate treatment of interfragment interactions, is brought down to a "semiempirical" level (to be parametrized). Nevertheless, the essence to strive for success of the Lifson-style force fields in biomolecular simulation is the possibility of careful parametrization of potential energy functions against experimental data.[7,15,17,18]

To this end, we have used the formalisms based on the neglect diatomic differential overlap (NDDO)[54] approximation along with the Dewar–Thiel multipole treatment of two-electron integrals for force field developments.[44] The present study, however, focuses on ab initio WFT and DFT, which will eventually be combined with the NDDO X-Pol force field[2,3,6] in a multilayer representation. Thus, the L3 approximation (for the formalism and parametrization of the Coulomb integrals between fragments) in the X-Pol method only concerns the use of Mulliken charges and the $E_{int,ed}^{AB}$ terms in this study.

For the exchange-repulsion and dispersion interaction between different fragments, although it is desirable to employ a general approach that treat the dependence of the $E_{int,ed}^{AB}[\rho^A, \rho^B, \{\zeta^{AB}\}]$ energy on fragment densities explicitly,[55] which will be considered in future work, to proceed, we use the Lennard-Jones potential to parametrically model these effects.[1,3]

$$E_{int,ed}^{AB} \approx \sum_a^A \sum_b^B 4\varepsilon_{ab}^{AB} \left[\left(\frac{\sigma_{ab}^{AB}}{R_{ab}} \right)^{12} - \left(\frac{\sigma_{ab}^{AB}}{R_{ab}} \right)^6 \right] \quad (15)$$

where the parameters $\{\zeta^{AB} \equiv \varepsilon_{ab}^{AB}, \sigma_{ab}^{AB}\}$ are obtained by standard combining rules from the corresponding atomic parameters such that $\varepsilon_{ab}^{AB} = (\varepsilon_a^A \varepsilon_b^B)^{1/2}$ and $\sigma_{ab}^{AB} = (\sigma_a^A \sigma_b^B)^{1/2}$.

2.4. Double Self-Consistent-Field (DSCF) and the Variational X-Pol Method. There are two ways of optimizing the wave function in eq 1, depending on the way the Fock matrices are constructed. First, if one considers each fragment as an isolated molecule embedded in the environment of the partial charges of the rest of the system, i.e., treating each fragment by the traditional QM/MM approach,[53] the Fock matrix for each fragment can be written as follows:[1–3]

$$\mathbf{F}_{HF,DSCF}^A = \mathbf{H}^A + \sum_i (2\mathbf{J}_i^A - \mathbf{K}_i^A) + \mathbf{I}^A \quad (16)$$

where $\mathbf{H}^A$ is the one-electron Hamiltonian matrix, $\mathbf{J}_i^A$ and $\mathbf{K}_i^A$ are the Coulomb and exchange integral matrices, and the last term $\mathbf{I}^A$ is the one-electron integral matrix due to the potential given in eq 14. The corresponding KS matrices are given as follows by replacing the exchange integral with the exchange-correlation potential, $v_{xc}^A(\mathbf{r})$:

$$\mathbf{F}_{KS,DSCF}^A = \mathbf{H}^A + \sum_i 2\mathbf{J}_i^A + \sum_i \mathbf{V}_i^A + \mathbf{I}^A \quad (17)$$

where

$$\mathbf{V}_i^A = \int v_{xc}^A(\mathbf{r}) |\psi_i^A(\mathbf{r})|^2 \, d\mathbf{r} \quad (18)$$

The total electronic energy of the system can be determined by a double self-consistent-field (DSCF) procedure.[1–3,5] Starting with an initial guess of the one-electron density matrix for each fragment, one loops over all fragments in the system and performs SCF optimization of the orbitals $\{\psi_i^A; i = 1,..., m^A\}$ for each fragment in the presence of the Mulliken charges of all other fragments. This is repeated until the change in total electronic energy or in electron density satisfies a predefined tolerance. The DSCF optimization procedure is straightforward and was the approach proposed in ref 1 and adopted in several subsequently implementations.[29,31] However, a major short coming of the DSCF approach is that the Fock matrix in eq 16 was not derived variationally with respect to a perturbation of the charge density on the total energy (eq 4). Although the total energy obtained by using the DSCF method has negligible deviation from the true minimum energy of the system, it imposes severe difficulty to obtain analytical gradients and one has to make use of the coupled-perturbed Hartree–Fock method to determine forces iteratively.

Alternatively, a set of variational equations for the X-Pol potential has been derived[5] in a way similar to that used by Roothaan in deriving the Hartree–Fock equations. The Fock matrix for the variational X-Pol energy has two additional terms compared with eq 16, resulting from the variation of Mulliken population charges (eq 12):

$$\mathbf{F}_{HF,Var}^A = \mathbf{H}^A + \sum_i (2\mathbf{J}_i^A - \mathbf{K}_i^A) + \frac{1}{2}\mathbf{I}^A -$$

$$\frac{1}{2}\sum_{B\neq A}(\sum_{pq}P_{pq}^B J_{pq}^{0,B} + \sum_{b\in B}L_b^{0,B})\sum_{a\in A}^{on\ a\ in\ A}\sum_\mu\sum_\eta S_{\mu\eta}^A \mathbf{T}^{(\eta,\mu)} \quad (19)$$

where the two terms in parentheses specify the interactions between electron *one* in fragment *A* and the charge densities and nuclear charges of fragment *B*, and the superscript "0" denotes that the matrix element is calculated by setting the charges on atom *a* in fragment *A* to +e (note that the Fock matrix will be multiplied by the density matrix in energy calculation).[5] $\mathbf{T}^{(\eta,\mu)}$ represents a matrix of delta functions for convenience in writing eq 19, and it has been fully described in ref 5. It is matrix element is defined by

$$T_{pq}^{(\eta,\mu)} = \delta_{p\eta}\delta_{\mu q} \quad (20)$$

where $\delta_{p\eta}$ and $\delta_{\mu q}$ are Kronecker deltas. It is interesting to notice that each fragment is fully polarized by the rest of the system, but half of the polarization comes from the Mulliken charges specified by the one-electron integral matrix $\mathbf{I}^A$, and the other half originates from the explicit charge density and nuclear charges in all other fragments.

The corresponding variational KS matrix for the X-Pol potential is

$$\mathbf{F}_{DFT,Var}^A = \mathbf{H}^A + \sum_i 2\mathbf{J}_i^A + \sum_i \mathbf{V}_i^A + \frac{1}{2}\mathbf{I}^A -$$

$$\frac{1}{2}\sum_{B\neq A}(\sum_{pq}P_{pq}^B I_{pq}^{0,B} + \sum_{b\in B}L_b^{0,B})\sum_{a\in A}^{on\ a\ in\ A}\sum_\mu\sum_\eta \lambda S_{\mu\eta}^A \mathbf{T}^{(\eta,\mu)} \quad (21)$$

Obviously, the DSCF optimization procedure can also be applied to eqs 19 and 21 for ab initio HF theory and DFT in the X-Pol potential. Nevertheless, we use the subscript DSCF (eqs 16 and 17) and Var (eqs 19 and 21) to indicate that the first approach is nonvariational and the latter is variational. In practice, it is more efficient to optimize all orbital coefficients simultaneously at each system-SCF iteration. There are also other uses of frozen or constrained fragmental electron density in DFT calculations; here we present a variational approach in which analytical gradient consistent with the potential energy can be obtained for dynamics simulations.

3. Computational Details

The computational procedure of the present X-Pol potential using ab initio HF theory and KS-DFT is identical to the algorithm described previously using a semiempirical NDDO method.[1−5] To illustrate the performance of the ab initio HF and DFT X-Pol method, we optimized hydrogen bonding interactions for a series of bimolecular complexes between water and small organic compounds at the HF and B3LYP levels of theory using the 6-31G(d) basis sets. The computed hydrogen bond geometries and interaction energies are compared with those obtained at the same level of theory by treating the entire system uniformly, and with results obtained using coupled cluster theory at CCSD(T)/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ. We decided not to optimize the geometries at the CCSD(T) level because DFT calculations generally yield excellent geometrical parameters and we do not expect noticeable differences would result from WFT optimizations. The charge scaling parameter

TABLE 1: Lennard-Jones Parameters Used in the X-Pol Potential with B3LYP and the 6-31G(d) Basis Set

atom	σ (Å)	ε (kcal/mol)
H	1.30	0.05
C	3.65	0.15
N	3.45	0.20
O (sp³)	3.35	0.15
O (sp²)	3.10	0.15
Cl⁻	4.25	0.21
Na⁺	2.35	0.30

λ in eq 14 was set to unity, i.e., without any modification of the Mulliken charges in the present ab initio HF and DFT treatment. The initial Lennard-Jones parameters were taken from those optimized for combined QM/MM calculations using the ab initio HF/3-21G method for QM and the OPLS potential[56] for MM,[57] and were slightly adjusted to be adopted in the X-Pol potential. The final parameters are listed in Table 1.

In each bimolecular complex, the monomer geometries are held fixed at the B3LYP optimized structure in X-Pol calculations. Thus, only the hydrogen bond length and angle are optimized. This procedure has been used in the development of the OPLS force field when bimolecular complexes are examined,[56,58] and it has also been used to generate a set of universal parameters in combined QM/MM calculations.[53,57,59]

In the following, we use the short-hand notation XP to specify calculations carried out using the X-Pol potential, followed by "@" to indicate the specific method with which X-Pol calculations are made. Thus, the notation XP@HF/6-31G(d)//B3LYP/6-31G(d) specifies an X-Pol calculation at the HF/6-31G(d) level of theory using the geometry optimized with B3LYP/6-31G(d). All calculations were performed using a locally modified version of the GAMESS package.[60]

4. Results and Discussion

The main goal of this article is to show that the X-Pol method can be conveniently developed to yield excellent results on intermolecular interactions at the ab initio HF and DFT levels of theory with a modest basis set in comparison with experimental and high level ab initio results. We first briefly describe the parametrization philosophy of the X-Pol potential, and present the potential energy profile for a water dimer complex. Then, we consider the results for a set of 14 bimolecular complexes of simple organic molecules with water.

4.1. Parametrization of Repulsive and Dispersive Interactions between Fragments. The approximation to write the total molecular wave function as a Hartree product of the antisymmetric wave functions of individual molecules tremendously reduces computational costs for large systems, which also enables the X-Pol potential to be conveniently parallelized in practical implementations. However, this is at the expense of neglecting short-range exchange repulsions and long-range dispersion attractions between different fragments. Thus, to retain the computational accuracy of the entire system that is treated as one fragment, one must introduce a formalism to remedy this difference. To this end, we have decided to use a purely empirical approach, making use of the Lennard-Jones potential to parametrically model the repulsive and dispersive interactions. Furthermore, it provides an opportunity to parametrize the X-Pol potential to yield results for intermolecular interactions in better agreement with experiment with the use of a modest level of electronic structure method.

Table 1 lists a set of Lennard-Jones parameters used in the present calculation, which were adjusted for the XP@B3LYP/
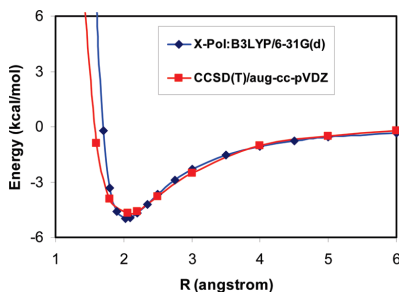
Figure 1. Interaction energy profile of water dimer as a function of the hydrogen bond distance from CCSD(T)/aug-cc-pVDZ (red) and XP@B3LYP/6-31G(d) optimizations.

6-31G(d) model, starting with those optimized in combined QM/MM calculations at the HF/3-21G level,[57] to reproduce the results for 14 bimolecular complexes obtained from CCSD(T)/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ. There is no attempt made here in choosing a training set and a test set of complexes for validation, but a thorough comparison will be made in a later publication. As it turns out, the X-Pol parameters are very similar to those of the ab initio QM/MM values. These parameters were employed in the XP@HF/6-31G(d) calculations without further alteration, although in principle a different set of parameters may be needed for different methods and basis functions. The aim of the present study is not to provide a set of optimized parameters for the X-Pol potential; we aim to illustrate that a set of atomistic parameters can yield excellent results for bimolecular interactions using the X-Pol potential, in better agreement with high level ab initio data than the full B3LYP/6-31G(d) and HF/6-31G(d) results.

4.2. Potential Profile for the Water Dimer. In Figure 1, we first examine the potential energy profile of the water dimer as a function of the hydrogen-bond distance using XP@B3LYP/6-31G(d) compared with that fully optimized at the CCSD(T)/aug-cc-pVDZ level. The computed interaction energy is 5.0 kcal/mol from the X-Pol potential, which is in good agreement with a value of 4.8 kcal/mol using CCSD(T). The corresponding hydrogen bond distances are, respectively, 1.98 and 1.96 Å, also in excellent accord. At short distance, the repulsive interactions increase more quickly than that of the CCSD(T) curve, indicating that a softer repulsive potential than that of the Lennard-Jones potential may be more appropriate. The attractive part of the potential profile is in excellent agreement between the XP@B3LYP and CCSD(T) calculations. Notice that the full B3LYP/6-31G(d) calculations overestimate the binding energy for the water dimer, whereas increasing the size of the basis function to aug-cc-pVDZ improves its agreement with high-level ab initio results (Table 2). The good performance of the X-Pol potential for bimolecular interactions is due to its capability to empirically adjust the repulsive and dispersive interactions. In a recent study, Fujimoto and Yang described a frozen-fragment interaction calculation,[61] in which the KS orbitals are polarized by the charge densities of other fragments similar to the method described here.[1−3] However, using B3LYP/6-31G(d), these authors reported an error of −33.1 kcal/mol for the water dimer over the full QM result at the same level of theory, which overestimates the interaction energy only by 2 kcal/mol than that from CCSD(T)/aug-cc-pVDZ. Thus, the net error from that implementation is more than 35 kcal/mol for $(H_2O)_2$, emphasizing the significance of an adequate treatment of the repulsive and attractive interactions.
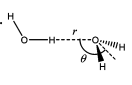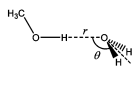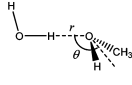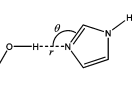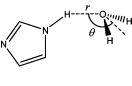
4.3. Bimolecular Complexes. We considered nine bimolecular complexes of water with water, methanol, formamide, and immidazole (Table 2), and five ion−water complexes from chloride, sodium, and acetate ions (Table 3) to illustrate the feasibility that the X-Pol potential implemented with ab initio HF and DFT approaches can yield adequate description of intermolecular interactions. Furthermore, we do not aim to simply reproduce the interaction energies of the corresponding HF/6-31G(d) or B3LYP/6-31G(d) data, the same levels of theory used in the X-Pol calculations, because the latter in fact has significantly large errors in comparison with CCSD(T) results. When larger basis functions are used, the B3LYP results improve considerably, but the use of a very large basis set is not desirable in the present X-Pol model for treating large systems including protein and nucleic acids. On the other hand, with an adjustment of the van der Waals parameters (Table 1), we found that it is possible to obtain good agreement between X-Pol and CCSD(T) calculations with a reasonable, but with the use of a modest basis set in X-Pol calculations.

