

© 2017

Haoyuan Chen

ALL RIGHTS RESERVED

MULTISCALE SIMULATION OF RNA CATALYSIS

By

Haoyuan Chen

A dissertation submitted to the
Graduate School—New Brunswick
Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Chemistry and Chemical Biology

Written under the direction of

Professor Darrin M. York

And approved by

New Brunswick, New Jersey

January, 2017

ABSTRACT OF THE DISSERTATION

Multiscale Simulation of RNA Catalysis

By Haoyuan Chen

Dissertation Director:

Professor Darrin M. York

Molecular simulations have become an essential part in modern chemistry research. Due to the wide variety and complex nature of practical chemistry problems, the multiscale simulation approach, which attempts to solve problems at different scales using different level of theories, has been widely employed. In this dissertation, several practical applications of multiscale simulation methods to the study of RNA catalysis as well as some method development work are presented. In Chapters 3 and 4, detailed investigations using quantum mechanical methods on the transition state structures in native and catalyzed model reactions that mimic RNA self-cleavage are presented, which serve as a baseline for the following mechanistic studies on real catalytic RNAs. Chapter 5 presents an example of a holistic approach on a mechanistic problem in a real catalytic RNA, which is the hammerhead ribozyme. Quantum mechanical methods, classical simulations and hybrid quantum-classical calculations are employed together to resolve the controversial role of residue G12 in hammerhead ribozyme. Finally, in Chapter 6, method developments on an advanced sampling method called the multi-dimensional replica exchange method is presented and applied to two prototypical problems in biochemistry. Taken together, the work presented here provide new insights into the molecular-level mechanism of RNA catalysis, which is of fundamental importance to our understanding of the underlying chemical principles in life.

Acknowledgments

First and foremost, I would like to thank my advisor, Professor Darrin M. York. As a talented theoretical chemist, he can always come up with cutting edge research projects and guide me to the right path of studying them. His wisdom, critical thinking and academic integrity make him an ideal role model from whom I can learn how to become a great scientist. He is always super supportive of my career, during which he has already helped me gathered several awards and secured a postdoc position with his letters of recommendation. Besides scientific research, his enthusiasm in integrating modern technology into traditional college teaching also inspires me a lot about what does a professor job really mean.

I would also like to thank my committee members Professor David A. Case, Professor Lu Wang and Professor Shantenu Jha for your help all along the way. I'm always more than respectful of the tremendous efforts led by Professor Case in developing a molecular simulation package which greatly benefited the entire scientific community. Professor Case is also very supportive whenever I need help from him. To Professor Wang, as compatriots, I fully understand how difficult it is for you as a foreigner to secure a faculty position in a major research university in United States and I'm motivated by your great success. To Professor Jha, although we have quite different backgrounds in science, the discussion with you is always so insightful and helps me a lot in thinking about interdisciplinary research. I'm also very grateful of the help from Professor Karsten Krogh-Jespersen, who used to be in my committee before his retirement. I've taken two of his courses and I really enjoyed them.

During my doctoral research, I've been involved in several collaborative projects with experimental chemists and I want to thank all of them for their great work and fruitful discussions with me. In particular, I want to thank Professor Michael E. Harris from Case Western Reserve University, Professor Joseph A. Piccirilli from University of Chicago and

Professor Barbara L. Golden from Purdue University. I also want to thank Antons Treikalis in Professor Shantenu Jha's group for the collaboration in the CDI project.

I feel privileged to pursue my Ph.D. in the York group, where I can work with a group of amazing people. The atmosphere in the group is really great and everyone is nice and glad to help others. In particular, I want to thank Dr. Tai-Sung Lee for helping me solve many problems in my research and giving me very helpful advice on my career, Dr. Timothy J. Giese for always being able to give me the best solution to my question every time, Dr. George M. Giambaşu for sharing his useful experiences in his career with me, Dr. Kin-Yiu Wong for inspiring me to think about scientific problems in innovative ways, Dr. Ming Huang for numerous help in both research and life, and Dr. Maria T. Panteva for always being so supportive in everything and introducing me to American culture.

When I first arrived Rutgers as an international student, I felt greatly welcomed by everybody in the department. So, I want to thank all the staff members who have helped me a lot, especially Melissa Grunberg, Janice Pawlo, Karen Fowler, Allison Larkin and Bill Abbott. I'm also grateful to be awarded the Rieman Award for outstanding teaching performance in 2014, the Reid Award for outstanding research accomplishments in 2016 and the Thomas Duff travel award in 2016, all from the chemistry department at Rutgers.

Last but not least, I always feel deeply indebted to my family. Without their tremendous love and support, I wouldn't be able to reach this point in my life. My mother Mrs. Jing Tang, besides introducing me to the world of chemistry as a high-school chemistry teacher, has been a wonderful mother in all aspects. She has brought me up almost all by herself and always gives me the best she could, with her incredible love for me. My wife Mrs. Jie Zhu, who is also a chemistry graduate student at Rutgers, always cares about me so much and tries to support my career. As long as you're by my side, I feel warm and blessed everyday.

Dedication

This dissertation is dedicated to my maternal grandmother Mrs. Qiufang Xu who passed away three months before my dissertation defense. She was a fantastic mother and grandma who was always sacrificing herself for her family. Although she lived through the most difficult time in contemporary China history, she was always optimistic about life and encouraging other people. I wouldn't been able to obtain my doctoral degree without her encouragements and support all the way along. May she rest in peace.

Table of Contents

Abstract	ii
Acknowledgments	iii
Dedication	v
List of Tables	viii
List of Figures	xii
1. Introduction	1
2. Background	3
2.1. Overview of Modern Molecular Simulations at Different Scales	3
2.2. Strategies in the Simulations of Enzymatic Reactions	23
2.3. General Principles in RNA Catalysis	33
3. Mechanistic Insights into RNA Transphosphorylation from Kinetic Iso- tope Effects and Linear Free Energy Relationships of Model Reactions . .	36
3.1. Introduction	37
3.2. Computational Methods	39
3.3. Results and Discussion	40
3.4. Conclusion	54
3.5. Supporting Information for: Mechanistic Insights into RNA Transphospho- rylation from Kinetic Isotope Effects and Linear Free Energy Relationships of Model Reactions	55

4. Effect of Zn²⁺ Binding and Enzyme Active Site on the Transition State for RNA 2'-O-transphosphorylation Interpreted Through Kinetic Isotope Effects	57
4.1. Introduction	58
4.2. Computational Methods	59
4.3. Results and Discussion	60
4.4. Conclusion	69
5. Metal Ion-induced Activation of the Catalytic General Base in Hammerhead Ribozyme Self-cleavage	71
5.1. Introduction	72
5.2. Computational Methods	76
5.3. Results and Discussion	80
5.4. Conclusion	91
5.5. Supporting Information for: Metal Ion-induced Activation of the Catalytic General Base in Hammerhead Ribozyme Self-cleavage	92
6. New Insights and Applications Enabled by Multi-dimensional Replica Exchange Molecular Dynamics Simulations	97
6.1. Introduction	97
6.2. Computational Methods	99
6.3. Results and Discussion	103
6.4. Conclusion	107
Bibliography	108

List of Tables

2.1.	Affordable system size and general accuracy of various <i>ab initio</i> quantum chemistry methods. Affordable system size is estimated assuming standard supercomputing resources are available. To achieve the listed accuracy, the basis set used in the calculations should be no worse than double-zeta with polarization functions.	14
2.2.	Summary of the suitable system sizes and application areas of molecular simulation methods at different scales.	23
3.1.	Comparison of KIE values for phosphate/phosphorothioate ester hydrolysis reactions. Most experimental results are from Ref. 1 while the $^{18}k_{\text{nb}}$ value for EtOpNPP ⁻ (0.9974) comes from Ref. 2. Structures and abbreviations of all reactants are shown in Figure 3.1. Different classes of mechanisms are illustrated in Figure 3.2. $^{18}k_{\text{nb}}$ values for the hydrolysis of (MeO) ₂ pNPPT are not applicable because there is no non-bridging oxygen in this reactant. Numbers in parentheses are the signed differences multiplied by 10 ⁴ . R is Pearson’s correlation coefficient, MSD and MAD stand for mean signed difference and mean absolute difference, respectively.	43
3.2.	Computed KIE values for LFER reactions and RNase A model reaction. Experimental p <i>K_a</i> s of the conjugate acids of different leaving groups are taken from IUPAC chemical data series (No. 23) ³ , except for HOCH ₂ CH ₂ OH and 2,3,5,6-F ₄ -C ₆ HOH, which are obtained from CRC Handbook ⁴ and Bourne <i>et al.</i> ⁵ , respectively. RNase A experimental KIEs were measured at 310.15K ⁶ instead of 298.15K for all other KIEs. Numbers in parentheses are the standard deviations for experimentally measured KIEs.	47

3.3. Key bond lengths (in Å) in the optimized transition state structures of phosphate/phosphorothioate ester hydrolysis reactions. R(P-Nu) and R(P-Lg) stand for the phosphorous-nucleophile oxygen and phosphorous-leaving group oxygen bond lengths, respectively.	55
3.4. Computed C2'-H and C5'-H H/D KIE values for the LFER reaction series and RNase A model reaction. Reaction coordinate ξ values in rate-limiting transition states are also shown here.	56
4.1. Comparison of calculated and experimental KIE values and the effect of catalysts for UpG dinucleotide 2'-O transphosphorylation model reactions in solution. All KIE values were measured/calculated at the temperature of 90° except for $^{18}k_{\text{NPO}}$ in the baseline models, where the only available experimental value was at 37° (from Ref. 7). Therefore, the corresponding calculations were also performed at 37°. $k^{\text{Cat}}/k^{\text{BL}}$ quantifies the effect of the catalyst on the KIEs, which is the ratio between KIEs in the catalyzed and uncatalyzed reactions, where the BL refers to the baseline model B1. . . .	62
4.2. Comparison of calculated KIEs and reaction coordinate ξ values in the TSs from models I to IX. Model numbering is the same as in Figure 4.2. Temperature is 90° for all calculations and experiment. Reaction coordinate ξ is defined as $\xi = R_2 - R_1$, where R_1 and R_2 are the P-O2' and P-O5' bond lengths, respectively. Percentage deviation %D is defined as $\%D = [(k^{\text{Cat}}/k^{\text{BL}})_{\text{Calc.}} - (k^{\text{Cat}}/k^{\text{BL}})_{\text{Expt.}}] \times 100\%$, where BL is the baseline model (B1) consistent with the series of Zn^{2+} calculations. The numbers in parentheses following the experimental values are the experimental uncertainties.	66

4.3. Comparison of calculated KIEs in RNase A catalysis using different methods. The AM1/d-PhoT QM/MM KIEs were calculated using the procedure described in the Computational Methods section while for the B3LYP QM/MM KIEs, the QM/MM equilibration and annealing steps were replaced by QM/MM geometry optimization due to the otherwise untractable computational cost. Basis set used in B3LYP QM/MM calculations (both geometry optimization and single point calculations) is 6-31+G(d). Temperature is 37° in all calculations.	69
5.1. Calculated and experimental pK _a shifts of guanine, chemically modified guanine and Mg ²⁺ -guanine complexes. The calculated pK _a shifts have been corrected for Mg ²⁺ -guanine binding affinity and Mg ²⁺ concentration (0.01 M) using Eq. 5.2. Error estimates are not available because the method used to obtain those values are deterministic and do not involve sampling.	83
5.2. Partial atomic charges of selected atoms in guanine and Mg ²⁺ -guanine complexes derived from NBO analysis.	84
5.3. Deprotonation free energies (kcal/mol), pK _a s and pK _a shifts (in pK _a units) of G12 in HHR from different sets of TI simulations and experiments.	84
5.4. Summary of predicted free energy costs in different steps for HHR:G12 and HHR · Mg ²⁺ :G12. (Relative) ΔG _{DEPR} values are taken from Table 5.3 while ΔG _{GBPT} values are from QM/MM free energy profiles as labeled in Figure 5.9. All numbers are in kcal/mol. Error estimates for ΔG _{DEPR} values are from Table 5.3 while for ΔG _{GBPT} are from bootstrapping (see supporting information for details). Standard errors of ΔG _{TOTAL} are obtained using the propagation rule.	90
5.5. Comparison of binding free energies (kcal/mol) between different divalent metal ions and guanine/6-thioguanine in both neutral and deprotonated forms. Numbers in the first column are converted from experimental binding affinities ^{8,9} . In each row, the numbers are normalized according to the corresponding experimental values in the first column.	91

5.6. M06-2X single point energies computed using different basis sets at the geometries optimized using M06-2X/6-31+G(d) level of theory. G and MG stands for guanine and Mg^{2+} -bound guanine, respectively. Absolute energies (column 2-5) are in hartree while relative energies (column 6) are in kcal/mol. 94

List of Figures

2.1.	(Left) Highest occupied molecular orbital (HOMO) and (right) lowest unoccupied molecular orbital (LUMO) of the benzene molecule from a HF/6-31G* calculation, which agrees with Hückel theory. This calculation only takes a few seconds on a desktop computer.	7
2.2.	Illustration of the “problem space” in enzymatic reactions.	24
2.3.	Comparison of the sampling efficiency between standard simulation and umbrella sampling in an analytical symmetric double-well potential $E = 10(x - 1)^2(x + 1)^2$. Standard simulation is 10^7 cycles of Monte-Carlo (MC) simulation starting at $x = -1.0$ while the umbrella sampling simulation has 10 windows with the centers of biasing potentials from -1.25 to 1.25 and each window is run for 10^6 cycles (so 10^7 cycles in total). With the same number of cycles, the standard MC is trapped in the local minimum while the umbrella sampling simulation manages to cross the barrier and give a reasonable global free energy profile.	26
2.4.	Illustration of the (temperature) replica exchange method. Circles in different colors represent different replicas.	27
2.5.	Illustration of the mechanism that metadynamics enhances the sampling.	29
2.6.	General scheme of the self-cleavage reaction in ribozymes. B and AH stand for the general/specific base and acid, respectively.	34
3.1.	List of structures and abbreviations of the 8 reactants in the phosphate ester hydrolysis reactions studied in this work. Nitrogen, bridging oxygen and non-bridging oxygen atoms which are colored in dark green, blue and red are used to calculate ^{15}k , $^{18}k_{\text{bridge}}$ and $^{18}k_{\text{nonbridge}}$ values, respectively.	40

3.2.	Illustration of different types of mechanisms in the phosphate ester hydrolysis reactions. A, B, C and D depict the mechanisms with respect to monoester dianionic, monoester monoanionic, diester and triester hydrolysis reactions. R = <i>p</i> -nitrobenzene, R' = methyl/ethyl, X = O/S.	41
3.3.	Correlations between computed KIEs using different methods [B3LYP & UFF (left), B3LYP & UAKS (middle) and M06-2X & UAKS (right)] and experimental KIEs in the 8 phosphate ester hydrolysis reactions. Circles in red, blue and green correspond to ^{15}k , $^{18}k_{\text{bridge}}$ and $^{18}k_{\text{nonbridge}}$, respectively.	44
3.4.	Illustration of the two types of mechanisms in the LFER series.	45
3.5.	(Top) Linear free energy relationships for early (TS1) and late (TS2) transition states in reverse alcoholysis of ethylene phosphate with different leaving groups. log <i>k</i> values are converted from the calculated reaction barrier and pK_a is the conjugate acid pK_a of the leaving group (data adopted from Ref. 10). The pK_a of the cross point can be used to determine whether the rate-limiting TS is early or late. (Bottom) Computed 2'-O and 5'-O KIE values for this set of reactions. Filled and unfilled symbols represent the values obtained from rate-limiting and non-rate-limiting transition states, respectively.	48
3.6.	Relationship between computed 2'-O and 5'-O KIEs and reaction coordinate ξ in the rate-limiting TSs in LFER model reactions and RNase A enzymatic model. The points for RNase A (in diamonds) are excluded in the linear fitting.	50
3.7.	(Top) Structure of RNase A transition state mimic, in which His12 stabilizes the negative charges on the non-bridging oxygens and His119 acts as a general acid to facilitate P-O5' bond cleavage. (Bottom) Reactant state (left) and transition state (right) structures in the model reaction. The two imidazole rings in the RS and TS structures are used to mimic His12 (left) and His119 (right). Key bond lengths (in Å) are labeled.	52

4.1. (Left) Schematic description of the reactant state and rate-limiting transition state in the RNA transphosphorylation reaction model. (Right) Illustration of three different Zn^{2+} binding sites (non-bridging oxygens in A, nucleophile oxygen in B, leaving group oxygen in C) and two interaction modes (direct coordination in A, B and C, indirect binding via a solvent water molecule in D).	59
4.2. TS structures of baseline models for un-catalyzed RNA transphosphorylation. Key bond lengths in Å are labeled.	61
4.3. TS structures located from all 9 Zn^{2+} binding models and comparison with the baseline model B1. Model IX matches best with experimental KIEs and has been highlighted. Key bond lengths in Å are labeled. All Zn^{2+} are saturated to hexacoordination by water but only key water molecules are shown for clarity.	64
4.4. Comparison of TS structures in baseline, Zn^{2+} -catalyzed and RNase A-catalyzed model reactions. The two imidazole rings in the RNase A model represents His12 (left) and His119 (right) residues in RNase A.	68
5.1. Illustration of the active site interactions in HHR and its self-cleavage mechanism. The N1 position of guanine in the putative general base G12 (blue) needs to be deprotonated before acting as a proton acceptor to deprotonate the 2'-OH in C17 (red), which will then act as the nucleophile to attack the phosphorous. Recent studies ^{11,12} indicate that there could be a Mg^{2+} directly bound at the Hoogsteen face of G12 ("G-site") to facilitate its deprotonation. Another Mg^{2+} is believed to play the role of activating the 2'-OH of the general acid G8 (green) by migrating from the binding site at N7 of G10.1 ("C-site") observed crystallographically into a bridging position ("B-site") with the scissile phosphate, in accord with thio/rescue effect experiments ^{13,14,15} . In this bridging position, the Mg^{2+} can coordinate the 2'-OH of G8, increasing its acidity, and facilitating proton transfer to the O5' leaving group in the general acid step of the reaction ¹⁶	74

5.2.	(Left) Canonical numbering of guanine nucleobase. (Right) Resonance structures of guanine deprotonated at N1 position with formal charge alternately on the N1 and the O6.	81
5.3.	Chemical structures of guanine and several chemically modified guanine molecules studied in this work. Experimental pK_a values at the N1 position (taken from Refs. 17,18) are shown.	81
5.4.	Optimized geometries of Mg^{2+} -guanine (A) and Mg^{2+} -deprotonated guanine (B) complexes . Selected bond lengths shown are in Å.	82
5.5.	Thermodynamic cycle used in TI calculations.	84
5.6.	Convergence of pK_a shift values from TI simulations. Filled squares connected by solid lines are the pK_a shift values evaluated using all available data at certain simulation time, with an increment of 100 ps per point. Dashed lines are drawn to help show the convergence.	85
5.7.	Representative active site conformation of the initial (left) and final (right) states from the free energy (TI) simulation of $HHR \cdot Mg^{2+}:G12$. The conformations from the simulation of $HHR:G12$ are very similar, and therefore not shown here. For clarity, water molecules and some other atoms/residues are not displayed. White, cyan, blue, red, pink and gold spheres stand for H, C, N, O, Mg and P atoms, respectively.	86
5.8.	Radial distribution function (RDF) of water oxygens around the O6 position of neutral and deprotonated G12 in HHR. Both sets of data are extracted from 1 ns of MD simulation of $HHR:G12$ with G12 in neutral and deprotonated forms, respectively. RDF data points are calculated using window size of 0.2 Å and are interpolated using the Akima spline.	87

5.9. Potential of mean force (PMF) for the general base proton transfer (GBPT) reaction in HHR · Mg ²⁺ :G12 (red) and HHR:G12 (black) generated by <i>ab initio</i> QM/MM umbrella sampling. The two chemical structures depict the reactant (left) and product (right) states. Reaction coordinate is the difference (R ₁ -R ₂) between the distance from C17:O2' to C17:HO2' (R ₁) and the distance from G12:N1 to C17:HO2' (R ₂). Free energies of the reactant and product states are marked.	89
5.10. Kinetic model used in the pK _a shift correction. pK _a values are directly converted from the free energy differences in DFT calculations while ΔG _{bind} is the experimentally measured binding affinity.	93
5.11. QM/MM free energy profile of the general base proton transfer using different basis sets.	95
5.12. All the free energy profiles generated in bootstrapping.	96
6.1. (Left) Illustration of the φ and ψ torsion angles and major conformers of alanine dipeptide. (Right) Decomposition of free energy into enthalpy and entropy terms. In all 3 subplots, the x and y axes correspond to φ and ψ torsion angles, respectively. All energies are in kcal/mol while each level in the contour corresponds to a 2 kcal/mol increment.	103
6.2. Free energy profile of alanine dipeptide backbone torsion at 6 different temperatures. In all 6 subplots, the x and y axes correspond to φ and ψ torsion angles, respectively. The range of energies is from 0 kcal/mol to 16 kcal/mol while each level in the contour corresponds to a 1 kcal/mol increment. . . .	104

6.3. (Top, side) Illustration of the DNA helix-junction-helix (HJH) system and the 3 pairs of marker thymine residues. (Top, middle) Averages and standard errors (from 4 runs) of distances between each pair of residues at different salt concentrations at 288 K. Dashed lines are obtained from least-square linear regression. Circles, triangles and squares represent the 2-28, 4-26 and 6-24 pairs, respectively. (Bottom) Distribution of the 3 different distances at different salt concentrations in 1 of the 4 runs. In all subplots, red, green, blue and orange colors correspond to the salt concentrations of 0.00 M, 0.05 M, 0.10 M and 0.15 M, respectively. 106

Chapter 1

Introduction

The focus of my doctoral research is the computational study of RNA catalysis using multiscale techniques which includes electronic structure calculations, classical molecular dynamics, hybrid quantum mechanical/molecular mechanical simulations and advanced free energy methods. Such a wide variety of theoretical tools are required because computer simulation of RNA catalysis is a complex task in which different types of results need to be interpreted/predicted using different level of theories. The ultimate goal is to take a holistic approach using multiscale simulations to truly bridge the gap between computational results and experimental observables, which will take the different complicating factors such as conformational switching, metal ion binding and acid-base equilibria into account. In this dissertation, the projects I've worked on during my doctoral research that attempt towards the ultimate goal will be presented.

In Chapter 2, a summary of the essential theoretical background of the state-of-the-art multiscale molecular simulation methods is provided and the specific strategies in simulating catalysis in biomacromolecules are discussed. In Chapter 3, which is partially reprinted with permission from Ref. 19, high-level electronic structure calculations on the non-enzymatic model reactions of RNA self-cleavage are performed to characterize the effects of different leaving groups on the reaction mechanism via kinetic isotope effects calculations. The results also correctly predict the relationship between leaving group acidity and transition state structure in the ribonuclease A enzyme. In Chapter 4 that is partially reprinted with permission from Ref. 20, the strategy is extended to predict the mechanism of Zn^{2+} -catalyzed non-enzymatic model reactions which could serve as a prototype of metal ion-facilitated RNA self-cleavage and shed light on the mechanism of many self-cleaving ribozymes. The predicted mechanism closely resembles the mechanism in ribonuclease A which also has

very similar kinetic isotope effects signatures as the Zn^{2+} -catalyzed reaction. In Chapter 5 which is partially reprinted with permission from Ref. 21, we move from non-enzymatic model reactions to a real RNA enzyme which is the hammerhead ribozyme. A wide variety of simulation methods has been applied to reconcile the inconsistencies between biochemical data and recent crystal structure on the activation mode of general base. Simulation results support the existence of an additional Mg^{2+} in the active site which is observed in the recent crystal structure and predict that the pK_a of the general base guanine to be shifted down by the additional Mg^{2+} to a value that is consistent with biochemical data. Finally, Chapter 6 presents the development and applications of a software framework that enables customizable and multi-dimensional replica exchange molecular dynamics simulations, which could be used to improve the description of complicated biophysical problems that are multi-dimensional in nature in molecular simulations.

Chapter 2

Background

2.1 Overview of Modern Molecular Simulations at Different Scales

Benefited from the rapid growth in the power of computers, computational modeling and simulations are now playing a huge role in almost all areas of science and engineering. For chemistry that primarily deal with molecules, a wide variety of molecular modeling and simulation methods have been developed. However, even within the field of chemistry, the scope of interesting applications is very diverse, from the hyperfine spectrum of astrochemical molecules which normally only have 2~10 atoms, to organic chemistry reaction mechanisms in solution that typically involves 20 to 100 atoms, to the folding process of a protein in solution that could have thousands of atoms and to the structures of lipid bilayers in the membrane that could even go up to hundreds of thousands of atoms. Problems at different scales require different level of accuracy and need to be solved with quite different approaches. In this section, we're going to briefly introduce and discuss the theoretical foundations and areas of suitable applications of the state-of-the-art molecular modeling/simulation techniques at different scales.

2.1.1 *Ab initio* Electronic Structure Theory: Atoms and Small Molecules

Chemistry is mostly about the behavior of electrons, which is governed by the laws of quantum mechanics. Unlike classical particles, the motion of quantum particles such as electrons are described by wavefunctions, which satisfy the Schrödinger equation

$$\hat{H}\Psi = i\hbar\frac{\partial\Psi}{\partial t} \quad (2.1)$$

In most chemistry problems, we could use the time-independent Schrödinger equation instead

$$\hat{H}\Psi = E\Psi \quad (2.2)$$

This equation, however, could be analytically solved *only* if there's only 1 electron in the system (H, He⁺, Li²⁺ etc.). For almost all chemical systems which are many-electron systems, approximations need to be made to solve the equation. Here, we'll introduce the most common quantum chemistry methods to solve this problem.

Hartree-Fock and Post-Hartree-Fock Methods

The Hartree-Fock (HF) theory, which solves the electronic Schrödinger equation of many-electron systems under the variational principle, is the foundation of most modern quantum chemistry methods²². To begin with, the electronic Schrödinger equation of electronic configuration $\{\mathbf{r}\}$ at a *fixed* set of nuclear coordinates $\{\mathbf{R}\}$ could be written as

$$\left[-\frac{1}{2} \sum_i \nabla_i^2 - \sum_{i,A} \frac{Z_A}{r_{iA}} + \sum_{A>B} \frac{Z_A Z_B}{R_{AB}} + \sum_{i>j} \frac{1}{r_{ij}} \right] \Psi(\mathbf{r}; \mathbf{R}) = E_{el} \Psi(\mathbf{r}; \mathbf{R}) \quad (2.3)$$

in which the four terms in the Hamiltonian on the l.h.s correspond to the electron kinetic energy, nuclear-electron potential energy, nuclear-nuclear potential energy and electron-electron potential energy, respectively. Note that the kinetic energy term for the nuclear is ignored here according to the Born-Oppenheimer approximation. The wave function $\Psi(\mathbf{r}; \mathbf{R})$ for a N-electron system could be written in the form of a Slater determinant

$$\Psi(\mathbf{r}; \mathbf{R}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \dots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \dots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \dots & \chi_N(\mathbf{x}_N) \end{vmatrix} \quad (2.4)$$

that satisfies the Pauli exclusion principle in which \mathbf{x}_i is the spin orbital (product of the spatial orbital and α/β spin function) of electron i. The Slater determinant is often written

in short as $|\chi_1\chi_2\dots\chi_N\rangle$. For the Hamiltonian, we further divide it into the one-electron operator

$$h(i) = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{r_{iA}} \quad (2.5)$$

and the two-electron operator

$$v(i, j) = \frac{1}{r_{ij}} \quad (2.6)$$

with which we could re-write the total electronic Hamiltonian as

$$\hat{H}_{el} = \sum_i h(i) + \sum_{i>j} v(i, j) + V_{NN} \quad (2.7)$$

in which V_{NN} is a constant (because $\{\mathbf{R}\}$ is fixed) therefore could be ignored. Now the electronic energy could be expressed as

$$E_{el} = \langle \Psi | \hat{H}_{el} | \Psi \rangle \quad (2.8)$$

and our goal is to find the optimal Ψ that minimizes E_{el} . Using a variational approach, we can obtain the Hartree-Fock equation

$$\left[h(\mathbf{x}_1) + \sum_j (J_j(\mathbf{x}_1) - K_j(\mathbf{x}_1)) \right] \chi_i(\mathbf{x}_1) = \epsilon_i \chi_i(\mathbf{x}_1) \quad (2.9)$$

where J_j is the Coulomb operator

$$J_j(\mathbf{x}_1) = \int d\mathbf{x}_2 |\chi_j(\mathbf{x}_2)|^2 r_{12}^{-1} \quad (2.10)$$

and K_j is the exchange operator

$$K_j(\mathbf{x}_1) \chi_i(\mathbf{x}_1) = \left[\int d\mathbf{x}_2 \chi_j^*(\mathbf{x}_2) r_{12}^{-1} \chi_i(\mathbf{x}_2) \right] \chi_j(\mathbf{x}_1) \quad (2.11)$$

The l.h.s of HF equation is also named as the Fock operator

$$f(\mathbf{x}_1) = h(\mathbf{x}_1) + \sum_j (J_j(\mathbf{x}_1) - K_j(\mathbf{x}_1)) \quad (2.12)$$

This scheme is, however, not easy to be solved by computers. Roothaan then introduced the matrix transformation form of HF equations which addressed the problem. First, we expand the spin orbitals as linear combinations of a set of basis functions $\tilde{\chi}$

$$\chi_i = \sum_{\mu}^K C_{\mu i} \tilde{\chi}_{\mu} \quad (2.13)$$

with which HF equation becomes

$$f(\mathbf{x}_1) \sum_{\nu} C_{\nu i} \tilde{\chi}_{\nu}(\mathbf{x}_1) = \epsilon_i \sum_{\nu} C_{\nu i} \tilde{\chi}_{\nu}(\mathbf{x}_1) \quad (2.14)$$

Multiplying both sides by $\tilde{\chi}_{\mu}^*(\mathbf{x}_1)$ and integrating, we get the Hartree-Fock-Roothaan (HFR) equation as

$$\sum_{\nu} F_{\mu\nu} C_{\nu i} = \epsilon_i \sum_{\nu} S_{\mu\nu} C_{\nu i} \quad (2.15)$$

where the elements of the Fock matrix \mathbf{F} is

$$F_{\mu\nu} = \int d\mathbf{x}_1 \tilde{\chi}_{\mu}^*(\mathbf{x}_1) f(\mathbf{x}_1) \tilde{\chi}_{\nu}(\mathbf{x}_1) \quad (2.16)$$

and the elements of the overlap matrix \mathbf{S} is

$$S_{\mu\nu} = \int d\mathbf{x}_1 \tilde{\chi}_{\mu}^*(\mathbf{x}_1) \tilde{\chi}_{\nu}(\mathbf{x}_1) \quad (2.17)$$

The HFR equation could also be written in a more compact form as

$$\mathbf{FC} = \mathbf{SC}\epsilon \quad (2.18)$$

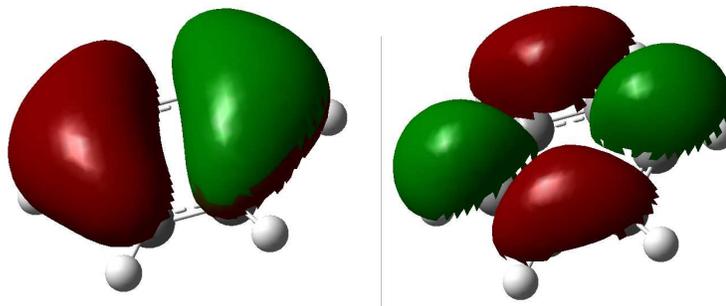


Figure 2.1: (Left) Highest occupied molecular orbital (HOMO) and (right) lowest unoccupied molecular orbital (LUMO) of the benzene molecule from a HF/6-31G* calculation, which agrees with Hückel theory. This calculation only takes a few seconds on a desktop computer.

This equation is equivalent to HF equation and solving this equation iteratively by starting with a initial guess of coefficient matrix \mathbf{C} and a set of atomistic basis functions (basis set) will yield the optimal \mathbf{C} that minimized E_{el} . Since this method is variational, the result will depend on the completeness of the basis set. With a complete basis set, the electronic energy E_{el} one could get by solving HF or HFR equation is called the Hartree-Fock limit. However, the Hartree-Fock limit is not the exact electronic energy because it ignores the electron correlation between opposite-spin electrons by nature, which leads to the underestimation of electron correlation energy. The electron correlation energy is defined as

$$E_{corr} = E_{exact} - E_{HFlimit} \quad (2.19)$$

which is usually about 1 % of the total E_{el} but is in the same order of magnitude as common chemical reaction energies. In practice, people did notice that HF had poor performances in predicting reaction energies and non-covalent interaction energies²³. Therefore, although HF calculations could now be routinely performed for relatively small (< 50 atoms) molecules even on personal desktop computers (see the Figure above for an example), the application of HF in solving real chemical problems is still fairly limited. A variety of HF-based methods that attempt to treat electron correlation properly have been developed

(often referred as the “post-HF” methods) and we’ll very briefly introduce some of them.

As stated above, the Hartree-Fock limit is still quite far away from the exact electronic energy, but that’s already the best we can get if just considering a single electronic configuration by using only one Slater determinant. In HF, only the ground state configuration is considered in the wavefunction expressed as a single Slater determinant. So, an immediate thought would be to also include excited states in the wavefunction, such as

$$\Psi = a_0\Psi_{HF} + \sum_i \sum_r^{occ. \ vir.} a_i^r \Psi_i^r + \sum_{i<j} \sum_{r<s}^{occ. \ vir.} a_{ij}^{rs} \Psi_{ij}^{rs} + \dots \quad (2.20)$$

in which indexes i, j go over all occupied (occ.) orbitals and indexes r, s go over all unoccupied/virtual (vir.) orbitals in the ground state. If we truncate after the third term in the equation above, we’ll have a wavefunction that contains the wavefunctions of the ground state, all single-excited states and all double-excited states and they’re all allowed to interact with each other. This is called the configuration interaction (CI) method. In this particular case, the method is called CISD where S and D stand for single-excited states and double-excited states, respectively. If triple-excited states are also included, it becomes CISDT. If all combinations of possible excited configurations are considered, we get the full-CI method which gives the exact electronic energy E_{exact} if using a complete basis set. However, the computational cost of full-CI becomes totally untractable even for molecules with only ~ 10 atoms. Therefore, full-CI is impractical for most chemical problems and is only used for extremely small molecules as benchmarks. Using the variational principle, the CI equation could be written as

$$\mathbf{Hc} = E\mathbf{Sc} \quad (2.21)$$

where

$$H_{ij} = \langle \Psi_i | \hat{H}_{el} | \Psi_j \rangle \quad (2.22)$$

$$S_{ij} = \langle \Psi_i | \Psi_j \rangle \quad (2.23)$$

Note that in CI, only the coefficients of all the *configuration* wavefunctions got variationally optimized, but not the coefficients of all the *spin orbitals* within each configuration wavefunction, which were optimized once using HF in the beginning of the CI calculation and were fixed in the remaining steps. Otherwise, it becomes the multi-configurational self-consistent field (MCSCF) method in which all the coefficients of the spin orbitals within each configuration wavefunction are also variational parameters. In general, CI methods could recover a significant portion ($> 80\%$) of the total electron correlation energy²² and in practice, CISD and CISD(T) (the triple-excited states are treated perturbatively) are the most commonly used truncated CI methods and could even be applied to relatively small molecules (< 50 atoms) if supercomputer is available. One problem of truncated CI methods is the size-consistency issue, which means in truncated CI, the energy of two non-interacting molecules calculated together is not equal to the sum of the energies of them calculated separately. Quadratic configuration interaction (QCI) methods have been developed to solve this issue.

Another class of post-HF methods is the Møller-Plesset (MP) perturbation theory, in which we first re-write the Hamiltonian as

$$\hat{H} = \hat{H}_0 + \lambda\hat{H}' \quad (2.24)$$

where the zeroth-order Hamiltonian \hat{H}_0 is the Fock operator and the perturbation term \hat{H}' is the difference between the real electron repulsion r_{12}^{-1} and \hat{H}_0 . By doing a Taylor expansion on the energy and wavefunction as

$$E_i = \sum_{j=0} \lambda^j E_i^{(j)} \quad (2.25)$$

$$|\Psi_i\rangle = \sum_{j=0} \lambda^j |\Psi_i^{(j)}\rangle \quad (2.26)$$

and by comparing coefficients, we could get the expressions for energy and wavefunction terms at all orders. The first three orders of energy terms are

$$E_0^{(0)} = \sum_i \epsilon_i \quad (2.27)$$

$$E_0^{(1)} = -\frac{1}{2} \sum_{ij} \langle ij||ij \rangle = E_0^{HF} - E_0^{(0)} \quad (2.28)$$

$$E_0^{(2)} = - \sum_{i < j, a < b} \frac{|\langle ab||ij \rangle|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \quad (2.29)$$

where $\langle ab||ij \rangle$ is defined as the difference between two two-electron integrals

$$\langle ab||ij \rangle = \langle ab|ij \rangle - \langle ab|ji \rangle \quad (2.30)$$

$$\langle ab|ij \rangle = \int d\mathbf{x}_1 d\mathbf{x}_2 \chi_a^*(\mathbf{x}_1) \chi_b^*(\mathbf{x}_2) r_{12}^{-1} \chi_i(\mathbf{x}_1) \chi_j(\mathbf{x}_2) \text{ (“physicists’ notation”)} \quad (2.31)$$

$$(ab|ij) = \int d\mathbf{x}_1 d\mathbf{x}_2 \chi_a^*(\mathbf{x}_1) \chi_b(\mathbf{x}_1) r_{12}^{-1} \chi_i^*(\mathbf{x}_2) \chi_j(\mathbf{x}_2) \text{ (“chemists’ notation”)} \quad (2.32)$$

The sum of the zeroth-order and first-order energies is equal to the HF energy, while the sum up to nth-order is called MPn energy. The most commonly used MP method is MP2, which could give results slightly worse than CISD (still much better than HF) but at a largely reduced computational cost. For molecules up to mid-size (< 100 atoms), MP2 could be routinely performed on supercomputers. MPn methods do not have size-consistency issue.

