

© 2016

Thakshila D. Dissanayake Rallage

ALL RIGHTS RESERVED

A MULTISCALE COMPUTATIONAL APPROACH TO STUDY RNASE A CATALYSIS

By

THAKSHILA D. DISSANAYAKE RALLAGE

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 Programs in Computational Biology and Molecular Biophysics &
Chemistry and Chemical Biology

Written under the direction of

Darrin M. York

And approved by

New Brunswick, New Jersey

January, 2016

ABSTRACT OF THE DISSERTATION

A multiscale computational approach to study RNase A catalysis

By THAKSHILA D. DISSANAYAKE RALLAGE

Dissertation Director:

Darrin M. York

Enzyme catalysis is an extremely important and complex process that is fundamental to biology. Experiments provide a wealth of valuable information about the function of enzymes; however, this information requires the use of computational models to establish a meaningful interpretation that can be used to guide design. Multiscale computational models, which integrate a hierarchy of theoretical methods to address complex biomolecular problems that span large spatial and temporal ranges, afford powerful tools to provide a detailed molecular level interpretation of a wide range of experimental data from which a consensus view of catalytic mechanism may emerge. In this dissertation, I detail my efforts to develop and apply multiscale methods to study the mechanisms of RNA backbone cleavage catalyzed by Ribonuclease A, an important archetype enzyme system, and the factors that regulate its activity. In the first phase of this research, I use molecular dynamics simulations to characterize the structure and dynamics of the active enzyme in solution at different stages along the reaction path. In this work, I demonstrate that the crystallographic structure represents an inactive, catalytically non-relevant state, and make predictions that a conformational change involving the flipping of the side chain of a conserved histidine residue (His12) is required to adopt a catalytically competent conformation. In the second phase of this research, I apply “constant pH molecular dynamics simulations” (CpHMD) to characterize the conditional probability of finding key active site residues in a protonation state that supports general acid-base catalysis. This allowed the prediction of pK_a shifts for His12, His119 and Lys41, and, for the first time, activity-pH profiles

for an enzyme system that can be compared directly with those measured in kinetic experiments. In the third phase of this research, I use combined quantum mechanical/molecular mechanical methods to study the catalytic chemical steps of transphosphorylation. Results of this work predict a free energy landscape for the reaction, from which the minimum free energy pathway that connects reactants and products allows a detailed molecular-level picture of mechanism. In the fourth phase of this research, I extend the CpHMD method to nucleic acid systems, to benchmark the method for the study of ribozymes that catalyze the same reaction as RNase A.

Acknowledgments

I would first like to thank my advisor, Dr. Darrin M. York, for his guidance and encouragement over the last five years. This thesis also would not have been possible without the work from our collaborators, including Dr. Jason Swails, Dr. Adrian Roitberg, (University of Florida), who supported developing constant-pH simulations in explicit solvent for AMBER program; Dr. Michael E. Harris (Case Western Reserve University), for providing the experimental insights in interpretation of the simulation data; and Dr. Ming Huang (Rutgers University, NJ) for supporting QMMM simulations; Dr. Tai-sung Lee, Dr. Tim Giese for technical and consulting support (Rutgers University, NJ). Nor would it have been possible without financial support from the NIH and NSF. I would also like to thank to former lab members for all the helpful discussions and valuable comments in this project, and Maria Panteva, Erich Kuechler, Dr. Timothy Giese and Dr. George Giambasu for all their help with the theoretical aspects of this work. I would also like to acknowledge Dr. Natalia Tretyakova (University of Minnesota) for introducing me to the theoretical computational field. I thank for the Biochemistry journal and the Methods in Enzymology journal for providing me the permission to fully or partially adapt the previous publications related to this dissertation. Finally, this work could not be accomplished without the support of my family; my husband, Uthpala and daughter, Resara for always making days so exciting and encouraging.

Dedication

This dissertation is dedicated to my mother and to my late father, who are my first teachers and to my loving husband, Uthpala and to my ever loving little princess, Resara.

Table of Contents

Abstract	ii
Acknowledgments	iv
Dedication	v
List of Tables	ix
List of Figures	xi
1. Introduction	1
2. Background	4
2.1. Enzyme catalysis and rate enhancement	4
2.2. Enzyme kinetics	4
2.2.1. Transition state theory (TST)	6
2.3. Theoretical aspects for enzyme catalysis	6
2.4. Modeling conformational changes	7
2.4.1. Theory of molecular dynamics simulations	8
2.5. Modeling protonation states	9
2.5.1. Constant-pH simulations	9
2.5.2. Theory of Constant-pH simulations	10
2.5.3. Constant-pH replica exchange	11
2.6. Modeling the reaction path and transition states	12
2.6.1. QM/MM simulations	12
2.6.2. QM/MM theory	13
2.6.3. Link atoms	14
2.6.4. Semiempirical AM1/d-PhoT	14
2.6.5. Free energy calculations	15

2.6.6. Umbrella sampling	15
2.7. Ribonuclease A	16
3. Catalytic mechanism of RNase A, 1: Characterization of the structure and dynamics in solution at different stages along the reaction path	21
3.1. Introduction	22
3.2. Methods	24
3.2.1. Determination of the “active state”	25
3.3. Results and Discussion	26
3.4. Conclusions	31
3.5. Supporting Information	31
3.5.1. Extended details of the simulated systems	31
3.5.2. Preparation of the simulation supercell	32
4. Interpretation of pH–Activity Profiles for Acid–Base Catalysis from Molecular Simulations	50
4.1. Methods	52
4.2. Results and Discussion	55
4.3. Supporting Information	63
4.3.1. Molecular Dynamics	63
4.3.2. pH-REMD simulations	64
4.3.3. pH-activity curves	64
4.3.4. Stability of cCMP in the active site	69
4.3.5. Convergence of fractions of protonations	69
4.3.6. Structural interpretation of the pK_a shifts of the active-site residues in cCMP-bound RNase A	69
4.3.7. The structural information of His119-Asp121 dyad	74
5. Catalytic mechanism of RNase A, 2: Prediction of reaction path of Ribonuclease A transphosphorylation from QM/MM simulations	76
5.1. Introduction	77
5.2. Materials and Method	78
5.2.1. Molecular dynamics simulations	78
5.2.2. QM/MM simulations	78

5.2.3. Umbrella sampling	78
5.2.4. Kinetic isotope effects calculations	79
5.3. Results and Discussion	80
5.3.1. Determination of initial structure for QMMM simulations	80
5.3.2. Obtaining the reaction paths	81
5.3.3. Impact of the selection of the QM region	81
5.3.4. Exploring the uncatalyzed reaction	82
5.3.5. Modeling uncatalyzed reaction	82
5.3.6. Reaction paths starting from His12 non-flipped conformation	83
5.3.7. Generating reaction paths for RNase A with His12 flipped conformation	85
5.4. Conclusions	88
6. Assessment of ionization of nucleic acids using constant-pH explicit solvent simulations	96
6.1. Introduction	96
6.2. Methods	97
6.2.1. Reference compounds	97
6.2.2. RNA systems	97
6.3. Results and discussion	98
6.4. Conclusions	104
Bibliography	107

List of Tables

3.1. Summary of the simulated systems, including protonation states of the active site residues and conformations of His12	26
3.2. Average key distances of the active site in the inactive state crystal simulation . . .	27
3.3. The percent occupancy for the in-line attack angle (θ_{Inl}), occupancies of DH-A distances (d_{DH--A}), D-H-A angles (θ_{D-H-A}) and co-occupancies of d_{DH--A} vs. θ_{Inl} for the reactant state models.	29
3.4. Summary of average distance of key active site interactions in the reactant state model simulations.	41
3.5. The average key distances in the active site of the activated precursor state models.	42
3.6. The comparison of the key average distances in the transition state consist of His12 flipped conformation (TS^F), to the distances in the ribonuclease A-uridine vanadate complex.	43
3.7. Key average distance in the active site of 2',3'-cyclic phosphate product state bound RNase A.	44
3.8. The average distance in the active site from the crystal simulation	48
4.1. Comparison of experimental and simulated macroscopic pK_a s for apo and cCMP-RNase A using the Hill equation and the "apparent pK_a " model.	56
4.2. Comparison of experimental and simulated microscopic pK_a s using the microscopic model illustrated in Figure 4.1.1 for apo and cCMP-bound RNase A.	57
4.3. The occupancy of the key contacts to the substrate (cCMP) at 5 Å cutoff along the pH ladder and the average distances. The numbers in parenthesis are standard deviations.	69
4.4. The occupancy of the key contacts to the substrate at 5 Å cutoff along the pH ladder and the average distance	70
4.5. The occupancy of H119:N(NE2/ND1)-D121:O(OD1/OD2) distance along the pH-ladder for apo RNase A at cutoff 5 Å	74
4.6. The occupancy of H119:N(NE2/ND1)-D121:O(OD1/OD2) distance along the pH-ladder for cCMP-bound RNase A at cutoff 5 Å	75