The CCSD(T) interaction energy fully optimized using the aug-cc-pVDZ basis set for the water dimer is 4.9 kcal/mol (Figure 1), which is in excellent agreement with the experimental value of 4.9 ± 0.9 kcal/mol determined from pressure broadening of near-IR spectra.[62,63] The results listed in Table 2 were obtained by partial geometry optimizations with the monomer geometry fixed,[56−59] and the corresponding CCSD(T) interaction energies were computed by single point calculations using the B3LYP/aug-cc-pVDZ geometries. Thus, the dimer interaction energy for water is slightly greater than that from full CCSD(T) optimizations. The computed interaction energies are 5.0 kcal/mol from both XP@HF/6-31G(d) and XP@B3LYP/6-31G(d) optimizations. For the methanol−water bimolecular complex, there are two possible hydrogen bonding interactions, depending on water or methanol as the hydrogen bond donor. The 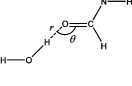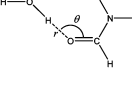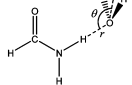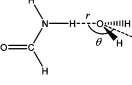XP@B3LYP model yields interaction energies of 5.2 and 4.8 kcal/mol, in favor of the complex in which water acts as the hydrogen bond donor. The same trend is reproduced for the full QM system at the CCSD(T) and B3LYP/aug-cc-pVDZ levels, whereas B3LYP/6-31G(d) optimizations yielded the opposite trend. An early study of the methanol−water system by Krischner and Woods yielded interaction energies of −5.7 and −4.9 kcal/mol using MP2/aug-cc-pVQZ//MP2/aug-cc-pVTZ,[64] in accord with the present X-Pol results.

For the immidazole−water complexes, we found that the immizaole ring is a better hydrogen bond acceptor from water than donating a hydrogen bond using the X-Pol potential. This is in agreement with high-level ab initio results in Table 2, although the difference is greater using CCSD(T) calculations. Again, B3LYP/6-31G(d) yields the opposite trend, suggesting that a much larger basis set than 6-31G(d) is needed in studies of enzymes. Four structures were considered in the formamide−water complex. In accord with experiment,[65] the structure with a water simultaneously donating a hydrogen bond to the carbonyl group and accepting one from the amide group is the global minimum. The computed XP@B3LYP/6-31G(d) interaction energy of −7.6 kcal/mol is in good accord with the CCSD(T) results. Overall, the carbonyl group forms strong hydrogen bonds with water than the amide unit donating hydrogen bonds to water.[66]

The interaction energies for chloride ion and sodium ion with water can be fitted exactly to the CCSD(T) or experimental results, the latter of which are −14.8 and +24 kcal/mol experimentally.[67,68] The ab initio results are slightly smaller, noting that the experimental values are enthalpies of interaction.

TABLE 2: Computed Geometries and Interaction Energies for Molecule−Water Bimolecular Complexes

Complex	Method	r (Å)	θ (degrees)	ΔE (kcal/mol)
	HF/6-31G(d)	2.03	129.3	-5.6
	B3LYP/6-31G(d)	1.93	109.1	-7.5
	B3LYP/aug-cc-pVDZ	1.95	129.8	-4.6
	CCSD(T)/aug-cc-pVDZ			-5.2
	XP@HF/6-31G(d)	1.99	145.0	-5.0
	XP@B3LYP/6-31G(d)	2.03	144.0	-5.0
	HF/6-31G(d)	2.02	132.1	-5.5
	B3LYP/6-31G(d)	1.91	112.2	-7.5
	B3LYP/aug-cc-pVDZ	1.96	130.0	-4.4
	CCSD(T)/aug-cc-pVDZ			-5.4
	XP@HF/6-31G(d)	2.01	123.0	-4.3
	XP@B3LYP/6-31G(d)	2.05	150.5	-4.8
	HF/6-31G(d)	2.02	130.4	-5.5
	B3LYP/6-31G(d)	1.91	118.9	-7.0
	B3LYP/aug-cc-pVDZ	1.92	134.9	-5.0
	CCSD(T)/aug-cc-pVDZ			-5.9
	XP@HF/6-31G(d)	1.98	130.0	-5.4
	XP@B3LYP/6-31G(d)	2.03	135.5	-5.2
	HF/6-31G(d)	2.12	115.8	-6.4
	B3LYP/6-31G(d)	2.01	117.9	-7.1
	B3LYP/aug-cc-pVDZ	1.96	120.1	-6.6
	CCSD(T)/aug-cc-pVDZ			-7.5
	XP@HF/6-31G(d)	1.97	117.0	-6.6
	XP@B3LYP/6-31G(d)	2.02	117.0	-6.9
	HF/6-31G(d)	2.04	202.7	-6.3
	B3LYP/6-31G(d)	1.93	132.9	-7.9
	B3LYP/aug-cc-pVDZ	1.99	170.5	-5.4
	CCSD(T)/aug-cc-pVDZ			-6.6
	XP@HF/6-31G(d)	1.97	180.0	-5.6
	XP@B3LYP/6-31G(d)	2.00	178.5	-6.6
	HF/6-31G(d)	2.02	116.8	-6.2
	B3LYP/6-31G(d)	1.93	112.1	-7.3
	B3LYP/aug-cc-pVDZ	1.91	116.2	-5.8
	CCSD(T)/aug-cc-pVDZ			-6.4
	XP@HF/6-31G(d)	1.85	125.0	-6.6
	XP@B3LYP/6-31G(d)	1.92	123.5	-6.3
	HF/6-31G(d)	1.98	111.8	-7.2
	B3LYP/6-31G(d)	1.88	104.8	-9.2
	B3LYP/aug-cc-pVDZ	1.86	110.4	-6.7
	CCSD(T)/aug-cc-pVDZ			-7.6
	XP@HF/6-31G(d)	1.84	110.0	-8.2
	XP@B3LYP/6-31G(d)	1.89	113.0	-7.6
	HF/6-31G(d)	2.09	111.9	-5.9
	B3LYP/6-31G(d)	1.97	103.1	-8.2
	B3LYP/aug-cc-pVDZ	2.04	115.0	-5.0
	CCSD(T)/aug-cc-pVDZ			-6.0
	XP@HF/6-31G(d)	1.99	146.0	-4.9
	XP@B3LYP/6-31G(d)	2.02	127.5	-5.7
	HF/6-31G(d)	2.09	181.4	-5.6
	B3LYP/6-31G(d)	1.97	130.2	-7.1
	B3LYP/aug-cc-pVDZ	2.04	167.7	-4.7
	CCSD(T)/aug-cc-pVDZ			-5.7
	XP@HF/6-31G(d)	2.00	183.0	-5.4
	XP@B3LYP/6-31G(d)	2.03	176.5	-5.9

Using the parameters in Table 1, the X-Pol interaction energies for these two ions with water are, respectively, −14.1 and −22.9 kcal/mol, in reasonable accord with the CCSD(T) results (Table 3). The most stable complex between acetate ion and water is the bidentate structure in which both hydrogen atoms of water donate hydrogen bonds to the two oxygen atoms of acetate ion. The interaction energy from CCSD(T) calculation is −18.3 kcal/mol, slightly stronger than the single-hydrogen bonded complex in the syn orientation (−18.0 kcal/mol) (Table 3).[58,67,69] The corresponding X-Pol results are −18.3 and −18.2 kcal/mol using the XP@B3LYP/6-31G(d) and are −18.9 and −18.4 from XP@HF/6-31G(d). The hydrogen-bond complex from the anti orientation is the least stable, with computed interaction energies of −16.2, −15.9, and −16.4 kcal/mol from CCSD(T),

TABLE 3: Computed Geometries and Interaction Energies for Ion−Water Complexes

Complex	Method	r (Å)	θ (degrees)	ΔE (kcal/mol)
	HF/6-31G(d)	3.26		-18.4
	B3LYP/6-31G(d)	3.12		-24.2
	B3LYP/aug-cc-pVDZ	3.16		-17.1
	CCSD(T)/aug-cc-pVDZ			-18.3
	XP@HF/6-31G(d)	3.15		-18.9
	XP@B3LYP/6-31G(d)	3.19		-18.3
	HF/6-31G(d)	1.88	102.3	-19.4
	B3LYP/6-31G(d)	1.80	101.3	-22.8
	B3LYP/aug-cc-pVDZ	1.76	110.0	-17.0
	CCSD(T)/aug-cc-pVDZ			-18.0
	XP@HF/6-31G(d)	1.78	106.0	-18.4
	XP@B3LYP/6-31G(d)	1.82	104.0	-18.2
	HF/6-31G(d)	1.82	130.0	-15.1
	B3LYP/6-31G(d)	1.74	123.0	-19.1
	B3LYP/aug-cc-pVDZ	1.71	127.9	-15.3
	CCSD(T)/aug-cc-pVDZ			-16.2
	XP@HF/6-31G(d)	1.73	134.0	-16.4
	XP@B3LYP/6-31G(d)	1.76	136.5	-15.9
	HF/6-31G(d)	2.21	179.9	-28.6
	B3LYP/6-31G(d)	1.19	180.0	-31.0
	B3LYP/aug-cc-pVDZ	2.21	179.9	-23.9
	CCSD(T)/aug-cc-pVDZ			-22.0
	XP@HF/6-31G(d)	2.38	180.0	-24.3
	XP@B3LYP/6-31G(d)	2.40	174.0	-22.9
	HF/6-31G(d)	2.38	157.8	-13.9
	B3LYP/6-31G(d)	2.23	161.7	-16.7
	B3LYP/aug-cc-pVDZ	2.19	168.8	-13.8
	CCSD(T)/aug-cc-pVDZ			-14.1
	XP@HF/6-31G(d)	2.32	160.5	-14.7
	XP@B3LYP/6-31G(d)	2.33	164.0	-14.1

XP@B3LYP, and XP@HF calculations. The experimental enthalpy of binding is about −16 kcal/mol and argues against the bidentate complex in the gas phase.[67]

The mean unsigned error from CCSD(T) results for the fourteen complexes considered is 0.4 kcal/mol for the XP@B3LYP/6-31G(d) method (Figure 2), whereas it is 0.9 kcal/mol from the XP@HF/6-31G(d) potential, a reflection that the Lennard-Jones parameters were not optimized for HF calculations in Table 1. In both cases, we have used the monomer structures optimized at the B3LYP/6-31G(d) level. Hydrogen bond distances from optimizations using XP@B3LYP/6-31G(d) are in good accord with those obtained from B3LYP/aug-cc-

pVDZ calculations. The mean unsigned error for the fourteen complexes considered is 0.08 Å. This is comparable to ab initio QM/MM calculations for similar bimolecular complexes.[57] The good agreement suggests that the X-Pol potential can be trained by introducing a set of atomistic parameters, associated with a given theory and basis function, to yield adequate results for the description of intermolecular interactions that are approximated by a Hartree product wave function.

5. Concluding Remarks

The explicit polarization (X-Pol) method was developed as a framework to develop next-generation force fields for biomolecular simulations based on electronic structure theory, and it has been demonstrated to be feasible for extended molecular dynamics simulation of a solvated protein. Importantly the X-Pol method is a general theory, which can be employed with any level of electronic structure theory. In principle, it is possible and desirable to treat the central region of interest by a high-level ab initio method or density functional theory and the remainder of the system is represented by an NDDO-type X-Pol force field. The present study examines the X-Pol approach using ab initio molecular orbital theory at the Hartree−Fock level and density functional theory using a hybrid functional. The computational results are illustrated by considering a set of bimolecular complexes of small organic molecules and ions with water. The computed interaction energies and hydrogen bond geometries are in good accord with CCSD(T) results and B3LYP/aug-cc-pVDZ optimizations.

The X-Pol potential complements the effective fragment potential (EFP) approach developed by Gordon and co-workers
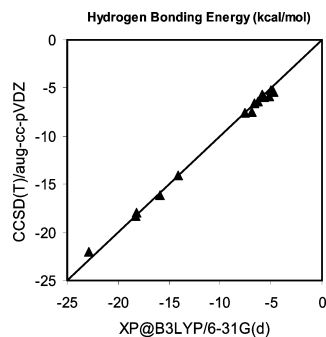

Figure 2. Comparison of the computed interaction distances for bimolecular complexes optimized using B3LYP/aug-cc-pVDZ and XP@B3LYP/6-31G(d) methods.

using ab initio MO theory and DFT. The main difference is that the electronic structure for each individual fragment is fully optimized under the electric field of the rest of the system for a given instantaneous geometrical configuration in the X-Pol, whereas the EFP is determined from a set of distributed multipole expansions to account for static and polarization interactions as well as charge penetration and transfer. The significance of and methods developed to treat charge penetration in the EFP model can be adopted for X-Pol calculations. On the other hand, the X-Pol potential can be designed as a force field parametrized to reproduce experimental liquid properties, and a combination of such an X-Pol force field and ab initio X-Pol treatment represents a general multilayer quantum mechanical approach for biomolecular systems.