The last class of post-HF methods I’ll introduce here is the coupled cluster (CC) theory, in which the wavefunction is expressed using the cluster operator

$$|\Psi_{CC}\rangle = e^{\hat{T}} |\Psi_0\rangle \quad (2.33)$$

$$= \left(\sum_{n=0} \frac{1}{n!} \hat{T}^n \right) |\Psi_0\rangle \quad (2.34)$$

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \dots \quad (2.35)$$

in which \hat{T}_1, \hat{T}_2 are excitation operators

$$\hat{T}_1 = \sum_{ia} t_i^a |\Psi_i^a\rangle \quad (2.36)$$

$$\hat{T}_2 = \sum_{ijab} t_{ij}^{ab} |\Psi_{ij}^{ab}\rangle \quad (2.37)$$

If $\hat{T} = \hat{T}_1 + \hat{T}_2$, then the method is coupled cluster with single and double excitations (CCSD). By adding a perturbative \hat{T}_3 term, the method becomes CCSD(T), which is generally accepted as the “gold standard” of quantum chemical methods at present. CCSD and CCSD(T) could give results within chemical accuracy (energy error < 1 kcal/mol) in most cases but are generally limited to small molecules (< 30 atoms) even on supercomputers.

Density Functional Theory

An alternative approach to solve the electronic Schrödinger equation for many-electron systems is called the density functional theory (DFT). Hohenberg and Kohn found that, the ground-state properties of a many-electron system *only* depends on the electron density. Also, the correct electron density of the system is the one that minimizes the total energy through a functional. These two theorems are the theoretical foundations of DFT. According to the theorems, it’s clear that finding a functional that could calculate the correct total energy at given electron density is the key to solving the problem. As known from HF and post-HF methods, the hardest part is to calculate electron-electron interaction energies. Here, we start from a non-interacting system where the electrons don’t interact with each other and add on the corrections for electron-electron interactions. The basic form of the functional could be written as

$$E[\rho(\mathbf{r})] = E_{kin,ele.}[\rho(\mathbf{r})] + E_{pot,nuc.-ele.}[\rho(\mathbf{r})] + E_{pot,ele.-ele.}[\rho(\mathbf{r})] \quad (2.38)$$

$$= \sum_i^{ele.} (\langle \chi_i | -\frac{1}{2} \nabla_i^2 | \chi_i \rangle - \langle \chi_i | \sum_k^{nuc.} \frac{Z_k}{|\mathbf{r}_i - \mathbf{r}_k|} | \chi_i \rangle) + \sum_i^{ele.} \langle \chi_i | \frac{1}{2} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r}' | \chi_i \rangle + E_{xc}[\rho(\mathbf{r})] \quad (2.39)$$

The first two terms on the r.h.s. of the second line correspond to the first two terms on the r.h.s. of the first line, which are the kinetic energy of the non-interacting electrons and the potential energy of the nuclear-electron interactions, respectively. The electron-electron interaction energy is divided into the Coulomb term and the exchange-correlation term (the last two terms on the r.h.s. of the second line). The form of the exact exchange-correlation functional $E_{xc}[\rho(\mathbf{r})]$ is the only unknown in this equation, which includes not only the exchange and correlation interactions between electrons but also the correction

to the electron kinetic energy from the non-interacting system to the real system. One should also remind that the “orbitals” χ_i in the equation only represent the single-particle wavefunctions in the non-interacting system (“Kohn-Sham orbitals”) therefore do not have direct chemical meanings as the orbitals in HF-based methods do.

Now we know that all problems remain in the form of E_{xc} . To begin with, it seems reasonable to just assume that E_{xc} only depends on the local electron density. With this approximation which is called the local density approximation (LDA), the exchange part in E_{xc} could be solved exactly based on the homogeneous electron gas (HEG) model, while the correlation part could be interpolated between the two exact solutions of HEG model at weak-correlation and strong-correlation limits²⁴.

$$E_{xc,LDA} = E_{x,LDA} + E_{c,LDA} \quad (2.40)$$

$$E_{x,LDA} = \frac{3}{4} \left(\frac{3}{\pi}\right)^{1/3} \int (\rho(\mathbf{r}))^{4/3} d\mathbf{r} \quad (2.41)$$

For open-shell systems, spin polarization effects could be added to LDA to give the local spin density approximation (LSDA). The most significant limitation of LDA is that when the electron density is far from homogeneous (which is the case in most molecular systems), the LDA could not properly describe the effect of the changing electron density in the environment. Therefore, LDA is seldomly used in modern computational chemistry applications although is still a useful method in solid-state physics. To improve over LDA, one could make the functional depend not only on the density but also on the gradient of the density. This leads to the generalized gradient approximation (GGA) functionals, which generally has the form of

$$E_{xc,GGA}[\rho(\mathbf{r})] = E_{xc,LDA}[\rho(\mathbf{r})] + \Delta E_{xc} \left[\frac{|\nabla\rho(\mathbf{r})|}{\rho^{4/3}(\mathbf{r})} \right] \quad (2.42)$$

in which the gradient of the density $\nabla\rho(\mathbf{r})$ is taken into account in the last term. If the second derivative of the density is also included, it becomes meta-GGA functionals. GGA and meta-GGA functionals offer a significantly better performance in chemical systems than LDA and are still very useful at present. The most widely-used GGA/meta-GGA functional

is the PBE functional²⁵, which is very useful in solid-state physics and could also give decent performance in chemical systems. Another way to improve the accuracy of the functionals, first introduced by Becke, is to mix in the Hartree-Fock exchange energy (which is the exact exchange). Those functionals are called hybrid functionals, based on the form of

$$E_{xc} = (1 - a)E_{xc}^{DFT} + aE_x^{HF} \quad (2.43)$$

which is also called the “adiabatic connection method” (ACM). In practice, the mixing parameter(s) between DFT exchange-correlation and HF exchange are often optimized by fitting to experimental or high-level post-HF calculation data such as atomisation energies, vibrational spectra and reaction energies. Of all the hybrid functionals, it’s almost no doubt that the most popular one is the B3LYP functional^{26,27}, which mixes Becke’s 3-parameter functional and the LYP correlation functional

$$E_{xc,B3LYP} = (1 - a)E_x^{LSDA} + aE_x^{HF} + b\Delta E_x^B + (1 - c)E_c^{LSDA} + cE_c^{LYP} \quad (2.44)$$

where $a = 0.20$, $b = 0.72$ and $c = 0.81$. The a, b, c parameters are directly taken from the original Becke’s 3-parameter functional and not optimized again in B3LYP. Hybrid functionals are the main-stream DFT methods used in the state-of-the-art computational chemistry applications because they could perform almost as well as MP2 in many chemical systems and they’re a lot faster. Molecular systems up to 200 atoms could be routinely treated with hybrid DFT. Other hybrid functionals besides B3LYP are also very popular such as PBE0^{25,28} which is the hybrid version of PBE, and the Minnesota functionals developed by Truhlar and co-workers. One care must be taken in doing DFT calculations is that different functionals might be fitted to different types of systems therefore a functional that performs well for certain types of systems might not give good results in other types of systems. One should always check relevant literatures to make sure choosing the right functional for a given chemical problem.

So far, we’ve very briefly gone over the atlas of modern *ab initio* quantum chemistry methods. Now, back to the theme of this chapter and the entire dissertation, which

method(s) should one use given a specific system and its chemical problem? Here is a table (modified from Ref. 23) that gives some basic ideas. To summarize, with *ab initio* quantum chemistry, we could treat systems up to ~ 300 atoms which is enough for most chemical systems but not for biochemical systems. For general geometry optimizations and reaction energy predictions, hybrid DFT should be able to give good results as long as one chooses the right functional for the system. For highly accurate energetics on small molecules, one might need to run expensive calculations such as CCSD(T) to calculate the single point energies, based on the geometry optimized by DFT or MP2. There're of course lots of other methods and scenarios that we haven't mentioned, such as multi-reference methods which are often necessary for systems with transition metals or radicals. Also, a variety of software packages have been developed for running quantum chemical calculations, such as Gaussian, Q-Chem, GAMESS, MOPAC, Molpro and ORCA (just to name a few). Different software packages often specialize in different methods and might have implemented the same method differently, which is also important to be kept in mind before running a calculation.

Table 2.1: Affordable system size and general accuracy of various *ab initio* quantum chemistry methods. Affordable system size is estimated assuming standard supercomputing resources are available. To achieve the listed accuracy, the basis set used in the calculations should be no worse than double-zeta with polarization functions.

Method	Affordable system size	Geometry	Energetics
HF	< 300 atoms	Okay	Poor
GGA/meta-GGA	< 200 atoms	Okay	Okay
Hybrid	< 200 atoms	Good	Good
MP2	< 100 atoms	Good	Good
CCSD(T)	< 30 atoms	Impractical	Excellent

2.1.2 Semi-empirical Electronic Structure Methods: Balancing Accuracy and Performance

Methods Based on Neglect of Differential Overlap

Back when a standard HF calculation on a small molecule was still computationally expensive, there had been thoughts on speeding up HF calculations by making reasonable approximations. One of the first efforts were made by Pople and coworkers, in which they came up with the complete neglect of differential overlap (CNDO) method²⁹. The first approximation in CNDO is to ignore all core orbitals and use only 1 STO for each valence orbital (often called the “minimal basis set” approximation), which largely reduces the dimension of the Slater determinant. Then, an approximation is made on the evaluation of the two-electron integrals, which is the most time-consuming part in HF, as

$$(\mu\nu|\lambda\sigma) = \delta_{\mu\nu}\delta_{\lambda\sigma}(\mu\mu|\lambda\lambda) \quad (2.45)$$

which will ignore all the two-electron integrals except for the ones that have identical μ and ν orbitals and also identical λ and σ orbitals. This reduces the computational cost of the two-electron integrals evaluation from N^4 to N^2 (N is number of orbitals). Moreover, the surviving two-electron integrals could be pre-calculated as parameters because they only depend on the identities of the two atoms where the orbitals belong to (one for μ and ν , the other for λ and σ , could be the same one as well). The original parametrization scheme in CNDO is to use certain combinations of atomic ionization potentials (IPs) and electron affinities (EAs). The one-electron integrals in CNDO are also approximated by empirical parameters including IPs and EAs and other adjustable parameters, which makes the entire Fock matrix being pre-calculated values and once again significantly reduces the computational cost. All those approximations, although make CNDO orders of magnitude faster than HF, lead to significant errors in many chemical systems. For example, as the parameters in CNDO depend only on atoms (elements) but not on different types of atomic orbitals, the repulsions between the two electrons on carbon in singlet and triplet methylenes becomes identical²³, which is obviously against chemical intuition. Therefore, CNDO is not seen in any “real” chemical applications now. Nonetheless, based on the

spirit of CNDO, a class of NDO-based semi-empirical methods have been developed, some of which are still widely used for certain problems. Among them, the neglect of diatomic differential overlap (NDDO) based methods form the basis of most state-of-the-art semi-empirical methods. NDDO, which is an improved version over NDO, also keeps the two-electron integrals $(\mu\nu|\lambda\sigma)$ in which μ and ν are different orbitals but belong to the same atom (same for λ and σ). Although this modification introduces much more integrals than in NDO, NDDO is still much faster than HF in which all possible two-electron integrals need to be evaluated. Also, by keeping those integrals, the accuracies of NDDO-based methods are significantly improved over NDO. Methods such as AM1³⁰, PM3³¹, PM6³² as well as their dispersion-corrected forms could give reasonable geometries and energetics for most organic molecules²³ therefore are still routinely used in some chemical applications. Also, Truhlar and coworkers had proposed the specific reaction parameter (SRP) approach³³ in which a specific parameter set is derived for a specific type of reactions and could be used in predicting mechanisms and energetics of similar reactions. For example, the AM1/d-PhoT Hamiltonian³⁴ that was specially parametrized for phosphoryl transfer reactions has been used to study the transphosphorylation in RNA which is a key step in the self-cleavage of RNA enzymes.

Density Functional Tight Binding Methods

Besides the class of semi-empirical methods based on the approximation to HF theory, approximated DFT methods have also been developed, one of which is the density functional tight binding (DFTB) method³⁵. DFTB was first developed and used for solid state material systems and have been extended to biomolecular simulations³⁶. The derivation of DFTB starts from DFT, where a reference density is chosen as the sum of the electron densities on each atom at their unperturbed (neutral) states (thus the name “tight binding”) over the space. The “real” density is then expanded up to the second order with respect to the reference density in the exchange-correlation functional.

$$\begin{aligned}
E^{DFTB} = & \sum_i^{occ.} \langle \Psi_i | \hat{H}_0 | \Psi_i \rangle - \frac{1}{2} \iint \frac{\rho_0 \rho_0'}{|r - r'|} \\
& + \frac{1}{2} \iint \left(\frac{1}{|r - r'|} + \frac{E_{xc}}{\delta \rho \delta \rho'} \right) \delta \rho \delta \rho' + E_{xc}[\rho_0] - \int V_{xc}[\rho_0] \rho_0 + E_{cc} \quad (2.46)
\end{aligned}$$

where \hat{H}_0 is the Kohn-Sham Hamiltonian constructed using the reference density ρ_0 , Ψ_i are the Kohn-Sham orbitals, E_{xc} and V_{xc} are the exchange-correlation energy and potential, and E_{cc} is the core repulsion energy. By expressing Ψ_i in minimal basis set, adopting a monopole approximation to the second term and using the two-center approximation to the integrals, we get the self-consistent field density functional tight binding (SCC-DFTB) energy expression in the following form³⁷

$$E^{SCC-DFTB} = \sum_{i\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}^0 + \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta} \Delta q_\alpha \Delta q_\beta + \frac{1}{2} \sum_{\alpha\beta} U[R_{\alpha\beta}; \rho_0^\alpha; \rho_0^\beta] \quad (2.47)$$

where c^i are orbital coefficients, Δq are the Mulliken charges on each atom, $\gamma_{\alpha\beta}$ is the pairwise interaction kernel between two spherical charge densities on atoms α and β and U is a pairwise repulsive term that includes E_{cc} . The last two pairwise terms could be pre-calculated once by fitting to DFT results and tabulated for any future calculations therefore greatly reduces the computational cost. Indeed, DFTB is about as fast as NDDO methods²³ therefore could be applied to fairly large molecules that normal DFT cannot handle. New method developments such as DFTB3 which further expands the density to the third order³⁸ as well as new parameter sets such as 3OB³⁹ have greatly enhanced the performance of DFTB. For organic/bioorganic molecules, DFTB3/3OB has shown a comparable performance to GGA in many properties with a much less computational cost^{39,40}. The SRP approach has also been adopted for DFTB to describe certain types of reactions with higher accuracy, such as the SRP for phosphate hydrolysis⁴¹.

Approaches Towards Linear Scaling

Electronic structure methods scale at least N^3 where N is the total number of basis functions, due to the required matrix diagonalization procedure to get eigenvalues and eigenvectors.

Many methods have even sharper scaling such as HF (N^4 , due to the construction of two-electron integrals in Fock matrix), MPn (N^{3+n}) and CCSD(T) (N^7). This makes the calculations of very large molecules prohibitive even not considering the prefactor which is quite large for HF, post-HF and DFT methods. To address this problem, a wide variety of methods that attempt linear scaling have been developed, such as the divide-and-conquer (DC) method⁴² and its modified version (mDC)⁴³, the explicit polarization (X-Pol) method⁴⁴, the fragment molecular orbital (FMO) method⁴⁵ and the effective fragment potential (EFP) method⁴⁶. The common strategy among those methods is to divide the large system, either the molecule itself or its Slater determinant, into many small subsystems and do the self-consistent field calculations on all subsystems but not the large system itself. Therefore, the computational cost will only increase almost linearly as the system size increases. Combined with proper descriptions of inter-subsystem interactions, the electronic structure of the large system could be reasonably reproduced with a significantly reduced computational cost. For example, using the mDC method with SCC-DFTB Hamiltonian for each subsystem, a single point energy evaluation of a system consists of 3000 water molecules (9000 atoms) could be finished in *less than a minute* on a single CPU, while the energy evaluation on the same system using a full SCC-DFTB treatment takes *several days* using the exact same CPU⁴³. Those methods could be used as a semi-empirical quantum chemistry method for large molecules or a quantum mechanical force field (QMFF) for dynamics simulations of biomacromolecules.

2.1.3 Classical Molecular Dynamics Simulations: Macromolecules and Beyond

Electronic structure methods have gain major success in the description of molecules and chemical reactions. However, for macromolecules (both biomacromolecules and organic polymers) that could have thousands of atoms, the computational cost of electronic structure methods even at semi-empirical level is often beyond the computational power at present. Moreover, people are often times more interested in the structure and dynamics of the macromolecules rather than their electronic structures. Therefore, the classical molecular dynamics simulation methods have been developed which approximate the atoms as classical particles to largely increase the speed and could describe the dynamical behavior

by propagating the system using Newtonian dynamics. This method has now been widely used in the simulations of biological systems as well as materials and have been implemented in many software packages such as Amber, CHARMM, NAMD, Gromacs and LAMMPS (just to name a few).

All-Atom Molecular Dynamics

In classical molecular dynamics (MD), the dynamics of the system is purely governed by classical mechanics and follows the Newton's 2nd law

$$-\nabla V(\mathbf{x}(t)) = m \frac{d^2 \mathbf{x}(t)}{dt^2} \quad (2.48)$$

in which $V(\mathbf{x}(t))$ is the potential energy of the system at the configuration $\mathbf{x}(t)$ at time t . In principle $V(\mathbf{x}(t))$ could be calculated by electronic structure methods but in MD simulations of macromolecular systems, due to the requirement of speed, $V(\mathbf{x}(t))$ is often evaluated by an empirical molecular mechanical (MM) "force field" that approximates the interactions between atoms by some simple functions. The functional form of the AMBER force field⁴⁷, which is one of the most widely used force fields in simulating proteins and nucleic acids, is

$$V_{AMBER}(\mathbf{x}) = \sum_{bonds} \frac{1}{2} k_b (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{torsions} \sum_n \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] \\ + \sum_i \sum_{j>i} \left\{ \frac{q_i q_j}{r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right\} \quad (2.49)$$

in which the first three terms on the r.h.s. are called the bonded terms and the last terms are the non-bonded terms. In the bonded terms, the harmonic force constants (k_b , k_a and V_n) and the equilibrational lengths/angles (r_0 , θ_0 and γ) are fitted to experimental values and quantum chemical calculation results. The non-bonded terms consists of the Coulomb electrostatic energy term and the Lennard-Jones (LJ) potential that describes van der Waals interactions. The partial atomic charges q_i are determined by fitting to the quantum mechanical electrostatic potential surfaces and the LJ parameters ϵ_{ij} and σ_{ij} are

fitted in a similar way as the bonded parameters. To expand the parameter space for more accurate results, “atom types” are introduced to distinguish among same elements under different chemical environments, such as sp^2 carbons and sp^3 carbons. Other force fields for biomolecular simulations such as CHARMM, GROMOS and OPLS all have very similar functional forms as AMBER. Recently, a new class of polarizable force fields have been developed in which the atomic charges are allowed to change upon polarization to eliminate the artifacts caused by the fixed-charge approximations, such as the AMOEBA force field⁴⁸ and force fields based on the Drude oscillator model⁴⁹.

With the force field, we could now numerically propagate the system in the phase space according to Newton’s 2nd law. There’re many different propagation schemes, of which the basic one is the Verlet algorithm⁵⁰

$$\mathbf{x}(t + \Delta t) = 2\mathbf{x}(t) - \mathbf{x}(t - \Delta t) + -\frac{\nabla V(\mathbf{x}(t))}{m}(\Delta t)^2 \quad (2.50)$$

in which the timestep Δt is often set to 1 fs or 2 fs. Variants such as the velocity Verlet algorithm⁵¹ are also commonly used. To maintain the system in certain ensembles such as the canonical (NVT) ensemble and the isothermal-isobaric (NPT) ensemble, thermostats/barostats are also required to regulate system variables, which functions by directly scaling the system variables or coupling the system to external environment/additional degrees of freedom. Finally, since the simulation system is normally a periodic box in which the biomacromolecule is solvated by water molecules and ions, the particle mesh Ewald method^{52,53} is often used to describe the long-range electrostatic interactions which should not be ignored by simply setting a “cut-off” value because it only decays as r^{-1} . Implicit solvation models that eliminates explicit water molecules have also been developed, such as the Generalized Born model⁵⁴.

In a statistical mechanical perspective, MD simulations, similar to the stochastic Monte-Carlo algorithm, are actually sampling the phase space to evaluate this integral for the partition function Q

$$Q = \iint e^{-\frac{E(\mathbf{p}, \mathbf{q})}{k_B T}} d\mathbf{p}d\mathbf{q} \quad (2.51)$$

with which the ensemble average of a property A is

$$\langle A \rangle = \frac{1}{Q} \iint A(\mathbf{p}, \mathbf{q}) e^{-\frac{E(\mathbf{p}, \mathbf{q})}{k_B T}} d\mathbf{p} d\mathbf{q} \quad (2.52)$$

Instead of calculating the *space average* in Monte-Carlo, MD is attempting to evaluate the *time average*, which should be equivalent to the *space average* assuming ergodicity. In practice, however, reaching ergodicity in a simulation is difficult, given the typical biomolecular system in MD with tens of thousands of degrees of freedom, which has become a big bottleneck of MD. A rule of thumb is, with 100 nanoseconds of MD simulation, the system could only cross energy barriers less than $10 k_B T$ (~ 6 kcal/mol) (Dr. Bernard Brooks, personal communication), which is smaller than many energy barriers in biomacromolecular systems. Therefore, many advanced sampling methods have been developed for MD simulations, which will be introduced in later sections.

MD simulations have been widely used in studying (bio)macromolecular behaviors, such as protein/nucleic acids folding, protein structure predictions/refinements and protein-ligand interactions. Currently, all-atom MD simulations of systems with thousands to hundreds of thousands of atoms are routinely performed for nanoseconds to microseconds (which correspond to 10^6 to 10^9 steps using 1 femtosecond timestep) on supercomputers. Recently, the utilization of graphic processing units (GPUs) has given a significant raise to the speed of MD simulations. According to the Amber 16 GPU benchmarks (<http://ambermd.org/gpus/benchmarks.htm>), running MD on a single GPU could be about 10 times faster than running on 20 CPUs in parallel. Moreover, with deeply-optimized codes and special hardwares, even millisecond-scale MD simulations have been achieved⁵⁵. Along with the rapid development of advanced sampling methods, MD simulations are now capable of covering the timescales for a wide variety of biophysical processes. For those problems that are extremely large in size and also requires considerable amount of sampling, a more practical approach is to use coarse-grained methods, which we're going to introduce in the following paragraphs.

Coarse-Grained Molecular Dynamics

All-atom MD simulations have become a powerful tool in many biophysical problems. However, in some cases especially for large-scale systems, we're not particularly interested in the atomistic details. Therefore, to reduce the computational cost, one could ignore a large amount of unnecessary atomistic details by integrating atoms into groups ("beads") therefore reduce the system size, which is called coarse-grained molecular dynamics (CG-MD). In a typical CG approach for protein⁵⁶, each amino acid residue is represented by two beads—one for the backbone and one for the sidechain. Also, every four water molecules are clustered into one "water" bead while an ion together with its hydration shell is represented by an "ion" bead. Those approximations could reduce the total degrees of freedoms in the simulation system by ~ 10 -fold. Moreover, CG-MD allows using much longer timesteps than all-atom MD because the high-frequency motions are eliminated. In all-atom MD, the timestep is often limited to 1 fs or 2 fs because the fastest motions such as C-H vibration have periods of ~ 10 fs. In CG-MD, since those modes no longer exist explicitly, the timestep could be as long as 25~50 fs, which gives another ~ 25 -fold speed-up to the simulation. Overall, CG-MD simulations are often 3 orders of magnitude faster than all-atom MD, which allow the simulation of much larger systems than all-atom MD at similar timescale. Currently, CG schemes and parameters have been developed for lipids⁵⁷, proteins⁵⁶, nucleic acids⁵⁸ and other systems and have become a useful technique in simulating those macromolecules that have up to millions of atoms.

Now we've briefly gone over different approaches for different problems in modern molecular simulations. Of course, there're other types of methods we did not cover due to space limitations such as quantum dynamics simulations which are used to simulate charge transfer processes and nuclear quantum effects, but the majority of the state-of-the-art molecular simulation methods have been discussed and a concise summary of them are organized in the following table.

Table 2.2: Summary of the suitable system sizes and application areas of molecular simulation methods at different scales.

Method	System Size	Applications
<i>Ab initio</i> quantum chemistry	$10^0 \sim 10^2$ atoms	Geometries, accurate energies
Semi-empirical quantum chemistry	$10^2 \sim 10^3$ atoms	Geometries, preliminary energies
All-atom molecular dynamics	$10^3 \sim 10^5$ atoms	Structure and dynamics
Coarse-grained molecular dynamics	$10^5 \sim 10^6$ atoms	Dynamics in large timescales

2.2 Strategies in the Simulations of Enzymatic Reactions

Enzyme-catalyzed (bio)chemical reactions play vital roles in many biological processes. The computational studies have become a powerful tool to aid in the interpretation of experimental results and guide the design of new functional inhibitors/enzymes. Nonetheless, reliable and accurate simulations of enzymatic reactions are still very challenging at present. First, the total degrees of freedom in an enzyme system could easily go beyond hundreds of thousands, which requires considerable amount of sampling to generate a converged ensemble. The timescale for an enzymatic process might also be quite long if conformational changes are needed to organize the active site prior to the chemical step. Therefore, multiple sets of simulations with different initial conditions such as conformations, protonation states and metal ion binding scenarios are often required⁵⁹ (see the Figure below). Also, although the entire enzyme could be treated with fast molecular mechanical (MM) force fields, residues in the active site that are involved in the chemical reaction must be treated with quantum mechanics (QM) because MM force fields cannot describe bond breaking and forming. This requires multiscale simulation techniques and careful handling of interactions between QM and MM regions. In the following subsections, we'll briefly go over some important techniques and strategies in the simulation of enzymatic reactions.

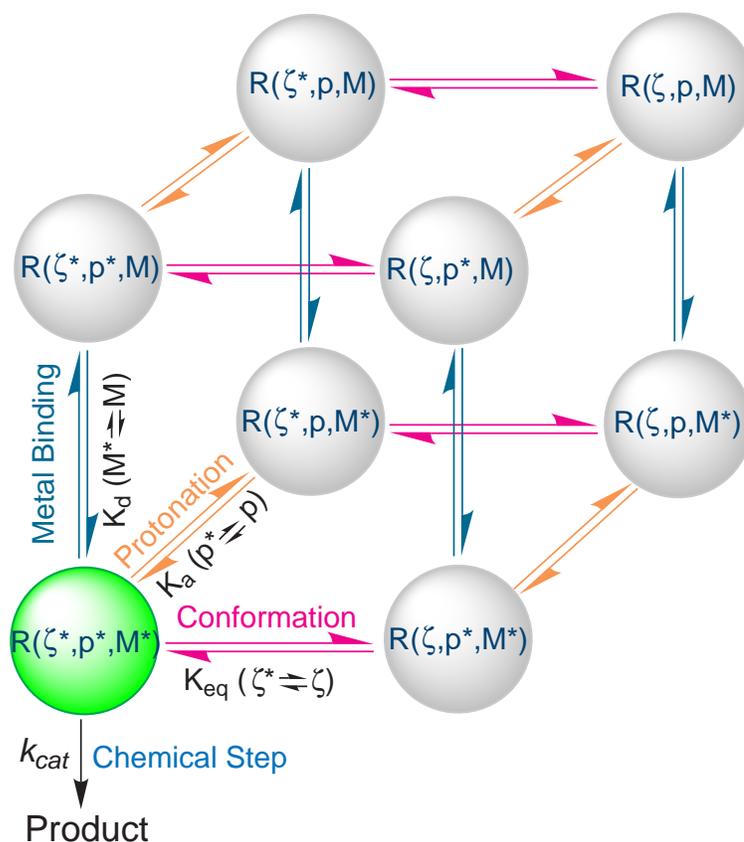


Figure 2.2: Illustration of the “problem space” in enzymatic reactions.

2.2.1 Employing Advanced Sampling Methods to Efficiently Explore Phase Space

As discussed above, one of the biggest bottlenecks of MD simulations is the sampling efficiency. Therefore, a wide variety of methods that aim at enhancing the sampling efficiency have been developed. The first method we’ll introduce here is the umbrella sampling (US) method, first proposed by Torrie and Valleau in 1977⁶⁰. The basic idea of US is, for the high-energy states that are hardly sampled by standard MD, we could add biasing potentials forcing the system to visit those states and get a reasonable amount of sampling on those states. After that, the biasing potentials need to be subtracted out from the resulting free energy profile (biased) to obtain the unbiased free energy profile. For example, if we’re interested in the free energy of a proton transfer reaction in an enzyme, we could define a “reaction coordinate” ξ which is the difference between donor-proton and acceptor-proton

bond lengths

$$\xi = R_{donor,proton} - R_{acceptor,proton} \quad (2.53)$$

Then, we define the harmonic biasing potential $U_{bias,i}(\xi)$ as

$$U_{bias,i}(\xi) = k_i(\xi - \xi_i)^2 \quad (2.54)$$

in which k_i is the force constant, ξ is the instantaneous reaction coordinate and ξ_i is the equilibrium value of reaction coordinate. Normally, a series of simulations will be performed in which each simulation (“window”) has its own biasing potential with different ξ_i value added to the total Hamiltonian (should be a hybrid QM/MM Hamiltonian, which will be introduced later) so all the high-energy areas could be covered and sampled. Each individual simulation will result in a biased potential of mean force (PMF) within a certain range of ξ

$$A'_i(\xi) = -k_B T \ln P'_i(\xi) - U_{bias,i}(\xi) + F_i \quad (2.55)$$

in which $A'_i(\xi)$, $P'_i(\xi)$ and F_i are the biased PMF in window i , the biased probability distribution in window i and a constant that appears because values in PMF are relative and depends on $U_{bias,i}(\xi)$. As the last step of analysis, we need to combine all those biased PMFs to generate the global, unbiased PMF. One of the analysis methods is the weighted histogram analysis method (WHAM)⁶¹, which bins the probability distribution into histograms and aims to find an optimal set of F_i values that give the best estimate of the global, unbiased probability distribution of ξ . The WHAM equations are

$$P(\xi) = \frac{\sum_i^{windows} n_i(\xi)}{\sum_i^{windows} N_i \exp([F_i - U_{bias,i}(\xi)]/k_B T)} \quad (2.56)$$

$$F_i = -k_B T \ln \left(\sum_j^{bins} P(\xi) \exp[-U_{bias,i}(\xi_j)/k_B T] \right) \quad (2.57)$$

in which $P(\xi)$ is the global, unbiased probability distribution which could be converted to the global, unbiased PMF and $n_i(\xi)$ is the counts in the histogram bin associated with

a certain ξ value. Because F_i and $P(\xi)$ depend on each other, the WHAM equations need to be solved iteratively. There're other methods that are also widely used such as the umbrella integration (UI) method⁶², the (multi-state) Bennett acceptance ratio ((M)BAR) method^{63,64} and the variational free energy profile (vFEP) method⁶⁵. The US method as well as the associated analysis methods could be extended to multiple dimensions.

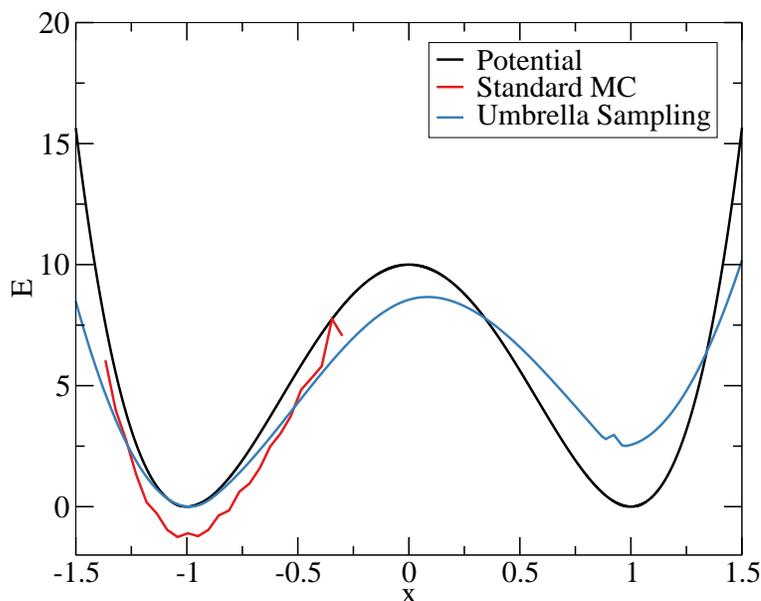


Figure 2.3: Comparison of the sampling efficiency between standard simulation and umbrella sampling in an analytical symmetric double-well potential $E = 10(x - 1)^2(x + 1)^2$. Standard simulation is 10^7 cycles of Monte-Carlo (MC) simulation starting at $x = -1.0$ while the umbrella sampling simulation has 10 windows with the centers of biasing potentials from -1.25 to 1.25 and each window is run for 10^6 cycles (so 10^7 cycles in total). With the same number of cycles, the standard MC is trapped in the local minimum while the umbrella sampling simulation manages to cross the barrier and give a reasonable global free energy profile.

Another approach to enhance the sampling in MD is the replica exchange molecular dynamics (REMD) method, which originates from the parallel tempering method⁶⁶ in Monte-Carlo simulations and was first applied to MD simulations by Sugita and Okamoto in 1999⁶⁷. Similar to US, REMD also employs a set of parallel simulations (“replicas”) but the difference is the replicas in REMD interact with each other while the windows in US

don't. The original and most common version of REMD is based on temperature exchange (T-REMD), in which several simulations are performed in parallel with exactly the same Hamiltonian but different temperatures. After a certain number of MD steps, replicas will attempt to exchange with others with a probability that satisfies detailed balance

$$P = \exp[(U_i - U_j)(\beta_i - \beta_j)] \quad (2.58)$$

in which U is the potential energy and $\beta = 1/k_B T$. Replicas with higher temperatures, which could overcome higher barriers more easily, could help the lower temperature replicas to cross those barriers through exchanges and explore the phase space faster. T-REMD is particularly useful in simulating protein folding⁶⁸ because the folding process normally could not be easily described by a few geometrical variables, which makes it difficult to be simulated using US. The exchange variables could also be pressure⁶⁹, pH⁷⁰ or geometrical biasing potentials (US-REMD)⁷¹, all of which need to adopt an exchange probability that satisfies detailed balance. The advantage of US-REMD over plain US is that the dimensions of phase space that are orthogonal to the reaction coordinate(s) could be better sampled through the exchanges in US-REMD, while in plain US some windows could be trapped in local minima in terms of those orthogonal degrees of freedom which might bias the global PMF. Multi-dimensional REMD that combines different exchange variables in one REMD simulation has also been developed^{72,73} and will be introduced in details in Chapter 6.

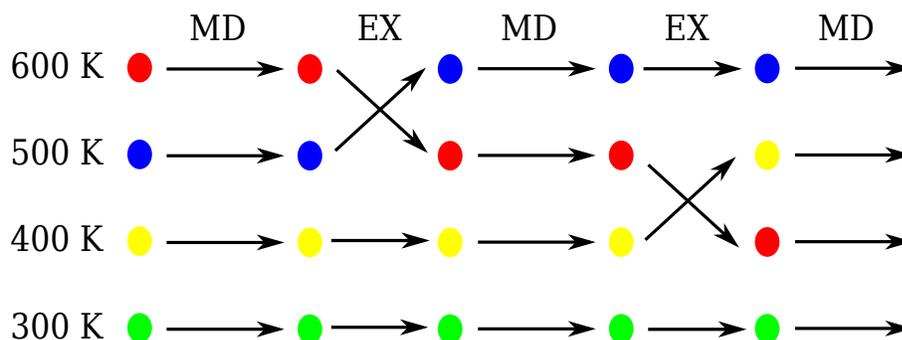


Figure 2.4: Illustration of the (temperature) replica exchange method. Circles in different colors represent different replicas.

The last advanced sampling method we'll discuss here is called "metadynamics" (MTD)⁷⁴,

which adopts a very different strategy from US and REMD. In MTD, with the selected reaction coordinates (also called “collective variables”), after every MD step, a small biasing potential that elevates the local potential energy surface is added to the total Hamiltonian. For a simple case with only one reaction coordinate, a Gaussian-type biasing potential takes the form

$$U_i(\xi) = k_i \exp[-(\xi - \xi_i)^2/\sigma^2] \quad (2.59)$$

in which k_i is the height of the Gaussian (might not be a constant, for example in the so-called “well-tempered” MTD where it decreases over time) and ξ_i is the current value of the reaction coordinate (note that MTD biasing potentials open downwards while US biasing potentials open upwards). In other words, every time the system visits a certain point, the potential energy at that point will be higher, trying to keep the system from visiting the same point again. At some point, the local minimum that the system was trapped in will be filled and flattened with the accumulating biasing potentials, which allows the system to overcome barriers and explore a larger portion of phase space. Eventually, the entire area of interest in the reaction coordinate space will be filled with biasing potentials and flattened, which will make the system evolving on a flat potential energy surface. When this is observed, the MTD simulation is considered as converged and if we do a summation over all the existing biasing potentials, we’ll get the free energy profile with the opposite sign.

$$A(\xi) = - \sum_i^{steps} U_i(\xi) \quad (2.60)$$

As shown in the figure below, MTD is often described metaphorically as “throwing sands into a valley”. Because MTD simulation depends on the history of itself, it could not be trivially parallelized among many CPUs like US and REMD, which limits the use and performance of this method.

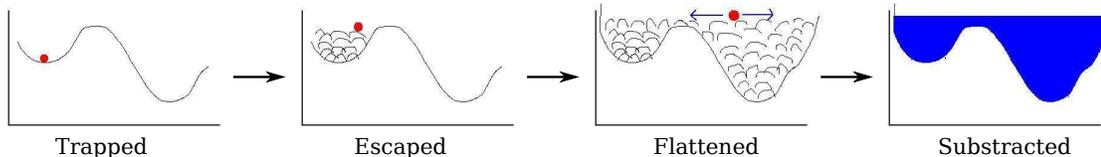


Figure 2.5: Illustration of the mechanism that metadynamics enhances the sampling.