5.1. The summary of QM/MM simulations and the QM regions	79
5.2. The energies of each state along the reaction path for the simulated systems	86
5.3. The average distances at each stages for each system of simulation.	87
5.4. Comparison of calculated Kinetic Isotope Effect values with experimental measurements	88
6.1. The reference pK_a values, TI energies and adjusted reference energy for reference compounds.	98
6.2. The calculated pK_a for dinucleotides	100
6.3. Comparison of calculated dihedral angles of dinucleotides with the experimental values	103
6.4. Comparison of calculated pK_a values of Branch-point helix (BPH) with experiment .	104
6.5. The distance between neighboring glycosidic nitrogen of dinucleotides	106

List of Figures

2.1. The figure illustrates the energetics of conversion of substrate (S) to product (P) in an enzymatic and non-enzymatic environment. E_a is the activation energy for the reaction which is the minimum energy required by the reactant species to overcome the reaction barrier in order to be converted into the product species. The nature of the substrate and product states remains unchanged while the activation energy is lowered by the catalyst.	5
2.2. The figure highlights the QM and MM regions used in the hybrid QM/MM simulations for the Ribonuclease A system. The QM region is defined as the active site of the enzyme where the biochemical reaction occurs. This region consists of a small number of atoms and is treated at the <i>ab-initio</i> level or with semi-empirical methods. The MM region is larger in size and the MM force field is applied on this region.	13
2.3. Active-site P1 and the sub-sites; B1,B2,B3,P0 and P2 of RNase A. The active site is made up of Gln11, His12, Lys41, His119 and Asp121. The P2 and P0 sub-sites are made up of Lys7, Arg10 and Lys66. The B1 and B2 sub-sites are defined by the Thr45, Aps83 residues and the Asn71, Glu111 residues, respectively ¹	19
2.4. Mutational effects of key active site residues. The red color represents the strongest effect. The orange color represents a medium effect and the green color represents a weak effect on catalytic rate.	20
3.1. The general interpretation of transphosphorylation and hydrolysis steps in RNase A catalytic mechanism. In the transphosphorylation (left), His12 acts as the general base to take the proton from O2' to promote a nucleophilic attack on the phosphate group, and His119 acts as the general acid to donate a proton to O5' to facilitate the cleavage of P-O5' bond. In the hydrolysis mechanism (right), neutral His119 acts as the general base to take the proton from a water molecule, promoting a nucleophilic attack on the phosphate group, and protonated His12 donates a proton back to O2' to facilitate the cleavage of P-O2' bond.	33

3.2.	The schematic of the active site orientations: A- The crystal structure (with added protons). B- The reactant state with His12 non-flipped side chain ($R(H12^\delta)$) state. C- The reactant state with His12 flipped side chain ($R^F(H12^\delta)$). D- The activated precursor state with His12 flipped sidechain (AP^F). E- The late-transition state mimic with His12 flipped conformation (TS^F), F- 2', 3'- cyclic phosphate product state (P). Hydrogen bonds are indicated in dotted lines.	34
3.3.	The average structures of active sites of each stage along the reaction path for, A- crystal structure, B- The reactant state model initiated from the coordinates of the crystal structure ($R(H12^\delta)$), C- "Active state" model, where His12 side chain is flipped and deprotonated at N ϵ site positioning to abstract the proton from O2' ($R^F(H12^\delta)$). His119 and Lys41 are in fully protonated state, D- Activated precursor state, His12 has accepted the proton from O2' (AP^F). The accumulating negative charge is stabilized by protonated Lys41. His119 is positioned towards the O5' of the leaving group, E- Late transition state, His12, His119 and Lys41 are protonated in the active site. His119 is position to donate the proton to O5' (TS^F), F- Product state where both His12 and Lys41 are fully protonated (P). His119 is deprotonated after the departure of the leaving group. The N δ site of His119 is likely to accept a proton from a water molecule in order to support the hydrolysis.	35
3.4.	Comparison of experimental and crystal simulation characteristics. Top: overlay of the active site of the refined structure (PDB:1RPG) depicted as colored licorice and the simulation average electron density (black mesh, contoured at 1sigma). Bottom: Comparison of C-alpha B factor values from the crystal simulation (red) with the experimental values (black) for all residues in the enzyme and the inactive substrate. The Pearson correlation coefficient is 0.67 and RMSD of the B-factor is 3.94 Å . . .	36
3.5.	A distribution of the d_{DH--A} vs. θ_{Inl} plots for 6 different active states. Left: the distributions for the reactant states where Lys41 acting as the general base. Right: the distributions for the reactant states where His12 acting as the general base. . . .	37
3.6.	The percent occupancies of the inline attack angle (θ_{Inl}), donor acceptor distance (d_{DH--A}) and D-H-A angle (θ_{D-H-A}) for six different reactant state mimics. The first and second three sets of the plot represent the reactant states with neutral Lys41 and neutral His12, respectively. The reactant state $R^F(H12^\delta)$, meet the criteria to be the best candidate for the "active state". The properties, that reached 80% or more of each criteria, is indicated in a green check mark; otherwise, indicated in a black 'x' mark.	37

3.7. Pseudorotation cycle for furanose conformations. The values and labels along the circumference indicates the phase angle of pseudorotation and the conformation of the sugar.	38
3.8. Polar plots of sugar puckering for cytosine ribose in the reactant state models: a) R(K41 ^ζ), b) R(H12 ^δ , K41 ^ζ), c) R(H12 ^ε ,K41 ^ζ), d) R(H12 ^δ), e) R(H12 ^ε), f) R ^F (H12 ^δ). Each conformations are shown with the analogous phase angle. The amplitude is shown along the coordinates of the radius. The sugar pucker of R ^F (H12 ^δ) state stays in C3'-endo conformation.	39
3.9. Polar plots of sugar puckering for adenosine ribose in the reactant state models: a) R(K41 ^ζ), b) R(H12 ^δ , K41 ^ζ), c) R(H12 ^ε ,K41 ^ζ), d) R(H12 ^δ), e) R(H12 ^ε), f) R ^F (H12 ^δ). Sugar pucker shows frequent transitions between C2'-endo and C3'-endo conformations.	40
3.10. K41:NZ–C125:C2' distance vs. simulation time plots and the distribution of K41:NZ–C125:C2' distance for both crystal unit cells. The results indicates two conformations of Lys41 side chain in the first unit cell. The conformation centered at distance 4.67 Å is lower in occupancy comparative to the conformation at distance 7.10 Å in the first unit cell. A broad conformational fluctuation is observed in the second unit cell and the average value is located at the 8.95 Å.	45
3.11. The plots of backbone RMSD of the reactant state models with respect to the crystal structure. a) The RMSD plots for the reactant state model when His12 act as the general base where His12 is in neutral state. b) The RMSD plots for the reactant state model when Lys41 is in deprotonated state.	45
3.12. The RMSD of the active site residues (His12, His119 and Lys41) in the reactant states models with respect to the crystal structure. a) The RMSD plots for the reactant state model when His12 act as the general base where His12 is in neutral state. b) The RMSD plots for the reactant state model when Lys41 is in deprotonated state.	46
3.13. RMSD of the substrate in comparison to the inactive substrate in the crystal structure is shown in the figure. a) The set of RMSD plots of the substrate, which represent the His12 deprotonated state in the reactant state model. b) RMSD plots of the substrate when Lys41 is in its deprotonated state.	46
3.14. RMSD of average structure for 100 ns of the crystal simulation (Xtal) is 0.34 Å for all backbone atoms and 0.70 Å for all heavy atoms.	47