References and Notes

(1) Gao, J. *J. Phys. Chem. B* **1997**, *101*, 657–663.
(2) Gao, J. *J. Chem. Phys.* **1998**, *109*, 2346–2354.
(3) Xie, W.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1890–1900.
(4) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Phys. Chem. B* **2008**, *112*, 14124–14131.
(5) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2008**, *128*, 234108/234101–234109.
(6) Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 459–467.
(7) Bixon, M.; Lifson, S. *Tetrahedron* **1967**, *23*, 769–784.
(8) Levitt, M. *Nat. Struct. Biol.* **2001**, *8*, 392–393.
(9) Hill, T. L. *J. Chem. Phys.* **1946**, *14*, 465.
(10) Westheimer, F. H.; Mayer, J. E. *J. Chem. Phys.* **1946**, *14*, 733–738.
(11) Hendrickson, J. B. *J. Am. Chem. Soc.* **1961**, *83*, 4537–4547.
(12) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, D.C., 1982.
(13) Levitt, M.; Lifson, S. *J. Mol. Biol.* **1969**, *46*, 269–279.
(14) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature (London)* **1977**, *267*, 585–590.
(15) For a special issue on the current development of polarizable force field, see Jorgensen, W. L. *J. Chem. Theory Comput.* **2007**, *3*, 1877.
(16) Vesely, F. J. *J. Comput. Phys.* **1977**, *24*, 361–371.
(17) Dinur, U.; Hagler, A. T. *Rev. Comput. Chem.* **1991**, *2*, 99–164.
(18) MacKerell, A. D., Jr. *Ann. Rep. Comput. Chem.* **2005**, *1*, 91–102.
(19) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
(20) Wierzchowski, S. J.; Kofke, D. A.; Gao, J. *J. Chem. Phys.* **2003**, *119*, 7365–7371.
(21) Yang, W.; Lee, T.-S. *J. Chem. Phys.* **1995**, *103*, 5674–5678.
(22) Dixon, S. L.; Merz, K. M., Jr. *J. Chem. Phys.* **1997**, *107*, 879–893.
(23) Van Der Vaart, A.; Gogonea, V.; Dixon, S. L.; Merz, K. M., Jr. *J. Comput. Chem.* **2000**, *21*, 1494–1504.
(24) Raha, K.; Merz, K. M., Jr. *J. Med. Chem.* **2005**, *48*, 4558–4575.
(25) Stewart, J. J. P. *Int. J. Quantum Chem.* **1996**, *58*, 133–146.
(26) Gao, J. *J. Comput. Chem.* **1997**, *18*, 1062–1971.
(27) Thompson, M. A. *J. Phys. Chem.* **1996**, *100*, 14492–14507.
(28) Kitaura, K.; Sawai, T.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *312*, 319–324.
(29) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701–706.
(30) Fukuzawa, K.; Mochizuki, Y.; Tanaka, S.; Kitaura, K.; Nakano, T. *J. Phys. Chem. B* **2006**, *110*, 16102–16110.
(31) Gascon, J. A.; Leung, S. S. F.; Batista, E. R.; Batista, V. S. *J. Chem. Theory Comput.* **2006**, *2*, 175–186.
(32) Field, M. J. *Mol. Phys.* **1997**, *91*, 835–845.
(33) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968–1986.
(34) Gordon, M. S.; Slipchenko, L.; Li, H.; Jensen, J. H. *Ann. Rep. Comput. Chem.* **2007**, *3*, 177–193.
(35) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714–4721.
(36) Amara, P.; Field, M. J.; Alhambra, C.; Gao, J. *Theor. Chem. Acc.* **2000**, *104*, 336–343.
(37) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 5454–5463.
(38) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632–650.
(39) Pu, J.; Gao, J.; Truhlar, D. G. *ChemPhysChem* **2005**, *6*, 1853–1865.
(40) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974–12980.
(41) Momany, F. A. *J. Phys. Cem.* **1978**, *82*, 592–601.
(42) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
(43) Alhambra, C.; Luque, E. J.; Orozco, M. *J. Comput. Chem.* **1994**, *15*, 12–22.
(44) Dewar, M. J. S.; Thiel, W. *Theoret. Chim. Acta* **1977**, *46*, 89–104.
(45) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233–239.
(46) Mulliken, R. S. *J. Chem. Phys.* **1964**, *61*, 20.
(47) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 2178–2182.
(48) Gao, J.; Luque, F. J.; Orozco, M. *J. Chem. Phys.* **1993**, *98*, 2975–2982.
(49) Hratchian, H. P.; Parandekar, P. V.; Raghavachari, K.; Frisch, M. J.; Vreven, T. *J. Chem. Phys.* 2008, *128*, 034107.
(50) Parandekar, P. V.; Hratchian, H. P.; Raghavachari, K. *J. Chem. Phys.* **2008**, *129*, 145101.
(51) Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. *J. Chem. Phys.* **2000**, *112*, 7300–7306.
(52) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, U.K., 1996.
(53) Gao, J.; Xia, X. *Science* **1992**, *258*, 631–635.
(54) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129–S135.
(55) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293–307.
(56) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
(57) Freindorf, M.; Gao, J. *J. Comput. Chem.* **1996**, *17*, 386–395.
(58) Jorgensen, W. L.; Gao, J. *J. Phys. Chem.* **1986**, *90*, 2174–2182.
(59) Gao, J. *ACS Symp. Ser.* **1994**, *569*, 8–21.
(60) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; et al. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
(61) Fujimoto, K.; Yang, W. *J. Chem. Phys.* **2008**, *129*, 054102.
(62) Fiadzomor, P. A. Y.; Keen, A. M.; Grant, R. B.; Orr-Ewing, A. J. *Chem. Phys. Lett.* **2008**, *462*, 188–191.
(63) Nakayama, T.; Fukuda, H.; Kamikawa, T.; Sakamoto, Y.; Sugita, A.; Kawasaki, M.; Amano, T.; Sato, H.; Sakaki, S.; Morino, I.; Inoue, G. *J. Chem. Phys.* **2007**, *127*, 134302.
(64) Kirschner, K. N.; Woods, R. J. *J. Phys. Chem. A* **2001**, *105*, 4150–4155.
(65) Lovas, F. J.; Suenram, R. D.; Fraser, G. T.; Gillies, C. W.; Zozom, J. *J. Chem. Phys.* **1988**, *88*, 722–729.
(66) Bende, A.; Suhai, S. *Int. J. Quantum Chem.* **2005**, *103*, 841–853.
(67) Meot-Ner, M. *J. Am. Chem. Soc.* **1988**, *110*, 3854–3858.
(68) Kebarle, P. *Annu. Rev. Phys. Chem.* **1977**, *28*, 445–476.
(69) Gao, J.; Garner, D. S.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1986**, *108*, 4784–4790.

JP902710A

CHAPTER 5

Kinetic Isotope Effects from Hybrid Classical and Quantum Path Integral Computations

JIALI GAO, KIN-YIU WONG, DAN T. MAJOR, ALESSANDRO CEMBRAN, LINGCHUN SONG, YEN-LIN LIN, YAO FAN AND SHUHUA MA

Department of Chemistry, Digital Technology Center and Supercomputer Institute, University of Minnesota, 207 Pleasant Street S.E., Minneapolis, MN 55455-0431, USA

5.1   Introduction

Proton, hydride and hydrogen-atom transfer reactions are ubiquitous in biological processes,[1] and because of their relatively small mass, zero-point energy and quantum tunnelling are significant in determining free-energy reaction barriers.[2,3] The incorporation of nuclear quantum effects (NQE) is also important for reactions involving heavy atoms since one of the most direct experimental assessment of the transition state and the mechanism of a chemical reaction is by measurements of kinetic isotope effects (KIE),[1] which are of quantum-mechanical origin. This is illustrated by the work of Schramm and coworkers,[4] who developed highly potent inhibitors to the enzyme purine nucleoside phosphorylase (PNP) based on the transition-state structure derived from measured KIEs. In principle, Schramm's approach can be applied to other enzymes, but in practice it is often limited by the lack of an adequate model to match computed and experimental KIEs. Therefore, it is of great

interest to develop practical computation methods to estimate KIEs for enzymatic reactions.

The challenge to theory is the difficulty to accurately determine the small difference in free energy of activation due to isotope replacements. This is further exacerbated by the complexity and size of an enzyme system that requires statistical averaging. The computational accuracy demands both an adequate treatment of the potential-energy surface that can properly describe the change of the bond orders of the primary and secondary sites and an efficient sampling procedure that can yield converged properties.[5] In this chapter, we describe an integrated path integral free-energy perturbation and umbrella sampling (PI-FEP/UM) approach in molecular-dynamics simulations, and an analytic, integration-free approach based on Kleinert's variational perturbation theory[6] that can systematically improve the accuracy and avoid numerical convergence problems,[7] to compute KIEs for chemical reactions.[8–10] In addition, we present a mixed molecular-orbital and valence-bond (MOVB) theory[11–14] that combines quantum mechanics with molecular mechanics (QM/MM)[15,16] to represent the potential surface of the reactive system. We show that this coupling of a combined QM-MOVB/MM potential-energy surface to describe the electronic structure with path integral–free energy simulation to model the nuclear quantum effects can yield remarkably accurate KIEs for reactions in solution and in enzymes.

Of course, a variety of methods have been developed to treat NQE for gas-phase reactions (see a recent review[17]). In principle, these techniques can be directly extended to condensed-phase systems; however, the size and complexity of these systems make it intractable computationally. Thus, a main goal is to develop new methods, or to extend gas-phase techniques to condensed phases or biomolecular systems. One method that has been successfully introduced to computational enzymology is the ensemble-averaged variational transition-state theory with QM/MM sampling (EA-VTST-QM/MM),[18,19] which has been applied to a number of enzyme systems.[2,20,21] Both primary and secondary KIEs can be computed using the EA-VTST-QM/MM method and the method includes contributions of multidimensional tunnelling. In another approach, Hammes-Schiffer and coworkers utilised a grid-based hybrid approach to model NQE in hydrogen-transfer reactions by numerically solving the vibrational wavefunction of the transferring hydrogen nucleus.[22] So far, only primary KIEs have been computed by this approach.

The discrete Feynman path integral method[23,24] has been used in a variety of applications since it offers an efficient and general approach for treating nuclear quantum effects in condensed-phase simulations.[25–31] In principle, centroid path integral simulations can be directly used to determine KIEs by carrying out two separate calculations for the heavy and light isotope, respectively, and indeed, this has been the approach in most applications; however, the convergence of the computed free-energy barrier from dynamics simulations is typically not sufficient to ensure the desired accuracy for KIE, especially when heavy isotopes and secondary effects are involved. To this end, we have developed a free-energy perturbation technique[10] by perturbing the atomic mass from light to heavy

isotopes in a bisection path integral sampling scheme,[8,9] and this has tremendously reduced the statistical uncertainty in the computed KIEs.[32]

The integrated path integral–free energy perturbation and umbrella sampling (PI-FEP/UM) method involves two computational steps.[10] First, classical molecular-dynamics simulation is carried out to obtain the potential of mean force along the reaction coordinate for a given reaction. Then, centroid path integral simulations are performed to determine the nuclear quantum effects. The most significant feature of these studies is that classical and quantum simulations are fully separated, making it particularly attractive and efficient for enzymatic reactions. This computational approach has been explored previously in the work of Sprik *et al.*[25] and in the quantised classical path (QCP) method by Warshel and coworkers.[33–35] The special feature in the PI-FEP/UM method is to use a free-energy perturbation scheme to obtain accurate KIEs for chemical reactions, by changing the atomic mass from one isotope into another in path integral sampling.[10]

Discretised path integral simulations often face numerical convergence difficulties, especially in view of the required accuracy is less than a fraction of a tenth of one kilocalorie per mole for computing KIEs. The analytic integration results based on Kleinert's variational perturbation theory does not have this problem, and the perturbation series has been shown to be convergent exponentially and uniformly,[6,36,37] making the second-order perturbation (KP2) sufficiently accurate for chemical applications.[7] We describe an automated, numerical integration-free centroid path integral method[7] for estimating KIEs for proton-transfer reactions to illustrate the computational power for potential applications to enzymatic reactions.

In the following we first summarise the theoretical background, the representation of the potential-energy surface, and the PI-FEP/UM computational details. Then, results and discussion are presented. The paper is concluded with highlights of main findings.

5.2   Theoretical Background

5.2.1   Path Integral Quantum Transition-State Theory

The theoretical framework in the present discussion is path integral quantum transition-state theory (QTST), which yields an expression of the quantum-mechanical rate constant. In the discrete Feynman path integral method, each quantised nucleus is represented by a ring of P quasiparticles called beads, whose coordinates are denoted as $\mathbf{r} = \mathbf{r}_i$; $i = 1, \cdots, P$.[23] The discrete paths are closed with $\mathbf{r}_{P+1} = \mathbf{r}_1$, in which each particle (bead) is connected harmonically with its neighbours, corresponding to the imaginary time slices $\tau_i = (i-1)\hbar\beta/P$. A key concept in QTST is the centroid variable in path integration,[27,38–42] defined as the geometrical centre of the quasiparticles:

$$\bar{\mathbf{r}}^{(n)} = \frac{1}{P}\sum_{i=1}^{P}\mathbf{r}_i^{(n)} \tag{5.1}$$

where the superscript (n) specifies the nth quantised atom. The discretisation parameter P is chosen to be sufficiently large such that the numerical results converge to the quantum limit. In this approach, the quantum-mechanical equilibrium properties are obtained from the classical averages for a fictitious system governed by the effective potential[23]

$$V_{\text{eff}}[\{r_i^{(n)}\}, \mathbf{S}] = \sum_{n=1}^{N} \frac{\pi P}{\beta \lambda_n^2} \sum_i^P (\mathbf{r}_i^{(n)} - \mathbf{r}_{i+1}^{(n)})^2 + \frac{1}{P} \sum_i^P U(\mathbf{r}_i^{(1)}, \cdots, \mathbf{r}_i^{(N)}, \mathbf{S}) \qquad (5.2)$$

where N is the number of quantised atoms, $\mathbf{S}$ represents the coordinates of all classical particles, $U(\mathbf{r}_i^{(1)}, \cdots, \mathbf{r}_i^{(N)}, \mathbf{S})$ is the potential energy of the system, and $\beta = 1/k_{\text{B}}T$ with k_{B} being Boltzmann's constant and T the temperature. In eqn (5.2), the de Broglie thermal wavelength of atom n with a mass of M_n is given by $\lambda_n = (2\pi\beta\hbar^2/M_n)^{1/2}$. Note that the dynamics generated by the effective potential of eqn (5.2) has no physical significance; it is merely used as a procedure to obtain the correct ensemble of configurations.[23,24]

QTST is derived by writing the rate expression analogous to classical TST, which includes a quantum activation term and a dynamical correction factor,[39,42,43] and the QTST rate constant is given by

$$k_{\text{QTST}} = \frac{1}{2} <|\dot{z}|>_{z^{\neq}} e^{-\beta w(z^{\neq})} \Big/ \int_{-\infty}^{z^{\neq}} \mathrm{d}z\, e^{-\beta w(z)} \qquad (5.3)$$

where $w(z)$ is the potential of mean force (PMF) as a function of the centroid reaction coordinate $z[\bar{\mathbf{r}}]$, $z^{\neq}$ is the value of $z[\bar{\mathbf{r}}]$ at the maximum of the PMF, and $<|\dot{z}|>_{z^{\neq}} = (2/\pi\beta M_{\text{eff}})^{1/2}$ is a dynamical frequency factor approximated by the velocity for a free particle of effective mass M_{eff} along the reaction coordinate $z[\bar{\mathbf{r}}]$ direction. The exact rate constant is obtained by multiplying the QTST rate constant by a correction factor or transmission coefficient γ_q:[43]

$$k = \gamma_q \cdot k_{\text{QTST}} \qquad (5.4)$$

Equations (5.3) and (5.4) have identical forms to that of the classical rate constant, but unlike classical variational transition-state theory, there is no variational upper bound in the QTST rate constant because the quantum transmission coefficient γ_{q} may be greater than or less than unity. Equation (5.3) was initially derived with the assumption of a planar dividing surface along a rectilinear reaction coordinate. Messina *et al.* described a generalisation of the dividing surface that may depend both on the centroid coordinates and on its momenta.[42]

There is no practical procedure to compute the quantum transmission coefficient γ_{q} in eqn (5.4). For a model reaction with a parabolic barrier along the reaction coordinate coupled to a bath of harmonic oscillators, the quantum

transmission coefficient is the Grote–Hynes (GH) classical transmission coefficient κ_{GH}.[39,44] Often, the classical γ_q is used to approximate the quantum transmission coefficient; however, there is no correspondence between classical and quantum dynamic trajectories and the effects of tunnelling may greatly affect reaction dynamics near the barrier top.