Besides the three methods we discussed, many other advanced sampling methods have been developed, such as accelerated MD (AMD)⁷⁵, steered MD (SMD)⁷⁶, adaptive biasing force (ABF)⁷⁷, locally enhanced sampling (LES)⁷⁸ and Wang-Landau sampling⁷⁹, which have all been applied to solve interesting problems in biophysical chemistry.

2.2.2 Using Hybrid QM/MM Simulations and Alchemical Methods to Calculate Free Energies

The core of enzymatic processes is chemical reaction, which could not be properly described by molecular mechanical force fields. Since fully quantum mechanical treatment of an enzyme system is impractical, hybrid quantum mechanical/molecular mechanical (QM/MM) approaches have been developed, in which the atoms involved in the chemical step are treated quantum mechanically while the other “environmental” atoms are described by regular force fields.

$$H_{QM/MM} = H_{QM} + H_{MM} + H_{QM-MM} \quad (2.61)$$

There’re multiple ways to describe the interactions H_{QM-MM} between QM and MM atoms. For the simplest case, assume there’re no covalent bonds that go across the boundary between QM and MM regions, the interaction Hamiltonian could be written as

$$H_{QM-MM} = \sum_i^{QM} \sum_j^{MM} \left[\frac{q_i q_j}{r_{ij}} + 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) \right] \quad (2.62)$$

in which q_i are partial charges of QM atoms mapped from the wavefunction and the Lennard-Jones parameters σ_{ij} and r_{ij} only depend on the MM atom types of atoms i and j . If the MM region is allowed to polarize QM region, then

$$H_{QM-MM} = \sum_k^{QMele.MM} \sum_j \frac{q_j}{r_{kj}} + \sum_i^{QMnuc.MM} \sum_j \left[\frac{Z_i q_j}{r_{ij}} + 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) \right] \quad (2.63)$$

With this, the electrostatic term between MM atoms and QM electrons will appear in the Fock matrix and polarize the QM region. This approach has gained decent success in describing organic reactions in aqueous solution where the solvent is treated by empirical rigid water models such as TIP3P and the solute is treated by semi-empirical QM Hamiltonians such as AM1²³. However, in enzyme systems, the active site residues are covalently bonded to other residues, which means that there will be covalent bonds across the QM-MM boundary. Careful treatment of those bonds is needed and currently several methods have been developed for this, for example the link-atom approach⁸⁰ and the generalized hybrid orbital (GHO) method⁸¹. The link-atom approach works by inserting specially parametrized atoms on the QM-MM boundary between QM and MM atoms to hold and mimic the normal behavior of the covalent bonds. In GHO method, four sp^3 hybrid orbitals (specially parametrized as well) are assigned to each MM boundary atom and one of the four orbitals that points towards the QM atom that bonds to the MM atom will be included in the SCF calculation while other three will only interact with the QM region in a static fashion⁸².

Most modern molecular simulation packages have the QM/MM functionality in which different choices of QM Hamiltonian, MM force fields and QM-MM interaction schemes are available. The limiting factor of the speed of QM/MM simulations is the SCF calculations of the QM region. Currently, both *ab initio* and semi-empirical (including SRPs) QM Hamiltonians are widely used in the QM/MM simulations of enzymatic reactions. With a normal size QM region of about 50 atoms, nanosecond-scale QM/MM simulation with semi-empirical QM Hamiltonian is generally affordable with supercomputers. In practice, QM/MM is often combined with advanced sampling methods such as US to characterize the free energies along the reaction path.

Besides the chemical step, another class of problems that often draws interests in the study of enzymatic systems is the free energy cost when a mutation is introduced, which also requires accurate calculation of free energies. For example, in the drug discovery community, people are often interested in knowing how much the binding affinity of a small molecule

inhibitor to an enzyme would change when, for example, a hydrogen atom in the inhibitor was substituted by a hydroxyl group. Another example is the prediction of the pK_a of certain residues in an enzyme, which would be different from the pK_a of those residues in solution and could provide important insights into the catalytic mechanism of the enzyme. These calculations are difficult to do with standard MD or QM/MM due to the different force field parameters and limited phase space overlap between the starting and end states. Here we'll briefly introduce two of the so-called "alchemical methods" that aim to tackle this problem, which are the free energy perturbation (FEP) method and the thermodynamic integration (TI) method.

FEP was first derived by Zwanzig⁸³, in which the free energy difference between states 0 and 1 could be evaluated by only simulating in state 0

$$\Delta F_{10} = \beta^{-1} \ln \langle e^{-\beta(U_1(\mathbf{q}) - U_0(\mathbf{q}))} \rangle_0 \quad (2.64)$$

where U_1 and U_0 are the Hamiltonians for states 1 and 0, respectively and the ensemble average is in the simulation ensemble at state 0. This method is rigorously derived from statistical mechanics and does not imply any approximations. However, it's not an efficient method by itself. Research has shown that FEP calculations are very difficult to converge unless the difference between U_1 and U_0 is very small (about 1 kcal/mol)⁸⁴. Therefore, in practice, a series of intermediate states between the two end states need to be constructed and interpolated (their potentials are usually linear combinations of U_1 and U_0) and FEP calculations ("legs") need to be done between every pair of two neighbor states. The total free energy will then be the summation of the free energies of all legs. This has become a very useful method in evaluating alchemical free energies.

The TI method works in a similar and different way as FEP. In TI, a set of intermediate states are also constructed and interpolated (characterized by the variable λ) between the starting and end states. Simulations at each λ will be performed to collect the time series of the derivative $\frac{dU(\lambda, \mathbf{q})}{d\lambda}$ and the total free energy difference is evaluated by the following equation

$$\Delta F_{10} = \int_0^1 \left\langle \frac{dU(\lambda, \mathbf{q})}{d\lambda} \right\rangle_{\lambda} d\lambda \quad (2.65)$$

Unlike FEP, TI is not an exact solution to the free energy problem because of the discretized λ values. Errors are introduced during the numerical integration. TI has also become a common method in evaluating alchemical free energies. Both FEP and TI could be coupled with REMD to further enhance the sampling efficiency.

2.2.3 Calculating Kinetic Isotope Effects to Validate Proposed Mechanism

One more complicating factor in the mechanistic studies of enzyme catalysis is that in many cases, there could be more than one plausible reaction pathways based on the interpretation of experimental and computational results. One way to distinguish between different pathways is to simulate those pathways and compare the free energy barriers. However, due to the complexity of enzymatic reactions and limited accuracy of free energy calculations, pathway with predicted lowest energy barrier might not be the correct one. To further validate the proposed mechanism, one way is to measure the kinetic isotope effects (KIEs), which is the ratio of reaction rate constants for the reactions involving the lighter (L) and heavier (H) isotopomers.

$$\text{KIE} = \frac{k_L}{k_H} \quad (2.66)$$

KIE is a result of the difference in zero point energies between isotopomers. Differences in bond stiffness on certain atoms between the reactant and (rate-controlling) transition states lead to different vibrational energy levels (which are mass-dependent) associated with those atoms therefore make the activation energy and rate constant mass-dependent⁸⁵. Decreased or increased stiffness in the bonding environment around a certain atom in the transition state compared to the reactant state leads to a normal (greater than unity) or inverse (less than unity) KIE, respectively, when this atom is substituted by its heavier isotope. Therefore, KIE could directly reflect changes in the chemical environment around certain atoms from reactant to transition state and has been widely used as a probe of transition

state structures which are very hard to be determined experimentally. In computational research, KIE values could normally be calculated as a by-product of vibrational analysis on optimized reactant and transition state geometries. Under the harmonic approximation, KIE could be expressed as the Bigeleisen equation⁸⁵

$$\text{KIE} = \frac{\frac{\Omega_L^{TS}}{\Omega_H^{TS}} \prod_{i=1}^{3N-7} \left(\frac{\omega_L^{TS}}{\omega_H^{TS}} \right)}{\prod_{i=1}^{3N-6} \left(\frac{\omega_L^{RS}}{\omega_H^{RS}} \right)} \times \frac{\prod_{i=1}^{3N-7} \left(\frac{\sinh(\beta \hbar \omega_H^{TS}/2)}{\sinh(\beta \hbar \omega_L^{TS}/2)} \right)}{\prod_{i=1}^{3N-6} \left(\frac{\sinh(\beta \hbar \omega_H^{RS}/2)}{\sinh(\beta \hbar \omega_L^{RS}/2)} \right)} \quad (2.67)$$

in which Ω and ω are the imaginary (should only appear in transition state and should be only one) and real vibrational frequencies, respectively. A special case is the hydrogen/deuterium KIE, which requires more sophisticated treatments beyond the harmonic approximation such as path-integral methods because the nuclear quantum effects are significant and could not be ignored. In the scenario discussed above where multiple possible reaction pathways exist, one could simulate and compute KIEs on each pathway and compare to experimental KIE values to gain more insights into the reaction mechanism^{19,20}.

2.3 General Principles in RNA Catalysis

In 1989, Sidney Altman and Thomas Cech were awarded the Nobel prize in chemistry for their contribution in discovering the catalytic properties of RNA. RNA enzymes (ribozymes), like protein enzymes, could catalyze specific biochemical reactions. There is a class of ribozymes that catalyzes the cleavage of themselves such as the hammerhead ribozyme, the hairpin ribozyme and the twister ribozyme, many of which seem to share a common theme in their catalytic mechanisms^{86,59}. As shown in the figure below, prior to the main step which is a 2'-O-transphosphorylation, the 2'-OH nucleophile has to be activated (deprotonated) by a general or specific base via proton transfer. After (or concerted with) the transphosphorylation, a general or specific acid will donate a proton to the 5'-O leaving group which carries a -1 charge. The transphosphorylation step is normally the rate-controlling step.

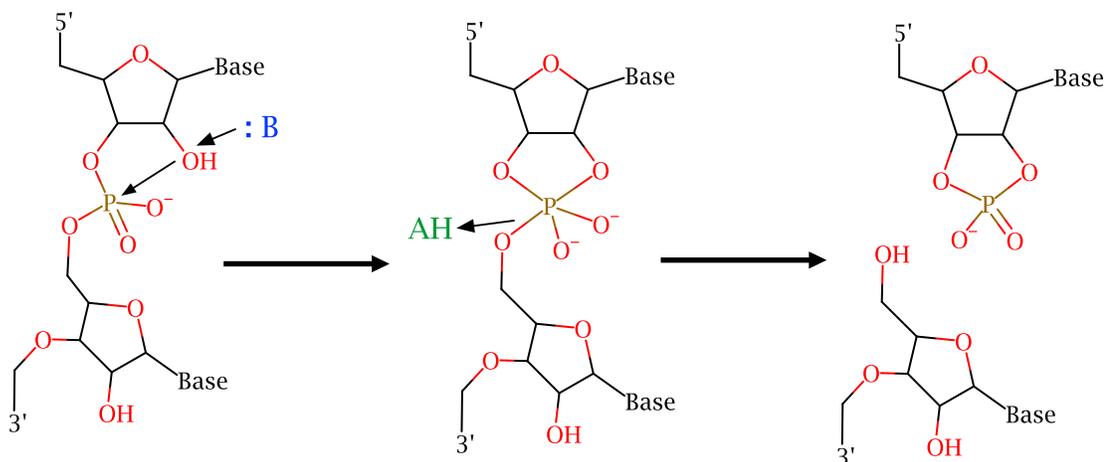


Figure 2.6: General scheme of the self-cleavage reaction in ribozymes. B and AH stand for the general/specific base and acid, respectively.

One of the most important factors in understanding the catalytic mechanism of a self-cleaving ribozyme is the identities of the general/specific acid and base¹⁸. Sometimes, a water molecule or even a hydroxyl anion will act as the specific base while a water molecule that coordinates to a metal ion (therefore has lower pK_a) could act as the specific acid. In many other cases, the nitrogen atoms in the heterocyclic nucleobases will play the roles of general acid/base. Adenine and cytosine have pK_a s as low as about 4 therefore could act as the general acid while guanine and uracil have pK_a s of 9~10 and could serve as the general base. Also, the ribozyme environment and metal ion binding could potentially tune the pK_a of acid/base towards ideal catalytic pK_a . Computational prediction of the pK_a and pK_a shifts in the ribozyme environment is very challenging but has been successfully applied to several ribozymes^{87,88,89,21}.

Another interesting feature in ribozyme catalysis is the stabilization of the negatively charged active site. Nucleic acids carry a lot of negative charges due to the phosphates therefore need to be held together by stabilizing factors such as metal ions. In the active site of self-cleaving ribozymes, especially in the transition state of the transphosphorylation step (see the figure above), the pentavalent phosphorane center has -2 net charge which is energetically highly unfavorable. Therefore, besides the hydrogen bonding network that stabilizes the negative charges, ribozymes often recruit metal ions in the active sites to

neutralize the electrostatic environment. Proper description of active site metal ion binding in molecular simulations of ribozymes is crucial because results such as conformations, pK_a and reaction energy barriers could change dramatically upon different metal ion binding scenarios. Traditional molecular mechanical force field parameters often fail in accurately describing the interactions between nucleic acids and divalent metal cations such as Mg^{2+} and Mn^{2+} ⁹⁰. By adding an extra attractive term in the Lennard-Jones potential⁹¹ and specific parametrising for the typical metal ion binding modes in nucleic acids⁹, the so-called “m12-6-4” model is able to properly describe the interactions between divalent metal ions and RNA⁹. For monovalent metal ions such as Na^+ and K^+ , a significant amount of sampling is required because they’re more mobile than divalents therefore might occupy different binding positions or even exchange with the diffusely bound cations in solution. Theoretical approaches such as the integral equation theory have been applied to study the distribution of monovalents in nucleic acids⁹².

Chapter 3

Mechanistic Insights into RNA Transphosphorylation from Kinetic Isotope Effects and Linear Free Energy Relationships of Model Reactions

Phosphoryl transfer reactions are ubiquitous in biology, and the understanding of the mechanisms whereby these reactions are catalyzed by protein and RNA enzymes is central to reveal design principles for new therapeutics. Two of the most powerful experimental probes of chemical mechanism involve the analysis of linear free energy relations (LFERs) and the measurement of kinetic isotope effects (KIEs). These experimental data report directly on differences in bonding between the ground state and the rate-controlling transition state, which is the most critical point along the reaction free energy pathway. However, interpretation of LFER and KIE data in terms of transition state structure and bonding optimally requires the use of theoretical models. In this work, we apply density-functional calculations to determine KIEs for a series of phosphoryl transfer reactions of direct relevance to the 2'-O-transphosphorylation that leads to cleavage of the phosphodiester backbone of RNA. We first examine a well-studied series of phosphate and phosphorothioate mono-, di- and triesters that are useful as mechanistic probes and for which KIEs have been measured. Close agreement is demonstrated between the calculated and measured KIEs, establishing the reliability of our quantum model calculations. Next, we examine a series of RNA transesterification model reactions with a wide range of leaving groups in order to provide a direct connection between observed Brønsted coefficients and KIEs with the structure and bonding in the transition state. These relations can be used for prediction or to aid in the interpretation of experimental data for similar non-enzymatic and enzymatic reactions. Finally, we apply these relations to RNA phosphoryl transfer catalyzed by ribonuclease A, and demonstrate the reaction coordinate-KIE correlation is reasonably preserved. A prediction

of the secondary deuterium KIE in this reaction is also provided. These results demonstrate the utility of building up knowledge of mechanism through the systematic study of model systems to provide insight into more complex biological systems such as phosphoryl transfer enzymes and ribozymes.

3.1 Introduction

The chemistry of phosphorus is central to many essential biological processes such as cell signaling, energy conversion, and gene regulation^{93,94,95,96}. Of interest here is the study of phosphoryl transfer reactions in RNA, and in particular, those reactions catalyzed by small prototype RNA and protein enzymes, including the hammerhead⁹⁷, hairpin⁹⁸, hepatitis delta virus⁹⁹, *glmS*¹⁰⁰ and Varkud satellite (VS)¹⁰¹ ribozymes and RNase A¹⁰². The mechanisms of phosphoryl transfer reactions, both in enzymatic and non-enzymatic systems, have been the focus of extensive experimental investigation. The comparison between non-enzymatic and enzymatic reaction mechanisms is essential since it reveals key information on the catalytic modes that enzymes achieve rate enhancement.

One important method used in those studies is linear free energy relationship (LFER) analysis, which quantifies the effect of changing the nucleophile or leaving group reactivity (via chemical modification) on the reaction rate^{103,2}. Brønsted coefficients¹⁰⁴ and Leffler indices¹⁰⁵ are valuable parameters that characterize the extent of bond formation/fission in the rate-controlling transition state (TS). Brønsted coefficients compare the effect of changes in nucleophile or leaving group reactivity (pK_a) on the reaction rate calibrated against the effect changing pK_a on reaction equilibria (β_{EQ}) in order to estimate charge development in the TS. In many instances LFER can be used to discern between pathways through the reaction free energy landscape^{106,107,5,108,109}. However, interpretation is limited by the accuracy of estimated β_{EQ} values, and can be complicated by effects on solvation and indirect effects due to differences in chemical structure².

Another widely used method in the mechanistic study is the measurement of kinetic isotope effects (KIEs). KIEs arise because heavier stable isotopes have lower zero point vibrational energies than their lighter counterparts. Differences in bond stiffness between

the ground state and TS result in differences in activation energy and consequently differences in rate constant (expressed as k_{light}/k_{heavy})^{85,110}. Decreased or increased stiffness in the bonding environment surrounding a certain atom in the TS compared to the reactant state leads to a normal (greater than unity) or inverse (less than unity) KIE, respectively, when this atom is substituted by its heavier isotope. Experiments that have been performed to measure KIEs in RNA transphosphorylation have greatly enhanced our understanding of the reaction mechanisms^{7,111,6}. However, observed KIEs necessarily reflect changes in *all* vibrational modes involving the substituted atom, including changes in protonation, reaction coordinate bonding and hybridization, which can make them difficult to interpret unambiguously. Thus, theoretical modeling is required in order to provide a detailed molecular-level interpretation of this data.

In the present work, we report results from quantum mechanical calculations of KIEs in a series of reactions that are closely related to RNA transphosphorylation. First, several computational methods for KIE prediction are tested for a set of benchmark phosphate/phosphorothioate ester hydrolysis reactions¹ which have well-established experimental results. Second, the validated method is applied to a series of RNA transphosphorylation model reactions in which LFERs have been calculated¹⁰ in order to form a quantitative connection with KIE data that can be used for prediction. Finally, model reactions that mimic the RNA phosphodiester backbone cleavage in solution and catalyzed by RNase A⁶ are studied. Experimental KIE results and coordinated computational simulations correspond well with the mechanistic predictions drawn from simulations of model reactions with different leaving group pKa. Observed KIEs for RNase A catalysis are generally consistent with the mechanistic signature for a late transition state, however, a significantly lower leaving group effect is observed that is attributable to stabilizing catalytic modes not present in the solution.

3.2 Computational Methods

3.2.1 Phosphate/phosphorothioate ester hydrolysis

DFT calculations were performed using both B3LYP^{27,26} and M06-2X¹¹² functionals to establish the appropriate level of theory. Reactant state (RS) and transition state (TS) geometries of the 8 phosphate ester hydrolysis reactions listed in Table 3 of Ref. 1 were optimized using those functionals with 6-31++G(d,p) basis set. PCM solvation model^{113,114} was used to address the solvent effects together with two sets of solute atomic radii, UFF¹¹⁵ and UAKS¹¹⁶. Harmonic vibrational analysis was performed to verify the nature of all stationary points. KIEs for those experimentally investigated isotopic substitutions in all 8 reactions were then computed using Bigeleisen equation^{117,85}. Temperatures in the KIE calculations were chosen to be consistent with experiments, where relevant.

3.2.2 LFER series

The reverse of dianionic in-line alcoholysis of ethylene phosphate was used in this work as a model for RNA phosphate transesterification reaction (see Figure 3.4), as was done in previous work^{111,10}. A series of 15 reactions with a wide spectrum of different 5'-O leaving groups (see Table 3.2) have been studied. The RS and TS geometries for all reactions were optimized using B3LYP/6-31++G(d,p) in PCM solvation with UAKS radii set. The B3LYP functional was chosen because the B3LYP results from the phosphate ester hydrolysis benchmark calculations described above show better agreement with experiment than those using M06-2X. Nucleophile oxygen (2'-O) and leaving group oxygen (5'-O) KIEs at 298.15K for all reactions were computed by the same method as described in the previous subsection.

3.2.3 RNase A model reactions

RS and TS geometries of the enzymatic model reaction which were optimized with B3LYP/6-31++G(d,p) in PCM implicit solvent using specialized atomic radii for RNase A catalysis, which we've adopted from previous work^{111,6}. An additional imidazole ring resembling His12 in RNase A, which was only used in TS in previous work 6, has been added to the

RS as well. KIEs of the 2'-O and 5'-O were calculated at 298.15K using the same protocol as described above. All electronic structure calculations were carried out in Gaussian 09 package¹¹⁸.

3.3 Results and Discussion

3.3.1 Validation and comparison of computed KIEs

Heavy-atom isotope effects, in most cases, have less than a few percent variation from unity¹¹⁰.

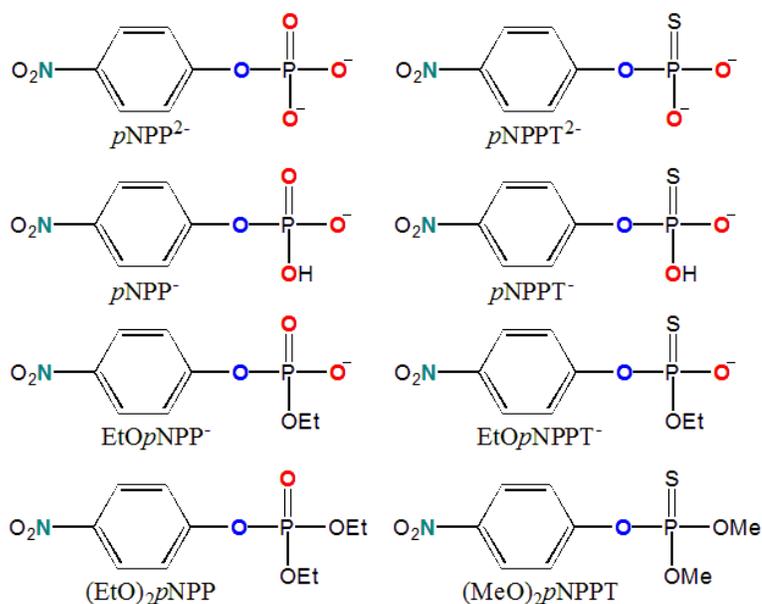


Figure 3.1: List of structures and abbreviations of the 8 reactants in the phosphate ester hydrolysis reactions studied in this work. Nitrogen, bridging oxygen and non-bridging oxygen atoms which are colored in dark green, blue and red are used to calculate ^{15}k , $^{18}k_{\text{bridge}}$ and $^{18}k_{\text{nonbridge}}$ values, respectively.

Therefore, it is important to establish a solid computational model which enables the reproduction and prediction of KIE values with satisfactory accuracy and precision. Here, we test three different methods on a related series of phosphate ester hydrolysis reactions¹ in which experimental KIEs have been well established. Figure 3.1 shows the structures of all 8 phosphate ester reactants and Figure 3.2 illustrates the different types of mechanisms

in the hydrolysis of those reactants. The KIEs calculated using the different computational protocols and their corresponding experimental values are listed in Table 3.1 and displayed in Figure 3.3.

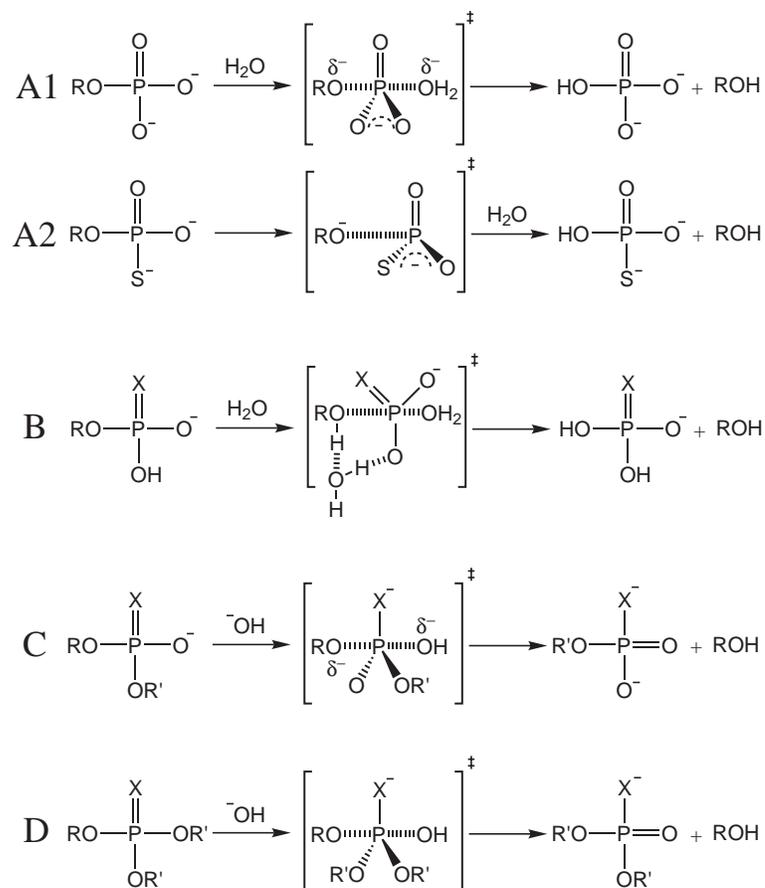


Figure 3.2: Illustration of different types of mechanisms in the phosphate ester hydrolysis reactions. A, B, C and D depict the mechanisms with respect to monoester dianionic, monoester monoanionic, diester and triester hydrolysis reactions. R = *p*-nitrobenzene, R' = methyl/ethyl, X = O/S.

In the hydrolysis reactions of monoester dianions (A1 & A2 in Figure 3.2), consistent with experiments, we obtain large normal KIEs on the bridging oxygens which suggests extensive cleavage of the phosphorous-bridging oxygen bond in TS^{110,119,120}. The KIEs and activation barriers of *p*NPP²⁻ hydrolysis has previously been investigated using DFT calculations¹²¹. The authors found excellent agreement with experiment upon inclusion of explicit water molecules in their solvation treatment¹²¹. We similarly found it necessary to

explicitly include a water molecule to locate the transition state, as illustrated in Figure 3.2 (A1). The larger $^{18}k_{\text{bridge}}$ value for the $p\text{NPPT}^{2-}$ than $p\text{NPP}^{2-}$ implies the former reaction has a larger degree of the P-O bond fission^{1,120}. The computed KIEs for these two reactions are in good agreement with experimental values in which B3LYP results are more consistent with experiment than M06-2X, while application of UAKS radii improves the correlation to experiment relative to UFF radii (Table 3.1).

For hydrolysis reactions of monoester monoanions (B in Figure 3.2) that occur in acidic conditions, experiments in Ref. 1 suggest an advanced but still incomplete proton transfer from the nonbridging oxygen to the bridging oxygen in the TS, which is supported by the normal $^{18}k_{\text{nonbridge}}$ values in our computational results. The $^{18}k_{\text{bridge}}$ values are significantly less normal than in dianionic reactions mainly due to the inverse isotope effect from the formation of the new O-H stretching mode^{122,1}. KIE values for these reactions given by the different computational models are all reasonably consistent with experiments (Table 3.1).

For alkaline hydrolysis of diesters and triesters (C & D in Figure 3.2), our calculations reveal an associative mechanism with a tight transition state which agree with various experimental and computational results^{123,124,1,108,125,2,126}. The less normal $^{18}k_{\text{bridge}}$ values compared to monoester dianionic cases confirm that leaving group bond fission is less advanced¹. The computed KIEs for all alkaline hydrolysis reactions show, in general, impressive correlation with experimental values.

Table 3.1: Comparison of KIE values for phosphate/phosphorothioate ester hydrolysis reactions. Most experimental results are from Ref. 1 while the $^{18}k_{\text{nb}}$ value for EtOpNPP^- (0.9974) comes from Ref. 2. Structures and abbreviations of all reactants are shown in Figure 3.1. Different classes of mechanisms are illustrated in Figure 3.2. $^{18}k_{\text{nb}}$ values for the hydrolysis of $(\text{MeO})_2\text{pNPPT}$ are not applicable because there is no non-bridging oxygen in this reactant. Numbers in parentheses are the signed differences multiplied by 10^4 . R is Pearson’s correlation coefficient, MSD and MAD stand for mean signed difference and mean absolute difference, respectively.

Reactant (Mechanism)	Sub.	B3LYP		M06-2X	Expt.
		UFF	UAKS	UAKS	
$p\text{NPP}^{2-}$ (95°C) (A1)	^{15}k	1.0033 (+5)	1.0050 (+22)	1.0049 (+21)	1.0028
	$^{18}k_{\text{br}}$	1.0229 (+40)	1.0122 (-67)	1.0142 (-47)	1.0189
	$^{18}k_{\text{nb}}$	0.9986 (-8)	0.9994 (0)	0.9906 (-88)	0.9994
$p\text{NPPT}^{2-}$ (50°C) (A2)	^{15}k	1.0047 (+20)	1.0045 (+18)	1.0099 (+72)	1.0027
	$^{18}k_{\text{br}}$	1.0273 (+36)	1.0246 (+9)	1.0352 (+115)	1.0237
	$^{18}k_{\text{nb}}$	1.0005 (-130)	1.0048 (-87)	0.9922 (-213)	1.0135
$p\text{NPP}^-$ (95°C) (B)	^{15}k	0.9998 (-6)	0.9997 (-7)	0.9996 (-8)	1.0004
	$^{18}k_{\text{br}}$	1.0008 (-79)	1.0004 (-83)	1.0016 (-71)	1.0087
	$^{18}k_{\text{nb}}$	1.0117 (-67)	1.0138 (-46)	1.0176 (-8)	1.0184
$p\text{NPPT}^-$ (30°C) (B)	^{15}k	1.0001 (-4)	1.0002 (-3)	0.9994 (-11)	1.0005
	$^{18}k_{\text{br}}$	1.0034 (-57)	1.0041 (-50)	1.0019 (-72)	1.0091
	$^{18}k_{\text{nb}}$	1.0077 (-144)	1.0100 (-121)	1.0153 (-68)	1.0221
EtOpNPP^- (95°C) (C)	^{15}k	1.0016 (+6)	1.0025 (+15)	1.0018 (+8)	1.0010
	$^{18}k_{\text{br}}$	1.0058 (+16)	1.0062 (+20)	1.0052 (+10)	1.0042
	$^{18}k_{\text{nb}}$	1.0007 (+33)	1.0010 (+36)	1.0018 (+44)	0.9974
EtOpNPPT^- (95°C) (C)	^{15}k	1.0014 (+4)	1.0019 (+9)	1.0015 (+5)	1.0010
	$^{18}k_{\text{br}}$	1.0033 (+13)	1.0031 (+11)	1.0037 (+17)	1.0020
	$^{18}k_{\text{nb}}$	1.0013 (-6)	1.0000 (-19)	1.0005 (-14)	1.0019
$(\text{EtO})_2\text{pNPP}$ (25°C) (D)	^{15}k	1.0010 (+3)	1.0014 (+7)	1.0012 (+5)	1.0007
	$^{18}k_{\text{br}}$	1.0037 (-23)	1.0035 (-25)	1.0033 (-27)	1.0060
	$^{18}k_{\text{nb}}$	1.0029 (-34)	1.0020 (-43)	0.9996 (-67)	1.0063
$(\text{MeO})_2\text{pNPPT}$ (30°C) (D)	^{15}k	1.0008 (+4)	1.0010 (+6)	1.0008 (+4)	1.0004
	$^{18}k_{\text{br}}$	1.0018 (-27)	1.0018 (-27)	1.0020 (-25)	1.0045

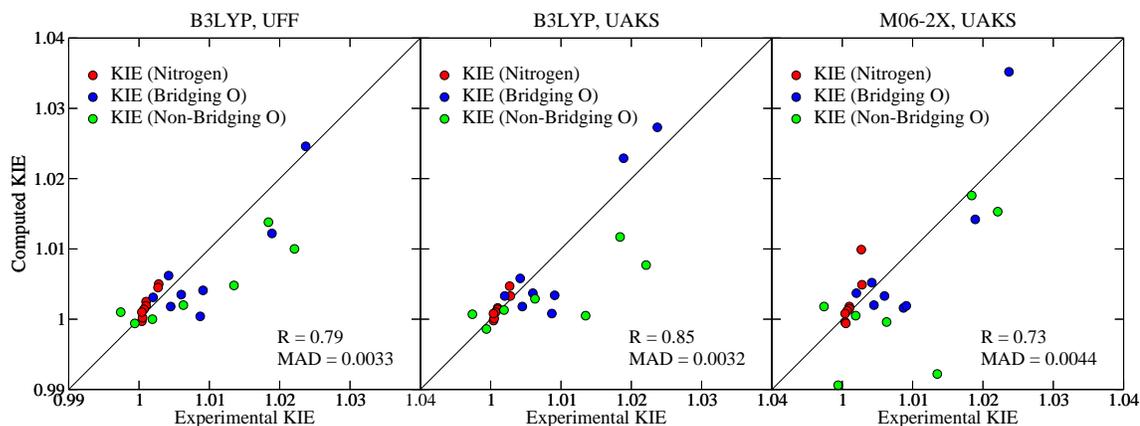


Figure 3.3: Correlations between computed KIEs using different methods [B3LYP & UFF (left), B3LYP & UAKS (middle) and M06-2X & UAKS (right)] and experimental KIEs in the 8 phosphate ester hydrolysis reactions. Circles in red, blue and green correspond to ^{15}k , $^{18}k_{\text{bridge}}$ and $^{18}k_{\text{nonbridge}}$, respectively.

Overall, both B3LYP and M06-2X functional give reasonable predictions of KIE values, but the B3LYP results are more consistent with experimental data (Figure 3.3, correlation coefficient 0.85 vs. 0.73, MAD 0.0032 vs. 0.0044). The KIE results using UAKS radii generally outperform those using UFF radii (correlation coefficient 0.85 vs. 0.79). Therefore, we choose to use B3LYP density functional with 6-31++G(d,p) basis set and UAKS atomic radii set for the computational model, which will be applied here after to the studies on RNA transphosphorylation model reactions.

3.3.2 KIEs and LFER in RNA transphosphorylation model reaction series

All of the model reactions studied here are initiated by attack of the 2'-O nucleophile on the phosphoryl group resulting in a pentavalent phosphorane species. There are two associative mechanisms as shown in Figure 3.4: a concerted mechanism that passes through a single transition state, and a stepwise mechanism that proceeds via two transition states separated by an intermediate.

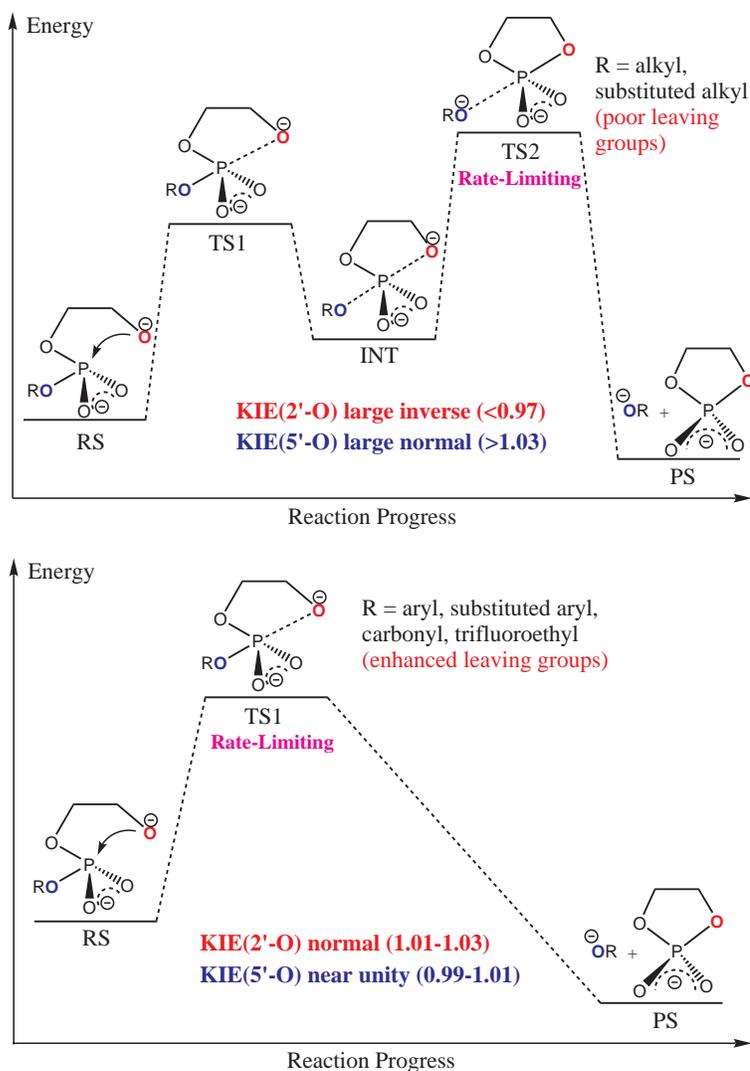


Figure 3.4: Illustration of the two types of mechanisms in the LFER series.