3.15.	The radial distribution functions ($g(r)$) of water molecules around N δ of His119, O2', O3', phosphoryl oxygens (OP1, OP2) and phosphate (P). a) The plots for the product state when neutral His119 is having a proton on N δ site b). The plots for the product state when neutral His119 is having a proton on N ϵ site	49
4.1.	The generally accepted view is that H119/H12 acts as the general acid/base pair in transphosphorylation, and their roles are reversed in hydrolysis. However, there has been some debate in the literature that alternatively, K41 might act as general base in transphosphorylation, although this is less widely accepted.	52
4.2.	Titration curves (lines) fitted to simulation data (points) with the Hill equation [eq 4.8]. The upper panel represents the apo enzyme and the lower panel represents the cCMP-bound enzyme. The Hill coefficients for His119, His12 and Lys41 are 0.94, 0.94, 0.98 for the apo enzyme and 1.04, 1.09, 1.10 for the cCMP-bound enzyme, respectively.	56
4.3.	<i>Microscopic pK_a model results with H12/H119 acting as the general acid and base.</i> The plots of the logarithm of protonated fractions, $\log(f)$, versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data for all fractions to the equations derived from the microscopic model (Figure 4.1.1) with His12/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively, as depicted in Figure 4.1. The simulation data fits well with RMS errors of 0.22 (apo) and 0.17 (cCMP) for the $\log(f)$ values. The $\log(f)$ maximums (-0.45 and -0.30) for the curve of the active fraction, $f_{(AH^+/B^-)}$, are at 6.1 (apo) and at 7.5 (cCMP), respectively.	58

4.4. *Apparent pK_a model results with H12/H119 acting as the general acid and base.* The plots of the protonated fractions, $\log(f)$, versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data to ONLY the ACTIVE FRACTION ($f_{(AH^+/B^-)}$, shown in red) with His12/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively, as depicted in Figure 4.1. The model assumes no coupling of the microstates, and although only the active fraction is considered in the fitting (as would be the case experimentally), the parameters nonetheless are able to determine the other fractions which can be compared with the simulation data. The upper panel represents the apo RNase A and the lower panel represents the cCMP bound RNase A. The RMS errors for the $\log(f)$ values are 0.25 (apo) and 0.18 (cCMP) respectively. The $\log(f)$ maximums (-0.43 and -0.30) for the curve of the active fraction, $f_{(AH^+/B^-)}$, are at 6.1 (apo) and at 7.5 (cCMP), respectively. 60

4.5. *Microscopic pK_a model results with K41/H119 acting as the general acid and base.* The plots of the logarithm of protonated fractions, $\log(f)$, versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data for all fractions to the equations derived from the microscopic model (Figure 4.1.1) with Lys41/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively. This is not the generally accepted mechanism depicted in Figure 4.1, but has nonetheless received some support in the literature^{2,3} and so is considered here for comparison. The RMS errors for the $\log(f)$ values are 0.31 (apo) and 0.44 (cCMP) respectively. The $\log(f)$ maximums (-2.09 and -0.05) for the curve of the active fraction, $f_{(AH^+/B^-)}$, are at 7.7 (apo) and at 8.4 (cCMP), respectively. 61

- 4.6. *Apparent pK_a model results with K41/H119 acting as the general acid and base.* The plots of logarithm of protonated fractions, $\log(f)$, versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data to ONLY the ACTIVE FRACTION ($f_{(AH^+/B^-)}$, shown in red) with Lys41/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively. This is not the generally accepted mechanism depicted in Figure 4.1, but has nonetheless received some support in the literature^{2,3} and so is considered here for comparison. The model assumes no coupling of the microstates, and although only the active fraction is considered in the fitting (as would be the case experimentally), the parameters nonetheless are able to determine the other fractions which can be compared with the simulation data. The upper panel represents the apo RNase A and the lower panel represents the cCMP bound RNase A. The RMS errors for the $\log(f)$ values are 0.33 (apo) and 0.48 (cCMP) respectively. The $\log(f)$ maximums (-2.69 and -0.11) for the curve of the active fraction, $f_{(AH^+/B^-)}$, are at 7.5 (apo) and at 8.4 (cCMP), respectively. 62
- 4.7. *Microscopic pK_a model results with H12/H119 acting as the general acid and base.* The plots of the probability versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data to the equations derived from the microscopic model with His12/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively. These plots were used in deriving microscopic pK_a parameters discussed in the main text (Table 5.3, Table 4.2). . . . 65
- 4.8. *Apparent pK_a model results with H12/H119 acting as the general acid and base.* The plots of the probability versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data to ONLY the ACTIVE FRACTION ($f_{(AH^+/B^-)}$, shown in red) with His12/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively. These plots were used in deriving apparent pK_a parameters discussed in the main text (Table 5.3, Table 4.2). 66

4.9. <i>Microscopic pK_a model results with K41/H119 acting as the general acid and base.</i>	
The plots of the probability versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data to the equations derived from the microscopic model with Lys41/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively. This is not the generally accepted mechanism. These plots were used in deriving microscopic pK_a parameters discussed in the main text (Table 5.3, Table 4.2).	67
4.10. <i>Apparent pK_a model results with K41/H119 acting as the general acid and base.</i>	
The plots of logarithm of probability versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data to ONLY the ACTIVE FRACTION ($f_{(AH^+/B^-)}$, shown in red) with Lys41/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively. This is not the generally accepted mechanism. These plots were used in deriving apparent pK_a parameters discussed in the main text (Table 5.3, Table 4.2).	68
4.11. The cumulative averaged protonated fractions of His12, Lys41 and His119 at each pH for apo-RNase A	70
4.12. The cumulative averaged protonated fractions of His12, Lys41 and His119 at each pH for cCMP bound RNase A	71
4.13. The running averages (1 ns window) of the fraction of protonated His12, Lys41 and His119 at each pH for apo RNase A	72
4.14. The running averages (1 ns window) of the fraction of protonated His12, Lys41 and His119 at each pH for cCMP-bound RNase A	73
5.1. The schematic of QM regions indicates both DNU and RDNU regions. The black circle represent the full QM region (DNU), where the whole dinucleotide and the active site residues His12, Lys41 His19 are included in the QM region. The red color represent the small QM region (RDNU) where the truncated form of the dinucleotide and the active site residues are included in the QM region.	80

5.2.	The reaction coordinates used in this study is represented in the figure. The protein residues (His12, His119) are colored in blue and the substrate is colored in black. His12 and His119 residues considered as the general base and the acid, respectively. The generalized coordinates are defined as the difference between the breaking bond and the forming bond; R1-R2, R4-R3 and R6-R5 where, R1 = O2'-OH2', R2= OH2'-H12:ND1, R3 = O2'-P, R4= P-O5', R5 = H119:HD1-O5', R6= H119:ND1-H119:HD1.	81
5.3.	The average structure of the reactant state model from 400 ns long simulation ("active state"), where His12 has its flipped conformation. Ne of His12 is locate at 2.02 Å away from the OH2' and poise to abstract the proton with 159° of average hydrogen bond angle of O2'-OH2'-NE:H12	82
5.4.	The departing structures for the QM/MM simulations. Left: the transition state mimic where His12 is in its flipped state, right: the transition state where His12 is in its non-flipped state.	83
5.5.	1D PMF of uncatalyzed transphosphorylation reaction of CpA. A minimum and a transition state is observed along the reaction path with a 20.80 kcal/mol of reaction barrier. R4=R _{P-O5'} and R3=R _{P-O2'}	84
5.6.	The structure of the minimum and the transition state for the uncatalyzed reaction. The minimum indicates a structure of a negatively charged activated precursor state and the transition state shows a negatively charged, late-like transition state.	85
5.7.	2D free energy profile of phosphoryl transfer vs. general acid in DNU-TS-NF _{Acid} state where QM region is fully occupied with the substrate CpA and active site residues His12, Lys41 and His119. Two minimum and one dianionic transition state is observed. The barrier of the reaction is 15.13 kcal/mol. The first minima indicates deprotonated Lys41 acting as the general base.	89
5.8.	2D free energy profile of phosphoryl transfer vs. general base for DNU-TS-NF _{Base} state where CpA substrate, His12, Lys41 and His119 are fully occupied in the QM region. Three minimum and two transition states are observed with 21.68 kcal/mol of overall reaction barrier from a reactant state of His12 acting as the general base.	90
5.9.	The 2D free energy profile for phosphoryl transfer vs. general acid in RDNU-TS-NF _{Acid} state where the QM region is fully occupied with His12, Lys41 and His119 and partially occupied with CpA substrate. Two minimums and a transition state are observed. The reactant state indicates Lys41 acting as the general base with 15.81 kcal/mol reaction barrier to the triester-like transition state.	91