As in classical TST, the PMF, $w(z)$, can be computed from the equilibrium average without any dynamical information, and it is defined by

$$e^{-\beta[w(z)-w(z_{\mathrm{R}})]} = e^{-\beta\Delta F(z)} = \frac{\langle \delta(z[\bar{\mathbf{r}}] - z) \rangle}{\langle \delta(z[\bar{\mathbf{r}}] - z_{\mathrm{R}}) \rangle} \tag{5.5}$$

where z_{R} is the minimum point at the reactant state in the PMF and the ensemble average $\langle \cdots \rangle$ is obtained by the effective potential of eqn (5.2). Equation (5.5) also serves as a definition of the path integral centroid free energy, $\Delta F(z)$, at z relative to that at the reactant state minimum. Note that the inherent nature of quantum mechanics is at odds with a potential of mean force as a function of a finite reaction coordinate. Nevertheless, the reaction coordinate function $z[\bar{\mathbf{r}}]$ can be evaluated from the path centroids $\bar{\mathbf{r}}$,[42–43,45] first recognised by Feynman and Hibbs as the most classical-like variable in quantum statistical mechanics.[23] Studies have shown that the activation free energy in the centroid path integral QTST "captures most of the tunnelling and quantisation effects"; which give rise to deviations from classical TST.[46–49]

It is also useful to rewrite eqn (5.3) in combination with eqn (5.5) as follows

$$k_{\mathrm{QTST}} = \frac{1}{\beta h} e^{-\beta\Delta F_{\mathrm{CPI}}^{\neq}} \tag{5.6}$$

where the centroid path integral free energy of activation $\Delta F_{\mathrm{CPI}}^{\neq}$ is defined by

$$\Delta F_{\mathrm{CPI}}^{\neq} = \Delta F(z^{\neq}) - F_{\mathrm{CPI}}^{\mathrm{R}} \tag{5.7}$$

and

$$F_{\mathrm{CPI}}^{\mathrm{R}} = -\frac{1}{\beta} \ln \frac{1}{\lambda_{\mathrm{eff}}} \int_{-\infty}^{z^{\neq}} \mathrm{d}z \, e^{-\beta[\Delta F(z)]} \tag{5.8}$$

$F_{\mathrm{CPI}}^{\mathrm{R}}$ corresponds to the free energy of the system in the reactant (R) state region relative to the lowest point (z_{R}), which may be interpreted as the entropic contributions or motions correlating with the progress coordinate z. λ_{eff} is the de Broglie thermal wavelength of the centroid reaction coordinate with an effective mass M_{eff} at the dividing surface, which is determined in the centroid path transition state ensemble.

5.2.2 Centroid Path Integral Simulations

In centroid path integral, the centroid position, $\bar{r}$, is used as the principle variable and the canonical QM partition function of a hybrid quantum and classical system, consisting of one quantised atom for convenience, can be written as follows:[23]

$$Q_P^{\text{qm}} = \frac{1}{\Omega} \int d\mathbf{S} \int d\mathbf{s} \left(\frac{P}{\lambda_{\text{M}}^2}\right)^{3P/2} \int d\mathbf{R}\, e^{-\beta V_{\text{eff}}(\{\mathbf{r}\},\mathbf{S})} \qquad (5.9)$$

where Ω is the volume element of classical particles, P is the number of quasiparticles, $V_{\text{eff}}(\{\mathbf{r}\},\mathbf{s})$ is the effective potential (eqn (5.2)), $\int d\mathbf{R} = \int d\mathbf{r}_1 \cdots \int d\mathbf{r}_P \delta(\bar{\mathbf{r}} = \mathbf{s})$ in which the delta function $\delta(\bar{\mathbf{r}} = \mathbf{s})$ is introduced for use in later discussion, and the centroid $\bar{\mathbf{r}}$ is defined in eqn (5.1). Importantly, eqn (5.9) can be rewritten exactly as a double average in eqn (5.10), which is the theoretical basis in the simulation approach of Sprik *et al.*, called the hybrid classical and path integral,[25] of Hwang and Warshel, called QCP,[34,35] and later, of Major and Gao, called PI-FEP/UM.[8–10]

$$Q_{P\to\infty}^{\text{qm}} = Q_{\text{cm}} \langle \langle e^{-\beta \Delta \bar{U}(\bar{r},\mathbf{S})} \rangle_{\text{FP},\bar{\mathbf{r}}} \rangle_U \qquad (5.10)$$

where the average $\langle \cdots \rangle_U$ is a purely classical ensemble average obtained according the potential $U(\bar{\mathbf{r}},\mathbf{S})$ (the potential U in eqn (5.2) with a single centroid position for the quantised particle), the average differential potential is given by

$$\Delta \bar{U}(\bar{\mathbf{r}},\mathbf{S}) = \frac{1}{P} \sum_i^P \left\{ U(\mathbf{r}_i,\mathbf{S}) - U(\bar{\mathbf{r}},\mathbf{S}) \right\} \qquad (5.11)$$

and the inner average $\langle \cdots \rangle_{\text{FP},\bar{\mathbf{r}}}$ represents a path-integral free-particle sampling, carried out without the external potential $U(\bar{\mathbf{r}},\mathbf{S})$:[10,34,35]

$$\langle \cdots \rangle_{\text{FP},\bar{\mathbf{r}}} = \frac{\int d\mathbf{R}\{\cdots\} e^{-(\pi P/\lambda_{\text{M}}^2) \sum_i^P (\Delta \mathbf{r}_i)^2}}{\int d\mathbf{R}\, e^{-(\pi P/\lambda_{\text{M}}^2) \sum_i^P (\Delta \mathbf{r}_i)^2}} \qquad (5.12)$$

where $\Delta \mathbf{r}_i = \mathbf{r}_i - \mathbf{r}_{i+1}$. In eqn (5.10), the factor Q_{cm} is the classical partition function defined in,[23]

$$Q_P^{\text{cm}} = \frac{1}{\Omega} \int d\mathbf{S} \int d\mathbf{s}\, e^{-\beta U(\mathbf{s},\mathbf{S})} \left(\frac{P}{\lambda^2}\right)^{3P/2} \int d\mathbf{R}\, e^{-(\pi P/\lambda_{\text{M}}^2) \sum_i^P (\Delta \mathbf{r}_i)^2} \qquad (5.13)$$

where we have defined the position of the quantised particle centroid to coincide with the coordinates of the corresponding classical particle $\mathbf{s} = \bar{\mathbf{r}}$.

251

We note that Warshel and coworkers have exploited this idea, which they called the quantised classical path (QCP) method, to estimate NQE for chemical reactions in solution and catalysed by enzymes.[34,35] In QCP, the classical simulations and quantum corrections are fully separated.[50,51] The expression of eqn (5.10) is particularly useful because the quantum free energy of the system can be obtained first by carrying out classical trajectories according to the classical distribution, $\exp[-\beta U(\bar{\mathbf{r}}, \mathbf{S})]$, and then, by determining the quantum contributions through free particle sampling based on the distribution $\exp[-\beta(\pi P/\beta\lambda_M^2)\sum_i^P(\Delta\mathbf{r}_i)^2]$. This double averaging yields the exact path integral centroid density,[10,25,34,35] which can be used to determine the centroid potential of mean force:

$$\mathrm{e}^{-\beta w(\bar{z})} = \mathrm{e}^{-\beta w_{\mathrm{cm}}(z)}\langle\delta(z=\bar{z})\langle\mathrm{e}^{-\beta\Delta\bar{U}(\bar{z}[\mathbf{r}],\mathbf{S})}\rangle_{\mathrm{FP},\bar{z}}\rangle_U \qquad (5.14)$$

where $w(\bar{z})$ and $w_{\mathrm{cm}}(z)$ are the centroid quantum-mechanical and the classical-mechanical PMF, respectively, and the average difference potential energy $\Delta\bar{U}(\bar{z}[\mathrm{r}], \mathbf{S})$ is given in eqn (5.11).

5.2.3 Kinetic Isotope Effects

We first present two algorithms for estimating KIEs using centroid path integral simulations in the context of QCP[34] or hybrid quantum and classical sampling.[25] Then, we present a third algorithm, making use of Kleinert variational perturbation (KP) theory[6] that determines the centroid potential of the mean force analytically without the need of performing discretised path integral sampling.[7]

5.2.3.1 Sequential Centroid Path Integral and Umbrella Sampling (PI/UM)

Using eqn (5.6), the kinetic isotope effects between a light isotope L and a heavy isotope H can be computed by

$$\mathrm{KIE} = \frac{k^L}{k^H} = \mathrm{e}^{-\beta[\Delta F^L(\bar{z}_L^{\neq}) - \Delta F^H(\bar{z}_H^{\neq})]}\mathrm{e}^{-\beta\{F_{\mathrm{CPI},L}^R(\bar{z}_L^R) - F_{\mathrm{CPI},H}^R(\bar{z}_H^R)\}} \qquad (5.15)$$

where $\bar{z}_L^{\neq}$, $\bar{z}_H^{\neq}$, $\bar{z}_L^R$, and $\bar{z}_H^R$ are, respectively, the values of the centroid reaction coordinate at the transition state and reactant state minimum for the light and heavy isotopes, and $F_{\mathrm{CPI},L}^R(\bar{z}_L^R)$ and $F_{\mathrm{CPI},H}^R(\bar{z}_H^R)$ are the free energies defined by eqn (5.8) for the two isotopes, respectively, which depend on the effective masses associated with the centroid reaction coordinate at the transition state. Equation (5.15) shows that KIEs may be determined by computing the potentials of mean force separately for the L and H isotopes,[9–10,34,35,52] and this can be achieved by first performing umbrella sampling using classical molecular dynamics and then by performing centroid path integral free-particle sampling with the constraints that the centroids of the quantised particles coincide with

the corresponding positions sampled in the classical trajectory. This is, indeed, what is typically done.[9,34,35,52] However, the statistical fluctuations in the actual simulation for computing the potential of mean force is typically as large as, or even greater than, the isotope effect itself, resulting in poor convergence.[10] To alleviate this difficulty, in refs 9 and 10, we have introduced a procedure such that the discretised beads positions for the heavy and light atoms are obtained from exactly the same sampling sequence. Thus, the heavy and light path integral beads distributions differ only by their relative spread from the centroid position, which is determined by the corresponding de Broglie wavelength.

To enhance convergence in these centroid path integral simulations, we have developed a bisection sampling technique for a ring of beads, called BQCP, by extending the original approach of Ceperley for free particle sampling in which the initial and final beads are not connected.[53,54] In our implementation, we first make the bisection sampling as originally proposed by Ceperley,[53,54] enforcing the first and last beads to be identical to enclose the polymer ring.[8,9] Then, we make rigid-body translation of the centroid position of the new beads configuration to coincide with the target (classical) coordinate. Since the free-particle distribution is known exactly at a given temperature, each ring-bead distribution is generated according to this distribution and thus 100% accepted.[54] Furthermore, in this construction, each new configuration is created independently, starting from a single initial bead position, allowing the new configuration to move into a completely different region of configurational space. This latter point is especially important in achieving convergence by avoiding being trapped in a local region of the classical potential in the path integral sampling. The BQCP method has been thoroughly tested[8,9] and applied to several condensed-phase systems.[9–10,55,56]

5.2.3.2   The PI-FEP/UM Method

The second algorithm is to obtain the ratio of the quantum partition functions (eqn (5.10)) for two different isotopes directly through free-energy perturbation (FEP) theory by perturbing the mass from the light isotope to the heavy isotope.[10] In other words, only one simulation of a given isotopic reaction is performed, while the ratio of the partition function, *i.e.* the KIE, to a different isotopic reaction, is obtained by FEP within this simulation. This is in contrast to algorithem 1, in which two separate simulations are performed, and this difference results in a major improvement in computation accuracy for KIE calculations.[10,32]

Specifically, eqn (5.15) for computing kinetic isotope effect is rewritten in terms of the ratio of the partial partition functions at the centroid reactant and transition state and is given by:

$$\text{KIE} = \frac{k^L}{k^H} = \left[\frac{Q_{\text{qm}}^L(\bar{z}_L^{\neq})}{Q_{\text{qm}}^H(\bar{z}_H^{\neq})}\right]\left[\frac{Q_{\text{qm}}^H(\bar{z}_H^R)}{Q_{\text{qm}}^L(\bar{z}_L^R)}\right]\text{e}^{-\beta\{F_{\text{CPI},L}^R(\bar{z}_L^R)-F_{\text{CPI},H}^R(\bar{z}_H^R)\}} \tag{5.16}$$

where the ratio of the partition function can be written as follows:

$$\frac{Q_{\text{qm}}^H(\bar{z})}{Q_{\text{qm}}^L(\bar{z})} = \frac{\langle\delta(z-\bar{z})\langle e^{-\frac{\beta}{P}\sum_i \Delta U_i^{L\to H}} e^{-\beta\Delta\bar{U}_L}\rangle_{\text{FP},L}\rangle_U}{\langle\delta(z-\bar{z}) e^{-\beta[F_L(\bar{z},\mathbf{S})-F_{FP}^o]}\rangle_U} \tag{5.17}$$

where the subscripts L specifies that the ensemble averages are done using the light isotope, $\Delta\bar{U}_L$ is defined by eqn (5.12), F_{FP}^o is the free energy of the free particle reference state for the quantised particles,[23] and $\Delta U_i^{L\to H} = U(\mathbf{r}_{i,H}) - U(\mathbf{r}_{i,L})$ represents the difference in "classical" potential energy at the heavy and light bead positions $\mathbf{r}_{i,H}$ and $\mathbf{r}_{i,L}$. In the bisection sampling scheme, the perturbed heavy isotope positions are related to the lighter ones by

$$\frac{\mathbf{r}_{i,L}}{\mathbf{r}_{i,H}} = \frac{\lambda_{M_L}\theta_i}{\lambda_{M_H}\theta_i} = \sqrt{\frac{M_H}{M_L}}; \quad i = 1, 2, \cdots, P \tag{5.18}$$

where $\mathbf{r}_{i,L}$ and $\mathbf{r}_{i,H}$ are the coordinates for bead i of the corresponding light and heavy isotopes, λ_{M_L} and λ_{M_H} are isotopic masses for the light and heavy nuclei, and θ_i is the position vector in the bisection sampling scheme that depends on the previous sequence of directions and has been fully described in ref. 8. Equation (5.18) indicates that the position vectors for the corresponding heavy and light isotope beads in the path integral simulation are identical, thereby, resulting in the relationship that bead positions are solely determined by the ratio of the square roots of masses.