A concerted mechanism can be classified as synchronous (having similar degrees of bonding to the nucleophile and leaving group in the transition state) or asynchronous (having differing degrees of bonding to the nucleophile and leaving group in the transition state). For either stepwise mechanisms or concerted asynchronous mechanisms, the transition states can be further designated as either “early” or “late”, depending on the location of the transition state along a reaction coordinate ξ that involves the difference in the leaving group (R_2) and nucleophile (R_1) distances with the reactive phosphorus:

$$\xi = R_2 - R_1 \quad (3.1)$$

Specifically, we denote a transition state as being “early” ($\xi < 0$) if it is characterized by a small degree of bond formation with the nucleophile (large R_2 value) and cleavage with the leaving group (small R_1 value). Conversely, we denote a transition state as being “late” ($\xi > 0$) if it involves a nearly fully formed bond with the nucleophile (small R_2 value) and a nearly cleaved bond with the leaving group (large R_1 value). Fitting parameters for Pauling’s model^{127,128}, which relates bond orders and bond lengths by an exponential model, have been established specifically for this reaction series in Ref. 10.

Table 3.2: Computed KIE values for LFER reactions and RNase A model reaction. Experimental pK_a s of the conjugate acids of different leaving groups are taken from IUPAC chemical data series (No. 23)³, except for HOCH₂CH₂OH and 2,3,5,6-F₄-C₆HOH, which are obtained from CRC Handbook⁴ and Bourne *et al.*⁵, respectively. RNase A experimental KIEs were measured at 310.15K⁶ instead of 298.15K for all other KIEs. Numbers in parentheses are the standard deviations for experimentally measured KIEs.

Leaving Group	Expt. pK_a	ξ (Å)	KIE(2'-O)	KIE(5'-O)
CH ₃ COO ⁻	4.46	-0.62	1.0256	0.9999
2,3,5,6-F ₄ -C ₆ HO ⁻	5.53	-0.57	1.0214	1.0090
4-NO ₂ -C ₆ H ₄ O ⁻	7.14	-0.65	1.0250	1.0032
4-CN-C ₆ H ₄ O ⁻	7.95	-0.60	1.0239	1.0032
3-CN-C ₆ H ₄ O ⁻	8.61	-0.56	1.0218	1.0036
C ₆ H ₅ O ⁻	9.95	-0.52	1.0233	1.0046
CF ₃ CH ₂ O ⁻	12.37	-0.44	1.0203	1.0057
HCCCH ₂ O ⁻	13.55	0.48	0.9672	1.0454
FCH ₂ CH ₂ O ⁻	14.2	0.54	0.9673	1.0481
ClCH ₂ CH ₂ O ⁻	14.31	0.49	0.9690	1.0451
HOCH ₂ CH ₂ O ⁻	15.07	0.48	0.9676	1.0441
CH ₃ O ⁻	15.54	0.63	0.9666	1.0649
CH ₃ CH ₂ O ⁻	16	0.60	0.9678	1.0489
CH ₃ CH ₂ CH ₂ O ⁻	16.1	0.59	0.9676	1.0465
CH ₃ CH ₃ CHO ⁻	17.1	0.55	0.9683	1.0475
RNase A Model		0.14	0.9973	1.0272
RNase A Expt.		N/A	0.994(2)	1.014(3)

Table 3.2 lists the computed 2'-O and 5'-O KIE values for those reactions as well as the key geometrical information in the rate-limiting TSs and the experimental pK_a s of the conjugate acid of those 15 different leaving groups. Figure 3.5 demonstrates the connection between LFER and KIEs in the characterization of the two classes of mechanisms. Figure 6 shows the correlation between the approximate reaction coordinate $\xi = R_2 - R_1$ (where in

RNA numbering R_1 and R_2 stand for the P-O2' and P-O5' bond lengths, respectively) and the KIE values, which will be addressed again in the later discussion on RNase A enzymatic models.

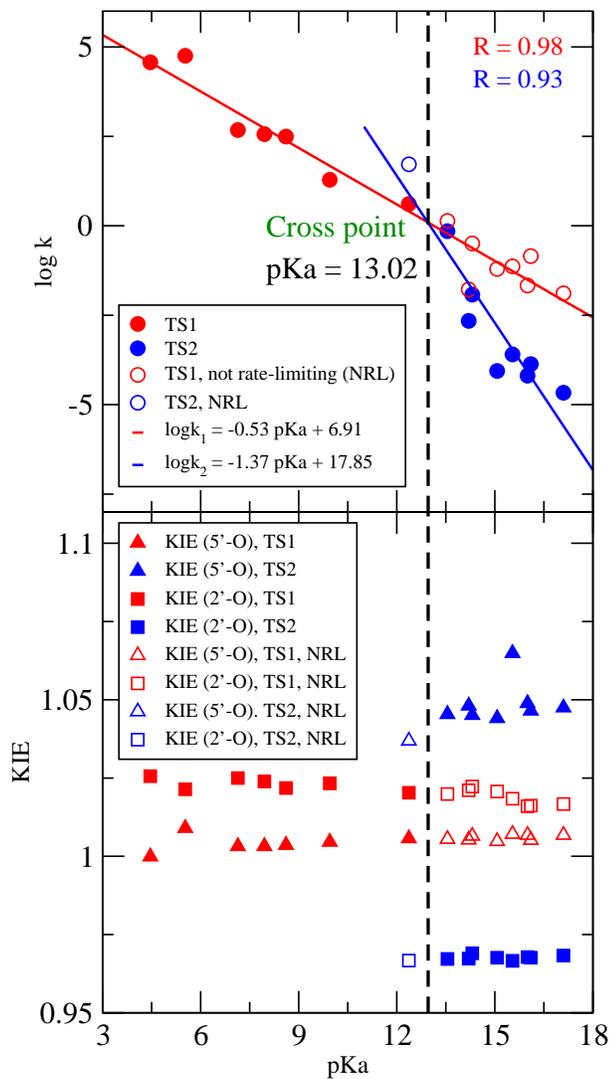


Figure 3.5: (Top) Linear free energy relationships for early (TS1) and late (TS2) transition states in reverse alcoholysis of ethylene phosphate with different leaving groups. $\log k$ values are converted from the calculated reaction barrier and pK_a is the conjugate acid pK_a of the leaving group (data adopted from Ref. 10). The pK_a of the cross point can be used to determine whether the rate-limiting TS is early or late. (Bottom) Computed 2'-O and 5'-O KIE values for this set of reactions. Filled and unfilled symbols represent the values obtained from rate-limiting and non-rate-limiting transition states, respectively.

As shown in Figure 3.5, Figure 3.6 and Table 3.2, two distinct groups of KIE values for both 2'-O and 5'-O clearly exist, which correspond to the two types of reaction mechanisms in Figure 3.4. For the reactants with poor leaving groups (conjugate acids have relatively high pK_a values, greater than ≈ 13), significantly large normal 5'-O KIEs ($k_5 > 1.03$) are observed together with the large inverse 2'-O KIEs ($k_2 < 0.97$). These numbers suggest that the P-O2' bonds are almost fully cleaved in the rate-limiting TSs while the P-O5' bonds are nearly fully formed. This is also demonstrated by the positive ξ values around 0.55 Å in the rate-limiting TSs (TS2). Early TSs (TS1) can also be located for these reactions but should not be used in the KIE predictions since they are not rate-controlling¹⁰. Phosphoryl transfer reactions for this type of reactants therefore occur via stepwise mechanisms with late rate-limiting TSs, which agrees with previous experimental and computational studies on closely related systems^{7,111,129}. As for those reactants with enhanced leaving groups (conjugate acids have lower pK_a values, less than ≈ 13), the 2'-O KIEs are always large normal while the 5'-O KIEs are mostly normal but much closer to unity. This indicates an early rate-limiting TS in which the P-O2' bond is still forming while the P-O5' bond remains almost uncleaved, which is again supported by the negative ξ . The transphosphorylation process is now shifted to a concerted fashion where the late TSs cannot be located for most reactions in this group as a result of the enhanced leaving groups.

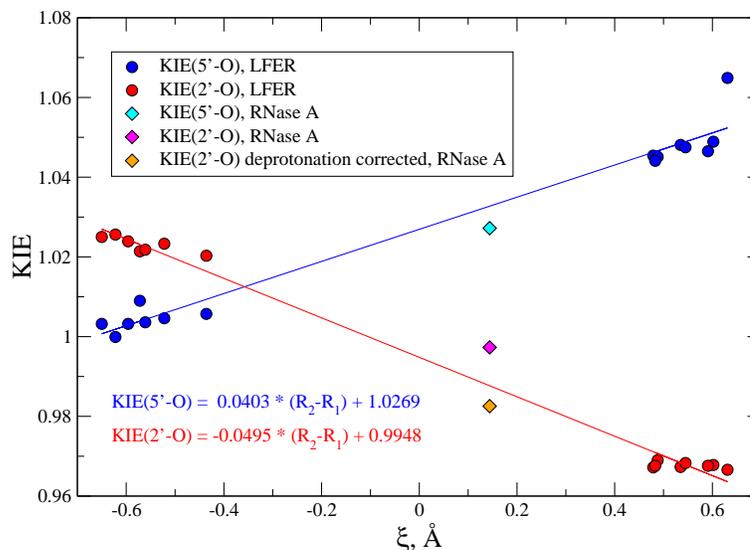


Figure 3.6: Relationship between computed 2'-O and 5'-O KIEs and reaction coordinate ξ in the rate-limiting TSs in LFER model reactions and RNase A enzymatic model. The points for RNase A (in diamonds) are excluded in the linear fitting.

In terms of LFER analysis, as seen in Figure 3.5, two distinct Brønsted coefficients (β_{lg} s) -1.37 and -0.53 were observed for the two groups of reactants with poor and enhanced and leaving groups, respectively. These two coefficients agrees excellently with experimental values reported by Lönnberg¹³⁰, which are -1.34 and -0.52. The β_{lg} value with a much greater magnitude -1.37 for those reactants with poor leaving groups suggests a later rate-limiting TS that involves more P-O5' bond cleavage motion, which makes the reaction rate more sensitive to the change of the leaving group¹⁰; the β_{lg} with a small magnitude -0.53 for the reactants with sufficiently enhanced leaving groups indicates a concerted mechanism through a single early TS involving mainly P-O2' bond formation motion, which reflects a diminished sensitivity of reaction rate to variation of the leaving group. The break point of pK_a between the two mechanisms is determined to be 13.02 (see Figure 3.5). From KIE results, we can observe from Table 3.2 that the pK_a of the break point should fall between 12.37 and 13.55, which coincides very well with the LFER results here and previous reported value 12.58 from Lönnberg¹³⁰. Hence, both the experimental and computational data are consistent with a change in overall mechanism as leaving group reactivity decreases. A direct connection between observed Brønsted coefficients, KIE values and the underlying

mechanisms of RNA transphosphorylation reactions is observed, whereby both experimentally observable parameters provide consistent signatures for the stepwise versus concerted reaction channels.

3.3.3 Application to RNase A enzymatic model

Recently, a combined experimental and theoretical investigation was carried out on the elucidation of RNase A catalytic mechanism⁶. A simplified reaction model was devised in that work to mimic the RNA 2'-O-transphosphorylation in the enzymatic environment, and the 2'-O and 5'-O KIEs were computed and shown to be consistent with the experimentally measured enzymatic KIEs. Here, the above relationship between ξ and KIE values are applied to this model. Optimized RS and TS structures in this model reaction are shown in Figure 3.7. Although the TS here is still a late one, the 5'-O KIE is less normal than those in the late TSs of LFER series while the 2'-O KIE becomes less inverse, which is well reflected in the more compact TS geometry and less advanced reaction coordinate ξ . Interestingly, computed 2'-O and 5'-O KIEs and ξ values in the rate-limiting TS fall near the fitted lines from the LFER series as shown in Figure 3.6. We can see that the LFER model can be used at least qualitatively to predict the geometrical details for TSs of more complex reactions.

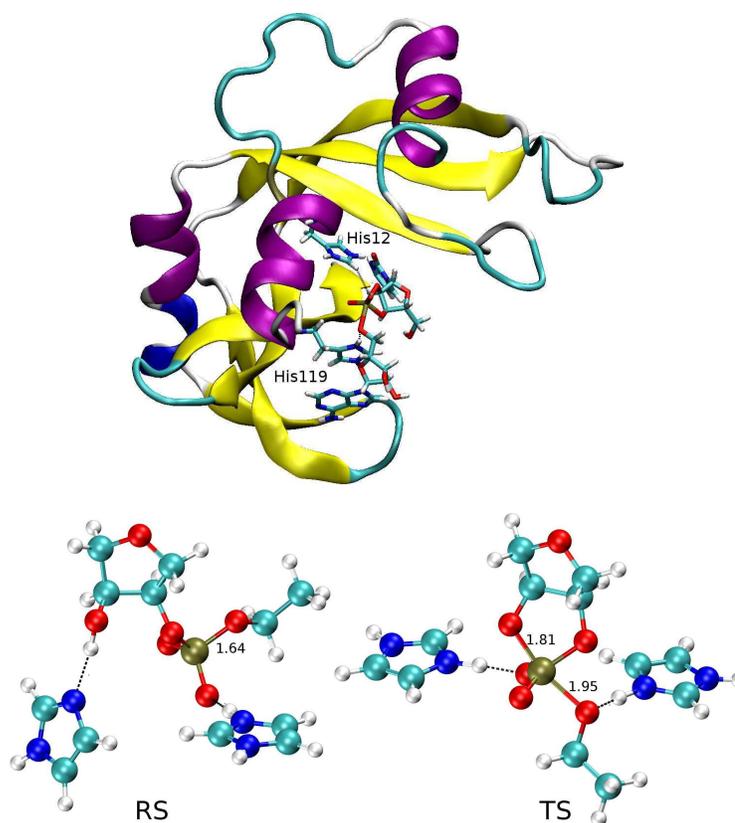


Figure 3.7: (Top) Structure of RNase A transition state mimic, in which His12 stabilizes the negative charges on the non-bridging oxygens and His119 acts as a general acid to facilitate P-O5' bond cleavage. (Bottom) Reactant state (left) and transition state (right) structures in the model reaction. The two imidazole rings in the RS and TS structures are used to mimic His12 (left) and His119 (right). Key bond lengths (in Å) are labeled.

The RNase A data points deviate from the model derived from the specific base-catalyzed nonenzymatic reactions in LFER series for several reasons that provide insight into enzyme mechanism. First, the RSs and TSs in the LFER series are all dianionic, under basic conditions in which the 2'-OH is deprotonated, while the RNase A reaction model was built to mimic ideal enzymatic condition at pH 7 so the 2'-OH remains protonated in the RS. Due to the large normal equilibrium isotope effect (EIE) on alcohol deprotonation (1.015) which offsets the inverse contribution from O-P bond formation, the observed 2'-O KIE value for the RNase A reaction is less inverse compared to reactions in LFER series with oxyanion reactant states^{110,1,131} (see Figure 3.6). Correcting for this difference in reactant

states between the RNase A and model reactions yields a value of 0.9826, which more closely corresponds to the predicted relationship between the observed nucleophile KIE and reaction coordinate progress (Figure 3.6). The presence of ribose sugar ring vibrational modes in the enzymatic model may also result in deviations from the trend interpreted from analyses of simpler intermolecular phosphoryl transfer model reactions. The 5'-O KIE for the enzyme reaction also corresponds in general to the relationship observed for the model reactions, however, the magnitude of this effect is influenced by general acid catalysis that is absent from the specific base catalyzed model reactions. Proton transfer from His119 creates a stiffer bonding environment^{132,133,134,6} for the 5'-O in the TS that leads to a decrease in the 5'-O KIE value, resulting in an overestimation of the observed value by the model. Thus, the KIE signature for RNase A matches expectations for the stepwise mechanism with a late TS drawn from comparison to model reactions. Importantly, deviation from predicted values for both the nucleophile and leaving group for the enzyme reaction are attributable primarily to proton transfer, either at equilibrium or in the transition state.

The success of the computational framework described here is best evaluated by the ability to identify and predict KIEs that provide insight into mechanism and, importantly, are amenable to subsequent experimental measurement. Previous studies showed that ionization of alcohols enhances the electron density of the alcohol oxygen, which effectively decreases the C_α-H bond strength due to hyperconjugation resulting in normal secondary deuterium isotope effects^{135,136,137,138}. Indeed, the secondary deuterium KIE on the 5' carbon for the model reactions are 1.15-1.2 which is near the EIE for ionization of an aliphatic alcohol (1.15) reflecting advanced leaving group bond cleavage (Table 3.3). The secondary deuterium effects on the 2' carbon of the nucleophilic alcohol are observed to be inverse (0.85) reflecting the loss of ionization upon going from an oxyanion to a phosphoester due to advanced O-P bond formation. Interestingly, the secondary deuterium KIEs at the C2'-H and C5'-H are predicted to be 1.0119 and 1.0291, respectively, for the RNase A catalyzed reaction. The significantly smaller leaving group effect provides another KIE signature for the general acid role in RNase A in which the proton transfer from His119 to the 5'-O leaving group largely offsets the accumulation of charge. Although these are secondary KIEs their magnitudes are large relative to primary ¹⁸O effects which have been analyzed

extensively. Thus, in addition to the framework for interpretation of primary KIEs and LFER results developed here, we identify secondary deuterium KIEs as an important indicator of transition state charge and provide predicted values useful for future experimental validation.

3.4 Conclusion

In this work, we present the results of quantum chemical studies on the KIEs in the phosphate ester hydrolysis reactions, non-enzymatic RNA transphosphorylation model reaction series and enzymatic model that represents RNA backbone cleavage catalyzed by RNase A. Benchmark KIE calculations have been performed on the experimentally well-studied phosphate ester hydrolysis systems to validate the computational methods for prediction of KIEs relevant to phosphoryl transfer in RNA. The method that yielded the most consistent agreement with experiments (B3LYP/6-31++G(d,p) with PCM/UAKS solvation) was identified and then applied to the prediction of KIEs in a model reaction series to establish a relationship between approximate reaction coordinate and 2'-O and 5'-O KIEs. Finally, KIEs in the RNase A catalysis model were computed and shown to be consistent with the trend observed in the LFER series between KIEs and reaction coordinate values. The present work demonstrates how LFER and KIE analysis provide complementary information from different measurements that, together with calculations, provide deep insight into molecular mechanism. The data presented in this work further serves as a useful benchmark and guide to the design and development of next-generation multiscale models for RNA catalysis mechanisms which are of great biological importance.

3.5 Supporting Information for: Mechanistic Insights into RNA Transphosphorylation from Kinetic Isotope Effects and Linear Free Energy Relationships of Model Reactions

Table 3.3: Key bond lengths (in Å) in the optimized transition state structures of phosphate/phosphorothioate ester hydrolysis reactions. R(P-Nu) and R(P-Lg) stand for the phosphorous-nucleophile oxygen and phosphorous-leaving group oxygen bond lengths, respectively.

Reactant	B3LYP+UFF		B3LYP+UAKS		M06-2X+UAKS	
	R(P-Nu)	R(P-Lg)	R(P-Nu)	R(P-Lg)	R(P-Nu)	R(P-Lg)
<i>p</i> NPP ²⁻	2.69	3.37	2.52	2.76	2.32	2.50
<i>p</i> NPPT ²⁻	N/A	3.89	N/A	3.53	N/A	4.60
<i>p</i> NPP ⁻	N/A	2.02	N/A	1.98	N/A	1.86
<i>p</i> NPPT ⁻	N/A	2.02	N/A	2.00	N/A	1.85
EtO <i>p</i> NPP ⁻	2.48	1.88	2.41	1.88	2.40	1.80
EtO <i>p</i> NPPT ⁻	2.55	1.83	2.49	1.84	2.48	1.78
(EtO) ₂ <i>p</i> NPP	2.69	1.70	2.74	1.70	2.74	1.67
(MeO) ₂ <i>p</i> NPPT	2.71	1.71	2.80	1.70	2.81	1.67

Table 3.4: Computed C2'-H and C5'-H H/D KIE values for the LFER reaction series and RNase A model reaction. Reaction coordinate ξ values in rate-limiting transition states are also shown here.

Leaving Group	ξ (Å)	KIE(C2'-H)	KIE(C5'-H)
CH ₃ COO ⁻	-0.62	0.9960	N/A
2,3,5,6-F ₄ -C ₆ HO ⁻	-0.57	0.9963	N/A
4-NO ₂ -C ₆ H ₄ O ⁻	-0.65	1.0001	N/A
4-CN-C ₆ H ₄ O ⁻	-0.60	0.9948	N/A
3-CN-C ₆ H ₄ O ⁻	-0.56	0.9893	N/A
C ₆ H ₅ O ⁻	-0.52	0.9911	N/A
CF ₃ CH ₂ O ⁻	-0.44	1.0005	0.9968
HCCCH ₂ O ⁻	0.48	0.9367	1.0419
FCH ₂ CH ₂ O ⁻	0.54	0.8489	1.1644
ClCH ₂ CH ₂ O ⁻	0.49	0.8544	1.1583
HOCH ₂ CH ₂ O ⁻	0.48	0.8550	1.1483
CH ₃ O ⁻	0.63	0.8568	1.2378
CH ₃ CH ₂ O ⁻	0.60	0.8484	1.2270
CH ₃ CH ₂ CH ₂ O ⁻	0.59	0.8532	1.2023
CH ₃ CH ₃ CHO ⁻	0.55	0.8555	1.1742
RNase A Model	0.14	1.0119	1.0291

Chapter 4

Effect of Zn^{2+} Binding and Enzyme Active Site on the Transition State for RNA 2'-O-transphosphorylation Interpreted Through Kinetic Isotope Effects

Divalent metal ions, due to their ability to stabilize high concentrations of negative charge, are important for RNA folding and catalysis. Detailed models derived from the structures and kinetics of enzymes and from computational simulations have been developed. However, in most cases the specific catalytic modes involving metal ions and their mechanistic roles and effects on transition state structures remains controversial. Valuable information about the nature of the transition state is provided by measurement of kinetic isotope effects (KIEs). However, KIEs reflect changes in all bond vibrational modes that differ between the ground state and transition state. QM calculations are therefore essential for developing structural models of the transition state and evaluating mechanistic alternatives. Herein, we present computational models for Zn^{2+} binding to RNA 2'-O-transphosphorylation reaction models that aid in the interpretation of KIE experiments. Different Zn^{2+} binding modes produce distinct KIE signatures, and one binding mode involving two zinc ions is in close agreement with KIEs measured for non-enzymatic catalysis by Zn^{2+} aquo ions alone. Interestingly, the KIE signatures in this specific model are also very close to those in RNase A catalysis. These results allow a quantitative connection to be made between experimental KIE measurements and transition state structure and bonding, and provide insight into RNA 2'-O-transphosphorylation reactions catalyzed by metal ions and enzymes.

4.1 Introduction

Divalent metal ions play critical roles in RNA folding and catalysis^{139,140,141,142,143,144,86,145}. The ability of divalent ions to stabilize high concentrations of negative charge in transphosphorylation reaction centers via electrostatic interactions, direct coordination or acid-base chemistry empowers them with potential mechanisms to assist in catalysis. However, unraveling the specific role of metal ions is extremely challenging due to the difficulty in discerning the catalytically active metal ion binding mode and its connection with the transition state (TS) structure and bonding^{140,141,142}, which also exists as the major barrier in the investigation of enzyme catalysis mechanisms.

A powerful strategy to resolve mechanistic ambiguity is to rationally design and study simplified model reaction systems using a joint experimental/theoretical approach. Perhaps the most sensitive experimental mechanistic probe is the measurement of kinetic isotope effects (KIEs) that compare the relative reaction rate constants between isotopologues. KIEs arise from subtle quantum effects associated with the changes in structure and bonding that occur in proceeding from the reactant state (RS) to rate-controlling TS^{85,146,110,147,2,148}. However, meaningful interpretation of KIE measurements requires the use of computational models. Computational modeling of KIEs has been extensively applied to study RNA transphosphorylation catalyzed by enzyme⁶, specifically designed metal catalyst^{149,150} and without catalyst^{7,111,19}. In a recent work¹⁵¹, Zhang *et. al.* measured the primary and secondary kinetic isotope effects for catalysis by Zn^{2+} ions and by specific base alone, and compared results with preliminary calculations. In the present work, we extend the scope of these calculations to explore 9 distinct, alternative Zn^{2+} ion binding modes (Figure 4.3) in the TS and characterize the resulting KIE signatures. Comparison across different model reactions and different methods of calculating KIEs are also performed and analyzed.

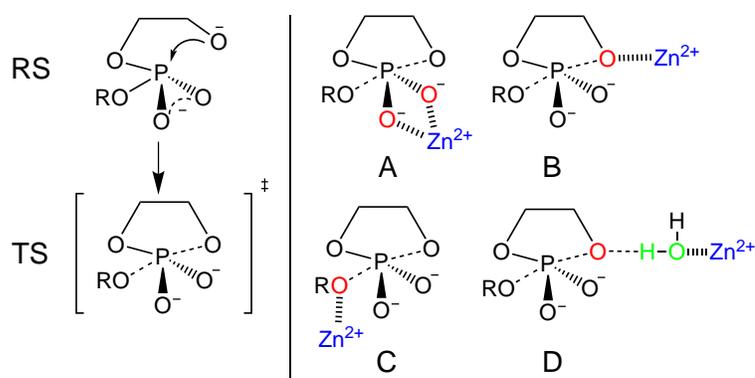


Figure 4.1: (Left) Schematic description of the reactant state and rate-limiting transition state in the RNA transphosphorylation reaction model. (Right) Illustration of three different Zn^{2+} binding sites (non-bridging oxygens in A, nucleophile oxygen in B, leaving group oxygen in C) and two interaction modes (direct coordination in A, B and C, indirect binding via a solvent water molecule in D).

4.2 Computational Methods

DFT calculations were performed using the B3LYP^{27,26} functional which has been demonstrated to be reliable for zinc complexes¹⁵². The 6-31+G(d) basis set was used for H, C, N, O and P, while the SDD effective core potential¹⁵³ was applied to Zn. Solvation effects were treated with the polarizable continuum model¹⁵⁴ (PCM) using specialized atomic cavity radii for RNA catalysis adopted from previous work^{111,6}. Water solvent with a dielectric constant of 78.4 is used in all PCM calculations. Kinetic isotope effects were calculated from the Bigeleisen equation⁸⁵ using the vibrational frequencies obtained from normal mode analysis of the optimized reactant and transition state geometries. All electronic structure calculations were carried out in Gaussian 09 package¹¹⁸.

QM/MM simulations of the RNA transphosphorylation catalyzed by RNase A were carried out using the Amber14¹⁵⁵ molecular dynamics package. The starting structures of the reactant and transition states of the enzyme-substrate complex was obtained from Ref. 6 in which more than 50 ns equilibration dynamics was performed. In the QM/MM simulations, the QM region including the RNA dinucleotide and the side chains of His 12 and His 119 was treated using the AM1/d-PhoT³⁴ semi-empirical Hamiltonian and

the MM region was described using AMBER-ff14SB force field^{47,156}. For both reactant and transition states, 20 ps of QM/MM equilibration simulations were first performed, with harmonic restraints that kept the reaction coordinates at their corresponding values according to the QM/MM free energy profiles (personal communications from Dr. Thakshila Dissanayake, to be published). Then, 200 ps of QM/MM annealing simulations with the restraints on was run to cool the system down to 0 K. From the 0 K structures of the reactant and transition states, +0.0001 Å and -0.0001 Å displacements in x, y and z directions were put on all QM atoms once in a time, resulting in 6N (N is number of QM atoms) displaced geometries. Single point QM/MM energies and forces were evaluated for all the 6N structures in Amber, which were then used to construct the numerical Hessian matrices and calculate the KIEs.

4.3 Results and Discussion

4.3.1 Building a baseline model for un-catalyzed RNA 2'-O-transphosphorylation

In order to understand the effect of Zn^{2+} binding on TS structure, it is necessary to first characterize the reaction mechanism and TS in the absence of Zn^{2+} . The transition states for a series of non-enzymatic baseline models (B1-B3) in the absence of Zn^{2+} are shown in Figure 4.2, and their calculated KIEs are compared with experimental values⁶ for a UpG dinucleotide in Table 4.1. As the models progress from the minimal model (B1) to the full dinucleotide (B3), the agreement between the calculated and experimental $^{18}k_{\text{LG}}$ values significantly improves, while for $^{18}k_{\text{NUC}}$ and $^{18}k_{\text{NPO}}$ it improves slightly in B3 but not B2. The notable decrease in the calculated $^{18}k_{\text{LG}}$ value from 1.0416 in B1 to 1.0358 in B2 mainly arises from the addition of a sugar ring to the leaving group, which enhances the leaving group activity since the $\text{p}K_a$ of tetrahydro-2-furanmethanol (14.68¹⁵⁷) is lower than that of ethanol (16.47¹⁵⁷). The addition of the full guanosine leaving group (B3) further reduces the $^{18}k_{\text{LG}}$ value to 1.0322 that is very close to the experimental value of 1.034. This is due to coupling of vibrational modes of the nucleobase, in addition to the overall greater effective mass of the leaving group that damps the frequency of certain key modes. It is noteworthy to mention that when the leaving group is a methoxide, which is

even lighter than the ethoxide group of B1, the calculated $^{18}k_{LG}$ value increases to 1.0649, despite having a pK_a value roughly 0.5 units lower¹⁹.

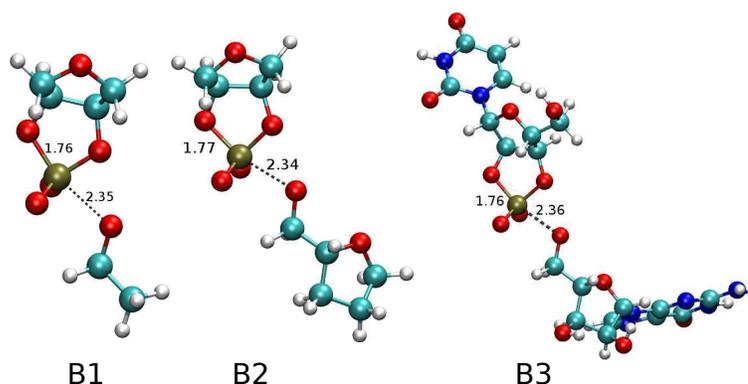


Figure 4.2: TS structures of baseline models for un-catalyzed RNA transphosphorylation. Key bond lengths in Å are labeled.

Although the full dinucleotide baseline model (B3) is in best agreement with experiment, it is too computationally intensive to be practical as a departure point from which to exhaustively explore multiple Zn^{2+} binding modes that add many more electrons to the quantum system and degrees of freedom to the optimization procedure. The goal of the present work is to determine the effect of Zn^{2+} binding on the TS structure of the UpG dinucleotide. As seen in Table 4.1, while the absolute values of the calculated KIEs for catalyzed and uncatalyzed reactions deviate modestly from the experimental values, their relative values are very consistent. Our comparison focuses on the relative KIEs for the catalyzed and uncatalyzed (baseline) models in order to maximize the cancellation of systematic errors in order to obtain quantitative agreement of the relative calculated and experimental KIE values. It has been suggested¹⁵⁸ to use isotope-effect-minus-one (KIE - 1) instead of KIE itself when comparing heavy atom isotope effects as these values are usually very close to unity. Here since we're focusing on the ratio between catalyzed and un-catalyzed KIEs, subtracting by one will make the magnitudes of both numerator and denominator much smaller, which will make the result a lot more sensitive to the computational errors and experimental uncertainties because both numerator and denominator are now in similar orders of magnitude with the uncertainties. It might also blow up because the denominator could be very close to zero. With an understanding of the deviations of the minimal baseline

model (B1) relative to the dinucleotide baseline model (B3), it is reasonable to expect that comparison of Zn^{2+} binding in the context of the minimal model (which is computationally tractable even with multiple hydrated Zn^{2+} ions bound) would be transferable to the dinucleotide. Consequently, in what follows, we use the minimal baseline model (B1) as a framework from which to calculate the effect of Zn^{2+} binding on the TS and KIE values.

Table 4.1: Comparison of calculated and experimental KIE values and the effect of catalysts for UpG dinucleotide 2'-O transphosphorylation model reactions in solution. All KIE values were measured/calculated at the temperature of 90° except for $^{18}k_{\text{NPO}}$ in the baseline models, where the only available experimental value was at 37° (from Ref. 7). Therefore, the corresponding calculations were also performed at 37° . $k^{\text{Cat}}/k^{\text{BL}}$ quantifies the effect of the catalyst on the KIEs, which is the ratio between KIEs in the catalyzed and uncatalyzed reactions, where the BL refers to the baseline model B1.

Condition	$^{18}k_{\text{LG}}$	$^{18}k_{\text{NUC}}$	$^{18}k_{\text{NPO}}$
Model B1 calc.	1.0416	1.0016	1.0029
Model B2 calc.	1.0358	1.0038	1.0032
Model B3 calc.	1.0322	1.0011	1.0025
Baseline expt. ⁷	1.034(3)	0.997(1)	0.999(1)
Zn model IX calc.	1.0276	0.9950	1.0028
Zn^{2+} -catalyzed expt. ¹⁵¹	1.015(2)	0.986(4)	1.0007(2)
$(k^{\text{Cat}}/k^{\text{BL}})_{\text{Calc.}}$	0.986	0.993	1.000
$(k^{\text{Cat}}/k^{\text{BL}})_{\text{Expt.}}$	0.982	0.989	1.002
RNase A calc. ¹⁹	1.0272	0.9973	1.0060
RNase A expt. ⁶	1.014(3)	0.994(2)	1.001(1)
$(k^{\text{Cat}}/k^{\text{BL}})_{\text{Calc.}}$	0.986	0.996	1.003
$(k^{\text{Cat}}/k^{\text{BL}})_{\text{Expt.}}$	0.981	0.997	1.002

4.3.2 Exploration of Zn^{2+} catalytic modes

Comparison of experimental KIEs for un-catalyzed (baseline) and Zn^{2+} -catalyzed reactions (Table 4.1) indicate that both primary $^{18}k_{\text{LG}}$ and $^{18}k_{\text{NUC}}$ values decrease considerably

(by 0.019 and 0.011) upon Zn^{2+} binding. Examination of the experimental ratio of the KIE values for Zn^{2+} -catalyzed and baseline reactions $[(k^{Cat}/k^{BL})_{Expt.}]$ indicates the largest deviation from unity occurs for the leaving group (1.8%). The large normal $^{18}k_{LG}$ value for the uncatalyzed reaction (1.034) suggests a late transition state characterized by a small bond order to the leaving group and high degree of accumulated charge at the O5' position. The significant reduction of the $^{18}k_{LG}$ value upon Zn^{2+} binding (1.015) suggests a TS that is not as late¹⁹, has greater bonding to the leaving group and less charge at the O5' position. The effect of Zn^{2+} binding on the $^{18}k_{NUC}$ value is also significant (1.1%), but less pronounced than for $^{18}k_{LG}$, and indicates a slightly higher degree of bond formation of the nucleophile to phosphorus for the Zn^{2+} -bound TS compared to the un-catalyzed reaction. There is little effect of Zn^{2+} binding on the secondary KIE ($^{18}k_{NPO}$) values (0.2%). The overall effect of Zn^{2+} binding is to produce a generally tighter TS bonding environment.

In order to establish a molecular electronic structure model that explains the effect of Zn^{2+} binding on the KIE values relative to the uncatalyzed reaction, we examined a series of 9 plausible Zn^{2+} binding modes (Figure 4.3), the results for which are shown in Table 4.2. Agreement between calculated and experimentally measured KIE values can be quantified by examination of the percent deviation (%D) in the KIE ratios defined as $\%D = [(k^{Cat}/k^{BL})_{Calc.} - (k^{Cat}/k^{BL})_{Expt.}] \times 100\%$. Further, the character of the TS can be quantified by a reaction coordinate ξ defined as $\xi = R_2 - R_1$, where R_1 and R_2 are the P-O2' and P-O5' bond lengths, respectively (Table 4.2). Negative values of ξ indicate an early TS, whereas positive values indicate a late one.

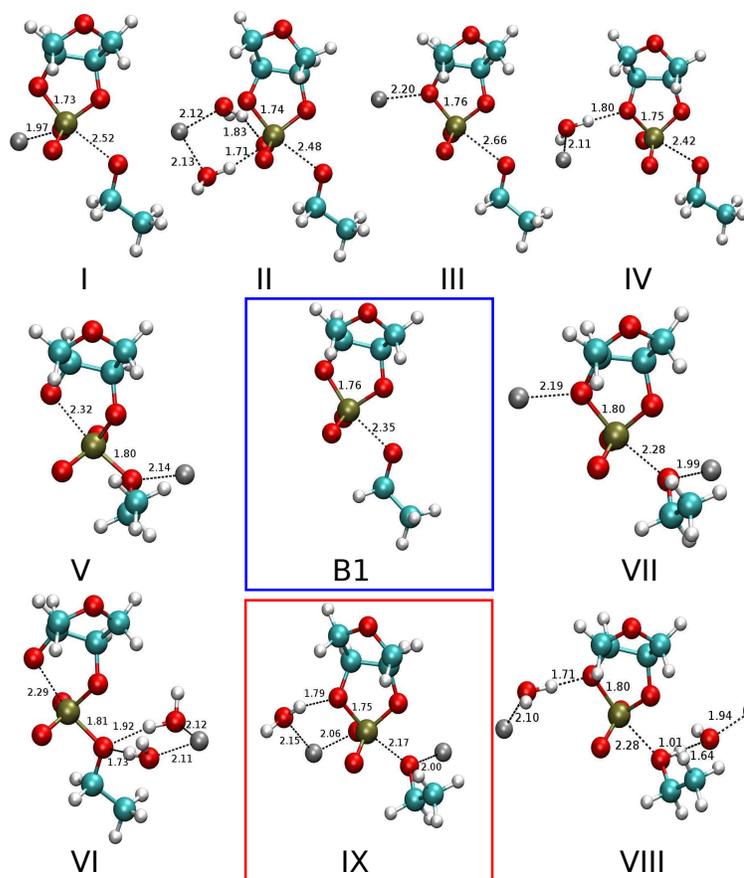


Figure 4.3: TS structures located from all 9 Zn^{2+} binding models and comparison with the baseline model B1. Model IX matches best with experimental KIEs and has been highlighted. Key bond lengths in \AA are labeled. All Zn^{2+} are saturated to hexacoordination by water but only key water molecules are shown for clarity.