5.10. The 2D free energy profile for the phosphoryl transfer vs. general base step in RDNU-TS-NF _{Base} state with the truncate QM region. A triester-like transition state is observed along the reaction path with 26.47 kcal/mol reaction barrier.	92
5.11. 2D free energy profile for the phosphoryl transfer vs. general base step when flipped His12, Lys41, His119 and CpA are occupied in the QM region (DNU-TS-F _{Base}). 21.91 kcal/mol of high barrier early transition state is observed.	93
5.12. The 2D free energy profile for phosphoryl transfer vs. general acid step when flipped His12 and CpA substrate fully occupied in the QM region (DNU-TS-F _{Acid}). The reaction barrier is 16.25 kcal/mol from the reactant state to the dianionic transition state.	94
5.13. 2D free energy profile for the phosphoryl transfer vs. acid step (upper panel) and the phosphoryl transfer vs. base step (lower panel) when His12 is in its flipped state and Lys41 is absence in the QM region.	95
6.1. The figure illustrates the ionization of A,C,G,U nucleobases. The pK _a of N1 site of adenosine and guanosine is 3.50 and 9.20, respectively. The pK _a of N3 site of cytidine and uridine is 4.20 and 9.20, respectively.	97
6.2. The running average, calculated for 1 ns time intervals, for 9 ns of the simulations for adenine nucleobase (A).	99
6.3. The running average, calculated for 1 ns time intervals, for 9 ns of the simulations for cytosine nucleobase (C).	100
6.4. The running average, calculated for 1 ns time intervals, for 9 ns of the simulations for guanine nucleobase (G).	101
6.5. The running average, calculated for 1 ns time intervals, for 9 ns of the simulations for uracil nucleobase (U).	101
6.6. The figure shows the titration curves for the reference compounds after applying the corrected reference energy. The upper left shows the curve for adenine protonation in AMP. Upper right shows the guanine titration curve in GMP. Lower left shows the cytosine titration curve in CMP. Lower right shows the uracil titration curve in UMP. All the computed values are closely reproduce the experimental pK _a values.	102
6.7. Occupancies of the stacking interactions between neighboring N atoms of each nucleobase is presented in the plot. The criteria for interaction is chosen as the distance between N atoms (R(NN)) < 5 Å. ApC shows highest stacking interactions at pH=7 and the lowest interactions are observed in GpG.	102

6.8. The calculated pK_a values and the titration curves for the dinucleotide systems. . .	105
6.9. The structural overlay of the average structures of ESS3 along the pH ladder of pH=2 to 13. The nucleobases of A7 and C21 residues show slight structural fluctuation due to titration of A7 residue. The deviation of the pH has no significant effect on the backbone of the structure.	106

Chapter 1

Introduction

This dissertation presents the application of multiscale computational methods to study the catalytic mechanism of Ribonuclease A (RNase A), and progress toward the development of improved models for biocatalysis simulations. Specifically, the dissertation describes the use of molecular dynamics simulations, constant-pH replica exchange simulations in explicit solvent and hybrid quantum mechanical and molecular mechanical simulations (QM/MM) to ultimately unravel the detailed mechanism of RNase A. These studies provide detailed, quantitative predictions about the roles of the active site residues and the nature of the transition state. Overall, the work presented here showcases the applicability of a multiscale computational method to predict the catalytic mechanisms of enzymes.

The Chapter 2 reviews the current understanding of the main principles of enzyme catalysis, rate enhancement and multiscale computational methods for exploring mechanistic details of catalyzed reactions. The first two sections focus on the details of enzyme catalysis, followed by enzyme kinetics. Section 2.3 discusses the important aspects of enzyme catalysis that can be explored by theoretical studies. The Sections 2.4-2.6 describe the problem space that emerges in identifying a detailed mechanism and discuss the theory of multiscale computational methods for addressing the problem space. Specifically, section 2.4 discusses the problem space of the conformational changes and the application and theory of molecular dynamics simulations to explore conformational changes of biomolecular catalysts. The section 2.5 discusses the problem space arising from considering the different protonation states of the enzyme residues and the use of constant-pH simulations to address the problem space. Sections 2.5 and 2.5.1 are reprinted from “*Methods in Enzymology, Computational Methods for Understanding Riboswitches, Volume 553, Maria T. Panteva, Thakshila Dissanayake, Haoyuan Chen, Brian K. Radak, Erich R. Kuechler, George M. Giambasu, Tai-Sung Lee, Darrin M. York, Multiscale methods for computational RNA enzymology, Pages 335–374, Copyright (2015), with permission from Elsevier.*” The section 2.6 focuses on the modeling of the reaction path of RNase A using hybrid quantum and molecular mechanics (QM/MM) simulations. Finally, the last section gives a broad overview of Ribonuclease A catalysis, including proposed mechanisms and a summary of the wealth of experimental evidence available.

Chapter 3 focuses on the dynamical behavior of RNase A and the active site conformations of the enzyme using molecular dynamics in conjunction with molecular mechanics force fields. The chapter discusses the applicability of this computational method in elucidating the mechanistic details by identifying relevant active states of the enzyme substrate complex. Further, the response of the enzyme environment to stabilize different key stages along the reaction coordinate is scrutinized by performing molecular dynamics along the reaction path. The work presented here identifies the conformational and protonation events that lead to a catalytically active state, setting the stage for further investigation of the detailed catalytic mechanism using QM/MM methods.

Chapter 4 discusses the application of a novel explicit solvent constant-pH replica exchange molecular dynamics (pH-REMD)⁴ method to investigate the populations of correlated protonation states of active site residues in Ribonuclease A, in order to provide information on the relation between pH and catalytic rate. The pH-rate profiles are a key piece of information used to infer mechanistic details of acid-base catalysis, allowing the assignment of the roles of the general base and the general acid. In this work, constant-pH MD simulations, together with a kinetic model, are used together for the first time to predict pH-activity curves, and microscopic and macroscopic pK_{as} ⁵. The chapter is “*reprinted with permission from Biochemistry journal, volume 54, 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, Page 1307-1313, 2015]. Copyright [2015] American Chemical Society*”

Chapter 5 explores a detailed mechanism of RNase A catalysis using hybrid quantum mechanical/molecular mechanical (QM/MM) simulations. The chapter is a direct follow up to the work conducted in Chapter 3 as here we depart from the presumed conformations and protonation states discovered from the MD simulations. These calculations were used to obtain two dimensional (2D) free energy profiles for the phosphoryl transfer reaction coordinate vs. general acid/base reaction coordinates to gain insight into the reaction mechanism. The nature of the minimum structures, transition states and the product states are also described in detail. Moreover, the dependence of the reaction barrier heights on the size of the QM region is briefly discussed in this chapter.

Chapter 6 focuses on the improvement of force field parameters for nucleic acid systems in order to be used in constant-pH replica exchange simulations of RNA systems. The set of optimized parameters and applications on a dinucleotide systems from larger systems are presented. The calculated pK_{as} and the structural properties are discussed for each system studied. This method will ultimately be applied on the ribozyme systems such as *glms*, hairpin and twister ribozymes to predict the pK_{as} as well as pH-rate profiles in the near future.

The work presented throughout this dissertation showcases the broad impact of application of

multiscale computational methods to identify the detailed mechanism of biological catalysts. Together with experimental studies, multiscale methods can be used to provide a robust interpretation of the mechanistic details of enzyme catalysis and the origin of the rate enhancement by enzyme catalysis.