In eqn (5.17), we obtain the free-energy (inner average) difference between the heavy and light isotopes by carrying out the bisection path integral sampling with the light atom and then perturbing the heavy isotope positions according to eqn (5.18). Then, the free-energy difference between the light and heavy isotope ensembles is weighted by a Boltzmann factor for each quantised configuration.

5.2.3.3 Kleinert's Variational Perturbation (KP) Theory

The canonical partition function Q_{QM} for a quantised particle in the bath of classical solvent can be written in terms of the effective centroid potential w, as a classical configuration integral:

$$Q_P^{\text{qm}} = \sqrt{\frac{Mk_{\text{B}}T}{2\pi\hbar^2}} \int_{-\infty}^{\infty} e^{-\beta w(\bar{\mathbf{r}})}d\bar{\mathbf{r}} \tag{5.19}$$

where M is the mass of the quantised particle, $\hbar$ is Planck's constant divided by 2π. In the 1980s, three groups proposed a variational method,[39,57,58] known as the Feynman and Kleinert (FK) variational method that yields an upper bound of $w(\bar{\mathbf{r}})$ and has been successfully used in a variety of applications.[59–62] Kleinert later showed that the FK approach in fact is just the first-order approximation to a general variation perturbation theory,[6,36] and the new Kleinert perturbation theory (KP) has been shown in several model systems to have very attractive features, including exponentially and uniformly convergent. For

example, the electronic ground-state energy of a hydrogen atom converges at accuracies of 85%, 95% and 98% in the first three orders of KP expansion. The KP theory does not suffer from the numerical convergence issues in discretised path integral simulations; however, it has so far only been applied to a few limited model cases beyond the first-order approximation, *i.e.* the Feynman–Kleinert approach.

We have shown recently that the KP theory can be adopted in chemical applications to determine KIEs for chemical reactions, and because of its fast convergence the second-order perturbation, denoted by KP2, can yield excellent results for a series of proton-transfer reactions.[7] We briefly summarise the key elements of our theoretical development based on the Kleinert variational perturbation theory.

The *n*th-order KP approximation $W_\Omega^n(\bar{\mathbf{r}})$ to the centroid potential $w(\bar{\mathbf{r}})$ is given by

$$
\begin{aligned}
-\beta W_\Omega^n(\bar{\mathbf{r}}) = {} & \ln Q_\Omega(\bar{\mathbf{r}}) - \frac{1}{\hbar}\left\langle V_p^{\bar{\mathbf{r}}}[\mathbf{r}(\tau_1)]\right\rangle_{\Omega,c}^{\bar{\mathbf{r}}} + \frac{1}{2!\hbar^2}\left\langle V_p^{\bar{\mathbf{r}}}[\mathbf{r}(\tau_1)]V_p^{\bar{\mathbf{r}}}[\mathbf{r}(\tau_2)]\right\rangle_{\Omega,c}^{\bar{\mathbf{r}}} \\
& + \cdots + \frac{(-1)^n}{n!\hbar^n}\left\langle \prod_{k=1}^{n} V_p^{\bar{\mathbf{r}}}[\mathbf{r}(\tau_k)]\right\rangle_{\Omega,c}^{\bar{\mathbf{r}}}
\end{aligned}
\tag{5.20}
$$

where the angular frequency Ω is a variational parameter, which is introduced to define the perturbation potential $V_p^{\bar{\mathbf{r}}}[\mathbf{r}] = U(\mathbf{r},S) - U_\Omega^{\bar{\mathbf{r}}}(\mathbf{r})$ about the reference state at $\bar{\mathbf{r}}$ with the harmonic potential $U_\Omega^{\bar{\mathbf{r}}}(\mathbf{r}) = (M/2)\Omega^2(\mathbf{r} - \bar{\mathbf{r}})^2$. The local quantum partition function of the harmonic reference state $Q_\Omega(\bar{\mathbf{r}})$ is given as follows:

$$
Q_\Omega(\bar{\mathbf{r}}) = \frac{\beta\hbar\Omega(\bar{\mathbf{r}})/2}{\sinh(\beta\hbar\Omega(\bar{\mathbf{r}})/2)}
\tag{5.21}
$$

The remaining terms in eqn (5.20) are the *n*th-order corrections to approximate the real system, in which the expectation value $\langle\cdots\rangle_{\Omega,c}^{\bar{\mathbf{r}}}$ is called the cumulant. The cumulants can be written in terms of the standard expectation value $\langle\cdots\rangle_\Omega^{\bar{\mathbf{r}}}$ by cumulant expansion. Kleinert and coworkers derived the expression for the expectation value of a function in terms of Gaussian smearing convolution integrals:[6]

$$
\left\langle \prod_{k=1}^{n} F[\mathbf{r}_k(\tau_k)]\right\rangle_\Omega^{\bar{\mathbf{r}}} = \frac{\prod_{j=1}^{n}\int_0^{\beta\hbar}\mathrm{d}\tau_j \prod_{k=1}^{n}\int_{-\infty}^{\infty}\mathrm{d}\mathbf{r}_k F(\mathbf{r}_k(\tau_k))}{\{(2\pi)^n\mathrm{Det}[a_{\tau_k\tau_{k'}}^2(\Omega)]\}^{3/2}}
$$
$$
\times \exp\left\{-\frac{1}{2}\sum_{k'=1}^{n}\sum_{k=1}^{n}(\mathbf{r}_k - \bar{\mathbf{r}})\cdot a_{\tau_k\tau_{k'}}^{-2}(\Omega)(\mathbf{r}_{k'} - \bar{\mathbf{r}})\right\}
\tag{5.22}
$$

where $\text{Det}[a^2_{\tau_k \tau_{k'}}(\Omega)]$ is the determinant of the $n \times n$ matrix consisting of the Gaussian width $a^2_{\tau_k \tau_{k'}}(\Omega)$, $a^{-2}_{\tau_k \tau_{k'}}(\Omega)$ is an element of the inverse matrix of the Gaussian width, which is given by

$$a^2_{\tau_k \tau_{k'}}(\Omega) = \frac{1}{\beta M \Omega^2} \left\{ \frac{\beta \hbar \Omega \cosh[(|\tau_k - \tau_{k'}| - \beta \hbar/2)\Omega]}{2 \sinh[\beta \hbar \Omega/2]} - 1 \right\} \tag{5.23}$$

As n tends to infinity, $W^n_\Omega(\bar{\mathbf{r}})$ becomes independent of the variational parameter Ω. At a given order of the KP theory, the optimal frequency is given by the least dependence of $W^n_\Omega(\bar{\mathbf{r}})$ on Ω, which is the solution to the equation of the lowest-order derivative of $W^n_\Omega(\bar{\mathbf{r}})$ setting to zero.[6] Thus, a self-consistent iterative procedure is carried out. Given an initial trial of Ω, the corresponding Gaussian widths are determined using eqn (5.23), with which a new $W^n_\Omega(\bar{\mathbf{r}})$ is obtained. This is then used to obtain a further value of Ω until it is converged.[7]

An especially attractive feature of eqn (5.20) is that if the real system potential U is expressed as a series of polynomials or Gaussian functions, analytic expressions of eqn (5.18) can be obtained,[7] making the computation extremely efficient because there is no need to perform the time-consuming Monte Carlo or molecular dynamics sampling of the path integrals. If the real potential is expanded to the mth-order polynomial and the KP theory is terminated at the nth order (eqn (5.20)), we denoted our results as KPn/Pm.

If the number of quantised particles is N, the angular frequency variational variable is a $3N \times 3N$ matrix, and this coupled with the $2n$-dimension integrals in eqn (5.22) makes the use of KP theory rather laborious and has been a major factor limiting its applications beyond the KP1 level, the original FK approach. To render the KP theory feasible for many-body systems with N nuclei for a given configuration $\{\bar{\mathbf{r}}_i; i = 1, \cdots, N\}$, we make use of instantaneous normal mode (INM) coordinates $\{q(\bar{\mathbf{r}}_i; i = 1, \cdots, N)\}^{3N}$ and assume their motions can be decoupled. Thus, system potential $U(\{\mathbf{r}\}, \mathbf{S})$ can be expanded in terms of the INM coordinates at $\{\bar{\mathbf{r}}_i; i = 1, \cdots, N\}$, and hence, the multidimensional potential is effectively reduced to $3N$ one-dimensional potentials along each normal coordinate. This approximation is particularly suited for the KP theory because of the exponentially decaying property of the Gaussian convolution integrations in eqn (5.22). With the INM approximation, the total effective centroid potential for N-quantised nuclei can be simplified as:

$$W^n_\Omega(\{\bar{\mathbf{r}}_i\}^{3N}) \approx U(\{\bar{\mathbf{r}}_i\}^{3N}, \mathbf{S}) + \sum_{i=1}^{3N} W^n_{\Omega_i}(\{q_i\}^{3N}, \mathbf{S}) \tag{5.24}$$

where $W^n_{\Omega_i}(\{q_i\}^{3N}, \mathbf{S})$ is the centroid potential for normal mode i of the quantised system in the classical bath coordinates of the remainder of the system. Although the INM approximation sacrifices some accuracy, in return, it allows analyses of quantum-mechanical vibration and tunnelling and their separate contributions to the total quantum effects. Note that positive and

negative values of $W_{\Omega_i}^n(\{q_i\}^{3N}, \boldsymbol{S})$ in eqn (5.24) raise (vibration) and lower (tunnelling) the classical potential $U(\{\bar{\mathbf{r}}_i\}^{3N}, \boldsymbol{S})$, respectively.

5.3   Potential-Energy Surface

Almost all enzyme reactions can be well described by the Born–Oppenheimer approximation, in which the sum of the electronic energy and the nuclear repulsion provides a potential-energy function, or potential-energy surface (PES), governing the interatomic motions. Therefore, the molecular modelling problem breaks into two parts: the PES and the dynamics simulations.

The potential-energy function describes the energetic changes as a function of the variations in atomic coordinates, including thermal fluctuations and rearrangements of the chemical bonds. The accuracy of the potential-energy function used to carry out molecular-dynamics simulations directly affects the reliability of the computed $\Delta F_{\text{TST}}^{\neq}$ and its nuclear quantum correction.[5,20] The accuracy can be achieved by the use of analytical functions fitted to reproduce key energetic, structural, and force constant data, from either experiments or high-level *ab initio* calculations. Molecular-mechanical (MM) potentials or force fields,[63,64] however, are not general for chemical reactions, and it requires reparameterisation of the empirical parameters for every new reaction, which severely limits its applicability. More importantly, often, little information is available in regions of the PES other than the stationary reactant and product states and the saddle point (transition state). On the other hand, combined quantum-mechanical and molecular-mechanical (QM/MM) potentials offer the advantages of both computational efficiency and accuracy for all regions of the PES.[65,66]

5.3.1   Combined QM/MM Potentials

In combined QM/MM potentials, the system is divided into a QM region and an MM region.[12,15,16,65,67–70,] The QM region typically includes atoms that are directly involved in the chemical step and they are treated explicitly by a quantum-mechanical electronic-structure method, whereas the MM region consists of the rest of the system and is approximated by an MM force field. The method of combining QM with MM was first developed by Warshel and Karplus for the treatment of conjugated polyenes with QM and the framework with a force field,[15] and it was subsequently applied to an enzyme by Warshel and Levitt in which electrostatic interactions between QM and MM were introduced.[16] The remarkable applicability of this approach that it enjoys today was in fact not appreciated until more than a decade later when molecular dynamics and Monte Carlos simulations using combined QM/MM potentials began to emerge.[65,68,71] The QM/MM potential is given by:[66,67]

$$U_{\text{tot}} = \langle \Psi(S) | H_{\text{qm}}^{\text{o}}(S) + H_{\text{qm/mm}}(S) | \Psi(S) \rangle + U_{\text{mm}} \qquad (5.25)$$

where $H_{\text{qm}}^{\text{o}}(S)$ is the Hamiltonian of the QM-subsystem (the substrate and key amino-acid residues) in the gas phase, U_{mm} is the classical (MM) potential

energy of the remainder of the system, $H_{qm/mm}(S)$ is the QM/MM interaction Hamiltonian between the two regions, and $\Psi(S)$ is the molecular wavefunction of the QM-subsystem optimised for $H_{qm}^{o}(S) + H_{qm/mm}(S)$.