We first explored a series of representative single Zn^{2+} binding modes to ascertain the effects on the predicted KIE values. In models I and II, Zn^{2+} are placed near the non-bridging oxygens to stabilize the negatively-charged phosphorane TS. However, the TSs are too late ($\xi \sim 0.7\text{-}0.8 \text{\AA}$, Table 4.2) and have $^{18}k_{\text{LG}}$ values that are considerably larger than the experimental value. The difference between models I and II involves direct versus indirect Zn^{2+} coordination and does not substantially alter the KIE values. For model III and IV in which Zn^{2+} binds directly and indirectly to the nucleophile $\text{O}2'$, the $^{18}k_{\text{NUC}}$ becomes slightly normal, and the $^{18}k_{\text{LG}}$ values remain large. Models V and VI explore direct and indirect Zn^{2+} binding to the leaving group oxygen, which leads to an early TS ($\xi \sim$

-0.5 Å) with considerably underestimated $^{18}k_{\text{LG}}$ and overestimated $^{18}k_{\text{NUC}}$ values. This can be explained by recognizing that this Zn^{2+} binding mode has a similar effect as that of an enhanced leaving group¹⁹ to shift the TS from late to early in character, with limited bond cleavage and charge accumulation at the O5' position.

The inability for the single Zn^{2+} binding models (I-VI) to reproduce the experimental KIEs lead us to explore dimetal Zn^{2+} binding modes. Models VII and VIII explore direct and indirect Zn^{2+} binding, respectively, at both the nucleophile and leaving group positions. Model VII produces a late transition state ($\xi = 0.47$ Å) that is much earlier than models I-IV and has a considerably improved $^{18}k_{\text{LG}}$ value (%D = 1.0), but has a normal $^{18}k_{\text{NUC}}$ value with greater deviation (%D = 1.5). Model VIII has a similar late transition state ($\xi = 0.48$ Å) to model VII, but has a nucleophile KIE value that is even more normal, and considerably underestimates the leaving group KIE, resulting from a partial proton transfer from a Zn^{2+} -coordinated water to leaving group oxygen.

Table 4.2: Comparison of calculated KIEs and reaction coordinate ξ values in the TSs from models I to IX. Model numbering is the same as in Figure 4.2. Temperature is 90° for all calculations and experiment. Reaction coordinate ξ is defined as $\xi = R_2 - R_1$, where R_1 and R_2 are the P-O2' and P-O5' bond lengths, respectively. Percentage deviation %D is defined as $\%D = [(k^{Cat}/k^{BL})_{Calc.} - (k^{Cat}/k^{BL})_{Expt.}] \times 100\%$, where BL is the baseline model (B1) consistent with the series of Zn^{2+} calculations. The numbers in parentheses following the experimental values are the experimental uncertainties.

Model	ξ (Å)	$^{18}k_{LG}$ (%D)	$^{18}k_{NUC}$ (%D)	$^{18}k_{NPO}$ (%D)
B1	0.59	1.0416	1.0016	1.0029
Expt. w/o Zn	N/A	1.034(3)	0.997(1)	0.999(1)
I	0.79	1.0466 (2.3)	0.9986 (0.8)	1.0017 (-0.3)
II	0.74	1.0441 (2.1)	0.9986 (0.8)	1.0029 (-0.2)
III	0.90	1.0517 (2.8)	1.0040 (1.3)	1.0007 (-0.4)
IV	0.67	1.0417 (1.8)	1.0012 (1.1)	1.0031 (-0.2)
V	-0.53	1.0063 (-1.6)	1.0463 (5.6)	1.0015 (-0.3)
VI	-0.49	1.0051 (-1.7)	1.0484 (5.8)	1.0029 (-0.2)
VII	0.47	1.0324 (1.0)	1.0054 (1.5)	1.0017 (-0.3)
VIII	0.48	1.0080 (-1.4)	1.0079 (1.7)	1.0005 (-0.4)
IX	0.42	1.0276 (0.5)	0.9950 (0.4)	1.0028 (-0.2)
Expt. w. Zn	N/A	1.015(2)	0.986(4)	1.0007(2)

Interestingly, each different Zn^{2+} binding model has a distinct set of predicted KIE values, however, only model IX corresponds closely with what is observed experimentally. Model IX involves one Zn^{2+} making direct coordination to the leaving group, and another that makes direct coordination to the non-bridge phosphoryl oxygen while maintaining indirect coordination with the nucleophile (Figure 4.3). In this model, the dimetal binding mode provides three distinct elements of TS stabilization—leaving group stabilization, negative charge redistribution and potentially assistance in proton transfer. The nucleophile and leaving group KIE deviations are 0.4% and 0.5%, respectively, a reduction in deviation by a factor of 2 with respect to the next smallest deviations in the series of models. The

$^{18}k_{\text{NUC}}$ value is slightly inverse and the $^{18}k_{\text{LG}}$ value normal. Excellent agreement is obtained between calculated and experimental ($k^{\text{Cat}}/k^{\text{BL}}$) ratios (Table 4.1) for the leaving group (0.986 and 0.982, respectively) and nucleophile (0.993 and 0.989, respectively). We've also expanded the leaving group in model IX with a sugar ring as in model B2, but the KIE values calculated from corresponding optimized structures are not significantly different from the original model IX ($^{18}k_{\text{LG}} = 1.0268$ vs 1.0276, $^{18}k_{\text{NUC}} = 0.9963$ vs 0.9950). Therefore we did not attempt to further expand the model to full UpG as in B3, as these are very large calculations that become more difficult to converge with more degrees of freedom. The general effect of Zn^{2+} binding is to create an earlier TS with an overall stiffer bonding environment that leads to a less pronounced normal leaving group KIE and slightly more inverse nucleophile KIE. The very close agreement of this model with recently measured KIE values, in stark contrast to that of a series of 8 other models tested, provides strong evidence that it can be used to provide an experimental interpretation of the TS structure and bonding for Zn^{2+} -catalyzed RNA 2'-O-transphosphorylation.

4.3.3 Comparison between Zn^{2+} catalysis and enzyme catalysis

The transition state for RNA 2'-O-transphosphorylation catalyzed by RNase A exhibits a primary KIE signature, both from experiment and computation, that is very close to that produced upon Zn^{2+} binding (Table 4.1). This is evident by analyzing the normalized $k^{\text{Cat}}/k^{\text{BL}}$ values for the Zn^{2+} and RNase A transition states which are all within 1% of one another. The baseline normalized computational values are very close to the experimental values, but particularly striking is the internal consistency for the $^{18}k_{\text{LG}}$ values between Zn^{2+} and RNase A systems as determined from either theory or experiment. The most straight forward interpretation is that the Zn^{2+} ions produce a local TS bonding environment that is similar to that of the RNase A active site: one Zn^{2+} ion stabilizes the negatively charged reaction center in transition state similar to a protonated His12 in RNase A, while another Zn^{2+} ion enhances the leaving group departure analogous to the role of His119 (Figure 4.4). This analysis sheds light on general principles involved in RNA catalysis.

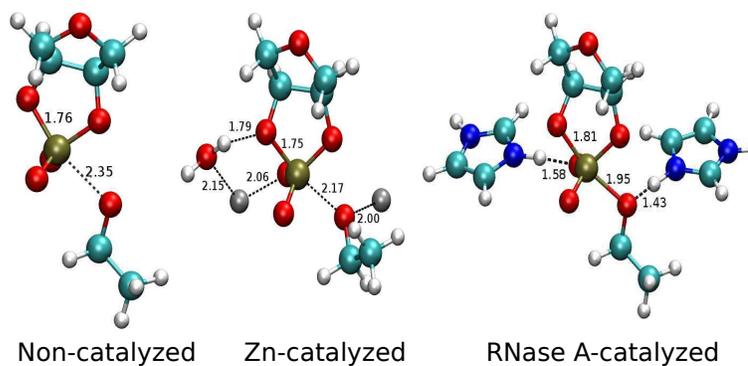


Figure 4.4: Comparison of TS structures in baseline, Zn^{2+} -catalyzed and RNase A-catalyzed model reactions. The two imidazole rings in the RNase A model represents His12 (left) and His119 (right) residues in RNase A.

4.3.4 Comparison of KIEs in RNase A catalysis from QM cluster calculations and QM/MM calculations

All the KIE calculations discussed above are based on the vibrational analysis of optimized reactant and transition state geometries. For the uncatalyzed and Zn^{2+} -catalyzed reactions, the size of the entire system is not too big to be treated quantum mechanically, therefore the geometry optimization and vibrational analysis could be done naturally within quantum chemistry packages. However, for the enzymatic reaction catalyzed by RNase A, one had to carefully carve out a truncated active site model from the QM/MM structure (the “QM cluster” approach) to be able to get the stable and correct stationary points through geometry optimization⁶ because it’s impossible to treat the entire system quantum mechanically. This method often requires manual adjustments to the model and might not be easy to be directly interpreted together with QM/MM results due to the incomplete chemical environment in the model. Therefore, here we attempt to directly calculate KIEs from QM/MM results of RNase A catalysis to generate KIE results that includes proper treatment of the enzyme environment and is consistent with the QM/MM simulations. As seen in Table 4.3, the QM/MM KIEs, even at the semi-empirical level, agree with the KIEs from previous QM cluster approach at DFT level and are in closer agreement with the experimental results while the higher-level (B3LYP) QM/MM KIEs are even closer to the experimental values. This suggests that the QM/MM approach might be able to give quantitatively correct

results and could be applied to study other enzymatic reactions. This approach could be further empowered by the recently developed, more efficient interface of *ab initio* QM/MM simulations in Amber that could significantly reduce the computational cost¹⁵⁹.

Table 4.3: Comparison of calculated KIEs in RNase A catalysis using different methods. The AM1/d-PhoT QM/MM KIEs were calculated using the procedure described in the Computational Methods section while for the B3LYP QM/MM KIEs, the QM/MM equilibration and annealing steps were replaced by QM/MM geometry optimization due to the otherwise untractable computational cost. Basis set used in B3LYP QM/MM calculations (both geometry optimization and single point calculations) is 6-31+G(d). Temperature is 37° in all calculations.

Method	$^{18}k_{\text{LG}}$	$^{18}k_{\text{NUC}}$	$^{18}k_{\text{NPO}}$
AM1/d-PhoT QM/MM	1.0179	0.9972	1.0017
B3LYP QM/MM	1.0122	0.9984	1.0008
B3LYP QM Cluster ¹⁹	1.0272	0.9973	1.0060
Expt. ⁶	1.014	0.994	1.001

4.4 Conclusion

In conclusion, we explored the effect of different Zn^{2+} binding modes on the ^{18}O kinetic isotope effects for Zn^{2+} -catalyzed RNA 2'O-transphosphorylation. Different Zn^{2+} binding modes yield distinct KIE signatures that can be connected to TS structure and bonding and used to aid in the interpretation of experimental measurements to give insight into mechanism. A unique binding mode was identified as being very closely aligned with recent experimental measurements. This mode involved two zinc ions, one directly coordinating the leaving group and the other directly coordinating a non-bridge phosphoryl oxygen while interacting with the nucleophile at solvent separation. This catalytic mode produces a KIE signature very close to that observed for the TS in RNase A, and leads to model TS structure that is also quite similar. We also identified the origin of the systematic overestimation of the $^{18}k_{\text{LG}}$ KIE value relative to experiment noted previously^{111,6,19,151} which herein was shown to be corrected by inclusion of more realistic leaving group models.

This work provides a predictive framework for the identification of Zn^{2+} ion binding modes in RNA 2'O-transphosphorylation reactions from KIE measurements that will advance our understanding of the role of divalent metal ions in mechanisms of RNA catalysis.

Chapter 5

Metal Ion-induced Activation of the Catalytic General Base in Hammerhead Ribozyme Self-cleavage

The hammerhead ribozyme is a well-studied nucleolytic ribozyme that catalyzes the self-cleavage of the RNA phosphodiester backbone. Despite numerous experimental and theoretical efforts, there remain key questions about details of the mechanism, particularly with regard to the activation of the nucleophile by the putative general base guanine (G12). One of the primary objections to the hypothesis that G12 acts as the general base involves the high pK_a value of guanine that would have to be considerably shifted by the ribozyme environment in order to be consistent with the interpretation of measured activity-pH profiles. Recent crystallographic and biochemical work has identified pH-dependent divalent metal ion binding at the N7/O6 position of G12 and suggested that the pK_a of G12 was shifted by that metal towards neutrality. In this work, we present results from quantum mechanical calculations and molecular simulations that unify the interpretation of available structural and biochemical data, and paint a detailed mechanistic picture of the general base step of the reaction. Electronic structure calculations are performed to quantify the magnitude of pK_a shifts induced by Mg^{2+} binding in several Mg^{2+} -guanine complexes. Molecular dynamics simulations are carried out using newly developed 12-6-4 parameters for divalent metal ion binding to nucleic acids to characterize the ribozyme active site environment and thermodynamic integration is used to evaluate the pK_a of G12 in HHR with and without the Mg^{2+} ion bound. The results show that Mg^{2+} is able to down-shift the pK_a of G12 by -1.2 units in accord with the apparent pK_a value determined from activity-pH measurements. In addition, *ab initio* quantum mechanical/molecular mechanical simulations are performed to explore the free energy profile for the general base step in the presence and

absence of Mg^{2+} . Taken together, these results are in quantitative agreement with available experimental data, and support a mechanism whereby Mg^{2+} serves to stabilize G12 in the functional, deprotonated form that can abstract a proton from the nucleophile in the general base step of the reaction. In this scenario, site-specific Mg^{2+} ion binding acts as a switch to activate the general base in HHR. Finally, experimentally-testable predictions are made on the mutational and rescue effects on G12, which will give further insights into the catalytic mechanism. These results contribute to our growing knowledge of the potential roles of divalent metal ions in RNA catalysis.

5.1 Introduction

The hammerhead ribozyme (HHR) is a prototype catalytic RNA system that has been extensively studied^{160,161,162,163,164,165,16,86} for almost three decades since it was first discovered by Uhlenbeck in 1987¹⁶⁶. The HHR catalyzes the cleavage transesterification of the RNA sugar-phosphate backbone¹⁶. In the generally accepted acid-base mechanism, the nucleophile (the 2'-hydroxyl of residue C17) is deprotonated by a general base to form an activated precursor that then goes on an inline attack to the adjacent scissile phosphate. Departure of the O5' leaving group is then facilitated by a general acid, leading to a 2',3'-cyclic phosphate.

Despite the wealth of structural^{167,168,169,170,11,12}, biochemical^{171,172,173,174,175,176,177,178,179,180,181,182,14,183,164,184,185,186,187,188,189,190,191,15,192} and computational^{16,162,193,194,195,196,197,198,133,199,143} data available, certain details of the catalytic mechanism remain unclear. In particular, the specific mechanism whereby the nucleophile becomes activated by a general base is not yet resolved. Biochemical studies^{183,184,189} have suggested that the guanine nucleobase in residue G12 may act as the general base in HHR, which is also supported by crystal structures^{169,11} in which the N1 site of G12 is well-positioned to interact with the 2'-hydroxyl group of C17 (Figure 5.1). Interestingly, guanine is observed to interact with the 2'-hydroxyl nucleophile in many other ribozyme systems⁸⁶, including the hairpin^{200,201}, glmS^{202,203}, Varkud satellite²⁰⁴ and twister^{205,89} ribozymes, and thus appears to be a common theme in RNA catalysis.

A conceptual objection to the hypothesis that G12 may act as the general base is that the apparent pK_a values derived from activity-pH profiles (around 8^{183,184,189,11}), are somewhat lower than the pK_a value of guanine in aqueous solution (9.2). In order to reconcile these observations, the most straight-forward interpretation is that the pK_a of G12 would need to be considerably down-shifted by the ribozyme active site. This prospect seems unlikely given the high degree of localized negative charge in the active site¹⁹⁸, the proximity of G12 to the scissile phosphate and the lack of nearby divalent metal ions in the active site of an earlier crystallographic data with resolved solvent structure¹⁹⁵.

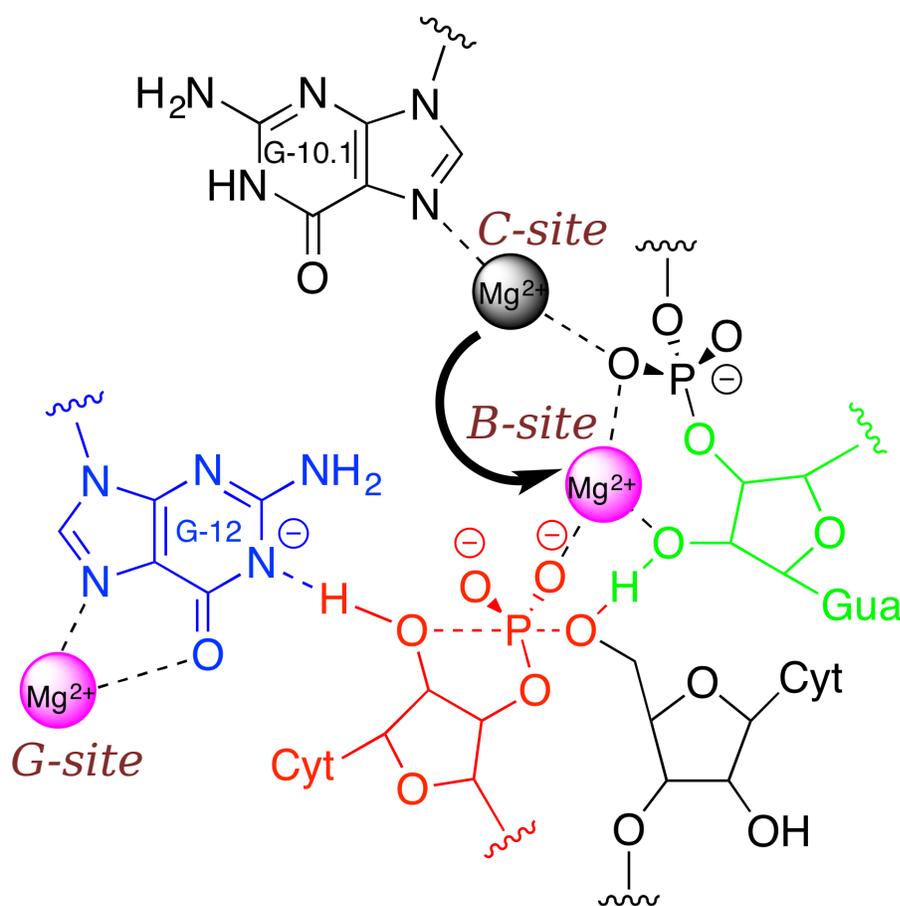


Figure 5.1: Illustration of the active site interactions in HHR and its self-cleavage mechanism. The N1 position of guanine in the putative general base G12 (blue) needs to be deprotonated before acting as a proton acceptor to deprotonate the 2'-OH in C17 (red), which will then act as the nucleophile to attack the phosphorous. Recent studies^{11,12} indicate that there could be a Mg^{2+} directly bound at the Hoogsteen face of G12 ("G-site") to facilitate its deprotonation. Another Mg^{2+} is believed to play the role of activating the 2'-OH of the general acid G8 (green) by migrating from the binding site at N7 of G10.1 ("C-site") observed crystallographically into a bridging position ("B-site") with the scissile phosphate, in accord with thio/rescue effect experiments^{13,14,15}. In this bridging position, the Mg^{2+} can coordinate the 2'-OH of G8, increasing its acidity, and facilitating proton transfer to the O5' leaving group in the general acid step of the reaction¹⁶.

Recently, a new set of crystal structures^{11,12} of a full length hammerhead ribozyme

(RzB) has been obtained by the Golden lab, which indicate a Mg^{2+} ion bound directly to the Hoogsteen face of G12 at pH 8. This binding mode is similar to the position of an electron density peak found in the previous crystal structures by and interpreted as a likely water molecule¹⁹⁵ or Na^+ ion¹⁷⁰. Divalent metal ions have been predicted to be capable of considerably shifting the pK_a of certain residues in ribozyme environment to facilitate catalysis^{206,141,207}. It is thus of considerable interest to understand whether this binding mode could represent an effective catalytic strategy for general base activation in nucleolytic ribozymes.

In this paper, we explore the feasibility of a Mg^{2+} ion, bound to G12, to shift the pK_a at the N1 position as required by its implicated role as general base in the reaction, and make quantitative predictions that are experimentally testable. Electronic structure calculations are performed to quantify the magnitude of pK_a shifts induced by Mg^{2+} binding in several Mg^{2+} -guanine complexes. Molecular dynamics (MD) simulations are carried out with newly developed force field parameters for divalent metal ion binding to nucleic acids⁹ to characterize the ribozyme active site environment and thermodynamic integration (TI) is used to evaluate the pK_a shift of G12 in HHR with and without the Mg^{2+} bound. *Ab initio* quantum mechanical/molecular mechanical (QM/MM) simulations with rigorous long-range electrostatic interactions¹⁵⁹ are performed to explore the free energy profile for the general base step in the presence and absence of Mg^{2+} bound at the newly identified position. Taken together, these results are in quantitative agreement with available experimental data, and support a mechanism whereby Mg^{2+} serves to stabilize G12 in an active (deprotonated) form that will ultimately facilitate deprotonation of the nucleophile in the general base step of the reaction. In addition, quantum mechanical calculations on the binding free energies between different divalent metal ions and guanine/6-thioguanine predict that a 6-thio substitution at G12 in HHR will decrease the catalytic reaction rate by knocking out the Mg^{2+} binding while replacing Mg^{2+} by Mn^{2+} or Cd^{2+} could have rescue effects, which could be tested by future experimental works.

5.2 Computational Methods

5.2.1 Electronic structure calculations

All electronic structure calculations were carried out in the Gaussian09 package¹¹⁸ using M06-2X¹¹² density functional with 6-311++G(3df,3pd) basis set. The PCM solvation model^{208,116,209,114} was used to treat the solvation effects, whose solvation cavity is constructed from the UFF radii set scaled by a factor of 1.1¹¹⁵. Harmonic vibrational analysis was performed on the optimized geometries using rigid rotor and harmonic oscillator approximations to verify the nature of the stationary points and obtain the thermal corrections to free energies. The relative pK_a s (i.e., the pK_a shifts) are calculated from:

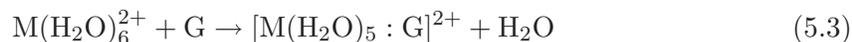
$$\Delta pK_a = (\Delta G' - \Delta G)/RT \ln(10) \quad (5.1)$$

where $\Delta G'$ and ΔG are the deprotonation reaction free energies of Mg^{2+} -guanine and guanine, respectively. Free energy difference between two deprotonation reactions in kcal/mol were converted to pK_a shift in pK_a units by dividing by $RT \ln(10)$. The pK_a shift is further corrected by taking the Mg^{2+} ion concentration and Mg^{2+} -guanine binding affinity into account. By considering a kinetic model that consists of deprotonation of guanine, Mg^{2+} -guanine binding equilibrium and deprotonation of Mg^{2+} -guanine complex, we obtained estimates for the corrected pK_a of Mg^{2+} -guanine complex as (see supporting information for details):

$$pK_a(Mg\text{-Gua,corr.}) = -\log_{10} \left(\frac{10^{-pK_a(\text{Gua})} + c_{Mg} \times e^{-\beta \Delta G_{bind}} \times 10^{-pK_a(\text{Mg:Gua,uncorr.})}}{1 + c_{Mg} \times e^{-\beta \Delta G_{bind}}} \right) \quad (5.2)$$

where c_{Mg} is the concentration of Mg^{2+} ion (0.01 M is used in this work) divided by the standard molar concentration 1.0 M (thus dimensionless), and ΔG_{bind} is the Mg^{2+} -guanine binding affinity which has been experimentally measured to be -0.34 kcal/mol^{8,9} (see Supporting Information for details). Electron population analysis was performed on the guanine and Mg^{2+} -guanine complex molecules (both neutral and deprotonated) at their optimized geometries using the natural bond orbital (NBO) analysis²¹⁰ method, as implemented in Gaussian09.

For the binding free energy calculations between different metal ions and guanine/6-thioguanine, the chemical equations that define the binding free energy are:



in which M could be Mg^{2+} , Mn^{2+} or Cd^{2+} and G could be guanine or 6-thioguanine. For Mn and Cd atoms, the SDD effective core potential¹⁵³ is used. The cavity radii of S atom was adjusted from 2.0175 (default as implemented in Gaussian09) to 2.6 which correctly reproduces the experimentally measured²¹¹ pK_a between guanine and 6-thioguanine.

Molecular dynamics simulations

All molecular dynamics simulations were carried out in the Amber14 package¹⁵⁵. Starting from the recent crystal structure of the RzB hammerhead ribozyme with divalent metal ion bound at G12^{11,12} (PDB code 5DI2), the RNA was solvated in a box of TIP4P-Ew²¹² water molecules with a 12 Å buffer. 67 Na^+ and 15 Cl^- ions were added to neutralize the system and solvate it with 0.14 M NaCl (roughly physiological salt conditions). The RNA was treated by the AMBER-ff14SB force field^{47,156} which includes the corrections for alpha/gamma conformers²¹³ and glycosidic torsions²¹⁴ in nucleic acids. The monovalent ions were described using the TIP4P-Ew-compatible parameters developed by Joung & Cheatham²¹⁵. To better characterize the interaction between Mg^{2+} and nucleic acids, we apply the recently developed m12-6-4 parameter set developed by Panteva, Giambaşu and York⁹ to accurately describe balanced interactions between Mg^{2+} and nucleic acids. This model was based on pioneering work by Li and Merz⁹¹ to include effects of ion-induced polarization, using a simple pairwise approximation that does not require further alteration of conventional molecular simulation force fields. This model provides outstanding agreement with structural, thermodynamic, kinetic and mass transport properties of Mg^{2+} in aqueous solution⁹⁰, and has recently been extended so as to give balanced interactions with nucleic acids through adjustment of specific pairwise parameters⁹. This model was

used to describe all 6 Mg^{2+} ions that were present in the crystal structure and in the simulations. The system was then slowly heated up to 300 K and equilibrated for a total of 30 ns simulation (referred as HHR · Mg^{2+} :G12). To justify the effect of the G12-bound Mg^{2+} , another system was built by removing the G12-bound Mg^{2+} and went through the same equilibration process (referred as HHR:G12). Also, a reference system was set up by solvating a single guanosine monophosphate residue (CH_3 -capped at the O3' position and CH_3O -capped at P position) in a box of TIP4P-Ew water with a 20 Å buffer along with 13 Na^+ and 12 Cl^- ions. This system was heated up to 300 K and equilibrated for a total of 5 ns simulation. All simulations employ an 8 Å nonbond cutoff and make use of Particle Mesh Ewald to account for the electrostatics beyond the cutoff.⁵³

5.2.2 Thermodynamic integration calculations

Starting from the equilibrated structures, all 3 systems (solvated guanosine, HHR:G12 and HHR · Mg^{2+} :G12) were prepared for thermodynamic integration (TI) calculations to compute the free energy of deprotonating guanine or G12. The net free energy change is: $\Delta G_{\text{DEPR}} = G_{\text{G}^-} - G_{\text{GH}}$, where G_{GH} is the free energy of the neutral guanine state (both in the reference system and at the G12 position in the HHR) and G_{G^-} is the free energy of the guanine deprotonated at the N1 position, which carries a net 1- charge. In order to stably compute the net free energy change from a reasonable amount of statistical sampling, it is advantageous to decompose the net free energy change as a sum of 3 stages²¹⁶ that pass through intermediate states:

$$\Delta G_{\text{DEPR}} = \Delta G_{\text{decharge}} + \Delta G_{\text{bond-removal}} + \Delta G_{\text{recharge}} \quad (5.5)$$

where $\Delta G_{\text{decharge}}$, $\Delta G_{\text{bond-removal}}$, and $\Delta G_{\text{recharge}}$ are referred to as the “decharging”, “bond-removal”, and “recharging” stages, respectively. The advantage of this decomposition is that less sampling is required to compute each stage than what would otherwise be necessary through a direct transformation between the two end states. In the decharging stage, the partial charge on H1 is removed. In the bond-removal stage, the bond, angle, torsion and Lennard-Jones interaction terms in the force field involving the H1 are removed. In the recharging stage, the partial charges on the entire residue are transformed to the charge set

for the deprotonated guanine derived from RESP charge fitting²¹⁷. The free energy of each stage is computed from TI:

$$\Delta G_{\text{stage}} = \int_0^1 \left\langle \frac{dH_{\text{stage}}(\lambda)}{d\lambda} \right\rangle_{\lambda} d\lambda \quad (5.6)$$

where $H_{\text{stage}}(\lambda)$ is a linear combination of Hamiltonians that define the end-states of the stage:

$$H_{\text{stage}}(\lambda) = (1 - \lambda)H_{\text{stage}}^{(0)} + \lambda H_{\text{stage}}^{(1)} \quad (5.7)$$

$H_{\text{stage}}^{(0)}$ and $H_{\text{stage}}^{(1)}$ are the Hamiltonians for the stage’s initial and final states, respectively. As an exception to Eq. 5.7, the removal of the Lennard-Jones interaction within the “bond-removal” stage is performed through a non-linear “soft-core” potential^{218,9}:

$$V_{\text{sc}}(\lambda) = 4\epsilon(1 - \lambda) \left[\frac{1}{[\alpha\lambda + (r/\sigma)^6]^2} - \frac{1}{\alpha\lambda + (r/\sigma)^6} \right] \quad (5.8)$$

where ϵ and σ are standard LJ parameters, r is atomic distance and α is an adjustable constant that was set to 0.5 by default²¹⁸. The free energy change of each stage is evaluated by performing 11 simulations (TI “windows”) corresponding to 11 evenly-spaced values of λ , and numerically integrating Eq. 5.6 from the trapezoid rule. Each TI window is equilibrated for 100 ps at $\lambda = 0.5$, and statistics are sampled from 1 ns of production simulation.

The TI calculations described above involves 33 TI window simulations per system (3 stages/system and 11 windows/stage) to compute a net free energy change. To ascertain the reliability of the result, we perform the procedure 3 times for each system (99 windows/system) to compute 3 estimates of the system’s net free energy change. For each system, we report (see Table 5.3) the average ΔG_{DEPR} value and its standard error from the 3 estimates.

5.2.3 Quantum mechanical/molecular mechanical simulations

After G12 is deprotonated, it is suited to act as a general base that activates the nucleophile by facilitating a proton transfer from the nucleophile C17:O2’ to G12:N1. *Ab initio* QM/MM umbrella sampling simulations were performed to obtain the free energy profile of the proton transfer reaction for both HHR:G12 and HHR · Mg²⁺:G12. The PBE0 hybrid density

functional^{25,28} with the 6-31+G(d) basis set was used to model the 47 atom (HHR:G12) or 48 atom (HHR · Mg²⁺:G12) QM region. The QM region includes the guanine base in the G12 residue, the Mg²⁺ bound at G12 (but not the coordinating water molecules), and the entire C17 residue (cytosine base, sugar, and phosphate). The simulations were performed with an 8 Å nonbond cutoff, and the Ambient-Potential Composite Ewald method¹⁵⁹ was used to describe the long-range electrostatics in the system.

The reaction coordinate ($\xi = R_1 - R_2$) is the difference between two distances. R_1 is the distance from the proton donor C17:O2' to the proton C17:HO2', and R_2 is the distance from the proton acceptor G12:N1 to the proton C17:HO2'. The umbrella biasing potential is:

$$U_{\text{bias}}(\xi) = k(\xi - \xi_0)^2 \quad (5.9)$$

where ξ_0 is the reference coordinate for the window, and $k = 50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ is a force constant. Each profile is generated from 16 simulation windows whose reference coordinate evenly samples the range $-1.6 \text{ \AA} \leq \xi_0 \leq 1.4 \text{ \AA}$. Each window is equilibrated for 2.5 ps, and statistics were drawn from 10 ps of production simulation. The statistics gathered from the biased simulations are analyzed using the variational free energy profile (vFEP) method⁶⁵ to generate the unbiased free energy profiles shown Figure 5.9.

5.3 Results and Discussion

Guanine nucleobases are often encountered in ribozyme active sites, and have been implicated as acting as the general base in catalysis. In order to act as a general base, the guanine nucleobase should be deprotonated at the N1 position (Figure 5.2). In the deprotonated state, resonance stabilization allows the negative charge to be distributed between the N1 and O6 positions. Divalent metal ion binding at the O6 position, therefore, would be expected to have a large influence on the microscopic pK_a at the N1 position by stabilizing the deprotonated form. In the following sections, we build up models for Mg²⁺ binding to guanine and related chemically modified analogs used in mechanistic experiments, and make predictions about the induced pK_a shift at the N1 position, and on the free energy of the general base step in the chemical reaction.

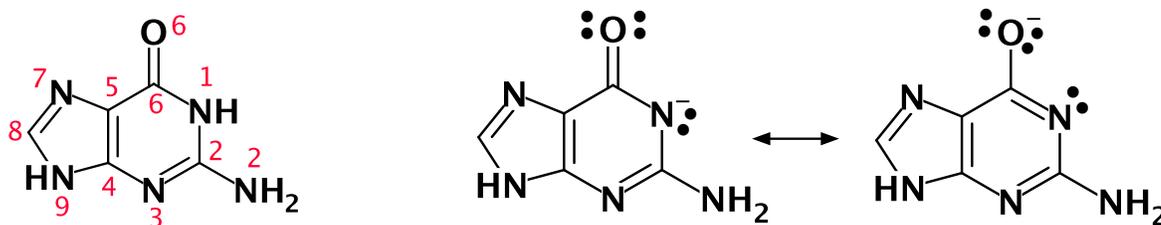


Figure 5.2: (Left) Canonical numbering of guanine nucleobase. (Right) Resonance structures of guanine deprotonated at N1 position with formal charge alternately on the N1 and the O6.

5.3.1 pK_a shifts of Mg^{2+} -guanine complexes

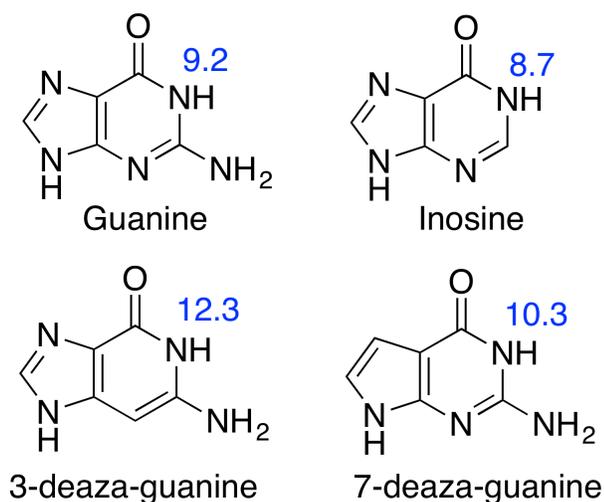


Figure 5.3: Chemical structures of guanine and several chemically modified guanine molecules studied in this work. Experimental pK_a values at the N1 position (taken from Refs. 17,18) are shown.

In order to get a more detailed picture of the Mg^{2+} -guanine interaction and its effect on the pK_a , we designed model compounds that mimic the Mg^{2+} -guanine complex in HHR crystal structure and performed density functional electronic structure calculations. To validate the computational protocol and quantum chemical model chemistry, we calculated the pK_a shifts of three chemically modified guanines: substitution of N3/N7 to CH which makes 3/7-deaza-guanine and replacement of exocyclic NH₂ with H which yields inosine (Figure 5.3).

From Table 5.1, the calculated pK_a shift of all three chemically modified guanines agree very well with experimental values (maximum error of 0.4 pK_a units).

Next, we examined several Mg^{2+} -guanine model complexes that mimicked the crystallographic metal ion binding mode in which the hexacoordinated Mg^{2+} has inner sphere contact with N7 and outer sphere coordination with O6. We also picked two of the three chemically modified species (3-deaza-guanine and inosine) and built complexes with Mg^{2+} in the same way. These two species were chosen because, unlike the 7-deaza modification, neither directly alters the chemical environment of the Hoogsteen face, therefore preserving the observed binding mode of the native guanine. For 7-deaza-guanine, it's expected that the Mg^{2+} binding would be considerably disrupted. As seen in Table 5.1, the magnitude of pK_a shifts induced by Mg^{2+} (≈ -7 pK_a units) is quite large, making the N1 position considerably more acidic. This large shift agrees with previous arguments^{206,207,219} that metal-ion binding could induce large pK_a shifts of nucleobase residues. The experimentally observed pK_a shift of G12 in the ribozyme environment is a much smaller shift toward neutrality (≈ -1.2 pK_a units). As will be discussed in more detail in the next section, the accumulated negative charge of the ribozyme active site in the hammerhead ribozyme leads to a considerable positive pK_a shift that requires divalent metal ion binding at the G site in order to offset. The main point of the current section is to establish a baseline for the expected pK_a shifts that a Mg^{2+} would be expected to induce if it were bound to an isolated guanine (or modified guanine) nucleobase in solution, before making predictions about the pK_a shifts in the more complex active site of the HHR.

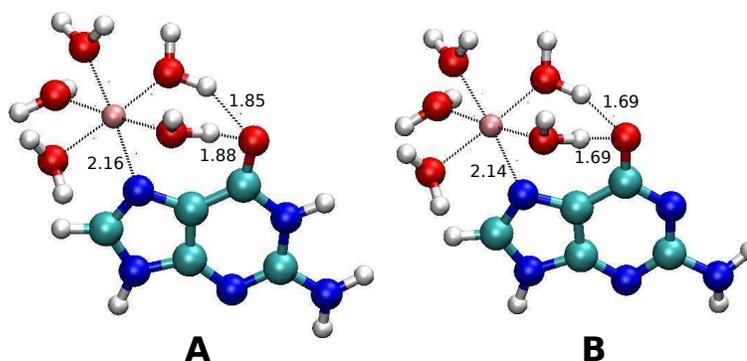


Figure 5.4: Optimized geometries of Mg^{2+} -guanine (A) and Mg^{2+} -deprotonated guanine (B) complexes. Selected bond lengths shown are in Å.