Chapter 2

Background

2.1 Enzyme catalysis and rate enhancement

Every second, an enormous amount of biochemical reactions are taking place inside a living cell. These reactions, mediated by enzymes, are essential parts of bodily functions including metabolism, degradation and formation of biomolecules, building of tissues, detoxification, food digestion, and a host of other life processes. Living systems produce enzymes that serve as catalysts for biochemical reactions. Unlike uncatalyzed reactions, which are generally non spontaneous and time consuming, catalyzed reactions often occur on the millisecond time scale, and support a smooth flow of biochemical tasks.

Figure 2.1 illustrates the energetics of catalyzed and uncatalyzed reaction⁶. A higher activation energy indicates a slower reaction and a catalyst helps to decrease the reaction barrier by lowering the activation energy. The process generates low barrier transition states, without altering the equilibrium between the reactants and products or consuming the catalyst⁶. The rate enhancement by enzyme catalysts is billion to trillion-fold⁷ larger compared to the *in vitro* reactions due to their higher substrate specificity and ability to functioning at body temperature and pressure.

2.2 Enzyme kinetics

Enzyme kinetics explore the time dependence of enzymatic reactions which provides valuable information about the detailed reaction mechanisms, and a measure of rate information that helps to establish a relationship between structure and reactivity. An enzymatic reaction is illustrated in two steps: (1) formation of the enzyme substrate complex (ES) and (2) formation of products (P) as shown in chemical equation 2.1^{6,8}. The process of formation of the enzyme substrate complex is a reversible equilibrium where K_s is the equilibrium constant. The formation of products (P) occur with a first order rate constant k_{cat} , also called the turnover number⁹.



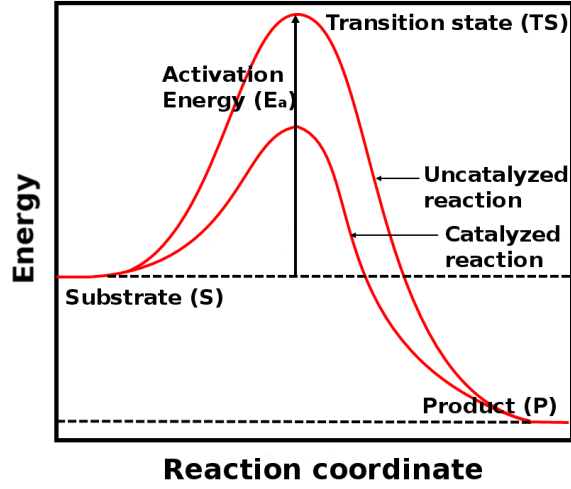


Figure 2.1: The figure illustrates the energetics of conversion of substrate (S) to product (P) in an enzymatic and non-enzymatic environment. E_a is the activation energy for the reaction which is the minimum energy required by the reactant species to overcome the reaction barrier in order to be converted into the product species. The nature of the substrate and product states remains unchanged while the activation energy is lowered by the catalyst.

Where K_s is binding constant and k_{cat} is first order rate constant (turnover number).

$$K_s = \frac{[E][S]}{[ES]}. \quad (2.2)$$

The rate of the reaction is given by,

$$\text{Rate} = k_{cat}[ES]. \quad (2.3)$$

The total enzyme concentration $[E]_0$,

$$[E]_0 = [E] + [ES] \quad (2.4)$$

where $[E]$ is the free enzyme concentration and $[ES]$ is the concentration of the enzyme substrate complex.

Combination of equation 2.2 and equation 2.4 yields,

$$[ES] = \frac{[E]_0[S]}{K_s + [S]}. \quad (2.5)$$

Thus the rate of the reaction is expressed by the Michaelis-Menten equation⁹,

$$\text{Rate} = \frac{[E]_0[S]k_{cat}}{K_s + [S]} \quad (2.6)$$

At high substrate concentration, all the enzyme converts into the enzyme-substrate complex ([ES]). Therefore, $[E]_0 = [ES]$. Consequently,

$$Rate = k_{cat}[E]_0. \quad (2.7)$$

At low substrate concentration, $[S] \ll K_s$ and

$$Rate = \frac{k_{cat}}{K_s}[E]_0[S]. \quad (2.8)$$

The equations 2.7 and 2.8 represent unimolecular and bi-molecular reactions, respectively.

2.2.1 Transition state theory (TST)

In transition state theory, a thermodynamic equilibrium is considered between the transition state ($[ES]^\ddagger$) and the ground state. By applying the thermodynamic relationship between Gibbs free energy (ΔG) and equilibrium constant (K),

$$\Delta G = RT \ln(K), \quad (2.9)$$

the relationship between the rate constant and the activation free energy is given by¹⁰,

$$k_{cat} = \left(\frac{kT}{h} \right) \exp \left(\frac{\Delta G_{cat}^\ddagger}{RT} \right). \quad (2.10)$$

T- absolute temperature, R- gas constant, k- Boltzmann constant h- Plank constant ΔG_{cat}^\ddagger - Gibbs energy of activation.

2.3 Theoretical aspects for enzyme catalysis

The topic of understanding efficiency of enzyme catalysis and achievement of extraordinary rate acceleration has been challenging over decades¹¹, for which, many hypotheses are proposed to identify the key contributing factors^{11,7}. The experimentally proposed factors are substrate binding, metal ion enzyme interactions and binding in the active site, protonation states and the conformational state of the active site residues, enzyme conformations¹² and reaction path and transition state stability^{13,10}. From theoretical perspective, these factors are identified as the “problem space” and are addressed using multiscale computational methods to explore catalytic pathways¹².

Multiscale simulations involve the integration of a hierarchy of models that, used together, can solve problems that span multiple spatial and/or temporal scales^{14,15,16,17,18,19,20}. In the context of biocatalysis, this implies consideration of the enzyme, along with its substrate and any required

cofactors, acting in a realistic condensed-phase environment. The active site, where chemical bond formation or cleavage occurs, generally requires a computationally intensive quantum mechanical (QM) description^{8,21,16,19,22}, and this problem is amplified by the need for different levels of phase space sampling to capture events that occur on vastly different timescales. Consequently, biocatalysis applications demand the use of multiscale models that can overcome these challenges in a fashion that is both reliable and computationally tractable.

2.4 Modeling conformational changes

Enzymes undergo conformational changes upon substrate binding, catalysis and product releasing events^{23,24}, on the microscopic or macroscopic scale. Loop motion, domain motions and secondary structure changes²⁵ are considered macroscopic events. The conformations of the active site residues and their interactions with the substrate, side chain rotations, conformational changes due to hydrogen bond breaking and forming, backbone conformational changes, substrate dynamics and sugar pucker changes fall in to the category of microscopic conformational events. These events can occur either before^{26,27,28,29} or after^{30,31} binding or releasing events of the substrate. Mostly, the changes that take place after binding events are related to the conversion of the enzyme substrate complex (ES) to the product state (P). These events could possibly trigger catalysis and the rate enhancement of the reaction. Therefore, identification of conformational changes are essential in the study of enzyme catalysis.

Conformational changes are studied by exploring the dynamics of the enzyme systems. In computational modeling, sampling techniques are used to address such problem to investigate the dynamical behavior and conformational fluctuations of the enzymes³². The most common sampling techniques are described below. Deterministic search algorithms are developed based on previous knowledge³² and can be used to eliminate some conformational changes in comparison to the existing data. This approach includes search for local minima of a small system, followed by and combining into larger fragments, self-consistent electrostatic field (SCEF) method³³, knowledgebased statistical methods, rotamer libraries, homology modeling. Stochastic hybrid methods, including Monte Carlo with minimization and self-consistent basin-to-deformed basin mapping. Genetic algorithms includes conformational space annealing. Heuristic search algorithms such as, Monte Carlo (MC) chain growing methods, Metropolis MC and molecular dynamics simulations³³. The large scale motions are addressed using induced-fit models or population-shift models to overcome the limitation of small-scale simulation methods. In the case of induced-fit models, the substrate is assumed to be induced to adopt conformational changes of the enzyme³⁴. In the case of population-shift

models, the enzyme adopts a specific conformations and the substrate selectively binds into the specific enzyme conformation. In order to explore the internal conformational fluctuations, now a days, modified versions of molecular dynamics (MD) simulations, initially introduced by Alder et. al^{35,36}, are widely used. The method, further provides time-dependent information of conformational changes by solving Newton's equations of motion or the Lagrange's equation to obtain the coordinates and the momenta of the atoms, resulting in time dependent trajectories. MD simulations generate microscopic level details which can be connected with statistical mechanics and allow us to obtain microscopic properties.