We have found that it is most convenient to rewrite eqn (5.25) as follows:[65,68]

$$U_{tot} = E_{qm}^{o}(S) + \Delta E_{qm/mm}(S) + U_{mm} \tag{5.26}$$

where $E_{qm}^{o}(S)$ is the energy of an isolated QM subsystem in the gas phase,

$$E_{qm}^{o}(S) = \langle \Psi^{o}(S)|H_{qm}^{o}(S)|\Psi^{o}(S)\rangle \tag{5.27}$$

In eqn (5.26), $\Delta E_{qm/mm}(S)$ is the interaction energy between the QM and MM regions, corresponding to the energy change of transferring the QM subsystem from the gas phase into the condensed phase, which is defined by:

$$\Delta E_{qm/mm}(S) = \langle \Psi(S)|H_{qm}^{o}(S) + H_{qm/mm}(S)|\Psi(S)\rangle - E_{g}^{o}(S) \tag{5.28}$$

In eqns (5.25)–(5.28), we have identified the energy terms involving electronic degrees of freedom by E and those purely empirical functions by U, the combination of which is also an empirical potential.

Equation (5.26) is especially useful in that the total energy of a hybrid QM and MM system is separated into two "*independent*" terms – the gas-phase energy and the interaction energy – which can now be evaluated using different QM methods. There is sometimes confusion about the accuracy of applications using semiempirical QM/MM potentials.[72] Equation (5.26) illustrates that there are two issues. The first is the intrinsic performance of the model, *e.g.*, the $E_{qm}^{o}(S)$ term, which is indeed not adequate using semiempirical models and that would require *extremely* high-level QM methods to achieve the desired accuracy. This is only possible by using CCSD(T), CASPT2 or well-tested density functionals along with a large basis set, none of which methods are tractable for applications to enzymes. When semiempirical methods are used, the PES for the $E_{qm}^{o}(S)$ term, is either reparameterised to fit experimental data, or replaced by high-level results. Only in rare occasions when a semiempirical model yields good agreement with experiment, are these methods directly used without alteration.[73,74]

The second issue on accuracy is in the calculation of the $\Delta E_{qm/mm}(S)$ term. It was recognised early on,[65,69,75] when explicit QM/MM simulations were carried out, that combined QM/MM potential is an empirical model, which contains empirical parameters and should be optimised to describe QM/MM interactions. By systematically optimising the associated van der Waals parameters for the "QM-atoms",[65,69,75] both semiempirical and *ab initio* (Hartree–Fock) QM/MM potentials can yield excellent results for hydrogen-bonding and dispersion interactions in comparison with experimental data. The use of semiempirical methods such as the Austin model 1 (AM1)[76] or parameterised model 3

(PM3)[77] in QM/MM simulations has been validated through extensive studies of a variety of properties and molecular systems, including computations of free energies of solvation and polarisation energies of organic compounds,[65,78] the free-energy profiles for organic reactions,[55,73,79] and the effects of solvation on molecular structures and on electronic transitions.[80,81]

The QM/MM PES combines the generality of quantum-mechanical methods for treating chemical processes with the computational efficiency of molecular mechanics for large molecular systems. The use of an explicit electronic-structure method to describe the enzyme-active site is important because understanding the changes in electronic structure along the reaction path can help to design inhibitors and novel catalysts. It is also important because the dynamic fluctuations of the enzyme and aqueous solvent system have a major impact on the polarisation of the species involved in the chemical reaction, which, in turn, affects the chemical reactivity.[65,82] Combined QM/MM methods have been reviewed in several articles.[66–67,83–85,]

5.3.2   The MOVB Potential

A novel combined QM/MM approach has been developed that utilises a mixed molecular-orbital and valence-bond (MOVB) theory to represent the reactive potential-energy surface. In this MOVB approach, molecular orbital theory based on a block-localised orbital scheme is used to define the Lewis resonance structures.[11,86] These Lewis resonance structures are called effective diabatic states, representing the reactant state, the product state or other states important for the transition state of the chemical reaction under investigation, which are coupled by configuration interaction (CI).[12–14] A closely analogous model is the empirical valence bond (EVB) potential that has been used by Warshel and coworkers,[87] who approximate these effective diabatic states by an empirical force field, making the computation cheap for simulation studies. Although it is perfectly reasonable to use empirical force field to represent these diabatic states, provided that the force field is adequately parameterised to reflect the true electronic structure for the entire system, in practice, it is often assumed that the atomic partial charges are invariant by taking values at the reactant state and product state, respectively. This is a major shortcoming because intramolecular charge polarisation is ignored, which leads to inconsistent representation of the bond orders for the "primary" and "secondary" sites when KIEs are being computed. Note that although it is possible to parameterise the free-energy barrier and its change from water to the enzyme by reproducing experimental results, there is no rigorous justification for the accuracy of the 3N-degree of freedom PES, casting doubt on results that require a knowledge of the gradient and Hessian of the potential surface, a prerequisite for computing KIEs.

Truhlar and coworkers developed a multiconfiguration molecular-mechanics (MCMM) method that involves a systematic parameterisation for the off-diagonal element.[88,89] The MCMM method is fitted to reproduce *ab initio*

259

energies and gradients. Thus, the MCMM potential as well as the MOVB theory are proper empirical and first-principle functions, respectively, for use in computing nuclear quantum effects and kinetic isotope effects.

5.4   Computational Details

In the path integral simulations using algorithms 1 and 2, we employ a combined QM/MM potential in molecular-dynamics simulations.[65,66] In these studies, the solute is represented explicitly by an electronic-structure method and the solvent is approximated by the three-point charge TIP3P model for water.[90] In the deprotonation of nitroethane by acetate ion, the standard semiempirical AM1 model[76] failed to yield adequate energetic results. Consequently, a set of specific reaction parameters (SRP) has been developed within the AM1 formalism to fit results from high-level *ab initio* theory as well as from experiments.[10,55] The performance of the SRP-AM1 model has been reported previously, and we focus here on the study of the kinetic isotope effects using the PI-FEP/UM method.

In the study of carbon acid deprotonation using the KP2 theory, the solute is treated by B3LYP/6-31 + G(d,p) density functional theory, whereas the solvent is represented by the polarisable continuum model with a dielectric constant of 78 for water. Note that the KP centroid path-integral method is numerical integration free; thus, high-level electronic-structure methods such as DFT with a relatively large basis set can be used. In these calculations, the potential along each normal mode direction for the reactant state, transition state and product state is determined at the B3LYP level by stepping in each direction at an interval of $0.1\,\text{Å}$, and the resulting potential is fitted to a 20th-order polynomial, whose Gaussian smearing integrals (eqn (5.22)) have been derived analytically at the KP2 level. Thus, our results are obtained at the KP2/P20 level for the proton-transfer reactions of a series of aryl-substituted α-methoxystyrene compounds to acetate ion.

For algorithms 1 and 2, the simulations were performed using periodic boundary conditions in the isothermal–isobaric (NPT) ensemble at $25\,°\text{C}$ and 1 atm. A total of 898 water molecules were included in a cubic box of about $30 \times 30 \times 30\,\text{Å}^3$. Nonbonded electrostatic interactions are described by the particle-mesh Ewald summation method for QM/MM simulations,[91] whereas van der Waals interactions are smoothed to zero at $9.5\,\text{Å}$ based on group–group separations. The bond lengths and angles of solvent water molecules are constrained by the SHAKE algorithm, and an integration step of 1 fs was used for all calculations.[92]

The potential of mean force (PMF) profile is obtained using the umbrella sampling technique.[93] For the proton-transfer reaction between nitroethane and acetate ion, the classical reaction coordinate is defined as the difference in distance for the proton between the donor (α carbon of nitroethane) and the acceptor (an oxygen of the acetate ion) atom: $Z_{PT} = r(C_\alpha - H) - r(H - O)$. A total of *ca.* 4 ns of simulations were performed (with time step of 1 fs).

The BQCP simulations employed 29 168 classical configurations for each isotope (^{1}H, ^{2}H, and ^{3}H; or H, D, and T), combined with 10 path-integral steps per classical step. For the deprotonation reaction the nitroethane Cα-atom, the abstracting acetate oxygen, the transferring proton, as well as the secondary hydrogen atom, are quantised. Each quantised atom has spawned into 32 beads.

To estimate statistical uncertainties in the computed KIEs, the entire path integral simulations have been divided into ten segments, each of which is treated as independent simulations. Standard uncertainties (±1σ) were determined from the total average and those from the ten separate blocks. All simulations employed the CHARMM program[94] and all path-integral simulations used a parallel version that efficiently distributes integral calculations for the quantised beads.[8–10]

5.5   Illustrative Examples

5.5.1   Proton Transfer between Nitroethane and Acetate Ion

The proton-abstraction reaction of nitroalkane by acetate ion is a classical example, exhibiting an unusual Brønsted relationship in water, known as the nitroalkane anomaly.[95–97] This process is also catalysed by the nitroalkane oxidase in the initial step of the oxidation of nitroalkanes.[98] Valley and Fitzpatrick reported a KIE of 9.2 ± 0.4 for the di-deuterated substrate ($1,1$-^{2}H$_2$) nitroethane in nitroalkane oxidase, and 7.8 ± 0.1 for uncatalysed reaction by an acetate ion in water.[99] We have previously studied the solvent effects and reported preliminary results of H/D kinetic isotope effects.[55] Then, the KIEs for all D and T primary and secondary isotope effects were determined.[32]

The computed PMF, in which the ''primary'' and ''secondary'' hydrogen atoms as well as the donor carbon and acceptor oxygen atoms are quantised for the deprotonation of nitroethane by acetate, reveals a computed barrier of 24.4 kcal/mol,[10,32,55] in reasonable accord with the experimental values of 24.8.[99] Without the quantum contribution the computed barrier would have been 27.4 kcal/mol, illustrating the importance of including nuclear quantum effects to accurately determine the free energy of activation for proton-transfer reactions.[100]

The computed primary and secondary KIEs for D and T substitutions are listed in Table 5.1. Figure 5.1 depicts an illustrative example of the computed ratio of the partial quantum partition functions as a function of the centroid path integral reaction coordinate. The kinetic isotope effects have been computed without including the free-energy difference given in eqn (5.16), which may introduce some errors in the present calculations. The computed H/D primary and secondary intrinsic KIEs are $k_{\mathrm{H}}^{\mathrm{H}}/k_{\mathrm{D}}^{\mathrm{H}} = 6.63 \pm 0.31$ and $k_{\mathrm{H}}^{\mathrm{H}}/k_{\mathrm{H}}^{\mathrm{D}} = 1.34 \pm 0.13$, respectively, whereas the total effect where both primary and secondary hydrogen atoms are replaced by a deuterium isotope is $k_{\mathrm{H}}^{\mathrm{H}}/k_{\mathrm{D}}^{\mathrm{D}} = 8.31 \pm 0.13$ (Table 5.1), which may be compared with the

Table 5.1   Computed and experimental primary and secondary kinetic isotope effects for the proton-transfer reaction between nitroethane and acetate ion in water at 25 °C.[a]

KIE	PI-FEP/UM	Expt.[b]
Primary KIE		
k_H^H/k_D^H	6.63 ± 0.31	
k_H^H/k_T^H	12.96 ± 0.98	
K_D^D/k_T^D	2.17 ± 0.04	
Secondary KIE		
k_H^H/k_H^D	1.340 ± 0.132	
k_H^H/k_H^T	1.375 ± 0.183	
K_D^D/k_D^T	1.096 ± 0.039	
Total KIE		
k_H^H/k_D^D	8.31 ± 0.13	7.8

[a]Kinetic isotope effects are determined by using the average value of the top two bins in the potential of the mean force for the ratio of the partial quantum partition function for the transition state, and the average value of the middle fifteen bins for the reactant state. The bin size used for data collection is 0.05 Å in the reaction coordinate, half of which may be considered as the error in the reaction coordinate value.
[b]Ref. [99].
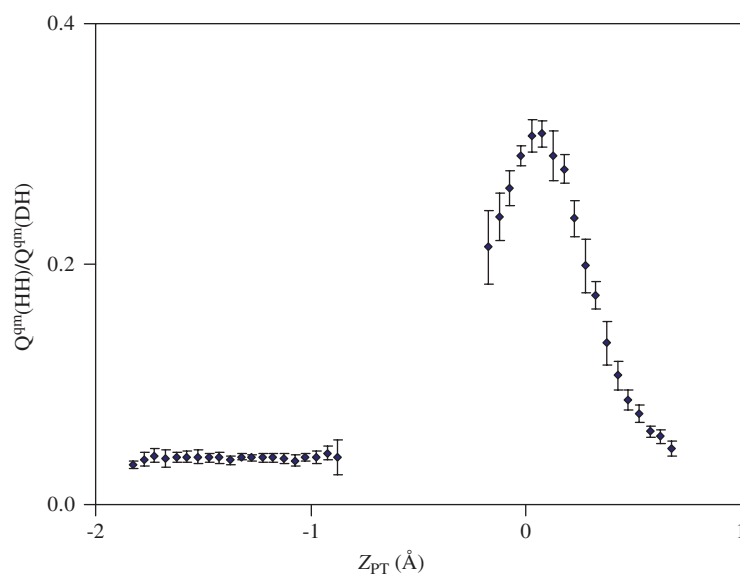
Figure 5.1   Computed ratio of the quantum-mechanical partition functions for the primary H/D kinetic isotope effects in which the secondary site is occupied by a hydrogen along the proton-transfer reaction coordinate. The PI-FEP/UM method was used.

experimental value of 7.8.[99] There are no experimental data for comparison with the results of single-site substitutions.

The computational results allow us to examine the rule of the geometric mean (RGM),[101] which is expressed as follows:

$$k_{\text{H}}^{\text{H}}/k_{\text{D}}^{\text{H}} = k_{\text{H}}^{\text{D}}/k_{\text{D}}^{\text{D}} = \frac{(k_{\text{H}}^{\text{H}}/k_{\text{D}}^{\text{D}})}{(k_{\text{H}}^{\text{H}}/k_{\text{H}}^{\text{D}})} \tag{5.29}$$

The RGM states that there is no isotope effect from a second site on the kinetic isotope effect of the first site.[102] The rule was originally derived at the high-temperature limit with small quantum tunnelling corrections,[101] and it has been shown to have negligible deviations on model systems using semiclassical transition-state theory.[103] However, deviations or the observations of RGM breakdown are often used as a measure of the extent of tunnelling in the system.[102] Using the RGM of eqn (5.29), we obtain an estimated value of 8.88 (6.63×1.34) for the total deuterium KIE if the free energies of the primary and secondary KIE were additive. This gives a ratio of 1.07 over the actual computed value (8.31). Another way of interpreting the results is that there is a secondary kinetic isotope effect of 1.07 on the primary kinetic isotope effects:

$$g_{\text{HD}}^{\text{HD}} = \frac{(k_{\text{H}}^{\text{H}}/k_{\text{D}}^{\text{H}})}{(k_{\text{H}}^{\text{D}}/k_{\text{D}}^{\text{D}})} = \frac{6.630}{(8.308/1.340)} = 1.07 \tag{5.30}$$

This result indicates that there is some correction in the motions between the secondary hydrogen and the primary hydrogen in the proton-transfer reaction between nitroethane and acetate ion in water.