Table 5.1: Calculated and experimental pK_a shifts of guanine, chemically modified guanine and Mg^{2+} -guanine complexes. The calculated pK_a shifts have been corrected for Mg^{2+} -guanine binding affinity and Mg^{2+} concentration (0.01 M) using Eq. 5.2. Error estimates are not available because the method used to obtain those values are deterministic and do not involve sampling.

Model	pK_a shift (calc.)	pK_a shift (expt.)
guanine (ref.)	–	–
inosine	-0.3	-0.5 ¹⁷
3-deaza-guanine	3.2	3.1 ¹⁸
7-deaza-guanine	1.5	1.1 ¹⁸
Mg^{2+} -(H_2O) ₅ -guanine	-7.0	–
Mg^{2+} -(H_2O) ₅ -inosine	-7.0	–
Mg^{2+} -(H_2O) ₅ -3-deaza-guanine	-4.4	–
HHR G12	–	-1.2 ¹¹

To examine the effect of Mg^{2+} binding on the electronic structure of guanine, we performed NBO analysis to calculate the partial atomic charges in the neutral and deprotonated species (Table 5.2). Of the key atoms listed, the largest negative charge is seen to reside at the O6 position in both protonation states and in the presence and absence of Mg^{2+} . The O6 position also exhibits the greatest increase in negative charge upon deprotonation, both in the presence and absence of Mg^{2+} . The O6 position is also seen to considerably increase in negative charge, for both neutral and deprotonated guanine, upon Mg^{2+} binding, but not as dramatically as at the N7 position where the ion has direct inner-sphere coordination. Overall, the charge transfer from guanine to the hydrated Mg^{2+} ion is relatively modest for both neutral and deprotonated (anionic) guanine (0.17 and 0.23 e , respectively). These results indicate that the interaction between Mg^{2+} and guanine is mostly electrostatic in nature, providing justification for the use of fixed-charge molecular mechanical (MM) force fields to model the Mg^{2+} -guanine interaction that will be introduced in the next subsection.

Table 5.2: Partial atomic charges of selected atoms in guanine and Mg^{2+} -guanine complexes derived from NBO analysis.

Model	N1	H1	N3	O6	N7	Guanine
Guanine	-0.663	0.472	-0.603	-0.687	-0.510	0.000
Mg^{2+} -(H_2O) ₅ -guanine	-0.635	0.484	-0.582	-0.743	-0.582	0.168
(Difference)	(0.028)	(0.012)	(0.021)	(-0.056)	(-0.071)	(0.168)
Guanine(-)	-0.716	-	-0.659	-0.792	-0.533	-1.000
Mg^{2+} -(H_2O) ₅ -guanine(-)	-0.651	-	-0.617	-0.830	-0.602	-0.767
(Difference)	(0.065)	-	(0.042)	(-0.038)	(-0.069)	(0.233)

5.3.2 pK_a shift of G12 in HHR

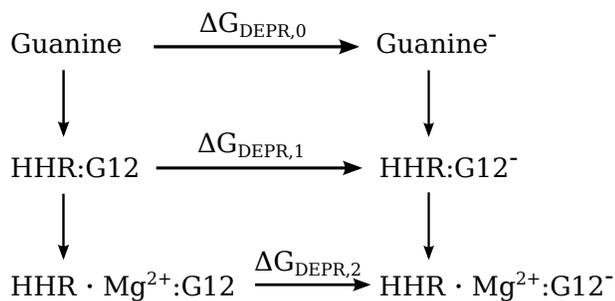


Figure 5.5: Thermodynamic cycle used in TI calculations.

Table 5.3: Deprotonation free energies (kcal/mol), pK_a s and pK_a shifts (in pK_a units) of G12 in HHR from different sets of TI simulations and experiments.

Model	ΔG_{DEPR}	pK_a	pK_a shift
Guanine (ref.)	-114.1 ± 0.5	9.2	-
HHR:G12	-108.9 ± 0.3	12.9 ± 0.2	3.7 ± 0.2
$\text{HHR} \cdot \text{Mg}^{2+}:\text{G12}$	-115.8 ± 0.5	8.0 ± 0.4	-1.2 ± 0.4
Experiments ¹¹	-	8.0	-1.2

To better characterize the effects of the complicated enzyme environment, molecular dynamics (MD) simulation on the full HHR was conducted and the pK_a shift of guanine N1 position in G12 was then evaluated using thermodynamic integration (TI) technique. The

reference system was the deprotonation of a single guanine nucleotide in solution which has known experimental pK_a of 9.2. For the HHR system, both HHR:G12 and HHR · Mg^{2+} :G12 were simulated. The two pK_a shift values were then determined using Eq. 5.1. This approach has been adopted to predict pK_a shifts in ribozymes induced by the enzyme environments^{87,89} and a closely-related protein enzyme system where the leading factor of pK_a shift was also a Mg^{2+} ²²⁰ in previous works and was shown to be giving reasonable results.

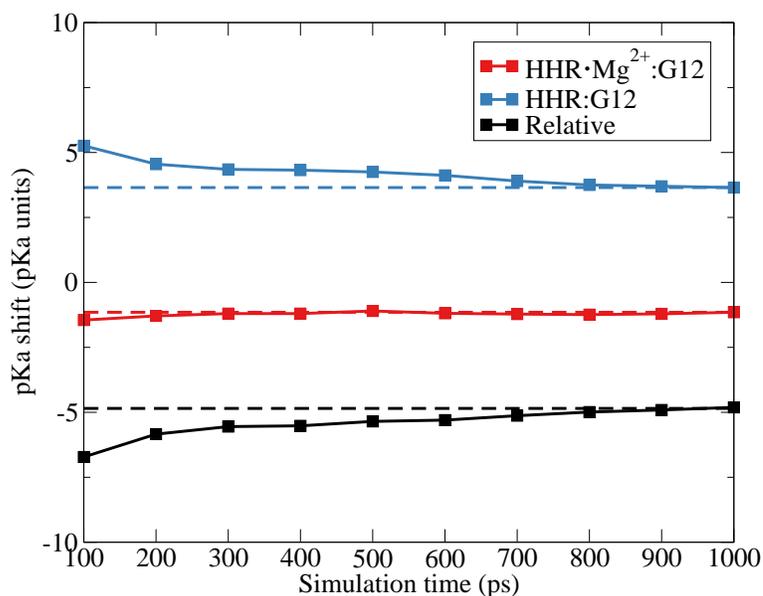


Figure 5.6: Convergence of pK_a shift values from TI simulations. Filled squares connected by solid lines are the pK_a shift values evaluated using all available data at certain simulation time, with an increment of 100 ps per point. Dashed lines are drawn to help show the convergence.

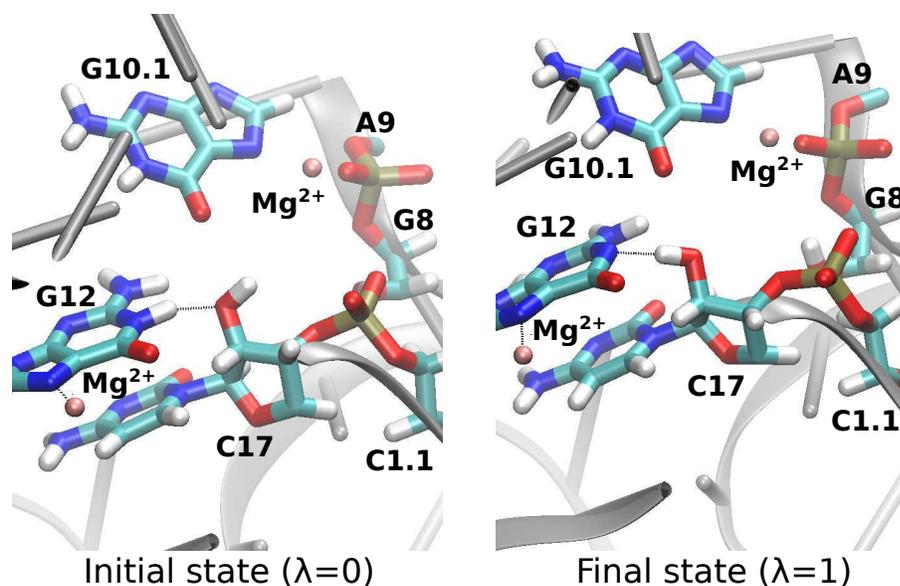


Figure 5.7: Representative active site conformation of the initial (left) and final (right) states from the free energy (TI) simulation of HHR · Mg²⁺:G12. The conformations from the simulation of HHR:G12 are very similar, and therefore not shown here. For clarity, water molecules and some other atoms/residues are not displayed. White, cyan, blue, red, pink and gold spheres stand for H, C, N, O, Mg and P atoms, respectively.

As seen in Table 5.3, the simulations predicted that in HHR · Mg²⁺:G12, the pK_a of N1 in G12 is shifted down by -1.2, while in HHR:G12, the pK_a is shifted up by 3.7 units. The convergence of the TI simulations in terms of the target observable pK_a shifts was measured (Figure 5.6) and showed that the simulations reached reasonable convergence after 1 ns, which was similar to the timescales investigated in related work^{87,220,89}. Experimental activity-pH profiles indicate that the apparent pK_a of the general base in HHR is around 8 in the presence of Mg²⁺¹¹. If guanine is assumed to be the general base, and the most straight forward interpretation of the apparent pK_a values are made, this would imply an expected shift of the pK_a of G12 by ~ -1.2 units.

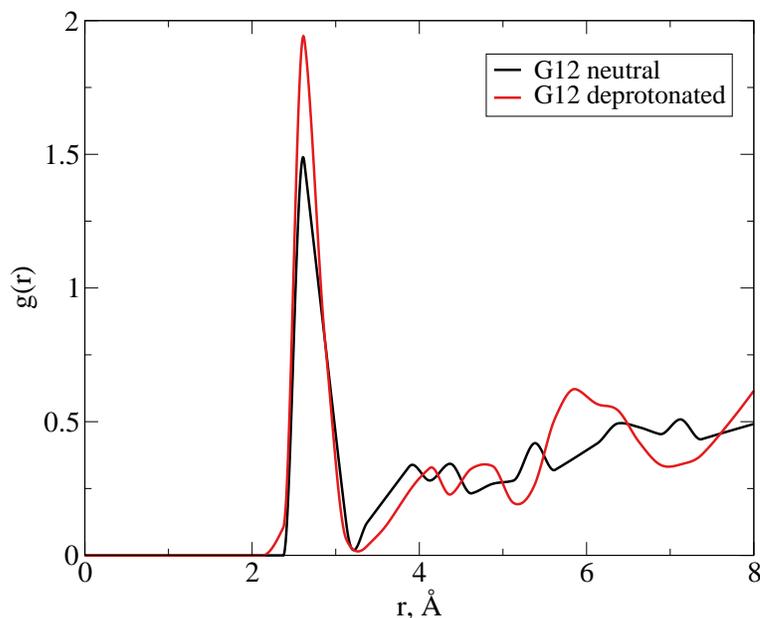


Figure 5.8: Radial distribution function (RDF) of water oxygens around the O6 position of neutral and deprotonated G12 in HHR. Both sets of data are extracted from 1 ns of MD simulation of HHR:G12 with G12 in neutral and deprotonated forms, respectively. RDF data points are calculated using window size of 0.2 Å and are interpolated using the Akima spline.

As mentioned above, an electron density peak was found in the vicinity of G12 in the previous crystal structure of HHR obtained at pH 6.5¹⁹⁵ (PDB code 2OEU) and was interpreted as a solvent water molecule. In the recent work¹¹, the G12-bound Mg²⁺ was found in the crystal obtained at pH 8.0 but not in the one obtained at pH 5.0. Here, we analyzed the solvent distribution around the O6 position of G12 from the HHR:G12 simulations by calculating the radial distribution functions (RDFs). As seen in Figure 5.8, there is a clear peak at around 2.6 Å in both RDFs, which is consistent with the previous structure in which the oxygen in the G12-bound water is 2.54 Å away from the O6 of G12. Also, in the RDF where G12 is deprotonated, the height of the peak is greater than in the other one, which suggests that the deprotonated G12, which is carrying a -1 charge, needs stronger electrostatic interactions to stabilize the negative charges that are building up on O6 and N7, than the neutral form. These results indicate that the Mg²⁺-G12 binding in HHR might have pH-dependence, which could be the reason why the G12-bound Mg²⁺ is

only found in high-pH crystal structures but not in the low-pH ones.

Taken together, the results reported here are consistent with a model whereby a Mg^{2+} ion binds at the G site and results in an overall pK_a shift of G12 toward neutrality by 1.2 pK_a units. In the absence of the metal ion binding at the G site, the pK_a of G12 is predicted to be considerably shifted to higher pK_a values (12.9), making it highly unlikely to be activated as the general base in the biologically relevant pH range. This scenario would contradict the interpretation of activity-pH data in terms of apparent pK_a values. As our calculated result is in striking agreement with experiment suggests that the apparent pK_a values observed experimentally may have this straight forward interpretation.

These results are consistent with previous calculations that indicate that HHR folds to form an electrostatically strained active site, which acts as an electronegative recruiting pocket for a threshold cationic charge required for efficient activity¹⁹⁸. If the local environment around G12 in HHR is overall negatively charged, then it's expected that the pK_a of G12 will be shifted up in HHR compared to in aqueous solution. By recruiting divalent metal ions, this shift is offset and even reversed. By comparing the pK_a shifts with/without Mg^{2+} in Table 5.3, we can estimate the net effect of Mg^{2+} on the pK_a of G12 to be $(-1.2) - 3.7 = -4.9$, which is in qualitative agreement with our DFT result -7.0. In the hepatitis delta virus (HDV) ribozyme, which has a similar divalent metal ion requirement under physiological conditions, a pK_a shift of similar magnitude (~ -4 units) on the 2'-hydroxyl nucleophile induced by Mg^{2+} was predicted by 3D-RISM calculations⁸⁸. Further, Amaro and co-workers have reported the Mg^{2+} -induced pK_a shift of lysine K82 in guanylyltransferase mRNA capping enzyme to be -4.2²²⁰ which is very close to the result reported here using a similar TI approach. Rosta and co-workers have also observed that Mg^{2+} plays the role of altering pK_a in dUTPase which catalyzes phosphate hydrolysis^{221,222}.

5.3.3 Free energy profile for the general base step

From TI-MD simulations, we found that the G12-bound Mg^{2+} in HHR could activate the general base G12 by down-shifting its pK_a to facilitate its deprotonation. Nonetheless, since the next step after general base activation is the proton transfer from nucleophile to general base, one might expect that having a Mg^{2+} bound at G12 could make this step

thermodynamically less favorable because Mg^{2+} binding to deprotonated G12 makes it a worse proton acceptor. One possibility is that the Mg^{2+} could “leave the scene” after the deprotonation of G12 to avoid the energy penalty in the following step. However, in the crystal structure of the vanadate transition state mimic¹² (PDB code 5EAQ) which mimics the state after general base proton transfer step is completed, the Mg^{2+} was still found to be bound at G12. This suggested that the Mg^{2+} might be bound at G12 all the time despite the energy penalty of the general base proton transfer step, which could potentially be compensated by providing electrostatic stabilization to the overall negatively-charged active site.

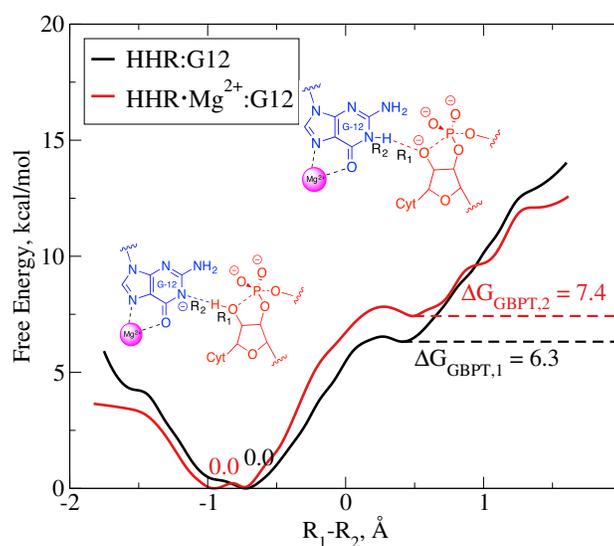


Figure 5.9: Potential of mean force (PMF) for the general base proton transfer (GBPT) reaction in $\text{HHR} \cdot \text{Mg}^{2+}:\text{G12}$ (red) and $\text{HHR}:\text{G12}$ (black) generated by *ab initio* QM/MM umbrella sampling. The two chemical structures depict the reactant (left) and product (right) states. Reaction coordinate is the difference ($R_1 - R_2$) between the distance from C17:O2' to C17:HO2' (R_1) and the distance from G12:N1 to C17:HO2' (R_2). Free energies of the reactant and product states are marked.

Table 5.4: Summary of predicted free energy costs in different steps for HHR:G12 and HHR · Mg²⁺:G12. (Relative) ΔG_{DEPR} values are taken from Table 5.3 while ΔG_{GBPT} values are from QM/MM free energy profiles as labeled in Figure 5.9. All numbers are in kcal/mol. Error estimates for ΔG_{DEPR} values are from Table 5.3 while for ΔG_{GBPT} are from bootstrapping (see supporting information for details). Standard errors of ΔG_{TOTAL} are obtained using the propagation rule.

Model	ΔG_{DEPR}	ΔG_{GBPT}	ΔG_{TOTAL}
HHR:G12	6.7 ± 0.6	6.3 ± 0.2	13.0 ± 0.6
HHR · Mg ²⁺ :G12	0.0 ± 0.0	7.4 ± 0.2	7.4 ± 0.2

To evaluate the energy penalty induced by Mg²⁺, we carried out *ab initio* QM/MM umbrella sampling simulations of the general base proton transfer step for both HHR:G12 and HHR · Mg²⁺:G12. As seen in Figure 5.9, the reaction in HHR · Mg²⁺:G12 did have a higher free energy cost. However, the decrease in the free energy cost of the G12 deprotonation induced by Mg²⁺ was more than enough to compensate for this (Table 5.4). Overall, the free energy required to get the system into the state that is ready for phosphoryl transfer is 5.6 kcal/mol less when a G12-bound Mg²⁺ is present, which again supports the existence and the catalytic effect of that Mg²⁺ in HHR.

5.3.4 Effect of 6-thio substitution on G12 and rescue effects

All the results and discussions above support the hypothesis that the G-site Mg²⁺ could facilitate the catalysis by down-shifting the pK_a of G12. To further test the hypothesis and make predictions that could be directly tested by experiments, we examine the effects of changing a guanine to a 6-thioguanine. The 6-thioguanine molecule (sometimes referred to as tioguanine) has been used in medications for the treatment of leukemia and other diseases, but has not, to our knowledge, been used extensively at a mechanistic probe in ribozymes. The pK_a of 6-thioguanine has been experimentally measured to be 8.3²¹¹, almost a full unit lower than the pK_a of guanine (9.2). In a mechanistic scenario whereby there is no divalent metal ion bound at the G-site, one would expect the lower pK_a of 6-thioguanine would lead to slightly increased activity at neutral pH due to the higher probability of being

in deprotonated form and able to accept a proton from the nucleophile in the general base step, and the activity-pH profile would be accordingly shifted. If, however, a divalent metal ion did bind at the G-site, thio substitution at the 6 position could create a thio/rescue effect scenario.

Table 5.5: Comparison of binding free energies (kcal/mol) between different divalent metal ions and guanine/6-thioguanine in both neutral and deprotonated forms. Numbers in the first column are converted from experimental binding affinities^{8,9}. In each row, the numbers are normalized according to the corresponding experimental values in the first column.

Metal	Guanine	Guanine ⁻	6-thioguanine	6-thioguanine ⁻
Mg ²⁺	-0.3	-11.7	6.4	5.0
Mn ²⁺	-0.6	-12.4	0.7	-0.7
Cd ²⁺	-2.0	-10.8	-3.2	-4.5

In order to test this, we employed electronic structure calculations to obtain the binding free energies of Mg²⁺, Mn²⁺ and Cd²⁺ ions to guanine and 6-thioguanine, both in neutral and deprotonated (at the N1 position) form. Results are shown in Table 5.5, including normalization adjustments such that binding free energies are relative to guanine in its normal protonation state at neutral pH, and the column corresponding guanine is taken from the experimental binding affinities^{8,9}. As one can see, the Mg²⁺-6-thioguanine binding is unfavorable relative to Mg²⁺-guanine binding, and suggests that the G-site Mg²⁺ could be knocked out upon 6-thio substitution at G12. Mn²⁺ and Cd²⁺, which are softer, more thiophilic ions, show more favorable binding to 6-thioguanine than Mg²⁺, which implies a partial rescue effect to the 6-thio substitution at G12 upon replacing Mg²⁺ with Mn²⁺ and particularly Cd²⁺. This is an experimentally testable prediction that could help to further reconcile the role of divalent metal ions in general base activation.

5.4 Conclusion

This work presents a series of quantum chemical calculations and molecular simulations to probe the activation mode of general base in hammerhead ribozyme (HHR). The hypothesis, motivated by recent crystallographic data, is that the presumed general base G12 is

activated by a Mg^{2+} ion causing a pK_a shift toward neutrality that would allow it to have increased probability of being in an ionized state where the N1 position is able to extract a proton from the 2'OH nucleophile. The electronic structure calculations on small model systems suggest that a Hoogsteen face-bound Mg^{2+} ion could greatly down-shift the pK_a of the N1 position in guanine, primarily due to electrostatic interactions. To consider the effects of the full ribozyme environment, free energy simulations coupled with thermodynamic integration have been performed, and strongly support a model whereby the pK_a of G12 in HHR is shifted by a Mg^{2+} ion bound at the experimentally observed G site by -1.2 units, in agreement with the apparent pK_a value derived from measured activity-pH profiles. QM/MM simulations of the general base proton transfer reaction further support the catalytic significance of this metal ion binding mode, and role of the G-site Mg^{2+} . Based on binding free energy calculations, 6-thio substitution on G12 is predicted to reduce the catalytic activity but could be rescued by replacing Mg^{2+} with more thiophilic metal ions, which could be tested by future experiments. The suggested role of metal ions in the activation of a guanine general base may represent a general catalytic strategy used by other RNA enzymes.

5.5 Supporting Information for: Metal Ion-induced Activation of the Catalytic General Base in Hammerhead Ribozyme Self-cleavage

5.5.1 Derivation of the correction for ion concentration and binding affinity in DFT calculations of Mg^{2+} -guanine model complexes

In the DFT calculations of the Mg^{2+} -induced pK_a shift of guanine in model complexes, the free energy differences between species need to be further corrected by accounting for the metal ion concentration and the metal-guanine binding affinity to yield the prediction of pK_a shift values. Consider the kinetic model below,

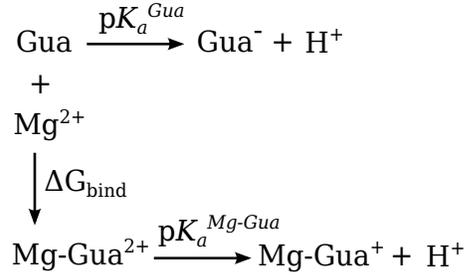


Figure 5.10: Kinetic model used in the $\text{p}K_a$ shift correction. $\text{p}K_a$ values are directly converted from the free energy differences in DFT calculations while ΔG_{bind} is the experimentally measured binding affinity.

we have

$$K_a(\text{Mg} - \text{Gua}, \text{corr.}) = \frac{c_{\text{Gua}} \times 10^{-\text{p}K_a(\text{Gua})} + c_{\text{Mg-Gua}} \times 10^{-\text{p}K_a(\text{Mg-Gua}, \text{uncorr.})}}{c_{\text{Gua}} + c_{\text{Mg-Gua}}} \quad (5.10)$$

in which all concentrations have been divided by the standard molar concentration 1.0 M and therefore are dimensionless. Since

$$c_{\text{Mg-Gua}} = c_{\text{Gua}} \times c_{\text{Mg}} \times e^{-\beta \Delta G_{\text{bind}}} \quad (5.11)$$

where $\beta = \frac{1}{k_B T}$, Eq. 5.10 becomes

$$\begin{aligned}
K_a(\text{Mg} - \text{Gua}, \text{corr.}) &= \frac{1}{c_{\text{Gua}} + c_{\text{Gua}} \times c_{\text{Mg}} \times e^{-\beta \Delta G_{\text{bind}}}} \times (c_{\text{Gua}} \times 10^{-\text{p}K_a(\text{Gua})} \\
&+ c_{\text{Gua}} \times c_{\text{Mg}} \times e^{-\beta \Delta G_{\text{bind}}} \times 10^{-\text{p}K_a(\text{Mg-Gua}, \text{uncorr.})})
\end{aligned} \quad (5.12)$$

where c_{Gua} could be eliminated from both numerator and denominator, which leads to

$$K_a(\text{Mg} - \text{Gua}, \text{corr.}) = \frac{10^{-\text{p}K_a(\text{Gua})} + c_{\text{Mg}} \times e^{-\beta \Delta G_{\text{bind}}} \times 10^{-\text{p}K_a(\text{Mg-Gua}, \text{uncorr.})}}{1 + c_{\text{Mg}} \times e^{-\beta \Delta G_{\text{bind}}}} \quad (5.13)$$

that gives Eq. 5.2 in the main text. The corrected Mg^{2+} -induced $\text{p}K_a$ shift on guanine is therefore

$$\Delta pK_a(Mg, corr.) = -\log_{10}\left(\frac{10^{-pK_a(Gua)} + c_{Mg} \times e^{-\beta\Delta G_{bind}} \times 10^{-pK_a(Mg-Gua, uncorr.)}}{1 + c_{Mg} \times e^{-\beta\Delta G_{bind}}}\right) - pK_a(Gua)$$
(5.14)

5.5.2 Choice of the basis set in QM/MM simulations

Table 5.6: M06-2X single point energies computed using different basis sets at the geometries optimized using M06-2X/6-31+G(d) level of theory. G and MG stands for guanine and Mg²⁺-bound guanine, respectively. Absolute energies (column 2-5) are in hartree while relative energies (column 6) are in kcal/mol.

Basis Set	E(G)	E(G ⁻)	E(MG ²⁺)	E(MG ⁺)	$\Delta\Delta E$
6-31G(d)	-542.3443	-541.7819	-1124.0730	-1123.7497	-150.07
6-31G(d,p)	-542.3580	-541.7923	-1124.1382	-1123.8118	-150.10
6-311G(d,p)	-542.4946	-541.9351	-1124.4169	-1124.0938	-148.34
6-31+G(d)	-542.3653	-541.8196	-1124.1038	-1123.7885	-144.61
6-31+G(d,p)	-542.3788	-541.8300	-1124.1669	-1123.8486	-144.65
6-311+G(d,p)	-542.5073	-541.9588	-1124.4328	-1124.1144	-144.40
6-31++G(d)	-542.3655	-541.8198	-1124.1047	-1123.7895	-144.65
6-31++G(d,p)	-542.3790	-541.8302	-1124.1677	-1123.8495	-144.70
6-311++G(d,p)	-542.5074	-541.9590	-1124.4332	-1124.1149	-144.45

The recent developed ambient potential Ewald method¹⁵⁹ that was implemented in Amber allows much more efficient *ab initio* QM/MM simulations. However, it's still important to carefully choose the level of theory to balance accuracy and performance since using *ab initio* calculations in dynamics is still slow in general. In particular, the choice of basis set for the QM calculations usually affects both accuracy and performance. Since our QM/MM simulations in this work were to estimate the free energy cost of a proton transfer reaction in which a deprotonated guanine (with/without Mg²⁺ bound) was the proton acceptor, we first performed some single point energy calculations to evaluate the proton affinity of

deprotonated guanine with/without Mg^{2+} bound. As seen in Table 1, the diffuse functions seem to play a much bigger role than the polarization and additional split-valence functions. In terms of the relative proton affinity $\Delta\Delta E$ value (which is really what we care about in the QM/MM simulations in this work), 6-31+G(d) basis set could already give very close answer comparing to the largest basis set 6-311++G(d,p). Therefore, we chose 6-31+G(d) as the basis set for QM/MM simulations.

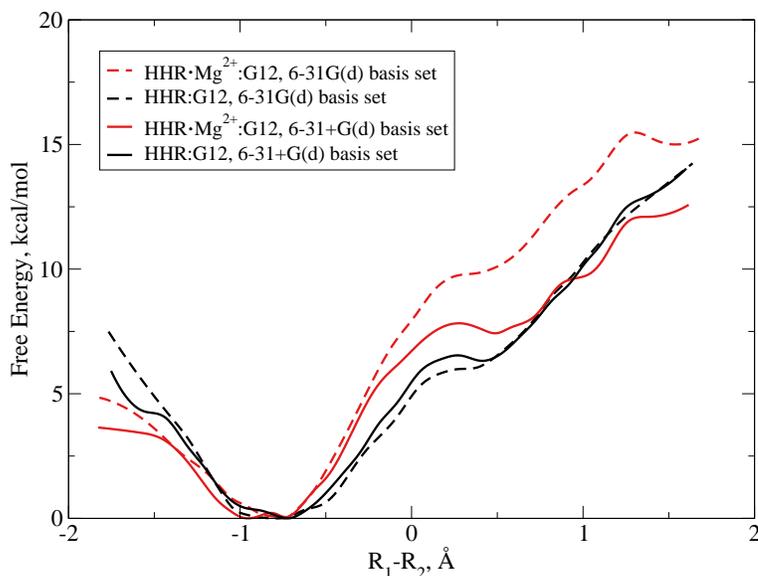


Figure 5.11: QM/MM free energy profile of the general base proton transfer using different basis sets.

We then conducted QM/MM simulations of the general base proton transfer reaction, using PBE0 density functional with 6-31+G(d) basis set and a smaller 6-31G(d) basis set for comparison. As seen in Figure 5.11, the 6-31G(d) profiles did not show clear minima for the product states while the 6-31+G(d) did. A possible explanation is that in the product state, the negative charge on the 2'O nucleophile could not be accurately modeled without diffuse functions. The overall free energies are also higher in the 6-31G(d) profiles, especially for the one with Mg^{2+} . Therefore, we used the data from 6-31+G(d) basis set to give the final evaluation of the free energy cost of the reaction. The simulation using 6-31+G(d) basis set is 2~3 times slower than using 6-31G(d) basis set.

5.5.3 Detail of error estimation in QM/MM free energy profiles

For both simulations (with/without Mg^{2+} bound at G12), bootstrapping method was used to give the error estimates. For each simulation, 1000 data points of the reaction coordinate value was saved in the 10 ps production run. In each bootstrapping step, one-fourth (250 points) of the data was randomly chosen and put in vFEP⁶⁵ to generate a free energy profile. The step was repeated for 100 times for both simulations. Standard error of the energy difference between the two states (reactant and product) were then estimated from the 100 samples for both profiles.

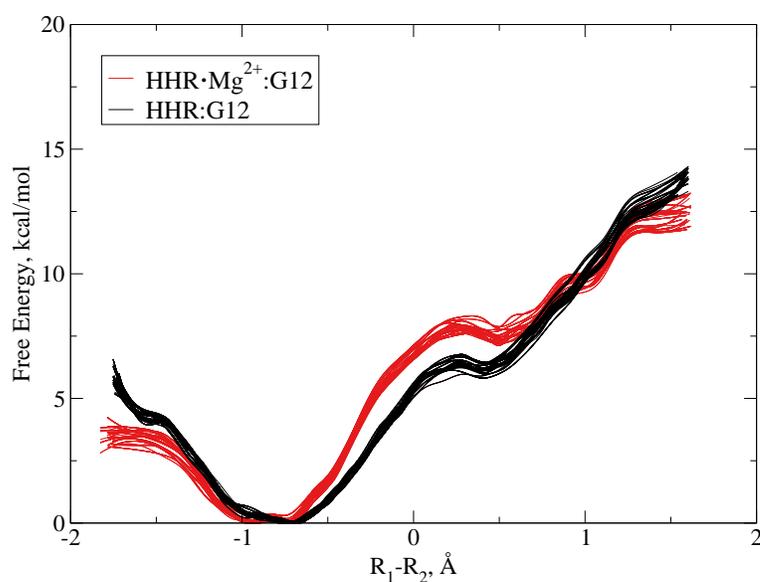


Figure 5.12: All the free energy profiles generated in bootstrapping.

Chapter 6

New Insights and Applications Enabled by Multi-dimensional Replica Exchange Molecular Dynamics Simulations

We demonstrate a new framework that enables highly flexible multi-dimensional replica exchange molecular dynamics (REMD) simulations with customized choice of order parameters in each dimension. This framework opens the door to new types of enhanced sampling applications that require non-local information about each order parameter in different simulations to be kept in equilibrium to achieve reliable, converged results. We apply the method to the thermodynamic decomposition of 2D free energy surfaces for the alanine dipeptide into enthalpic and entropic contributions, and the prediction of the effect of salt concentration on the conformational ensemble for a model DNA helix-junction-helix system.

Customizable multi-dimensional replica exchange molecular dynamics simulations enabled by recently-developed software package RepEx is introduced. Investigations of backbone torsion free energy profiles in alanine dipeptide and ionic strength-dependent conformational changes in joint DNA helices are discussed as proof-of-concept scientific applications.

6.1 Introduction

Replica exchange molecular dynamics (REMD) has evolved into one of the most powerful enhanced sampling techniques used in modern biomolecular simulations. Since the original work by Sugita and Okamoto⁶⁷ that introduces replica exchange in the temperature dimension, the scope of REMD simulations has been extended to include generalized coordinates in umbrella sampling REMD^{71,223,224}, thermodynamic coupling parameters in FEP/alchemical Hamiltonian REMD^{225,226,227,228,229,230}, system variables such as

pressure⁶⁹ and pH^{231,70,232,233,234}, as well as others^{66,235,236}. At the same time, great effort has been made to overcome technical challenges associated with generalizing REMD to 2D^{72,237,238,239,240,234} and even 3D⁷³.

There are emerging examples of important biomolecular problems that involve multi-state equilibria, and for which the interpretation of experiments requires scanning control variables such as temperature, ionic conditions, and pH in addition to geometrical or Hamiltonian order parameters^{72,241,242,243,238,237,244,239,240,73,234,245,246}. These applications have the added challenge that sampling along the space of the order parameters needs to be statistically converged at all points. Here, the REMD method offers the added advantage that equilibrium between simulations is enforced through the exchange sampling. An illustrative example is the “problem space” associated with biocatalysis whereby conformational equilibria, metal ion binding and protonation events lead to an active state that is able to catalyze the chemical steps of the reaction⁵⁹. Thus, these applications require not only the elucidation of the free energy landscape of the chemical reaction itself^{247,248,249}, but also the characterization of the probability of finding the system in the catalytically active state as a function of system variables²⁵⁰. To address these novel applications requires a flexible and efficient multi-dimensional REMD framework that can be used for both system control variables and generalized coordinates.

In this paper, we outline a new implementation of customized multi-dimensional REMD in the RepEx package²⁵¹ and demonstrate its use in novel applications. First, the 2D conformational landscape for the alanine dipeptide is studied as a function of temperature so as to allow the thermodynamic free energy decomposition into enthalpic and entropic contributions. Second, the effect of the ion atmosphere on the conformational ensembles of DNA helices in a model helix-junction-helix system is examined, and the Au-SAXS distance profile for different Au particle positions is predicted. These examples highlight the new insights and applications that are enabled by multi-dimensional REMD.

6.2 Computational Methods

6.2.1 Implementation of customized multi-dimensional REMD

Multi-dimensional REMD is a potentially powerful technique for enhanced sampling in complicated systems. However, its current implementation in most MD packages is highly limited. Here we present our development of the RepEx package²⁵¹ (<https://github.com/radical-cybertools/radical.repex>) as a platform that allows efficient execution of customized REMD simulations on high-performance computing clusters. With partial inheritance from the ASyncRE package²⁵², RepEx separates exchange steps from MD runs such that it can support different MD engines (currently Amber¹⁵⁵ and NAMD²⁵³), and introduces several new features. First, our implementation allows any number of RE dimensions, and has been recently demonstrated in 3D-REMD simulation⁷³ of the conformational space of solvated uridine using the ASyncRE package. Second, we enhance flexibility by enabling a range of order parameters, including temperature (T), umbrella sampling (U) and salt concentration (S). The modular software design facilitates the extension to new order parameters such as pH or Hamiltonian coupling parameters. In terms of usability, several features have also been designed and implemented into RepEx to improve it. First, all the parameters that control the simulation are contained in a single JSON file, which is easy to read and to modify. The generation of different input parameters for different replicas such as optimal temperature spacing in temperature-REMD and files that specify biasing potentials for Hamiltonian-REMD, will be automated by the program according to the user input in the JSON file. Also, the program will take care of submitting the job, monitoring the status of the job and organizing the output files for each individual replica in a systematic manner, which could reduce the workload of the user.

Replica exchange is normally implemented in a way that every replica only attempts to exchange with its neighbors. However, in RepEx, we adopt the Gibbs sampling approach of infinite swapping²²³ which attempts to sample from the permutations of the entire state index vector instead of just the swapping between neighboring pairs. It has been shown that independence sampling scheme could improve sampling efficiency over traditional nearest-neighbor exchange methods^{254,255,223,256}. For multi-dimensional REMD in RepEx, exchange

attempts are proposed within one dimension at a time.

6.2.2 Computational details

Backbone torsion free energy profile of alanine dipeptide

For each replica, the physical system was built by capping a single alanine residue with acetyl and N-methylamide groups on the N- and C-terminal positions, respectively. The compound was described by AMBER-ff14SB force field^{47,257,213,214,156} and was solvated in a cubic water box composed of 953 TIP3P²⁵⁸ water molecules. Long-range electrostatics were treated with particle mesh Ewald method^{52,53}. A 8 Å cutoff was applied to both real-space electrostatics and Lennard-Jones potentials. Langevin dynamics with a 5 ps⁻¹ collision frequency was performed to approximate the canonical ensemble.