2.4.1 Theory of molecular dynamics simulations

The molecular dynamics simulations are based on solving Newton's equation of motion (second law),

$$F_i = m_i a_i, \quad (2.11)$$

numerically for each atom in the system. The force exerted on the particles is given by the negative gradient of the potential energy function with respect to the position of the atoms (eq 2.12).

$$F_i = -\frac{\partial U}{\partial r_i}. \quad (2.12)$$

The acceleration, positions and velocities of the particles are evaluated by integration of the equation of motion. The exerted force is determined using a predefined potential energy function (U). Generally, the potential energy function is a combination of bonded, non-bonded, angle, torsion and electrostatic energy terms, which depend on several parameters which can be obtained from either experimental studies or quantum mechanical studies. The potential energy function and the associated parameters are called a "force field", and contain independent additive terms. The potential energy function (V_{AMBER}) of the AMBER³⁷ force field is stated below.

$$\begin{aligned} V_{AMBER} = & \sum_{bonds} k_b(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] \\ & + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right] + \sum_{i < j} \left[\frac{q_i q_j}{\epsilon R_{ij}} \right] \end{aligned} \quad (2.13)$$

Where k_b , k_θ , V_n are bond lengths, bond angles and dihedral force constants. r , θ are the current bond length and the angle at specific time point. r_0 , θ_0 are the equilibrium bond lengths and angles, q_i and q_j are partial atomic charges, ϵ is the well depth, n is the dihedral periodicity, ϕ is the dihedral angle, and γ is the phase angle.

2.5 Modeling protonation states

The second dimension in the computational RNA enzymology “problem space” involves examination of protonation states. Within the context of catalysis, the specific protonation states of key residues are a critical requirement of the catalytically active state. For general acid/base catalysis, such as in RNase A and HDVr, the general base must be in the deprotonated form (able to accept a proton) whereas the general acid must be in the protonated form (able to donate a proton) for catalytic activity. In the analysis of kinetic data as a function of pH, the simplest mechanistic assumption is that the catalytic rate is directly proportional to the probability of the general acid and base being in the active state³⁸. If this assumption is true, and if other protonation events within the pH range of interest are benign with respect to affecting the rate, then general acid/base catalysts will give a classic bell-shaped profile that can be fit to a simple kinetic model where the parameters are the “apparent pK_a ” values of the general acid and base³⁹. Agreement between the kinetic “apparent pK_a ” values and macroscopic pK_a values from direct measurements (such as NMR) of the postulated general acid and base is usually considered as evidence in support of their roles in acid/base catalysis. An underlying assumption in this interpretation, however, is that the protonation states of the general acid and base are not correlated^{40,41}, i.e., the pK_a of the general acid is independent of the protonation state of the general base and vice versa. The validity of this assumption, which is system dependent, can be tested computationally in order to gain a more fundamental understanding of catalytic mechanism for a particular system⁵.

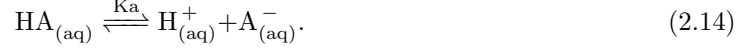
2.5.1 Constant-pH simulations

Constant-pH molecular dynamics simulations (CpHMD) and pH replica exchange molecular dynamics (pH-REMD) have emerged as powerful computational tools for deriving pH-rate profiles for general acid-base catalysts. CpHMD is a method where protonation states are sampled dynamically from a Boltzmann distribution at a fixed pH^{42,43,44,45,46}. We adopt a discrete protonation state model that employs Metropolis Monte Carlo (MC) exchange attempts between different protonation states throughout the course of the MD simulation,^{43,46} which has been recently implemented in the AMBER software suite⁴⁶ for proteins. Unlike free energy perturbation and thermodynamic integration calculations, CpHMD intrinsically takes into account the correlated effects of residue protonation states for a fixed value of pH. The use of pH-REMD allows multiple simulations to be performed over a range of discrete pH values, and is used to enhance sampling and ensure that simulations are in equilibrium with one another. The result is that complete atomic-level simulation data, including conditional probabilities for different protonation states (including tautomers), are

generated over a range of pH values. From these data, titration and pH-activity curves can be predicted and used to aid in the interpretation of experimental data.

2.5.2 Theory of Constant-pH simulations

The ionization equilibria of titratable residues is given by



where, A- is the conjugate base. K_a is the acid dissociation constant. K_a is given by

$$K_a = \frac{[A^-][H^+]}{[HA]} \quad (2.15)$$

The pK_a is defined as

$$pK_a = -\log K_a \quad (2.16)$$

The free energy change for the deprotonation is related the pK_a of the residue and the pH of the solvent by the following equation⁴⁷:

$$\Delta G = 2.303(pH - pK_a) \quad (2.17)$$

The equation 2.15 can be rearranged to the Henderson-Hasselbalch equation for non-interacting residues as follows:

$$pK_a = pH - \log \left(\frac{[A^-]}{[HA]} \right) \quad (2.18)$$

The re-written equation in the presence of the interacting residues is then⁴⁸

$$pK_a = pH - n \log \left(\frac{[A^-]}{[HA]} \right) \quad (2.19)$$

The fraction of deprotonation $f(d)$ can be defined as,

$$f(d) = \frac{[A^-]}{[HA] + [A^-]} \quad (2.20)$$

Hence,

$$f(d) = \frac{1}{1 + 10^{n(pK_a - pH)}} \quad (2.21)$$

where n is the Hill coefficient, $n > 1$ represents cooperative behavior of the titratable residues while $n < 1$ indicates anti-cooperativity^{49,41}. This equation is frequently applied in predicting titration behavior and fitting data. The Henderson-Hasselbalch equation, or variants of the equation, is used

in experimental and theoretical studies to obtain the ionization constants of a residue of interest. In computational studies, the concentrations of protonated and deprotonated state are obtained using time dependent fractions of each state, assuming ergodic behavior is preserved in the system. In this dissertation, the AMBER software³⁷ has been used to perform all simulations. The constant-pH algorithm⁴⁶ in AMBER is based on the discrete protonation states of the titratable residues, where, the protonation sites are described by a vector n_i , an integer representing the protonation state. According to equation 2.21, it is necessary to develop an equilibrium between the protonated and deprotonated states in order to calculate the pK_a values of the residues. In the discrete protonation model, the populations are sampled using the Monte Carlo (MC) method. At each MC step, a random protonation state for the titratable residue is chosen and the transition free energy is calculated according to equation 2.22 in order to accept or reject the transitions:

$$\Delta G = k_B T (pH - pK_{a,ref}) \ln 10 + \Delta G_{elec} - \Delta G_{elec,ref} \quad (2.22)$$

ΔG is the transition free energy, k_B is the Boltzmann constant, T is temperature, pH refers to the solvent pH, $pK_{a,ref}$ is the pK_a of the reference compound, G_{elec} is the electrostatic component of the free energy of the titratable group and is equal to the difference between the electrostatic potential of the current and proposed protonation states. $G_{elec,ref}$ is the electrostatic component of the transition free energy of the reference compound. The $G_{elec,ref}$ is calculated using thermodynamic integration calculations for each reference compound and adjusted to give equal populations of the protonated and deprotonated fractions when the reference pK_a equal to solution pH.