Primary and secondary tritium kinetic isotope effects are also given in Table 5.1, which have values of $k_{\text{H}}^{\text{H}}/k_{\text{T}}^{\text{H}} = 12.96 \pm 0.98$ and $k_{\text{H}}^{\text{H}}/k_{\text{H}}^{\text{T}} = 1.37 \pm 0.18$. These effects are greater than the deuterium KIEs because of its larger mass. Employing the rule of the geometric mean, an estimated total tritium KIE of $k_{\text{H}}^{\text{H}}/k_{\text{T}}^{\text{T}} = 17.8$ is obtained.

The single site Swain–Schaad exponents[104] is expressed below using the notation of Huskey,[102]

$$n_{\text{HD}} = \frac{\ln(k_{\text{H}}^{\text{H}}/k_{\text{T}}^{\text{H}})}{\ln(k_{\text{H}}^{\text{H}}/k_{\text{D}}^{\text{H}})} \tag{5.31}$$

and

$$n_{\text{DT}} = \frac{\ln(k_{\text{H}}^{\text{H}}/k_{\text{T}}^{\text{H}})}{\ln(k_{\text{D}}^{\text{H}}/k_{\text{T}}^{\text{H}})} \tag{5.32}$$

These equations assume that the isotope effects are determined solely by the use of a one-frequency model with contributions only from the zero-point energy without tunnelling. Studies have shown that the value of n_{HD} for primary KIEs is typically in the range of 1.43–1.45.[102] Deviations from these values are

thought to be indications of contributions from tunnelling.[105] Using the data in Table 5.1, we obtained a single-site Swain–Schaad exponent of $n_{\mathrm{HD}}^{(1)} = 1.35$ for the primary KIE, and of $n_{\mathrm{HD}}^{(2)} = 1.09$ for the secondary KIE. The exponents, $n_{\mathrm{DT}}^{(1)}$ and $n_{\mathrm{DT}}^{(2)}$, for D/T ratios are 3.82 and 12.3, respectively. These values show significant deviations from the semiclassical limits, particularly on secondary KIEs, which can have greater computational errors because of the small free-energy difference. The deviations may be attributed to too large a secondary H/D effect.

Mixed isotopic Swain–Schaad exponent is often used to assess tunnelling:[104]

$$n_{\mathrm{DT}}^{\mathrm{DD}} = \frac{\ln(k_{\mathrm{H}}^{\mathrm{H}}/k_{\mathrm{T}}^{\mathrm{H}})}{\ln(k_{\mathrm{D}}^{\mathrm{D}}/k_{\mathrm{T}}^{\mathrm{D}})} \tag{5.33}$$

and

$$n_{\mathrm{DD}}^{\mathrm{DT}} = \frac{\ln(k_{\mathrm{H}}^{\mathrm{H}}/k_{\mathrm{H}}^{\mathrm{T}})}{\ln(k_{\mathrm{D}}^{\mathrm{D}}/k_{\mathrm{D}}^{\mathrm{T}})} \tag{5.34}$$

The first equation is the primary Swain–Schaad exponent, which describes the relationship between H/T primary KIE when the secondary position is occupied by a hydrogen atom with the D/T primary KIE when the secondary position is occupied by a deuterium isotope. The second equation describes a similar relationship for the secondary kinetic isotope effects. Values of the mixed Swain–Schaad exponents significantly greater than 3.3 are typically attributed to contributions from tunnelling,[105] and experimental studies showed that the secondary exponent is more sensitive than the primary exponent in this type of analysis.

For the proton-transfer reaction between nitroethane and acetate ion in water, we obtain a primary KIE Swain–Schaad exponent of $n_{\mathrm{DT}}^{\mathrm{DD}} = 3.31$ and a secondary exponent of $n_{\mathrm{DT}}^{\mathrm{DD}} = 3.47$. These results are close to the semiclassical limit, suggesting that tunnelling contributions are not significant for this reaction in aqueous solution. It will be interesting to examine the effects of the enzyme-active site on tunnelling and the Swain–Schaad exponents.[98,99]

5.5.2   The Decarboxylation of N-Methyl Picolinate

The primary and secondary heavy atom kinetic isotope effects for the decarboxylation of N-methyl picolinate have been determined by Rishavy and Cleland.[106] QM/MM simulations were carried out for N-methyl picolinate, treated by the AM1 Hamiltonian, in a cubic box ($30 \times 30 \times 30\,\text{Å}^3$) of 888 water molecules, described by the TIP3P potential.[10] As usual, long-range electrostatic interactions were treated by the PME method for QM/MM potentials.[91] 2 ns of classical dynamics simulations were performed followed by path integral PI-FEP/UM simulations, employing a total of 97 800 classical configurations

for each isotope (^{12}C, ^{13}C, ^{14}N, and ^{15}N), combined with 10 path-integral steps per classical configuration. Each quantised atom was described by 32 beads.
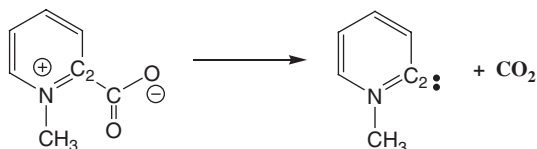

Solvent effects are significant, increasing the free-energy barrier by 15.2 kcal/mol to a value of 26.8 kcal/mol, which is accompanied by a net free energy of reaction of 24.7 kcal/mol. The large solvent effect is due to the presence of a positive charge on the pyridine nitrogen, which is annihilated in the decarboxylation reaction.

Both the $^{12}C/^{13}C$ primary KIE and the $^{14}N/^{15}N$ secondary KIE have been determined (Table 5.2), with the immediate adjacent atoms about the isotopic substitution site quantised as well. To our knowledge, we are not aware of any such simulations prior to our work for a condensed-phase reaction with converged secondary heavy isotope effects.[10] This demonstrate the applicability and accuracy of the PI-FEP/UM method.[10] Figures 5.2 and 5.3 depict the difference between the ^{12}C and ^{13}C quantum effects along the reaction path to illustrate the computational sensitivity using algorithem 1 (PI/UM) and algorithem 2 (PI-FEP/UM). First, the nuclear quantum effects are non-negligible even for bond cleavage involving two carbon atoms, which reduce the free-energy barrier by 0.45 kcal/mol (Figure 5.2). The computed intrinsic ^{13}C primary KIE, without including the reactant state quantum free energy term in eqn (5.16), is 1.0345 ± 0.0051 at 25 °C (Table 5.1). To emphasise the sensitivity of the computational result, the computed KIE is equivalent to a free-energy difference of merely 0.0187 kcal/mol, which is feasible by the use of free-energy perturbation/umbrella sampling techniques. For comparison, the experimental value is 1.0281 ± 0.0003 at 25 °C. For the secondary ^{15}N KIE, the PI-FEP/UM simulation yields an average value of 1.0083 ± 0.0026, which may be compared with experiment (1.0070 ± 0.0003).[106] The agreement between theory and experiment is excellent, which provides support for a unimolecular decarboxylation mechanism in this model reaction.

Table 5.2 Computed and experimental primary $^{12}C/^{13}C$ and secondary $^{14}N/^{15}N$ kinetic isotope effects for the decarboxylation of N-methyl picolinate at 25 °C in water.

	$^{12}k/^{13}k$	$^{14}k/^{15}k$
Exp (120 °C)	1.0212 ± 0.0002	1.0053 ± 0.0002
Exp (25 °C)	1.0281 ± 0.0003	1.0070 ± 0.0003
PI/UM	1.035 ± 0.877	1.007 ± 0.886
PI-FEP/UM	1.034 ± 0.005	1.008 ± 0.003
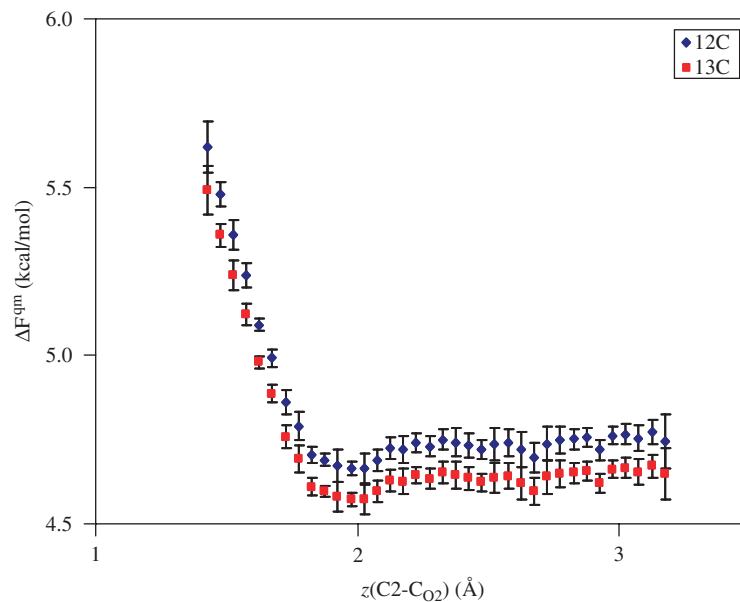
Figure 5.2 Nuclear quantum-mechanical free-energy corrections for the decarboxylation reaction of N-methyl picolinate in aqueous solution from PI/UM calculations. Two separate path integral simulations are performed in this method.
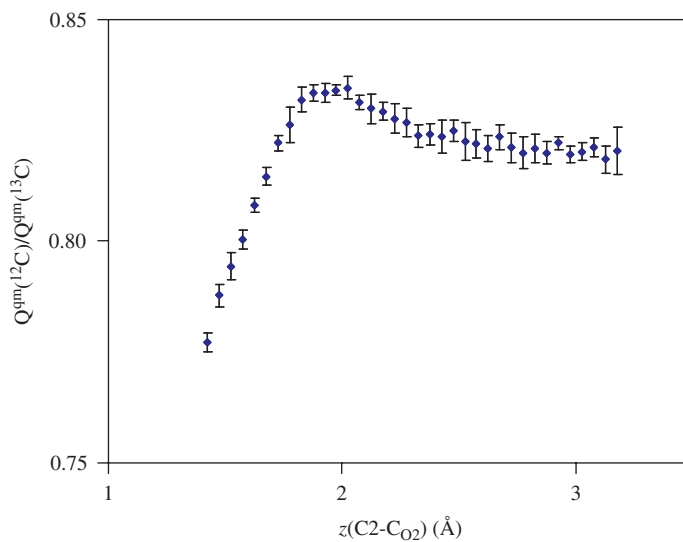
Figure 5.3 The ratio of the quantum-mechanical partition functions between ^{12}C and ^{13}C isotopic substitutions at the carboxyl carbon position from the PI-FEP/UM method. A single path integral simulation was performed.
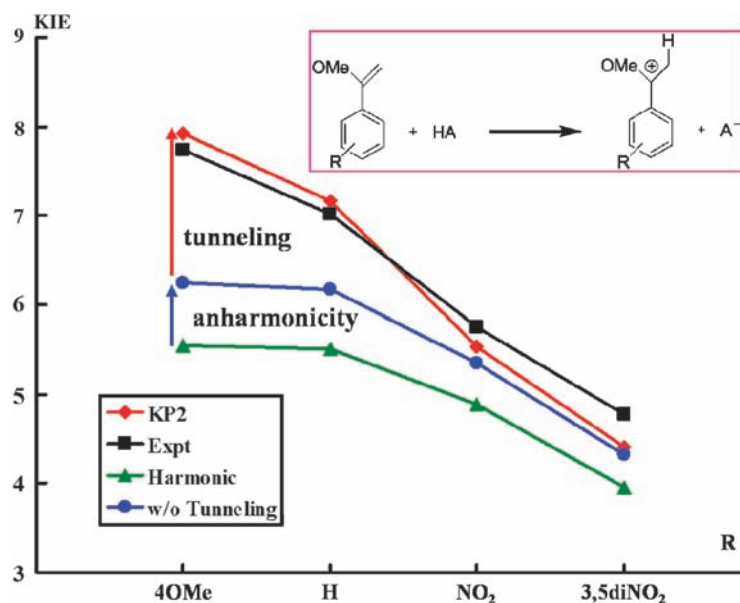
Figure 5.4   Computed primary (protium/deuterium) kinetic isotope effects for the proton-transfer reaction from chloroacetic acid to substituted methoxystyrenes (colour lines) in water represented by a polarizable dielectric continuum model using the KP2/P20 theory ($1 = 4$-methoxy, $2 = H$, $3 = 4$-Nitro, and $4 = 3,5$-dinitro groups).

5.5.3   Proton Transfer between Chloroacetic Acid and Substituted α-Methoxystyrenes

Richard and coworkers reported the product isotope effects of a series of interesting reactions involving the formation of carbocations from aryl-substituted (R) α-methoxystyrenes by proton transfer from chloroacetic acid.[107,108] These product isotope effects have been converted into KIEs, and the effects of substituents on the observed KIEs have been determined by the KP2/P20 theory.[7] Figure 5.4 shows the computed KIEs using only harmonic frequencies (zero-point energies) and the KP2/P20 values that include tunnelling and anharmonic corrections. Clearly, although the harmonic values correctly reproduced the experimental trend as a function of substituent electron-withdrawing power, the absolute values are significantly underestimated in comparison with experiment. Inclusion of anharmonicity for the description of the individual potential along each normal mode coordinate and tunnelling near the barrier top is critical to obtaining quantitative results for the computed KIEs.

5.6   Concluding Remarks

Centroid path integral methods have been presented for computing kinetic isotope effects for chemical reactions in aqueous solution. Three algorithms are

described: the first two algorithms involve discretised centroid path integral simulations to make quantum corrections to the classical free-energy path, and the third algorithm employs Kleinert variational perturbation theory at the second order in which the path integrals have been integrated analytically. In computing kinetic isotope effects, the first discretised simulation method is similar to the quantised classical path (QCP) approach developed by Hwang and Warshel, but we have developed a fast-converging sampling scheme, namely the bisection sampling method BQCP, by extending the method originally used by Ceperley and coworkers. In the second simulation algorithm, free-energy perturbation is employed by perturbing the light isotope mass into a heavier one in one single simulation, thereby avoiding the need of subtracting two quantum free energies with equally large fluctuations from two separate simulations. Thus, the accuracy in the computed KIEs is significantly improved. In Kleinert variational perturbation theory, analytical results are obtained when the potential surface is represented as a polynomial function, avoiding numerical convergence problems altogether. These methods are illustrated by computing the primary and secondary KIEs, analysing the results of the rule of geometric mean, and elucidating substituents effects. Three reactions have been considered: the proton abstraction of nitroethane by acetate ion, the proton-transfer reaction between substituted methoxystyrenes and chloracetic acid, and the decarboxylation of N-methyl picolinate, all of which are in water.