3D-REMD was performed using order parameters of temperature (T) and umbrella sampling (U) with biasing potentials of the ϕ and ψ torsion angles (as shown in Figure 6.1). In the T dimension, 11 windows were chosen from 273K to 373K by geometrical progression. In both U dimensions, 8 windows were chosen uniformly between 0° and 360° where each window corresponds to a harmonic restraint centered on it with a force constant of 0.002 kcal·mol⁻¹·degree⁻². The total number of replicas is therefore 11×8×8=704. Each replica was previously equilibrated for 1 ns. In the production run, we set the exchange attempt interval (cycle) to be 10000 steps (20 ps) and in a 8-hour run with 720 cores (45 nodes) on the Stampede cluster, the simulation finished 75 cycles (1.5 ns). The average acceptance ratios of exchange attempts are 10.1 % for T dimension and 23.5 % for U dimensions. Free energy profiles were then generated from the last 1 ns of production data (repeated for 4 times) using the maximum likelihood approach implemented in the vFEP package^{65,249}. In the thermodynamic decomposition analysis, only every other windows (6 out of 11) in the T space were selected to improve the accuracy of finite difference calculations.

Effect of salt concentration on the conformational ensemble of DNA helix-junction-helix

For each replica, the physical system DNA helix-junction-helix (HJH) model was built by inserting one polyethylene glycol nonamer (PEG9) tether after the 16th residues of both strands of a 32-base pair B-DNA double helix. The sequence is

5'-CAGATACACGAACGAG-PEG9-GCACGAGTCTATGTAC-3',

3'-GTCTATGTGCTTGCTC-PEG9-CGTGCTCAGATACATG-5'.

Experimentally, Au nanoparticles were attached to different pairs of thymine residues in the system to probe the structure and dynamics²⁵⁹. Here, we use 3 experimentally relevant pairs (T2-T28, T4-T26 and T6-T24, Daniel Herschlag, personal communication) and pick the distance between the two methyl carbon atoms in each pair as the umbrella sampling variable in 3 sets of independent simulations. The entire system has 2072 atoms and was described using AMBER ff14SB force field^{47,257,213,214,156}. The Hawkins, Cramer, Truhlar pairwise generalized Born (GB) implicit solvent model^{260,261,262} was adopted to treat solvation effects as it's recommended for nucleic acids simulations²⁶³. Langevin thermostat with a 5 ps^{-1} collision frequency was applied to the system while the SHAKE algorithm²⁶⁴ was used to constraint bonds involving hydrogens.

3D-REMD was performed using order parameters of temperature (T), salt concentration (S) and umbrella sampling (U) with biasing potentials along the distance (T2-T28, T4-T26, T6-T24) coordinates. In the T dimension, 8 windows were chosen from 288K (the experimental condition) to 400K to enhance sampling. In the S dimension, 4 salt concentration values were chosen uniformly between 0.0 M and 0.15 M. In the U dimension, 12 windows were placed evenly in the range of interest (70 to 97.5 Å for T2-T28, 60 to 87.5 Å for T4-T26, 50 to 77.5 Å for T6-T24) with a 2.5 Å interval where each window corresponds to a harmonic restraint centered on it with a force constant of $0.2 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$. The total number of replicas is therefore 384. The average acceptance ratios of exchange attempts are 3.6 % for T dimension, 47.2 % for S dimension and 17.6 % for U dimension. For each set of simulations, after 2 ns of equilibration, we performed production simulation (repeated for 4 times) of 1.5 ns with exchange interval of 20 ps and used the final 1 ns data for the free

energy calculation.

Decomposition of free energy into enthalpy and entropy

In this work, we adopt the finite difference (FD) approach to decompose free energy into enthalpy and entropy contributions, which has been used previously^{265,266,267} and has recently been applied to analyze REMD simulations²⁶⁸. In the FD approach, the potential of mean force (PMF) along the reaction coordinate ξ is first calculated using the probability distribution from the simulation ensemble:

$$W(\xi) = -k_B T \log(P(\xi)) \quad (6.1)$$

Then, entropy is calculated as the negative derivative of free energy with respect to temperature through FD:

$$S(\xi) = -\frac{W(\xi, T + \delta T) - W(\xi, T - \delta T)}{2\delta T} \quad (6.2)$$

The Enthalpy is then simply:

$$H(\xi) = W(\xi) + TS(\xi) \quad (6.3)$$

6.3 Results and Discussion

6.3.1 New insight: Backbone torsion enthalpy and entropy of alanine dipeptide

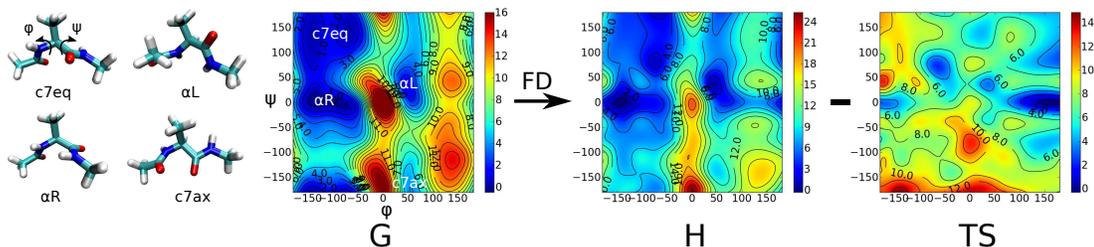


Figure 6.1: (Left) Illustration of the ϕ and ψ torsion angles and major conformers of alanine dipeptide. (Right) Decomposition of free energy into enthalpy and entropy terms. In all 3 subplots, the x and y axes correspond to ϕ and ψ torsion angles, respectively. All energies are in kcal/mol while each level in the contour corresponds to a 2 kcal/mol increment.

Free energy of backbone torsion in alanine dipeptide has been extensively used as a benchmark for testing enhanced sampling methods^{269,270,271,272,273,274,249,275}. However, much less effort has been put into the decomposition of the free energy into enthalpy and entropy components. Recently, Kamenik *et al.* have employed accelerated MD (aMD) to calculate the dihedral entropies in alanine dipeptide and larger protein systems from 1-dimensional free energy profiles²⁷⁶. Here, we'll take advantage of our multi-dimensional REMD tool and attempt to generate the enthalpy and entropy maps in two dimensions (Ψ , Ψ). As shown in Figure 6.2, we've obtained free energy profiles at 6 different temperatures that range from 273 K to 373 K, and from these have derived enthalpy and entropy profiles from finite difference calculations.

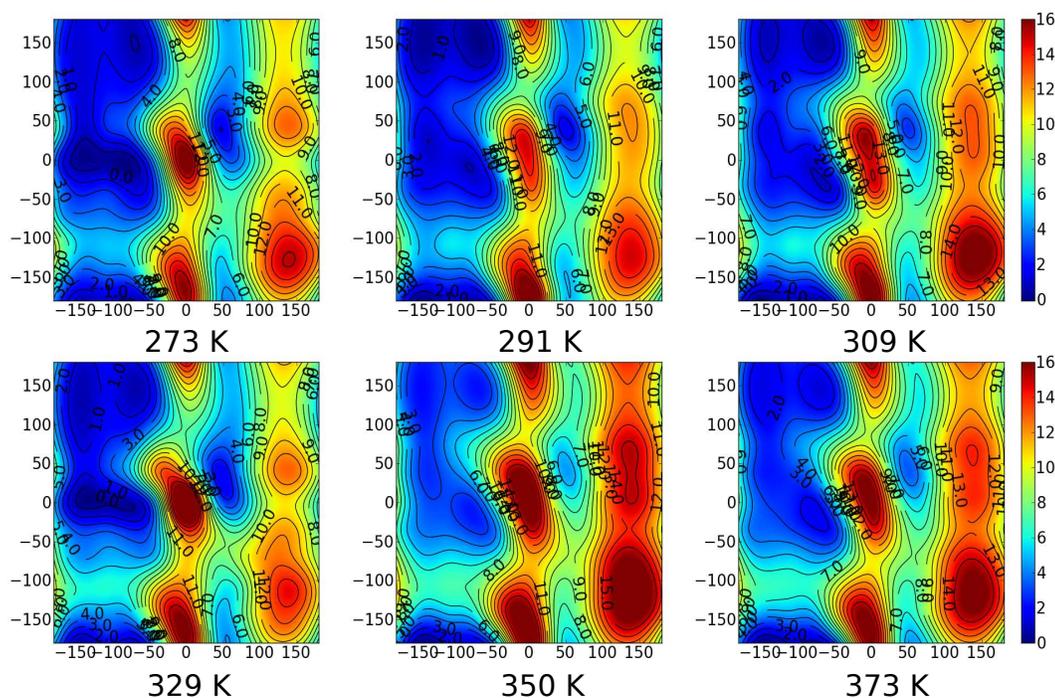


Figure 6.2: Free energy profile of alanine dipeptide backbone torsion at 6 different temperatures. In all 6 subplots, the x and y axes correspond to ϕ and ψ torsion angles, respectively. The range of energies is from 0 kcal/mol to 16 kcal/mol while each level in the contour corresponds to a 1 kcal/mol increment.

As a validation of the results here, we calculated the difference in enthalpy between the c7eq and α R states and compared to the experimental results where Takekiyo *et al.*²⁷⁷ have used Raman spectroscopy at different pressures to determine the enthalpy differences. Our calculated value of $\Delta H(\alpha R - c7eq)$ is 1.5 ± 0.8 kcal/mol from 4 independent simulations, which agrees with the experimental reference 1.1 ± 0.5 kcal/mol. More accurate enthalpy and entropy evaluations might require much longer simulations. Our approach here to map out enthalpy and entropy of peptide/protein backbone torsions in multi-dimensional space could be useful in the study of protein flexibility and folding. It could also be extended to predict reaction enthalpy and entropy from free energy profiles of chemical reactions simulated using QM/MM or other methods.

6.3.2 New application: Effect of salt concentration on DNA conformation

Recently, gold nanoparticle small-angle X-ray scattering (Au-SAXS) technique²⁷⁸ has been applied to study the conformational ensemble of a B-DNA duplex²⁵⁹ and bulge system²⁷⁹. In those experiments, two Au nanoparticles are attached to two thymine nucleobases via rigid linkers. Scattering patterns of Au are measured and post-processed to generate the radial distribution function of Au-Au distances, which then can be used to inform the conformational ensemble. In the present example, two B-DNA duplexes of same length are connected using polyethylene glycol (PEG) tethers to create a model system where the structure and dynamics of the rigid helices connected by a flexible tether can be systematically studied. Since DNA helices are highly negatively charged, salt concentration in the solution are assumed to have a significant effect on the conformational ensemble of the system via electrostatic screening. Here, we take advantage of the flexibility of RepEx to perform multi-dimensional REMD with salt concentration as one of the exchange variables to predict the salt effect.

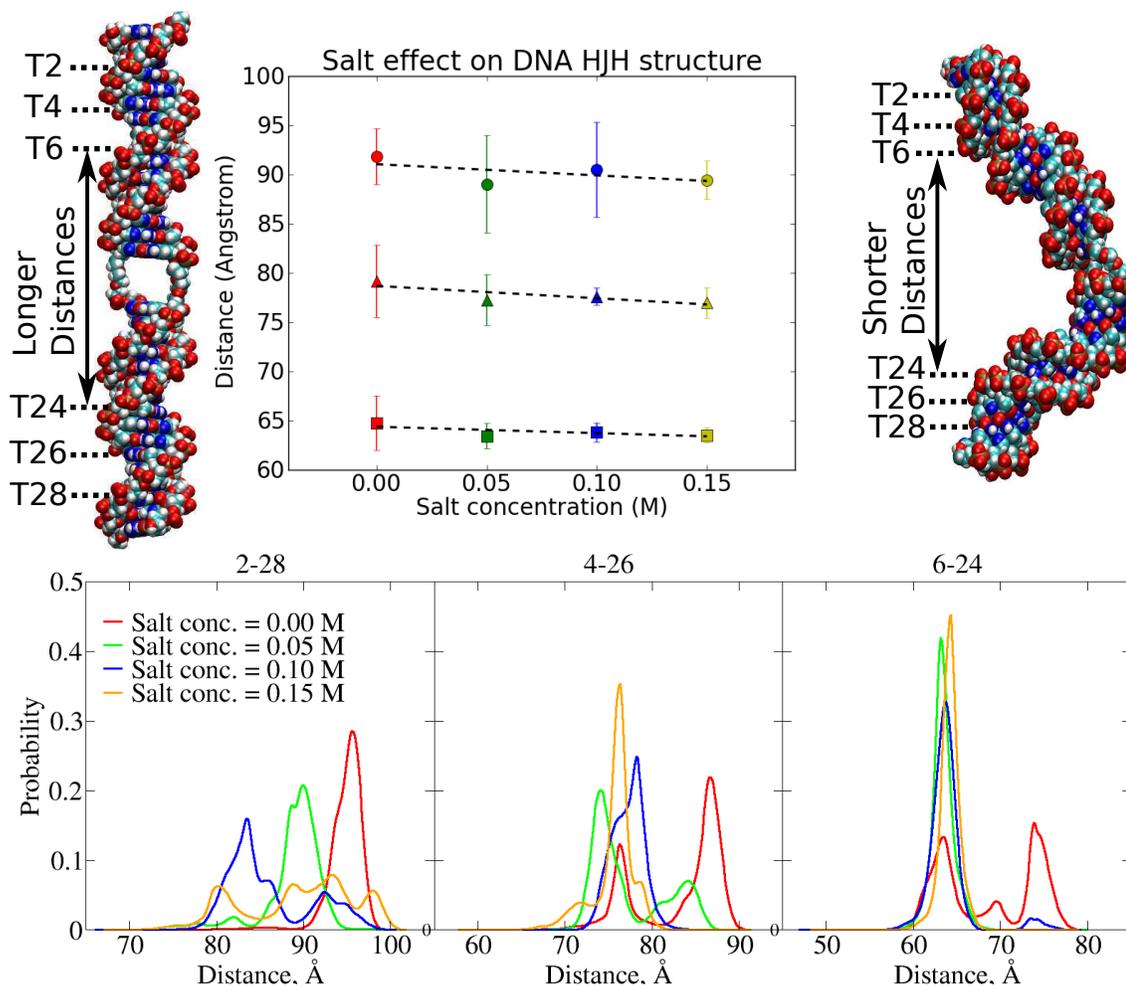


Figure 6.3: (Top, side) Illustration of the DNA helix-junction-helix (HJH) system and the 3 pairs of marker thymine residues. (Top, middle) Averages and standard errors (from 4 runs) of distances between each pair of residues at different salt concentrations at 288 K. Dashed lines are obtained from least-square linear regression. Circles, triangles and squares represent the 2-28, 4-26 and 6-24 pairs, respectively. (Bottom) Distribution of the 3 different distances at different salt concentrations in 1 of the 4 runs. In all subplots, red, green, blue and orange colors correspond to the salt concentrations of 0.00 M, 0.05 M, 0.10 M and 0.15 M, respectively.

For all 3 pair positions (T2-T28, T4-T26 and T6-T24), we used the distance between the two methyl carbons as the structural descriptor of the system instead of explicitly modeling the two big Au₆₉ nanoparticles, their thio-glucose shells and the linkers in order to

reduce system size and complexity, based on the assumption that the linkers between Au and the methyl groups in thymines are rigid and will not perturb DNA structure²⁵⁹. As shown in Figure 6.3, the overall trend of all the 3 distances is decreasing as the salt concentration increases, which agrees with our expectation because larger salt concentration introduces stronger electrostatic screening, which will compensate for the increasing unfavorable Coulomb interaction as the system folds along the middle, which corresponds to a smaller distance between the two residues in a pair.

6.4 Conclusion

Multi-dimensional REMD simulations offer a potentially powerful tool for the investigation of complex chemical/biophysical problems, but require efficient, highly-flexible and scalable implementation. Here, we introduce the RepEx package, in which we developed and implemented multi-dimensional REMD with customized choice of order parameters. We have chosen two examples to illustrate new insight and applications that are enabled by this technology. In the first example, we performed 3D-REMD to study the enthalpic and entropic contributions to the backbone torsion free energies of the alanine dipeptide, a classic benchmark system. Two-dimensional replica exchange umbrella sampling on the two torsion angles yield 2D free energy surfaces, while temperature replica exchange, as the 3rd dimension, not only improves sampling by allowing low-temperature replicas to exchange with high-temperature ones, but also provides useful thermodynamic information from which we can decompose the free energy into enthalpic and entropic contributions. Decent agreement with available experimental data validates our approach. In the second example, we investigated the effect of salt concentration on the conformation of a DNA helix-junction-helix model system. REMD with both salt concentration and geometrical variables as order parameters was performed to generate conformational ensembles of the system and predict the Au-Au distance profiles measurable by Au-SAXS. The capability of introducing customized, non-traditional order parameters into REMD simulations here is shown to enable new scientific applications. This new technology sets the stage for the investigation of a rich set of multi-dimensional biological problems.

Bibliography

- [1] Irina E. Catrina and Alvan C. Hengge. Comparisons of Phosphorothioate with Phosphate Transfer Reactions for a Monoester, Diester, and Triester: Isotope Effect Studies. *J. Am. Chem. Soc.*, 125:7546–7552, 2003.
- [2] Jonathan K. Lassila, Jesse G. Zalatan, and Daniel Herschlag. Biological phosphoryl-transfer reactions: Understanding mechanism and catalysis. *Annu. Rev. Biochem.*, 80:669–702, 2011.
- [3] E. P. Serjeant and B. Dempsey. *Ionisation Constants of Organic Acids in Aqueous Solution*. Pergamon Press, New York, 1979.
- [4] David R. Lide, editor. *CRC handbook of chemistry and physics*. CRC Press LLC, Boca Raton, FL, 90 edition, 2010.
- [5] Nicholas Bourne, Edwin Chrystiuk, Andrew M. Davis, and Andrew Williams. A single transition state in the reaction of aryl diphenylphosphinate esters with phenolate ions in aqueous solution. *J. Am. Chem. Soc.*, 110:1890–1895, 1988.
- [6] Hong Gu, Shuming Zhang, Kin-Yiu Wong, Brian K. Radak, Thakshila Dissanayake, Daniel L. Kellerman, Qing Dai, Masaru Miyagi, Vernon E. Anderson, Darrin M. York, Joseph A. Piccirilli, and Michael E. Harris. Experimental and computational analysis of the transition state for ribonuclease A-catalyzed RNA 2'-O-transphosphorylation. *Proc. Natl. Acad. Sci. USA*, 110:13002–13007, 2013.
- [7] Michael E. Harris, Qing Dai, Hong Gu, Daniel L. Kellerman, Joseph A. Piccirilli, and Vernon E. Anderson. Kinetic isotope effects for RNA cleavage by 2'-O-transphosphorylation: Nucleophilic activation by specific base. *J. Am. Chem. Soc.*, 132:11613–11621, 2010.

- [8] Roland K O Sigel and Helmut Sigel. A stability concept for metal ion coordination to single-stranded nucleic acids and affinities of individual sites. *Acc. Chem. Res.*, 43(7):974–984, Jul 2010.
- [9] Maria T. Panteva, George M. Giambasu, and Darrin M. York. Force field for Mg^{2+} , Mn^{2+} , Zn^{2+} , and Cd^{2+} ions that have balanced interactions with nucleic acids. *J. Phys. Chem. B*, 119:15460–15470, 2015.
- [10] Ming Huang and Darrin M. York. Linear free energy relationships in RNA transesterification: theoretical models to aid experimental interpretations. *Phys. Chem. Chem. Phys.*, 16:15846–15855, 2014.
- [11] Aamir Mir, Ji Chen, Kyle Robinson, Emma Lendy, Jaclyn Goodman, David Neau, and Barbara L. Golden. Two Divalent Metal Ions and Conformational Changes Play Roles in the Hammerhead Ribozyme Cleavage Reaction. *Biochemistry*, 54:6369–6381, 2015.
- [12] Aamir Mir and Barbara L. Golden. Two Active Site Divalent Ions in the Crystal Structure of the Hammerhead Ribozyme Bound to a Transition State Analogue. *Biochemistry*, 55:633–636, 2016.
- [13] Shenglong Wang, Katrin Karbstein, Alessio Peracchi, Leonid Beigelman, and Daniel Herschlag. Identification of the Hammerhead Ribozyme Metal Ion Binding Site Responsible for Rescue of the Deleterious Effect of a Cleavage Site Phosphorothioate. *Biochemistry*, 38(43):14363–14378, 1999.
- [14] E. M. Osborne, J. E. Schaak, and V. J. Derose. Characterization of a native hammerhead ribozyme derived from schistosomes. *RNA*, 11:187–196, 2005.
- [15] W Luke Ward and Victoria J Derose. Ground-state coordination of a catalytic metal to the scissile phosphate of a tertiary-stabilized Hammerhead ribozyme. *RNA*, 18:16–23, 2012.
- [16] Tai-Sung Lee, Kin-Yiu Wong, George M. Giambasu, and Darrin M. York. *Bridging the*

gap between theory and experiment to derive a detailed understanding of hammerhead ribozyme catalysis, volume 120, chapter 2, pages 25–91. 2013.

- [17] Timothy J. Wilson and David M. Lilley. A Mechanistic Comparison of the Varkud Satellite and Hairpin Ribozymes. *Prog. Mol. Biol. Transl. Sci.*, 120:93–121, 2013.
- [18] Ramanarayanan Krishnamurthy. Role of pKa of Nucleobases in the Origins of Chemical Evolution. *Acc. Chem. Res.*, 45(12):2035–2044, 2012.
- [19] Haoyuan Chen, Timothy J. Giese, Ming Huang, Kin-Yiu Wong, Michael E. Harris, and Darrin M. York. Mechanistic Insights into RNA Transphosphorylation from Kinetic Isotope Effects and Linear Free Energy Relationships of Model Reactions. *Chem. Eur. J.*, 20:14336–14343, 2014.
- [20] Haoyuan Chen, Joseph A. Piccirilli, Michael E. Harris, and Darrin M. York. Effect of Zn^{2+} binding and enzyme active site on the transition state for RNA 2'-O-transphosphorylation interpreted through kinetic isotope effects. *Biochim. Biophys. Acta.*, 1854:1795–1800, 2015.
- [21] Haoyuan Chen, Timothy J. Giese, Barbara L. Golden, and Darrin M. York. Metal ion-induced activation of the catalytic general base in hammerhead ribozyme self-cleavage. *Biochemistry*, submitted.
- [22] Attila Szabo and Neils S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Publications, Inc., New York, 1st edition, 1996.
- [23] Christopher J. Cramer. *Essentials of Computational Chemistry: Theories and Models*. John Wiley & Sons, Chichester, England, 2nd edition, 2004.
- [24] R. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford University Press, New York, 1989.
- [25] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, 1996.

- [26] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the Colle-Savetti Correlation energy formula into a functional of the electron density. *Phys. Rev. B.*, 37:785–789, 1988.
- [27] Axel D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98(7):5648–5652, 1993.
- [28] Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: the PBE0 model. *J. Chem. Phys.*, 110:6158–6170, 1999.
- [29] J. A. Pople and G. A. Segal. Approximate Self-Consistent Molecular Orbital Theory. II. Calculations with Complete Neglect of Differential Overlap. *J. Chem. Phys.*, 43(10):S136–S151, 1965.
- [30] Michael J. S. Dewar, Eve Zebisch, Eamonn F. Healy, and James J. P. Stewart. 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, 1985.
- [31] James J. P. Stewart. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.*, 10:209–220, 1989.
- [32] James J. P. Stewart. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.*, 13:1173–1213, 2007.
- [33] I. Rossi and D. G. Truhlar. Parameterization of NDDO wavefunctions using genetic algorithms. An evolutionary approach to parameterizing potential energy surfaces and direct dynamics calculations for organic reactions. *Chem. Phys. Lett.*, 233:231–236, 1995.
- [34] Kwangho Nam, Qiang Cui, Jiali Gao, and Darrin M. York. Specific reaction parametrization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O, and P atoms. *J. Chem. Theory Comput.*, 3:486–504, 2007.

- [35] D. Porezag, Th. Frauenheim, Th. Köhler, G. Seifert, and R. Kaschner. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Phys. Rev. B*, 51:12947–12957, 1995.
- [36] M. Elstner, Th. Frauenheim, E. Kaxiras, G. Seifert, and S. Suhai. A Self-Consistent Charge Density-Functional Based Tight-Binding Scheme for Large Biomolecules. *Phys. Status Solidi. B*, 217:357–376, 2000.
- [37] Marcus Elstner. The SCC-DFTB method and its application to biological systems. *Theor. Chem. Acc.*, 116(1-3):316–325, 2006.
- [38] Michael Gaus, Qiang Cui, and Marcus Elstner. DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J. Chem. Theory Comput.*, 7(4):931–948, 2011.
- [39] Michael Gaus, Albrecht Goez, and Marcus Elstner. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.*, 9:338–354, 2013.
- [40] Qiang Cui and Marcus Elstner. Density functional tight binding: values of semi-empirical methods in an ab initio era. 16:14368–14377, 2014.
- [41] Yang Yang, Haibo Yu, Darrin York, Marcus Elstner, and Qiang Cui. Description of phosphate hydrolysis reactions with the self-consistent-charge density-functional-tight-binding (SCC-DFTB) theory. 1. Parameterization. *J. Chem. Theory Comput.*, 4:2067–2084, 2008.
- [42] Weitao Yang. Electron density as the basic variable: a divide-and-conquer approach to the ab initio computation of large molecules. *J. Mol. Struct.*, 255:461–479, 1992.
- [43] Timothy J. Giese, Haoyuan Chen, Thakshila Dissanayake, George M. Giambasu, Hugh Heldenbrand, Ming Huang, Erich R. Kuechler, Tai-Sung Lee, Maria T. Panteva, Brian K. Radak, and Darrin M. York. A Variational Linear-Scaling Framework to Build Practical, Efficient Next-Generation Orbital-Based Quantum Force Fields. *J. Chem. Theory Comput.*, 9:1417–1427, 2013.

- [44] Jiali Gao. Toward a Molecular Orbital Derived Empirical Potential for Liquid Simulation. *J. Phys. Chem.*, 101(4):657–663, 1997.
- [45] Dmitri G. Fedorov, Ryan M. Olson, Kazuo Kitaura, Mark S. Gordon, and Shiro Koseki. A new hierarchical parallelization scheme: Generalized distributed data interface (GDDI), and an application to the fragment molecular orbital method (FMO). *J. Comput. Chem.*, 25(6):872–880, 2004.
- [46] Mark S. Gordon, Mark A. Freitag, Pradipta Bandyopadhyay, Jan H. Jensen, Visvaldas Kairys, and Walter J. Stevens. The Effective Fragment Potential Method: A QM-Based MM Approach to Modeling Environmental Effects in Chemistry. *J. Phys. Chem. A*, 105(2):293–307, 2001.
- [47] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, Kenneth M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.
- [48] Pengyu Ren and Jay W. Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B*, 107(24):5933–5947, 2003.
- [49] Victor M. Anisimov, Guillaume Lamoureux, Niu Vorobyov, Igor V. Huang, Benoît Roux, and Alexander D. MacKerell, Jr. Determination of electrostatic parameters for a polarizable force field based on the classical drude oscillator. *J. Chem. Theory Comput.*, 1:153–168, 2005.
- [50] Loup Verlet. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159(1):98–103, July 1967.
- [51] William C. Swope, Hans C. Andersen, Peter H. Berens, and Kent R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.*, 76:637–649, 1982.

- [52] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.
- [53] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Lee Hsing, and Lee G. Pedersen. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103(19):8577–8593, 1995.
- [54] Donald Bashford and David A. Case. Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.*, 51:129–152, 2000.
- [55] David E. Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, Michael P. Eastwood, Joseph A. Bank, John M. Jumper, John K. Salmon, Yibing Shan, and Willy Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, 2010.
- [56] Amy Shih, Anton Arkhipov, Peter Freddolino, and Klaus Schulten. Coarse grained protein-lipid model with application to lipoprotein particles. 110:3674–3684, 2006.
- [57] Siewert Marrink, Alex de Vries, and Alan Mark. Coarse Grained Model for Semi-quantitative Lipid Simulations. 108:750–760, 2004.
- [58] T. Knotts, N. Rathore, D. Schwartz, and J. de Pablo. A coarse grain model for DNA. 126:084901, 2007.
- [59] Maria T. Panteva, Thakshila Dissanayake, Haoyuan Chen, Brian K. Radak, Erich R. Kuechler, George M. Giambaşu, Tai-Sung Lee, and Darrin M. York. *Multiscale Methods for Computational RNA Enzymology*, chapter 14. Elsevier, 2015.
- [60] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23:187–199, 1977.
- [61] Shankar Kumar, Djamel Bouzida, Robert H. Swendsen, Peter A. Kollman, and John M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, 13:1011–1021, 1992.

- [62] Johannes Kästner and Walter Thiel. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: “Umbrella integration”. *J. Chem. Phys.*, 123:144104, 2005.
- [63] Charles H. Bennett. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.*, 22:245–268, 1976.
- [64] Michael R. Shirts and John D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.*, 129:124105, 2008.
- [65] Tai-Sung Lee, Brian K. Radak, Anna Pabis, and Darrin M. York. A new maximum likelihood approach for free energy profile construction from molecular simulations. *J. Chem. Theory Comput.*, 9:153–164, 2013.
- [66] Robert H. Swendsen and Jian-Sheng Wang. Replica Monte Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.*, 57:2607–2609, 1986.
- [67] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314:141–151, 1999.
- [68] Harold A. Scheraga, Mey Khalili, and Adam Liwo. Protein-folding dynamics: overview of molecular simulation techniques. *Annu. Rev. Phys. Chem.*, 58:57–83, 2007.
- [69] Takaharu Mori, Jaewoon Jung, and Yuji Sugita. Surface-Tension Replica-Exchange Molecular Dynamics Method for Enhanced Sampling of Biological Membrane Systems. *J. Chem. Theory Comput.*, 9:5629–5640, 2013.
- [70] Satoru G. Itoh, Ana Damjanović, and Bernard R. Brooks. pH replica-exchange method based on discrete protonation states. *Proteins*, 79:3420–3436, 2011.
- [71] Yuji Sugita, Akio Kitao, and Yuko Okamoto. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.*, 113:6042–6051, 2000.
- [72] Y. Sugita and Y. Okamoto. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem. Phys. Lett.*, 329(3-4):261–270, 2000.

- [73] Brian K. Radak, Melissa Romanus, Tai-Sung Lee, Haoyuan Chen, Ming Huang, Antonis Treikalis, Vivekanandan Balasubramanian, Shantenu Jha, and Darrin M. York. Characterization of the Three-Dimensional Free Energy Manifold for the Uracil Ribonucleoside from Asynchronous Replica Exchange Simulations. *J. Chem. Theory Comput.*, 11(2):373–377, 2015.
- [74] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA*, 99:12562–12566, 2002.
- [75] Donald Hamelberg, John Mongan, and J. Andrew McCammon. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.*, 120:11919–11929, 2004.
- [76] Barry Isralewitz, Mu Gao, and Klaus Schulten. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.*, 11:224–230, 2001.
- [77] Eric Darve and Andrew Pohorille. Calculating free energies using average force. *J. Chem. Phys.*, 115(20):9169–9183, 2001.
- [78] Adrian Roitberg and Ron Elber. Modeling side chains in peptides and proteins: Application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. 95:9277, 1991.
- [79] F. Wang and D. Landau. Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. 86:2050, 2001.
- [80] M. J. Field, P. A. Bash, and M. Karplus. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.*, 11:700–733, 1990.
- [81] Jiali Gao, Patricia Amara, Cristobal Alhambra, and Martin J. Field. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A*, 102:4714–4721, 1998.
- [82] Hai Lin and Donald G. Truhlar. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.*, 117:185–199, 2007.

- [83] Robert W. Zwanzig. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.*, 22:1420–1426, 1954.
- [84] Michael R. Shirts and Vijay S. Pande. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J Chem Phys*, 122:144107, 2005.
- [85] Jacob Bigeleisen and Max Wolfsberg. Theoretical and experimental Aspects of Isotope Effects in Chemical Kinetics. *Adv. Chem. Phys.*, 1:15–76, 1958.
- [86] W Luke Ward, Kory Plakos, and Victoria J. DeRose. Nucleic acid catalysis: metals, nucleobases, and other cofactors. *Chem. Rev.*, 114(8):4318–4342, 2014.
- [87] Yao Xin and Donald Hamelberg. Deciphering the role of glucosamine-6-phosphate in the riboswitch action of glmS ribozyme. *RNA*, 16(12):2455–2463, Dec 2010.
- [88] Brian K. Radak, Tai-Sung Lee, Michael E. Harris, and Darrin M. York. Assessment of metal-assisted nucleophile activation in the hepatitis delta virus ribozyme from molecular simulation and 3D-RISM. *RNA*, 21(9):1566–1577, September 2015.
- [89] Colin S. Gaines and Darrin M. York. Ribozyme Catalysis with a Twist: Active State of the Twister Ribozyme in Solution Predicted from Molecular Simulation. *J. Am. Chem. Soc.*, 138(9):3058–3065, 2016.
- [90] Maria T. Panteva, George M. Giambaşu, and Darrin M. York. Comparison of structural, thermodynamic, kinetic and mass transport properties of Mg^{2+} ion models commonly used in biomolecular simulations. *J. Comput. Chem.*, 36:970–982, 2015.
- [91] Pengfei Li and Kenneth M. Merz, Jr. Taking into account the ion-induced dipole interaction in the nonbonded model of ions. *J. Chem. Theory Comput.*, 10:289–297, 2014.
- [92] George M. Giambaşu, Tyler Luchko, Daniel Herschlag, Darrin M. York, and David A. Case. Ion counting from explicit-solvent simulations and 3D-RISM. *Biophys. J.*, 106: 883–894, 2014.

- [93] F. H. Westheimer. Why nature chose phosphates. *Science*, 235:1173–1178, 1987.
- [94] W. Wallace Cleland and Alvan C. Hengge. Enzymatic mechanisms of phosphate and sulfate transfer. *Chem. Rev.*, 106:3252–3278, 2006.
- [95] Jesse C. Cochrane and Scott A. Strobel. Catalytic strategies of self-cleaving ribozymes. *Acc. Chem. Res.*, 41:1027–1035, 2008.
- [96] Shina C. L. Kamerlin, Pankaz K. Sharma, Ram B. Prasad, and Arieh Warshel. Why nature really chose phosphate. *Q. Rev. Biophys.*, 46:1–132, 2013.
- [97] William G. Scott, James B. Murray, John R. P. Arnold, Barry L. Stoddard, and Aaron Klug. Capturing the structure of a catalytic RNA intermediate: The Hammerhead Ribozyme. *Science*, 274:2065–2069, 1996.
- [98] Jamal M. Buzayan, Wayne L. Gerlach, and George Bruening. Satellite tobacco ringspot virus RNA: A subset of the RNA sequence is sufficient for autolytic processing. *Proc. Natl. Acad. Sci. USA*, 83:8859–8862, 1986.
- [99] L. Sharmeen, M. Y. Kuo, G. Dinter-Gottlieb, and J. Taylor. Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J. Virol.*, 62:2674–2679, 1988.
- [100] Wade C. Winkler, Ali Nahvi, Adam Roth, Jennifer A. Collins, and Ronald R. Breaker. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, 428:281–286, 2004.
- [101] Barry J. Saville and Richard A. Collins. A site-specific self-cleavage reaction performed by a novel RNA in neurospora mitochondria. *Cell*, 61:685–696, 1990.
- [102] Ronald T. Raines. Ribonuclease A. *Chem. Rev.*, 98:1045–1065, 1998.
- [103] Mikko Oivanen, Satu Kuusela, and Harri Lönnberg. Kinetics and Mechanisms for the Cleavage and Isomerization of the Phosphodiester Bonds of RNA by Brnsted Acids and Bases. *Chem. Rev.*, 98:961–990, 1998.
- [104] J. N. Bronsted. Acid and Basic Catalysis. *Chem. Rev.*, 5:231–338, 1928.

- [105] John E. Leffler. Parameters for the Description of Transition States. *Science*, 117: 340–341, 1953.
- [106] David A. Jencks and William P. Jencks. On the Characterization of Transition States by Structure-Reactivity Coefficients. *J. Am. Chem. Soc.*, 99:7948–7960, 1977.
- [107] Andrew Williams. Effective charge and Leffler’s index as mechanistic tools for reactions in solution. *Acc. Chem. Res.*, 17(12):425–430, 1984.
- [108] Ikenna Onyido, Krzysztof Swierczek, Jamie Purcell, and Alvan C. Hengge. A Concerted Mechanism for the Transfer of the Thiophosphinoyl Group from Aryl Dimethylphosphinothioate Esters to Oxyanionic Nucleophiles in Aqueous Solution. *J. Am. Chem. Soc.*, 127:7703–7711, 2005.
- [109] A. C. Hengge and I. Onyido. Physical organic perspectives on phospho group transfer from phosphates and phosphinates. *Curr. Org. Chem.*, 9:61–74, 2005.
- [110] Alvan C. Hengge. Isotope effects in the study of phosphoryl and sulfonyl transfer reactions. *Acc. Chem. Res.*, 35:105–112, 2002.
- [111] Kin-Yiu Wong, Hong Gu, Shuming Zhang, Joseph A. Piccirilli, Michael E. Harris, and Darrin M. York. Characterization of the reaction path and transition states for RNA transphosphorylation models from theory and experiment. *Angew. Chem. Int. Ed.*, 51:647–651, 2012.
- [112] Yan Zhao and Donald G. Truhlar. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.*, 120:215–241, 2008.
- [113] Jacopo Tomasi, Benedetta Mennucci, and Roberto Cammi. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.*, 105:2999–3093, 2005.