2.5.3 Constant-pH replica exchange

Sander, an AMBER module has the ability to perform constant-pH simulations with replica exchange in and explicit solvent medium⁴. Using the Monte Carlo method, the adjacent pH values of replicas are exchanged. Considering M number of non-interacting replicas of i number of replicas, ($i = 1, \dots, M$), the q_i , m_i and pH_i (coordinates, momenta and pH) at temperature T . The detailed balance conditions for exchanging replica i at pH_l and replica j at pH_m is⁵⁰ summarized below:

$$P(X_i^l)P(X_j^m)w(X_i^l, X_j^m \rightarrow X_i^m, X_j^l) = P(X_i^m)P(X_j^l)w(X_i^m, X_j^l \rightarrow X_i^l, X_j^m) \quad (2.23)$$

Where P is the equilibrium probability and w is the transition probability and the exchange between replicas is described as follows:^{51,50}

$$\begin{aligned} X_i^l &= (q_i, p_i, N_i^p, pH_l) \rightarrow X_i^m = (q_i, p_i, N_i^p, pH_m) \\ X_j^m &= (q_j, p_j, N_j^p, pH_m) \rightarrow X_j^l = (q_j, p_j, N_j^p, pH_l) \end{aligned} \quad (2.24)$$

The equilibrium probability is given by

$$P(X_i^l) = \frac{1}{\Xi} \exp[-\beta(H(q_i, p_i, N_i^p) - kT \ln(10) \cdot pH_i \cdot N_i^p)] \quad (2.25)$$

2.6 Modeling the reaction path and transition states

Modeling the reaction path is the last critical step involved in identifying the rate enhancement and details of the reaction mechanism. Both kinetics and thermodynamics provide invaluable information on this subject. In theoretical studies, the reaction path is built using a potential energy surface (PES) described by the Born-Oppenheimer approximation. The PES depends on the potential energy of the nuclear coordinates⁵². The local minima, transition states and intermediates can be identified along the PES. The minima are identified where first order derivatives are zero and the eigenvalues are positive. The transition states have one eigenvalue from the Hessian-matrix. The minimum energy path (MEP) on the PES is identified as the reaction path. The MEP connects two equilibrium states along the path. Frequently, in the case of large biomolecules, the free energies are calculated using bias sampling techniques such as umbrella sampling^{53,54,55}, in order to sample the equilibrium states along the reaction path. This technique is most often used in QM/MM simulations⁵⁶. Several ways to obtaining the minimum free energy path have been summarized in the literature⁵⁶. One way of connecting each state along the reaction path is to use the chain-of-states method. In this method, the number of states are connected on the PES to obtain the reaction path. Line integral methods, nudged elastic band (NEB), zero temperature string (ZTS), growing string methods, conjugate peak refinement (CPR), and reference path methods are most commonly used methods and belong to the category of the chain-of-states method.

2.6.1 QM/MM simulations

Electron rearrangements are common in reaction mechanisms, which often involve bond forming, bond breaking and charge transfer processes. The molecular mechanics (MM) method is not capable of describing the electronic structure of a system. Therefore, the quantum mechanics (QM) approach is required in modeling those chemical processes in order to accurately describe the electronic structure and explore the bond forming and bond breaking events^{57,58,59}. However, the slow QM approach is impossible on larger systems contains thousands of atoms, as QM methods are only capable of handling system containing a few hundred atoms⁵⁹. Therefore, a combination of QM and MM methods, called the “hybrid QM/MM” method, developed by Warshel et. al^{60,61} is a convenient and widely used method for application to large bio-molecules. In QM/MM simulations,

the chemically active region is described in the QM region which can be evaluated with *ab-initio*, DFT or semi-empirical methods. The *ab-initio* or DFT levels of theory are much more expensive and the maximum simulation length that can be achieved is on the order of 100 ps. On the other hand, the semi-empirical methods such as AM1,⁶² AM1/d,⁶³ PM3,^{64,65,66} SCC-DFTB⁶⁷ reach 100 times longer time scales. The MM force field, which solves Newton's equation of motion is sufficient for treating the rest of the system. Figure 2.2 illustrates the QM and MM region for a larger protein, Ribonuclease A. The QM region is defined as the active site of the system where the phosphoryl transfer reaction occurs.

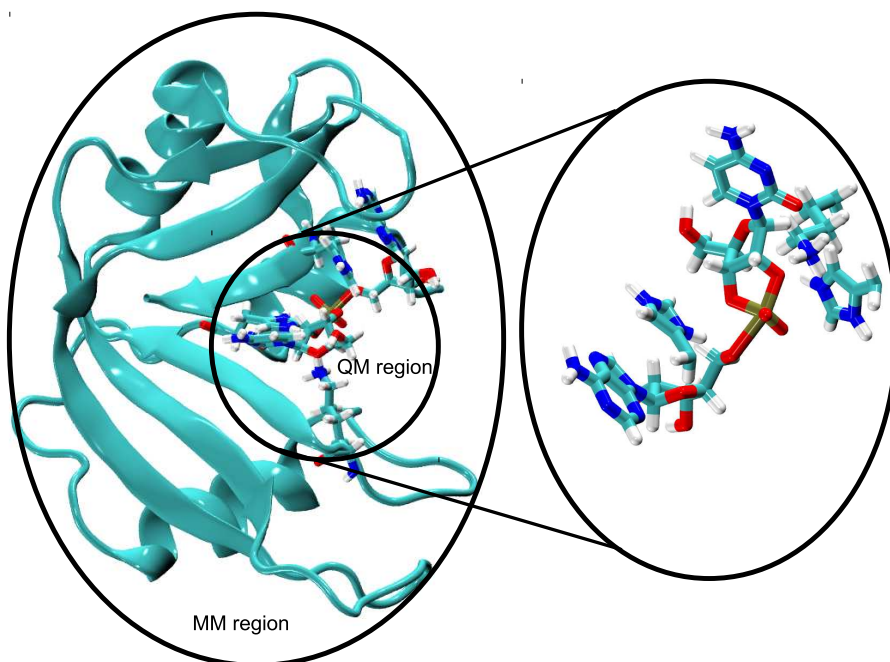


Figure 2.2: The figure highlights the QM and MM regions used in the hybrid QM/MM simulations for the Ribonuclease A system. The QM region is defined as the active site of the enzyme where the biochemical reaction occurs. This region consists of a small number of atoms and is treated at the *ab-initio* level or with semi-empirical methods. The MM region is larger in size and the MM force field is applied on this region.

2.6.2 QM/MM theory

The total energy of hybrid QM/MM simulations consists of three types of interactions: 1) The interactions between QM atoms (E_{QM}), the interactions between MM atoms (E_{MM}), and the interactions between QM and MM atoms ($E_{QM/MM}$). Therefore, the total energy of the system is given by:

$$E_{system} = E_{QM} + E_{MM} + E_{QM/MM} \quad (2.26)$$

The interactions between QM and MM regions can be described using subtractive or additive schemes^{59,68}. The total energy, E_{system} in a subtractive scheme can be derived by summing the energy of both the QM and MM regions together at the MM level ($E_{MM}(MM/QM)$), with the energy of the QM region evaluated at the QM level ($E_{QM}(QM)$), and subtracting the energy of the QM region evaluated at MM level ($E_{MM}(QM)$) as follows:

$$E_{system} = E_{MM}(MM/QM) + E_{QM}(QM) - E_{MM}(QM) \quad (2.27)$$

In the case of an additive scheme, E_{system} is evaluated by

$$E_{system} = E_{QM}(QM) + E_{MM}(MM) + E_{QM-MM}(QM + MM) \quad (2.28)$$

The coupling energy between the QM and MM regions, $E_{QM-MM}(QM+MM)$ consists of electrostatic energy ($E_{ES}(QM/MM)$), van der Waals interactions ($E_{vdw}(QM/MM)$) and MM bonded interactions ($E_{bonded}(QM/MM)$)⁶⁹. There are several ways to evaluate these coupling terms when using the additive scheme, as described below. The electrostatic interactions are modeled with different types of schemes such as mechanical embedding, electrostatic embedding and polarized embedding models⁷⁰. The van der Waals terms are expressed as Lennard-Jones potentials.

$$E_{vdw}(QM/MM) = \sum_{i \in QM} \sum_{j \in MM} 4\epsilon_{ij}[(\sigma_{ij}/R_{ij})^{12} - (\sigma_{ij}/R_{ij})^6] \quad (2.29)$$

Where R_{ij} is the distance between QM atom i and MM atom j and ϵ_{ij} and σ_{ij} are the pairwise Lennard-Jones parameters⁶⁹. Finally, the covalent bonds across the QM-MM boundary are treated with link-atom schemes, boundary-atom schemes or localized-orbital schemes.