These examples illustrate the shortcomings of the two-separate simulation strategy of algorithem 1, typically used in the literature in path integral-based calculations of KIEs, and improved accuracy in the integrated centroid path integral–free energy perturbation and umbrella sampling scheme (algorithem 2) that uses a single simulation of a single isotope by perturbing its mass in the context of bisection sampling. Finally, the Kleinert variational perturbation theory gives an automated, numerical integration-free method for computing KIEs. The KP theory has the feature of exponentially and uniformly convergent such that the second-order perturbation (KP2) is sufficient to provide accurate results. An instantaneous normal mode approximation was made to extend the KP2 theory to multidimensional systems.

Acknowledgments

References

1. A. Kohen, H. H. Limbach (Eds.) *Isotope Effects in Chemistry and Biology*, Taylor & Francis Group, CRC Press, New York, 2005.
2. J. Pu, J. Gao and D. G. Truhlar, *Chem. Rev.*, 2006, **106**, 3140–3169.
3. J. Gao, S. Ma, D. T. Major, K. Nam, J. Pu and D. G. Truhlar, *Chem. Rev.*, 2006, **106**, 3188–3209.

4. V. L. Schramm, *Acc. Chem. Res.*, 2003, **36**, 588–596.
5. J. Gao and D. G. Truhlar, *Ann. Rev. Phys. Chem.*, 2002, **53**, 467–505.
6. H. Kleinert, *Path Itegrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets*, 3rd edn ed. Singapore, World Scientific; 2004.
7. (a) K.-Y. Wong and J. Gao, *J. Chem. Phys.*, 2007, **127**, 211103. (b) K.-Y. Wong and J. Gao, *J. Chem. Theory Comput.*, 2008, **4**, 1409–1422.
8. D. T. Major and J. Gao, *J. Mol. Graph. Model*, 2005, **24**, 121–127.
9. D. T. Major, M. Garcia-Viloca and J. Gao, *J. Chem. Theory Comput.*, 2006, **2**, 236–245.
10. D. T. Major and J. Gao, *J. Chem. Theory Comput.*, 2007, **3**, 949–960.
11. Y. Mo and J. Gao, *J. Comput. Chem.*, 2000, **21**, 1458–1469.
12. Y. Mo and J. Gao, *J. Phys. Chem. A*, 2000, **104**, 3012–3020.
13. J. Gao and Y. Mo, *Prog. Theor. Chem. Phys.*, 2000, **5**, 247–268.
14. J. Gao, M. Garcia-Viloca, T. D. Poulsen and Y. Mo, *Adv. Phys. Org. Chem.*, 2003, **38**, 161–181.
15. A. Warshel and M. Karplus, *J. Am. Chem. Soc.*, 1972, **94**, 5612–5625.
16. A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227–249.
17. A. Fernandez-Ramos, J. A. Miller, S. J. Klippenstein and D. G. Truhlar, *Chem. Rev.*, 2006, **106**, 4518–4584.
18. C. Alhambra, J. Gao, J. C. Corchado, J. Villa and D. G. Truhlar, *J. Am. Chem. Soc.*, 1999, **121**, 2253–2258.
19. C. Alhambra, J. Corchado, M. L. Sanchez, M. Garcia-Viloca, J. Gao and D. G. Truhlar, *J. Phys. Chem. B*, 2001, **105**, 11326–11340.
20. D. G. Truhlar, J. Gao, C. Alhambra, M. Garcia-Viloca, J. Corchado, M. L. Sanchez and J. Villa, *Acc. Chem. Res.*, 2002, **35**, 341–349.
21. M. Garcia-Viloca, J. Gao, M. Karplus and D. G. Truhlar, *Science*, 2004, **303**, 186–195.
22. S. Hammes-Schiffer, *Curr. Opin. Struct. Biol.*, 2004, **14**, 192–201.
23. R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals,* McGraw-Hill, New York, 1965.
24. D. Chandler and P. G. Wolynes, *J. Chem. Phys.*, 1981, **74**, 4078–4095.
25. M. Sprik, M. L. Klein and D. Chandler, *Phys. Rev. B*, 1985, **31**, 4234–4244.
26. B. J. Berne and D. Thirumalai, *Ann. Rev. Phys. Chem.*, 1986, **37**, 401–424.
27. G. A. Voth, *Adv. Chem. Phys.*, 1996, **93**, 135–218.
28. N. Chakrabarti, T. Carrington Jr. and B. Roux, *Chem. Phys. Lett.*, 1998, **293**, 209–220.
29. D. Marx, M. E. Tuckerman and G. J. Martyna, *Comp. Phys. Commun.*, 1999, **118**, 166–184.
30. M. E. Tuckerman and D. Marx, *Phys. Rev. Lett.*, 2001, **86**, 4946–4949.
31. M. H. M. Olsson, P. E. M. Siegbahn and A. Warshel, *J. Biol. Inorg. Chem.*, 2004, **9**, 96–99.
32. J. Gao, K.-Y. Wong and D. T. Major, *J. Comput. Chem.*, 2008, **29**, 514–522.
33. J. K. Hwang, Z. T. Chu, A. Yadav and A. Warshel, *J. Phys. Chem.*, 1991, **95**, 8445–8448.

34. J. K. Hwang and A. Warshel, *J. Phys. Chem.*, 1993, **97**, 10053–10058.

35. J.-K. Hwang and A. Warshel, *J. Am. Chem. Soc.*, 1996, **118**, 11745–11751.

36. H. Kleinert, *Phys. Rev. D*, 1998, **57**, 2264.

37. W. Janke and H. Kleinert, *Phys. Rev. Lett*, 1995, **75**, 2787.

38. M. J. Gillan, *Phys. Rev. Lett.*, 1987, **58**, 563–566.

39. G. A. Voth, D. Chandler and W. H. Miller, *J. Chem. Phys.*, 1989, **91**, 7749–7760.

40. J. Cao and G. A. Voth, *J. Chem. Phys.*, 1994, **101**, 6184–6192.

41. J. N. Gehlen, D. Chandler, H. J. Kim and J. T. Hynes, *J. Phys. Chem.*, 1992, **96**, 1748–1753.

42. M. Messina, G. K. Schenter and B. C. Garrett, *J. Chem. Phys.*, 1993, **98**, 8525–8536.

43. G. A. Voth, *J. Phys. Chem.*, 1993, **97**, 8365–8377.

44. R. F. Grote and J. T. Hynes, *J. Chem. Phys.*, 1981, **74**, 4465–4475.

45. M. J. Gillan, *Philos. Mag. A*, 1988, **58**, 257–283.

46. D. E. Makarov and M. Topaler, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1995, **52**, 178–188.

47. M. Messina, G. K. Schenter and B. C. Garrett, *J. Chem. Phys.*, 1995, **103**, 3430–3435.

48. G. Mills, G. K. Schenter, D. E. Makarov and H. Jonsson, *Chem. Phys. Lett.*, 1997, **278**, 91–96.

49. S. Jang, G. A. Voth, *J. Chem. Phys.* 2000, **112**, 8747–8757. Erratum: 2001. **8114**, 1944.

50. I. Feierberg, V. Luzhkov and J. Aqvist, *J. Biol. Chem.*, 2000, **275**, 22657–22662.

51. J. Villa and A. Warshel, *J. Phys. Chem. B*, 2001, **105**, 7887–7907.

52. M. H. M. Olsson, P. E. M. Siegbahn and A. Warshel, *J. Am. Chem. Soc.*, 2004, **126**, 2820–2828.

53. E. L. Pollock and D. M. Ceperley, *Phys. Rev. B*, 1984, **30**, 2555–2568.

54. D. M. Ceperley, *Rev. Mod. Phys.*, 1995, **67**, 279–355.

55. D. T. Major, D. M. York and J. Gao, *J. Am. Chem. Soc.*, 2005, **127**, 16374–16375.

56. D. T. Major and J. Gao, *J. Am. Chem. Soc.*, 2006, **128**, 16345–16357.

57. R. P. Feynman and H. Kleinert, *Phys. Rev. A*, 1986, **34**, 5080.

58. R. Giachetti and V. Tognetti, *Phys. Rev. Lett.*, 1985, **55**, 912.

59. T. D. Hone, J. A. Poulsen, P. J. Rossky and D. E. Manolopoulos, *J. Phys. Chem. B.*, 2008, **112**, 294–300.

60. J. A. Poulsen, G. Nyman and P. J. Rossky, *J. Chem. Theory Comput.*, 2006, **2**, 1482–1491.

61. B. Palmieri and D. Ronis, *Phys. Rev. E*, 2006, **74**, 061105.

62. D. F. Coker, S. Bonella, *Springer Series in Chemical Physics*, 2007, **83**, 321.

63. A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, *et al.*: *J. Phys. Chem. B* 1998, **102**, 3586–3616.

64. J. W. Ponder and D. A. Case, *Adv. Protein Chem.*, 2003, **66**, 27–85.

65. J. Gao and X. Xia, *Science*, 1992, **258**, 631–635.

66. J. Gao, in *Rev. Comput. Chem.* ed. K. B. Lipkowitz, D. B. Boyd: VCH, 1995, 119–185. **vol 7**.

67. M. J. Field, P. Bash and A. M. Karplus, *J. Comput. Chem.*, 1990, **11**, 700–733.

68. J. Gao, *J. Phys. Chem.*, 1992, **96**, 537–540.

69. M. Freindorf and J. Gao, *J. Comput. Chem.*, 1996, **17**, 386–395.

70. I. H. Hillier, *Theochem.*, 1999, **463**, 45–52.

71. P. A. Bash, M. J. Field and M. Karplus, *J. Am. Chem. Soc.*, 1987, **109**, 8092–8094.

72. A. Warshel, *Ann. Rev. Biophys. Biomol. Struct.*, 2003, **32**, 425–443.

73. J. Gao, *J. Am. Chem. Soc.*, 1995, **117**, 8600–8607.

74. N. Wu, Y. Mo, J. Gao and E. F. Pai, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 2017–2022.

75. J. Gao, *ACS Symp. Ser.*, 1994, **569**, 8–21.

76. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**, 3902–3909.

77. J. J. P. Stewart, *J. Comp. Chem.*, 1989, **10**, 209–220.

78. M. Orozco, F. J. Luque, D. Habibollahzadeh and J. Gao, *J. Chem. Phys.*, 1995, **103**, 9112.

79. J. Gao, *Acc. Chem. Res.*, 1996, **29**, 298–305.

80. J. Gao, *J. Am. Chem. Soc.*, 1994, **116**, 9324–9328.

81. J. Gao and K. Byun, *Theor. Chem. Acc.*, 1997, **96**, 151–156.

82. M. Garcia-Viloca, D. G. Truhlar and J. Gao, *J. Mol. Biol.*, 2003, **327**, 549–560.

83. J. Aaqvist, M. Fothergill and A. Warshel, *J. Am. Chem. Soc.*, 1993, **115**, 631–635.

84. J. Gao and M. A. Thompson, *Combined Quantum Mechanical and Molecular Mechanical Methods,* **vol 712***,* American Chemical Society, Washington, DC, 1998.

85. A. J. Mulholland, *Theor. Comput. Chem.*, 2001, **9**, 597–653.

86. Y. Mo, Y. Zhang, J. Gao, *J. Am. Chem. Soc.* 1999, **121**, 5737–5742.

87. A. Warshel and R. M. Weiss, *Ann. N. Y. Acad. Sci.*, 1981, **367**, 370–382.

88. Y. Kim, J. C. Corchado, J. Villa, J. Xing and D. G. Truhlar, *J. Chem. Phys.*, 2000, **112**, 2718–2735.

89. T. V. Albu, J. C. Corchado and D. G. Truhlar, *J. Phys. Chem. A*, 2001, **105**, 8465–8487.

90. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

91. K. Nam, J. Gao and D. M. York, *J. Chem. Theory Comput.*, 2005, **1**, 2–13.

92. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, Oxford, 1987.

93. J. P. Valleau, G. M. Torrie, in *Modern Theoretical Chemistry,* vol 5, ed. B. J. Berne, Plenum, 1977, 169–194.

94. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187.

95. C. F. Baernascconi, *Acc. Chem. Res.*, 1992, **25**, 9.

96. F. G. Bordwell and W. J. Boyle Jr., *J. Am. Chem. Soc.*, 1972, **94**, 3907.

97. A. Kresge, *Can. J. Chem.*, 1974, **52**, 1897.

98. M. P. Valley and P. F. Fitzpatrick, *Biochemistry*, 2003, **42**, 5850–5856.

99. M. P. Valley and P. F. Fitzpatrick, *J. Am. Chem. Soc.*, 2004, **126**, 6244–6245.

100. M. Garcia-Viloca, C. Alhambra, D. G. Truhlar and J. Gao, *J. Chem. Phys.*, 2001, **114**, 9953–9958.

101. J. Bigeleisen, *J. Chem. Phys.*, 1955, **23**, 2264.

102. Huskey, in *Hydrogen-Transfer Reactions*, vol 4, ed. J. T. Hynes, J. P. Klinman, H. H. Limbach, R. L. Schowen, Wiley-VCH Verlag GmbH & Co. KgaA, 2006, 1285.

103. W. H. Saunders Jr, *J. Am. Chem. Soc.*, 1985, **107**, 164–169.

104. C. G. Swain, E. C. Stivers, J. F. Reuwer and L. J. Schaad, *J. Am. Chem. Soc.*, 1958, **80**, 558.

105. Y. Cha, C. J. Murray and J. P. Klinman, *Science*, 1989, **243**, 1325–1330.

106. M. A. Rishavy and W. W. Cleland, *Biochemistry*, 2000, **39**, 4569–4574.

107. W.-Y. Tsang and J. P. Richard, *J. Am. Chem. Soc.*, 2007, **129**, 10330–10331.

108. J. P. Richard and K. B. Williams, *J. Am. Chem. Soc.*, 2007, **129**, 6952–6961.