- [114] Giovanni Scalmani and Michael J. Frisch. Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J. Chem. Phys.*, 132:114110–114124, 2010.
- [115] Anthony K. Rappé, Carla J. Casewit, K. S. Colwell, William A. Goddard III, and W. Mason Skiff. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.*, 114:10024–10035, 1992.
- [116] Vincenzo Barone, Maurizio Cossi, and Jacopo Tomasi. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *J. Chem. Phys.*, 107(8):3210–3221, 1997.
- [117] Jacob Bigeleisen and Maria G. Mayer. Calculation of Equilibrium Constants for Isotopic Exchange Reactions. *J. Chem. Phys.*, 15(5):261–267, 1947.
- [118] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, Vreven T., J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09, Revision A.02. Gaussian, Inc., Wallingford, CT, 2009.
- [119] Marco Klähn, Edina Rosta, and Arieh Warshel. On the mechanism of hydrolysis of phosphate monoesters dianions in solutions and proteins. *J. Am. Chem. Soc.*, 128:15310–15323, 2006.
- [120] Kerensa Sorensen-Stowell and Alvan C. Hengge. Thermodynamic Origin of

- the Increased Rate of Hydrolysis of Phosphate and Phosphorothioate Esters in DMSO/Water Mixtures. *J. Org. Chem.*, 71:7180–7184, 2006.
- [121] Lidong Zhang, Daiqian Xie, Dingguo Xu, and Hua Guo. Supermolecule density functional calculations suggest a key role for solvent in alkaline hydrolysis of p-nitrophenyl phosphate. *Chem. Commun.*, 16:1638–1640, 2007.
- [122] Alvan C. Hengge and Robert A. Hess. Concerted or Stepwise Mechanisms for Acyl Transfer Reactions of p-Nitrophenyl Acetate? Transition State Structures from Isotope Effects. *J. Am. Chem. Soc.*, 116:11256–11263, 1994.
- [123] Adam G. Cassano, Vernon E. Anderson, and Michael E. Harris. Evidence for Direct Attack by Hydroxide in Phosphodiester Hydrolysis. *J. Am. Chem. Soc.*, 124:10964–10965, 2002.
- [124] Xabier Lopez, Darrin M. York, Annick Dejaegere, and Martin Karplus. Theoretical Studies on the Hydrolysis of Phosphate Diesters in the Gas Phase, Solution, and RNase A. *Int. J. Quantum Chem.*, 86:10–26, 2002.
- [125] Violeta López-Canut, Javier Ruiz-Pernía, Iñaki Tuñón, Silvia Ferrer, and Vicent Moliner. Theoretical Modeling on the Reaction Mechanism of p-Nitrophenylmethylphosphate Alkaline Hydrolysis and its Kinetic Isotope Effects. *J. Chem. Theory Comput.*, 5:439–442, 2009.
- [126] Guanhua Hou and Qiang Cui. QM/MM Analysis Suggests That Alkaline Phosphatase (AP) and Nucleotide Pyrophosphatase/Phosphodiesterase Slightly Tighten the Transition State for Phosphate Diester Hydrolysis Relative to Solution: Implication for Catalytic Promiscuity in the AP Superfamily. *J. Am. Chem. Soc.*, 134:229–246, 2012.
- [127] Linus Pauling. Atomic Radii and Interatomic Distances in Metals. *J. Am. Chem. Soc.*, 69:542–553, 1947.
- [128] K. N. Houk, Susan M. Gustafson, and Kersey A. Black. Theoretical secondary kinetic isotope effects and the interpretation of transition state geometries. 1. the cope rearrangement. *J. Am. Chem. Soc.*, 114:8565–8572, 1992.

- [129] Kin-Yiu Wong Yuqing Xu and Darrin M. York. Ab Initio Path-Integral Calculations of Kinetic and Equilibrium Isotope Effects on Base-Catalyzed RNA Transphosphorylation Models. *J. Comput. Chem.*, 35:1302–1316, 2014.
- [130] Harri Lönnberg, Roger Strömberg, and Andrew Williams. . *Org. Biomol. Chem.*, 2: 2165–2167, 2004.
- [131] Yun Liu, Brent A. Gregersen, Alvan Hengge, and Darrin M. York. Transesterification Thio Effects of Phosphate Diesters: Free Energy Barriers and Kinetic and Equilibrium Isotope Effects from Density-Functional Theory. *Biochemistry*, 45:10043–10053, 2006.
- [132] Daniel Herschlag. Ribonuclease revisited: Catalysis via the classical general acid-base mechanism or a triester-like mechanism? *J. Am. Chem. Soc.*, 116(26):11631–11635, 1994.
- [133] Fabrice Leclerc. Hammerhead Ribozymes: True Metal or Nucleobase Catalysis? Where Is the Catalytic Power from? *Molecules*, 15:5389–5407, 2010.
- [134] Elena Formoso, Jon M Matxain, Xabier Lopez, and Darrin M York. Molecular dynamics simulation of bovine pancreatic ribonuclease A-CpA and transition state-like complexes. *J. Phys. Chem. B*, 114(21):7371–7382, Jun 2010.
- [135] Ewa Gawlita, Marily Lantz, Piotr Paneth, Alasdair F. Bell, Peter J. Tonge, and Vernon E. Anderson. H-Bonding in Alcohols Is Reflected in the C α -H Bond Strength: Variation of C-D Vibrational Frequency and Fractionation Factor. *J. Am. Chem. Soc.*, 122:11660–11669, 2000.
- [136] S. Jarmelo, N. Maiti, V. Anderson, P. R. Carey, and R. Fausto. C α -H Bond Frequency in Alcohols as a Probe of Hydrogen-Bonding Strength: A Combined Vibrational Spectroscopic and Theoretical Study of n-[1-D]Propanol. *J. Phys. Chem. A*, 109:2069–2077, 2005.
- [137] Nakul C. Maiti, Yuping Zhu, Ian Carmichael, Anthony S. Serianni, and Vernon E. Anderson. 1JCH Correlates with Alcohol Hydrogen Bond Strength. *J. Org. Chem.*, 71:2878–2880, 2006.

- [138] Uttam Pal, Sudeshna Sen, and Nakul C. Maiti. C α -H Bond Frequency in Alcohols as a Probe of Hydrogen-Bonding Strength: A Combined Vibrational Spectroscopic and Theoretical Study of n-[1-D]Propanol. *J. Phys. Chem. A*, 118:1024–1030, 2014.
- [139] T. A. Steitz and J. A. Steitz. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA*, 90:6498–6502, 1993.
- [140] Victoria J. DeRose. Metal ion binding to catalytic RNA molecules. *Curr. Opin. Struct. Biol.*, 13:317–324, 2003.
- [141] Roland K. O. Sigel and Anna Marie Pyle. Alternative Roles for Metal Ions in Enzyme Catalysis and the Implications for Ribozyme Chemistry. *Chem. Rev.*, 2007:97–113, 2007.
- [142] Marcello Forconi and Daniel Herschlag. Metal ion-based RNA cleavage as a structural probe. *Methods Enzymol.*, 468:91–106, 2009.
- [143] Zdenek Chval, Daniela Chvalova, and Fabrice Leclerc. Modeling the RNA 2'OH Activation: Possible Roles of Metal Ion and Nucleobase as Catalysts in Self-Cleaving Ribozymes. *J. Phys. Chem. B*, 115:10943–10956, 2011.
- [144] Sebastian M. Fica, Nicole Tuttle, Thaddeus Novak, Nan-Sheng Li, Jun Lu, Prakash Koodathingal, Qing Dai, Jonathan P. Staley, and Joseph A. Piccirilli. RNA catalyses nuclear pre-mRNA splicing. *Nature*, 503:229–234, 2013.
- [145] Daniel L. Kellerman, Darrin M. York, Joseph A. Piccirilli, and Michael E. Harris. Altered (transition) states: mechanisms of solution and enzyme catalyzed RNA 2'-O-transphosphorylation. *Curr. Opin. Chem. Biol.*, 21:96–102, aug 2014.
- [146] Lars C. S. Melander and William H. Saunders. *Reaction rates of isotopic molecules*. Wiley, 1980.
- [147] Max Wolfsberg, W. Alexander Van Hook, and Piotr Paneth. *Isotope Effects in the Chemical, Geological and Bio Sciences*. Springer, 2009.
- [148] Katarzyna Swiderek and Piotr Paneth. Binding Isotope Effects. *Chem. Rev.*, 113: 7851–7879, 2013.

- [149] Tim Humphry, Subashree Iyer, Olga Iranzo, Janet R. Morrow, John P. Richard, Piotr Paneth, and Alvan C. Hengge. Altered transition state for the reaction of an RNA model catalyzed by a dinuclear zinc(II) catalyst. *J. Am. Chem. Soc.*, 130:17858–17866, 2008.
- [150] Hui Gao, Zhuofeng Ke, Nathan J. DeYonker, Juping Wang, Huiying Xu, Zongwan Mao, David Lee Phillips, and Cunyuan Zhao. Dinuclear Zn(II) Complex Catalyzed Phosphodiester Cleavage Proceeds via a Concerted Mechanism: A Density Functional Theory Study. *J. Am. Chem. Soc.*, 133:2904–2915, 2011.
- [151] Shuming Zhang, Hong Gu, Haoyuan Chen, Emily Strong, Edward W. Ollie, Daniel Kellerman, Danni Liang, Masaru Miyagi, Vernon E. Anderson, Joseph A. Piccirilli, Darrin M. York, and Michael E. Harris. Isotope effect analyses provide evidence for an altered transition state for RNA 2'-O-transphosphorylation catalyzed by Zn^{2+} . *Chem. Commun.*, 52:4462–4465, 2016.
- [152] Ulf Ryde. Carboxylate binding modes in zinc proteins: A theoretical study. *Biophys. J.*, 77:2777–2787, 1999.
- [153] M. Dolg, U. Wedig, H. Stoll, and H. Preuss. Energy-sdjusted ab initio pseudopotentials for the first row transition elements. *J. Chem. Phys.*, 86(2):866–872, 1987.
- [154] Maurizio Cossi, Nadia Rega, Giovanni Scalmani, and Vincenzo Barone. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.*, 24:669–681, 2003.
- [155] D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D.S Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Götz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, Tai-Sung Lee, S. Le Grand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, and P. A. Kollman. *AMBER 14*. University of California, San Francisco, San Francisco, CA, 2014.

- [156] James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.*, 11: 3696–3713, 2015.
- [157] David S. Wishart, Timothy Jewison, An Chi Guo, Michael Wilson, Craig Knox, Yifeng Liu, Yannick Djoumbou, Rupasri Mandal, Farid Aziat, Edison Dong, Souhaila Bouatra, Igor Sinelnikov, David Arndt, Jianguo Xia, Philip Liu, Faizath Yallou, Trent Bjorndahl, Rolando Perez-Pineiro, Roman Eisner, Felicity Allen, Vanessa Neveu, Russ Greiner, and Augustin Scalbert. HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res.*, 41:801–807, 2013.
- [158] Paul F. Cook. *Enzyme Mechanism from Isotope Effects*. CRC Press, 1991.
- [159] Timothy J. Giese and Darrin M. York. Ambient-Potential Composite Ewald Method for ab Initio Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulation. *J. Chem. Theory Comput.*, 12:2611–2632, 2016.
- [160] William G. Scott. RNA catalysis. *Curr. Opin. Struct. Biol.*, 8(6):720–726, 1998.
- [161] Robert G. Kuimelis and Larry W. McLaughlin. Mechanisms of Ribozyme-Mediated RNA Cleavage. *Chem. Rev.*, 98:1027–1044, 1998.
- [162] De-Min Zhou and Kazunari Taira. The Hydrolysis of RNA: From Theoretical Calculations to the Hammerhead Ribozyme-Mediated Cleavage of RNA. *Chem. Rev.*, 98: 991–1026, 1998.
- [163] Yasuomi Takagi, Yutaka Ikeda, and Kazunari Taira. Ribozyme Mechanisms. *Top. Curr. Chem.*, 232:213–251, 2004.
- [164] Kenneth F. Blount and Olke C. Uhlenbeck. The Structure-Function Dilemma of the Hammerhead Ribozyme. *Annu. Rev. Biophys. Biomol. Struct.*, 34:415–440, 2005.
- [165] David M. Lilley and Fritz Eckstein, editors. *Ribozymes and RNA Catalysis*. RSC Biomolecular Series. RSC Publishing, Cambridge, 2008.

- [166] O.C. Uhlenbeck. A small catalytic oligoribonucleotide. *Nature*, 328:596–600, 1987.
- [167] H. W. Pley, K. M. Flaherty, and D. B. McKay. Three-dimensional structure of a hammerhead ribozyme. *Nature*, 372:68–74, 1994.
- [168] William G. Scott, John T. Finch, and Aaron Klug. The Crystal Structure of an All-RNA Hammerhead Ribozyme: A Proposed Mechanism for RNA Catalytic Cleavage. *Cell*, 81:991–1002, 1995.
- [169] Monika Martick and William G. Scott. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell*, 126(2):309–320, 2006.
- [170] Michael Anderson, Eric P. Schultz, Monika Martick, and William G. Scott. Active-Site Monovalent Cations Revealed in a 1.55- Å-Resolution Hammerhead Ribozyme Structure. *J. Mol. Biol.*, 425:3790–3798, 2013.
- [171] SueAnn C. Dahm and Olke C. Uhlenbeck. Role of divalent metal ions in the hammerhead RNA cleavage reaction. *Biochemistry*, 30:9464–9469, 1991.
- [172] T. Tuschl, M. M. Ng, W. Pieken, F. Benseler, and F. Eckstein. Importance of exocyclic base functional groups of central core guanosines for hammerhead ribozyme activity. *Biochemistry*, 32(43):11658–11668, Nov 1993.
- [173] Thomas Tuschl and Fritz Eckstein. Hammerhead ribozyme: Importance of stem-loop II for activity. *Proc. Natl. Acad. Sci. USA*, 90:6991–6994, 1993.
- [174] T. Tuschl, C. Gohlke, T. M. Jovin, E. Westhof, and F. Eckstein. A three-dimensional model for the hammerhead ribozyme based on fluorescence measurements. *Science*, 266:785–789, 1994.
- [175] Robert G. Kuimelis and Larry W. McLaughlin. Hammerhead ribozyme-mediated cleavage of a substrate analogue containing an internucleotidic bridging 5'-phosphorothioate: implications for the cleavage mechanism and the catalytic role of the metal cofactor. *J. Am. Chem. Soc.*, 117:11019–11020, 1995.

- [176] Robert G. Kuimelis and Larry W. McLaughlin. Ribozyme-Mediated Cleavage of a Substrate Analogue Containing an Internucleotide-Bridging 5'-Phosphorothioate: Evidence for the Single-Metal Model. *Biochemistry*, 35:5308–5317, 1996.
- [177] Narayan Baidya and Olke C. Uhlenbeck. A kinetic and thermodynamic analysis of cleavage site mutations in the hammerhead ribozyme. *Biochemistry*, 36:1108–1114, 1997.
- [178] Edmund C. Scott and Olke C. Uhlenbeck. A re-investigation of the thio effect at the hammerhead cleavage site. *Nucleic Acids Res.*, 27(2):479–484, 1999.
- [179] Susan R. Morrissey, Thomas E. Horton, Christopher V. Grant, Charles G. Hoogstraten, R. David Britt, and Victoria J. DeRose. Mn²⁺-Nitrogen Interactions in RNA Probed by Electron Spin-Echo Envelope Modulation Spectroscopy: Application to the Hammerhead Ribozyme. *J. Am. Chem. Soc.*, 121:9215–9218, 1999.
- [180] Thomas E. Horton and Victoria J. DeRose. Cobalt Hexamine Inhibition of the Hammerhead Ribozyme. *Biochemistry*, 39:11408–11416, 2000.
- [181] Frank Seela, Harald Debelak, Lori Andrews, and Leonid Beigelman. Synthesis and enzymic hydrolysis of oligoribonucleotides incorporating 3-deazaguanosine: the importance of the nitrogen-3 atom of single conserved guanosine residues on the catalytic activity of the hammerhead ribozyme. *Helv. Chim. Acta*, 86:2726–2740, 2003.
- [182] Ken-ichi Suzumura, Yasuomi Takagi, Masaya Orita, and Kazunari Taira. NMR-Based Reappraisal of the Coordination of a Metal Ion at the Pro- R_p Oxygen of the A9/G10.1 Site in a Hammerhead Ribozyme. *J. Am. Chem. Soc.*, 126(47):15504–15511, 2004.
- [183] Joonhee Han and John M. Burke. Model for General Acid-Base Catalysis by the Hammerhead Ribozyme: pH-Activity Relationships of G8 and G12 Variants at the Putative Active Site. *Biochemistry*, 44:7864–7870, 2005.
- [184] Manami Roychowdhury-Saha and Donald H. Burke. Extraordinary rates of transition metal ion-mediated ribozyme catalysis. *RNA*, 12(10):1846–1852, 2006.

- [185] Manami Roychowdhury-Saha and Donald H. Burke. Distinct reaction pathway promoted by non-divalent-metal cations in a tertiary stabilized hammerhead ribozyme. *RNA*, 13:841–848, 2007.
- [186] Rita Przybilski and Christian Hammann. The tolerance to exchanges of the WatsonCrick base pair in the hammerhead ribozyme core is determined by surrounding elements . *RNA*, 13:1625–1630, 2007.
- [187] Jennifer L Boots, Marella D Canny, Ehsan Azimi, and Arthur Pardi. Metal ion specificities for folding and cleavage activity in the schistosoma hammerhead ribozyme. *RNA*, 14(10):2212–2222, Oct 2008.
- [188] Young-In Chi, Monika Martick, Monica Lares, Rosalind Kim, William G Scott, and Sung-Hou Kim. Capturing hammerhead ribozyme structures in action by modulating general base catalysis. *PLoS Biol*, 6(9):e234, Sep 2008.
- [189] Jason M. Thomas and David M. Perrin. Probing General Base Catalysis in the Hammerhead Ribozyme. *J. Am. Chem. Soc.*, 130:15467–15475, 2008.
- [190] Jason M Thomas and David M Perrin. Probing general acid catalysis in the hammerhead ribozyme. *J Am Chem Soc*, 131(3):1135–1143, Jan 2009.
- [191] Edith M Osborne, W. Luke Ward, Max Z Ruehle, and Victoria J DeRose. The identity of the nucleophile substitution may influence metal interactions with the cleavage site of the minimal hammerhead ribozyme. *Biochemistry*, 48(44):10654–10664, 2009.
- [192] Eric P. Schultz, Ernesto E. Vasquez, and William G. Scott. Structural and catalytic effects of an invariant purine substitution in the hammerhead ribozyme: implications for the mechanism of acid-base catalysis. *Acta. Crystallogr. D*, 70:2256–2263, 2014.
- [193] Fabrice Leclerc and Martin Karplus. Two-metal-ion mechanism for hammerhead-ribozyme catalysis. *J. Phys. Chem. B*, 110(7):3395–3409, 2006.
- [194] Tai-Sung Lee, Carlos Silva-Lopez, Monika Martick, William G. Scott, and Darrin M. York. Insight into the role of Mg^{2+} in hammerhead ribozyme catalysis from x-ray

- crystallography and molecular dynamics simulation. *J. Chem. Theory Comput.*, 3: 325–327, 2007.
- [195] Monika Martick, Tai-Sung Lee, Darrin M. York, and William G. Scott. Solvent structure and hammerhead ribozyme catalysis. *Chem. Biol.*, 15:332–342, 2008.
- [196] Tai-Sung Lee, Carlos Silva Lopez, George M. Giambaşu, Monika Martick, William G. Scott, and Darrin M. York. Role of Mg^{2+} in hammerhead ribozyme catalysis from molecular simulation. *J. Am. Chem. Soc.*, 130(10):3053–3064, 2008.
- [197] Tai-Sung Lee and Darrin M. York. Origin of mutational effects at the C3 and G8 positions on hammerhead ribozyme catalysis from molecular dynamics simulations. *J. Am. Chem. Soc.*, 130(23):7168–7169, 2008.
- [198] Tai-Sung Lee, George M. Giambaşu, Carlos P. Sosa, Monika Martick, William G. Scott, and Darrin M. York. Threshold Occupancy and Specific Cation Binding Modes in the Hammerhead Ribozyme Active Site are Required for Active Conformation. *J. Mol. Biol.*, 388:195–206, 2009.
- [199] Kin-Yiu Wong, Tai-Sung Lee, and Darrin M. York. Active participation of the Mg^{2+} ion in the reaction coordinate of RNA self-cleavage catalyzed by the hammerhead ribozyme. *J. Chem. Theory Comput.*, 7(1):1–3, 2011.
- [200] Philip C. Bevilacqua. Mechanistic considerations for general acid-base catalysis by RNA: Revisiting the mechanism of the hairpin ribozyme. *Biochemistry*, 42:2259–2265, 2003.
- [201] Hugh Heldenbrand, Pawel A. Janowski, George Giambaşu, Timothy J. Giese, Joseph E. Wedekind, and Darrin M. York. Evidence for the role of active site residues in the hairpin ribozyme from molecular simulations along the reaction path. *J. Am. Chem. Soc.*, 136:7789–7792, 2014.
- [202] Daniel J. Klein, Michael D. Been, and Adrian R. Ferré-D’Amaré. Essential role of an active-site guanine in *glmS* ribozyme catalysis. *J. Am. Chem. Soc.*, 129(48):14858–14859, 2007.

- [203] Sixue Zhang, Abir Ganguly, Puja Goyal, Jamie L. Bingaman, Philip C. Bevilacqua, and Sharon Hammes-Schiffer. Role of the active site guanine in the *glmS* ribozyme self-cleavage mechanism: Quantum mechanical/molecular mechanical free energy simulations. *J. Am. Chem. Soc.*, 137:784–798, 2015.
- [204] Nikolai B. Suslov, Saurja DasGupta, Hao Huang, James R. Fuller, David M J. Lilley, Phoebe A. Rice, and Joseph A. Piccirilli. Crystal structure of the Varkud satellite ribozyme. *Nature Chem. Biol.*, 11:840–846, 2015.
- [205] Yijin Liu, Timothy J. Wilson, Scott A. McPhee, and David M J. Lilley. Crystal structure and mechanistic investigation of the twister ribozyme. *Nature Chem. Biol.*, 10:739–744, 2014.
- [206] Bernhard Lippert. *Progress in Inorganic Chemistry*, volume 54, chapter Alterations of Nucleobase pK_a Values upon Metal Coordination: Origins and Consequences, pages 385–445. John Wiley and Sons, 2005.
- [207] Bernhard Lippert. Ligand–pK_a shifts through metals: potential relevance to ribozyme chemistry. *Chem. Biodivers.*, 5(8):1455–1474, 2008.
- [208] Maurizio Cossi, Vincenzo Barone, Roberto Cammi, and Jacopo Tomasi. Ab initio study of solvated molecules: a new implementation of the polarizable continuum model. *Chem. Phys. Lett.*, 255:327–335, 1996.
- [209] Maurizio Cossi, Giovanni Scalmani, Nadia Rega, and Vincenzo Barone. New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *J. Chem. Phys.*, 117:43–54, 2002.
- [210] Frank Weinhold and John E. Carpenter. *The Natural Bond Orbital Lewis Structure Concept for Molecules, Radicals, and Radical Ions*, pages 227–236. Springer US, Boston, MA, 1988. ISBN 978-1-4684-7424-4.
- [211] Senem Sanli, Yuksel Altun, and Gulsen Guven. Solvent Effects on pK_a Values of Some Anticancer Agents in Acetonitrile-Water Binary Mixtures. *J. Chem. Eng. Data*, 59:4015–4020, 2014.

- [212] Hans W. Horn, William C. Swope, Jed W. Pitera, Jeffrey D. Madura, Thomas J. Dick, Greg L. Hura, and Teresa Head-Gordon. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.*, 120(20):9665–9678, 2004.
- [213] Alberto Pérez, Iván Marchán, David Svozil, Jiri Sponer, Thomas E. Cheatham III, Charles A. Loughton, and Modesto Orozco. Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.*, 92:3817–3829, 2007.
- [214] Marie Zgarbová, Michal Otyepka, Jiří Šponer, Arnošt Mládek, Pavel Banáš, Thomas E. Cheatham III, and Petr Jurečka. Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.*, 7:2886–2902, 2011.
- [215] In Suk Joung and Thomas E. Cheatham III. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, 112:9020–9041, 2008.
- [216] Thomas Steinbrecher, InSuk Joung, and David A. Case. Soft-Core Potentials in Thermodynamic Integration: Comparing One- and Two-Step Transformations. *J. Comput. Chem.*, 32:3253–3263, 2011.
- [217] Christopher I. Bayly, Piotr Cieplak, Wendy D. Cornell, and Peter A. Kollman. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.*, 97:10269–10280, 1993.
- [218] Thomas Steinbrecher, David L. Mobley, and David A. Case. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J. Chem. Phys.*, 127:214108, 2007.
- [219] David Octavio Corona-Martinez, Paola Gomez-Tagle, and Anatoly K. Yatsimirsky. Electrophilic assistance to the cleavage of an rna model phosphodiester via specific and general base-catalyzed mechanisms. *J. Org. Chem.*, 77:9110–9119, 2012.

- [220] Robert V. Swift, Chau D. Ong, and Rommie E. Amaro. Magnesium-Induced Nucleophile Activation in the Guanylyltransferase mRNA Capping Enzyme. *Biochemistry*, 51:10236–10243, 2012.
- [221] Anna Lopata, Pablo G. Jambrina, Pankaz K. Sharma, Bernard R. Brooks, Judit Tóth, Beata G. Vértessy, and Edina Rosta. Mutations Decouple Proton Transfer from Phosphate Cleavage in the dUTPase Catalytic Reaction. *ACS Catal.*, 5:3225–3237, 2015.
- [222] Gergely N. Nagy, Reynier Suardíaz, Anna Lopata, Olivér Ozohanics, Károly Vékey, Bernard R. Brooks, Ibolya Leveles, Judit Tóth, Beata G. Vértessy, and Edina Rosta. Structural Characterization of Arginine Fingers: Identification of an Arginine Finger for the Pyrophosphatase dUTPases. *J. Am. Chem. Soc.*, 138:15035–15045, 2016.
- [223] John D. Chodera and Michael R. Shirts. Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. *J. Chem. Phys.*, 135:194110, 2011.
- [224] Danial Sabri Dashti and Adrian E. Roitberg. Optimization of Umbrella Sampling Replica Exchange Molecular Dynamics by Replica Positioning. *J. Chem. Theory Comput.*, 9:4692–4699, 2013.
- [225] Mikolai Fajer, Donald Hamelberg, and J Andrew McCammon. Replica-Exchange Accelerated Molecular Dynamics (REXAMD) Applied to Thermodynamic Integration. *J. Chem. Theory Comput.*, 4(10):1565–1569, October 2008.
- [226] Mikolai Fajer, Robert V. Swift, and J. Andrew McCammon. Using multistate free energy techniques to improve the efficiency of replica exchange accelerated molecular dynamics. *J. Comput. Chem.*, 30:1719–1725, 2009.
- [227] Wei Jiang, Milan Hodoseck, and Benoît Roux. Computation of absolute hydration and binding free energy with free energy perturbation distributed replica-exchange molecular dynamics. *J. Chem. Theory Comput.*, 5:2583–2588, October 2009.

- [228] Wei Jiang and Benoît Roux. Free energy perturbation Hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations. *J. Chem. Theory Comput.*, 6:2559–2565, 2010.
- [229] Yilin Meng, Danial Sabri Dashti, and Adrian E. Roitberg. Computing alchemical free energy differences with Hamiltonian replica exchange molecular dynamics (H-REMD) simulations. *J. Chem. Theory Comput.*, 7:2721–2727, 2011.
- [230] Mehrnoosh Arrar, Cesar Augusto F. de Oliveira, Mikolai Fajer, William Sinko, and J. Andrew McCammon. w-REXAMD: A Hamiltonian replica exchange approach to improve free energy calculations for systems with kinetically trapped conformations. *J. Chem. Theory Comput.*, 9:18–23, 2013.
- [231] Jason A. Wallace and Jana K. Shen. Continuous constant pH molecular dynamics in explicit solvent with pH-based replica exchange. *J. Chem. Theory Comput.*, 7:2617–2629, 2011.
- [232] Danial Sabri Dashti and Adrian E. Roitberg. pH-replica exchange molecular dynamics in proteins using a discrete protonation method. *J. Phys. Chem. B*, 116:8805–8811, 2012.
- [233] Jason M. Swails, Darrin M. York, and Adrian E. Roitberg. Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: Implementation, testing, and validation. *J. Chem. Theory Comput.*, 10:1341–1352, 2014.
- [234] Juyong Lee, Benjamin T. Miller, Ana Damjanovic, and Bernard R. Brooks. Enhancing Constant-pH Simulation in Explicit Solvent with a Two-Dimensional Replica Exchange Method. *J. Chem. Theory Comput.*, 11:2560–2574, 2015.
- [235] Christopher J. Woods, Jonathan W. Essex, and Michael A. King. The development of replica-exchange-based free-energy methods. *J. Phys. Chem. B.*, 107:13703–13710, 2003.

- [236] Ayori Mitsutake, Yoshiharu Mori, and Yuko Okamoto. Enhanced sampling algorithms. *Methods Mol. Biol.*, 924:153–195, 2013.
- [237] Wei Jiang, Yun Luo, Luca Maragliano, and Benoît Roux. Calculation of free energy landscape in multi-dimensions with Hamiltonian-exchange umbrella sampling on petascale supercomputer. *J. Chem. Theory Comput.*, 8:4672–4680, 2012.
- [238] Christina Bergonzo, Niel M. Henriksen, Daniel R. Roe, Jason M. Swails, Adrian E. Roitberg, and Thomas E. Cheatham III. Multidimensional replica exchange molecular dynamics yields a converged ensemble of an RNA tetranucleotide. *J. Chem. Theory Comput.*, 10:492–499, 2014.
- [239] Junchao Xia, William F. Flynn, Emilio Gallicchio, Bin W. Zhang, Peng He, Zhiqiang Tan, and Ronald M. Levy. Large-scale asynchronous and distributed multidimensional replica exchange molecular simulations and efficiency analysis. *J. Comput. Chem.*, 36:1772–1785, 2015.
- [240] Emilio Gallicchio, Junchao Xia, William F. Flynn, Baofeng Zhang, Sade Samlalsingh, Ahmet Mentesh, and Ronald M. Levy. Asynchronous replica exchange software for grid and heterogeneous computing. *Comput. Phys. Commun.*, 196236–246, 2015.
- [241] Asim Okur, Daniel R. Roe, Guanglei Cui, Viktor Hornak, and Carlos Simmerling. Improving Convergence of Replica-Exchange Simulations through Coupling to a High-Temperature Structure Reservoir. *J. Chem. Theory Comput.*, 3:557–568, 2007.
- [242] Adrian E. Roitberg, Asim Okur, and Carlos Simmerling. Coupling of replica exchange simulations to a non-Boltzmann structure reservoir. *J. Phys. Chem. B*, 111(10):2415–2418, 2007.
- [243] Rozita Laghaei, Normand Mousseau, and Guanghong Wei. Effect of the Disulfide Bond on the Monomeric Structure of Human Amylin Studied by Combined Hamiltonian and Temperature Replica Exchange Molecular Dynamics Simulations. *J. Phys. Chem. B*, 114:7071–7077, 2010.

- [244] Robert D. Malmstrom, Christopher T. Lee, Adam T. Van Wart, and Rommie E. Amaro. Application of Molecular-Dynamics Based Markov State Models to Functional Proteins. *J. Chem. Theory Comput.*, 10(7):2648–2657, 2014.
- [245] M. Olivia Kim, Patrick G. Blachly, Joseph W. Kaus, and J. Andrew McCammon. Protocols Utilizing Constant pH Molecular Dynamics to Compute pH-Dependent Binding Free Energies. *J. Phys. Chem. B*, 119:861–872, 2015.
- [246] M Olivia Kim, Patrick G. Blachly, and J Andrew McCammon. Conformational Dynamics and Binding Free Energies of Inhibitors of BACE-1: From the Perspective of Protonation Equilibria. *PLoS Comput. Biol.*, 11(10):1004341, October 2015.
- [247] Bernd Ensing, Marco De Vivo, Zhiwei Liu, Preston Moore, and Michael L. Klein. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Acc. Chem. Res.*, 39(2):73–81, 2006.
- [248] Eric Vanden-Eijnden. Some recent techniques for free energy calculations. *J. Comput. Chem.*, 30(11):1737–1747, 2009.
- [249] Tai-Sung Lee, Brian K. Radak, Ming Huang, Kin-Yiu Wong, and Darrin M. York. Roadmaps through free energy landscapes calculated using the multidimensional vFEP approach. *J. Chem. Theory Comput.*, 10:24–34, 2014.
- [250] Thakshila Dissanayake, Jason M. Swails, Michael E. Harris, Adrian E. Roitberg, and Darrin M. York. Interpretation of pH-Activity Profiles for Acid-Base Catalysis from Molecular Simulations. *Biochemistry*, 54:1307–1313, 2015.
- [251] Antons Treikalis, Andre Merzky, Haoyuan Chen, Tai-Sung Lee, Darrin M. York, and Shantenu Jha. RepEx: A Flexible Framework for Scalable Replica Exchange Molecular Dynamics Simulations. *arxiv:1601.05439 [cs.DC]*, 2016.
- [252] Brian K. Radak, Melissa Romanus, Emilio Gallicchio, Tai-Sung Lee, Ole Weidner, Nan-Jie Deng, Peng He, Wei Dai, Darrin M. York, Ronald M. Levy, and Shantenu Jha. *A Framework for Flexible and Scalable Replica-Exchange on Production Distributed CI*, pages 26:1–26:8. XSEDE '13. 2013. ISBN 978-1-4503-2170-9.

- [253] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kaleé, and Klaus Schulten. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.*, 26:1781–1802, 2005.
- [254] Edina Rosta and Gerhard Hummer. Error and efficiency of replica exchange molecular dynamics simulations. *J. Chem. Phys.*, 131:165102, 2009.
- [255] Edina Rosta and Gerhard Hummer. Error and efficiency of simulated tempering simulations. *J. Chem. Phys.*, 132:034102, 2010.
- [256] Nuria Plattner, J. D. Doll, and Markus Meuwly. Overcoming the Rare Event Sampling Problem in Biological Systems with Infinite Swapping. *J. Chem. Theory Comput.*, 9(9):4215–4224, 2013.
- [257] Junmei Wang, Piotr Cieplak, and Peter A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic biological molecules. *J. Comput. Chem.*, 21(12):1049–1074, 2000.
- [258] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.
- [259] Xuesong Shi, Daniel Herschlag, and Pehr A. B. Harbury. Structural ensemble and microscopic elasticity of freely diffusing DNA by direct measurement of fluctuations. *Proc. Natl. Acad. Sci. USA*, 110:1444–1451, 2013.
- [260] Gregory D. Hawkins, Christopher J. Cramer, and Donald G. Truhlar. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.*, 100:19824–19839, 1996.
- [261] Christopher J. Cramer and Donald G. Truhlar. Polarization of the nucleic acid bases in aqueous solution. *Chem. Phys. Lett.*, 198:74–80, 1992.
- [262] Vickie Tsui and David A. Case. Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers*, 56:275–291, 2001.

- [263] Thomas Gaillard and David A. Case. Evaluation of DNA Force Fields in Implicit Solvation. *J. Chem. Theory Comput.*, 7:3181–3198, 2011.
- [264] Shuichi Miyamoto and Peter A. Kollman. SETTLE: An analytic version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comput. Chem.*, 13:952–962, 1992.
- [265] David E. Smith and A. D. J. Haymet. Free energy, entropy, and internal energy of hydrophobic interactions: Computer simulations. *J. Chem. Phys.*, 98:6445–6454, 1993.
- [266] Justin L. MacCallum and D. Peter Tieleman. Computer Simulation of the Distribution of Hexane in a Lipid Bilayer: Spatially Resolved Free Energy, Entropy, and Enthalpy Profiles. *J. Am. Chem. Soc.*, 128:125–130, 2006.
- [267] Riccardo Baron, Piotr Setny, and J. Andrew McCammon. Water in Cavity-Ligand Recognition. *J. Am. Chem. Soc.*, 132:12091–12097, 2010.
- [268] Sunhwan Jo, Christophe Chipot, and Benoit Roux. Efficient Determination of Relative Entropy Using Combined Temperature and Hamiltonian Replica-Exchange Molecular Dynamics. *J. Chem. Theory Comput.*, 11:2234–2244, 2015.
- [269] Jason L. Smart, Tami J. Marrone, and J. Andrew McCammon. Conformational Sampling with Poisson-Boltzmann Forces and a Stochastic Dynamics/ Monte Carlo Method: Application to Alanine Dipeptide. *J. Comput. Chem.*, 18:1750–1759, 1997.
- [270] Christian Bartels and Martin Karplus. Multidimensional adaptive umbrella sampling: applications to main chain and side chain peptide conformations. *J. Comput. Chem.*, 18:1450–1462, 1997.
- [271] Pu Liu, Byungchan Kim, Richard A. Friesner, and B. J. Berne. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. USA*, 102:13749–13754, 2005.

- [272] M. Bonomi, A. Barducci, and M. Parrinello. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.*, 30(11):1615–1621, 2009.
- [273] Johannes Kästner. Umbrella integration in two or more reaction coordinates. *J. Chem. Phys.*, 131:034109, 2009.
- [274] Satoru G. Itoh and Hisashi Okumura. Hamiltonian replica-permutation method and its applications to an alanine dipeptide and amyloid- β (29-42) peptides. *J. Comput. Chem.*, 34(29):2493–2497, 2013.
- [275] Jeremy Esque and Marco Cecchini. Accurate calculation of conformational free energy differences in explicit water: the confinement-solvation free energy approach. *J. Phys. Chem. B*, 119(16):5194–5207, April 2015.
- [276] Anna S. Kamenik, Ursula Kahler, Julian E. Fuchs, and Klaus R. Liedl. Localization of Millisecond Dynamics: Dihedral Entropy from Accelerated MD. *J. Chem. Theory Comput.*, 9:3449–3455, 2016.
- [277] Takahiro Takekiyo, Takashi Imai, and Yoshihiro Taniguchi. Temperature and Pressure Effects on Conformational Equilibria of Alanine Dipeptide in Aqueous Solution. *Biopolymers*, 73:283–290, 2004.
- [278] Xuesong Shi, Steve Bonilla, Daniel Herschlag, and Pehr Harbury. Quantifying Nucleic Acid Ensembles with X-ray Scattering Interferometry. *Methods Enzymol.*, 558:75–97, 2015.
- [279] Xuesong Shi, Kyle A. Beauchamp, Pehr B. Harbury, and Daniel Herschlag. From a structural average to the conformational ensemble of a DNA bulge. *Proc. Natl. Acad. Sci. USA*, 111(15):1473–1480, April 2014.