2.6.3 Link atoms

The link atom method is commonly used in QM/MM studies^{71,72}. In this method, a new atom is simply introduced, in most cases a hydrogen atom,^{71,73,74,75} at the boundary between the QM and MM regions. There are some problems associated with this method like, the link atom can introduce a new degree of freedom to the system, the newly introduced atom can over polarize the atoms nearby the “frontier” MM atom and the link atom is electronically different from the initial structure.

2.6.4 Semiempirical AM1/d-PhoT

Hypervalent phosphates are involved in transition states and intermediates in transphosphorylation reactions and thus d orbital representation is required^{76,77,78} to model these reactions. The parameters for H, P and O were introduced into the AM1/d Hamiltonian and the new Hamiltonian is called

AM1/d-PhoT. Although other semiempirical methods are available such as MNDO,^{79,80} AM1,⁶² PM3,⁶⁴, these are unlikely to accurately model phosphoryl transfer reactions. The AM1/d-PhoT Hamiltonian has been successfully applied to ribozyme systems such as hairpin⁸¹ to predict the detailed reaction mechanisms.

2.6.5 Free energy calculations

QM/MM simulations are used to perform free energy calculations to obtain the free energy surfaces for a reaction coordinate of interest. The energetics associated with conversion of the reactant (R) species to the product (P) species can be calculated. Unlike MD simulations, the populations of each state can be accurately obtained with QM/MM simulations. The free energy difference, $\Delta G_{R \rightarrow P}$ for going from the reactant to product state is given by⁵⁷:

$$\Delta G_{R \rightarrow P} = -k_B T \ln \left[\frac{P_P}{P_R} \right]. \quad (2.30)$$

Where, k_B is the Boltzmann constant, T is the temperature, P_R is the population of the reactant state and P_P is the population of the product state. In order to sample each state in a chemical reaction, especially rare events such as transition states, biasing potentials are commonly used with a technique called “umbrella sampling”.

2.6.6 Umbrella sampling

This section gives a general overview of the umbrella sampling technique⁵⁴.

The following integral is evaluated to give the canonical partition function, Q :

$$Q = \int \exp[-\beta E(r)] d^N r \quad (2.31)$$

Where E is potential energy, N is number of degrees of freedom and $\beta = 1/k_B T$. The free energy A is defined as

$$A = -\frac{1}{\beta} \ln Q \quad (2.32)$$

Where A is the Helmholtz free energy (NVT conditions). Under NPT conditions, the free energy is called the Gibbs free energy (G). In the case of a reaction coordinate, the free energy difference between two states along the reaction coordinate (ζ) is interesting. The probability distribution is obtained by integrating all the degrees of freedom except ζ .

$$Q(\zeta) = \frac{\int \delta[\zeta(r) - \zeta] \exp[(-\beta E)] d^N r}{\int \exp[(-\beta E)] d^N r} \quad (2.33)$$

The free energy along the reaction coordinate ($A(\zeta)$) can be calculated using equations 2.35 and 2.36 and this is called the potential of mean force (PMF). In computer simulations, according to the ergodic principle, the time average ($P(\zeta)$) is equal to the ensemble average ($Q(\zeta)$). The time average can be written as

$$P(\zeta) = \frac{1}{t} \lim_{t \rightarrow +\infty} \int_0^t p[\zeta(t')] dt' \quad (2.34)$$

The free energy barriers are useful in calculating the reaction rate using transition state theory^{82,83,84}. One way of calculating the free energy along a reaction coordinate is by using biased umbrella sampling simulations. Therefore, the technique is frequently used in potential of mean force (PMF) calculations. An artificial biasing potential is added along the reaction coordinate in order to access the thermodynamic states along the path. All the states on the potential energy surface are covered by adding biasing potentials on selected range of points along the reaction coordinate and a series of simulations are carried out. The biasing potential usually is a harmonic potential energy function. The new potential function can be written as

$$E'(r) = E(r) + W(\zeta) \quad (2.35)$$

Where $E'(r)$ is the modified potential function, $E(r)$ is the initial potential and $W(\zeta)$ is the added weight in the form of a harmonic potential.

$$W_i(\zeta) = k(\zeta - \zeta^{ref})^2 \quad (2.36)$$

The biased system results in a biased distribution along the reaction coordinate over all windows. Finally the windows are combined and unbiased using the weighted histogram analysis method (WHAM) or the vFEP method in order to obtain the PMF.

2.7 Ribonuclease A

Key biological processes such as repair, replication, energy transduction, signal transduction and protein/nucleic acids synthesis in living organisms are manipulated by phosphoryl-transfer reactions^{85,86}. In particular, it is important to know how phosphodiester bonds in nucleic acids are cleaved and formed. These reactions can be catalyzed by both proteins and RNA molecules, known as ribozymes⁸⁷, both of which show great importance in biotechnology. The reaction pathway is known to be sensitive to different factors including pH and pK_a of the nucleophilic and leaving groups. Therefore, it is of fundamental importance to understand the different catalytic strategies

taken by enzymes and ribozymes to catalyze this reaction. The prototype protein enzyme that catalyzes this reaction is Ribonuclease A (RNase A). The Insight into the mechanism of Ribonucleases provides important information in the design of new drugs and inhibitors as well as molecular level understanding of the origin of mutational effects. Therefore, the study of RNase A will lead us to important future discoveries.

Ribonuclease is one large super family, consisting of a number of Ribonucleases such as RNase A, RNase 1, RNase P, etc. in different organisms⁸⁸. Although the primary sequence of the RNase super family have diverged⁸⁹, it shares the sequence homology, tertiary structure, catalytic elements and hydrolysis ability of polymeric RNA⁹⁰. Ribonucleases catalyze the cleavage of RNA in two steps; transesterification and hydrolysis, which shares similarity with RNA catalysis conducted by ribozymes including hairpin, HDV, hammerhead and VS. Bovine Pancreatic Ribonuclease A (RNase A), highly active in pancreas of the ruminants¹, is one of the most extensively studied members of the Ribonuclease family in the literature as a result of its smaller size, relative thermostability and abundance^{90,91}.

RNase A consists of 124 amino acids corresponding to a molecular mass of 13,686 Da¹. The secondary structure is made up of four long stranded anti-parallel β -sheets and three short α -helices¹. At physiological pH, the active site (P1) and the sub-sites (B1,B2,B3,P0 and P2 (Figure 2.3)) of the enzyme interacts with DNA and RNA strands. P1 is made up of Gln11, His12, Lys41, His119 and Asp121 as shown in Figure 2.3. His12 and His119 are the proposed general base and acid in the first step of the catalytic mechanism. His12 abstracts the proton from the nucleophilic O2'-oxygen by facilitating attack on the phosphorous atom. His119, the proposed general acid, protonates O5' to facilitate the cleavage of the P-O5' bond to form a 2',3'-cyclic phosphate product (Figure 3.1). The cyclic product hydrolyzes into a 3'-nucleotide by the reverse reaction with the His12 and His119 residues acting as general acid and base, respectively.

The detailed mechanism of the reaction was proposed in 1962 and identified as a general acid-base mechanism involving two histidine residues^{92,93,94,95,96}. Other alternative mechanisms were proposed by Witzel⁹⁷, Wang⁹⁸, Breslow⁹⁹ and Wladkowski^{2,100}. The Witzel mechanism involves only one histidine residue and the O2 of the pyrimidine. The Wang mechanism was similar to that proposed for chymotrypsin. Breslow proposed the existence of an intermediate and second transition state. Finally, Wladkowski proposed the involvement of Lys41 as the general base in the mechanism. Although all of these proposed mechanisms still stand, the Mathias and Rabin mechanism has been most widely accepted in the literature.

Several experimental studies has been carried out to probe the detailed reaction mechanism. The descending order of substrate specificity for the NpN' strand of 3'-pyrimidine^{97,1} and 5'-N': A >

$G > C > U$ ⁹⁷ was observed. The involvement of histidine residues in the reaction mechanism was further supported by the experimentally determined classical bell-shaped pH-rate profiles^{97,92,101}. The linear and pseudo-rotation mechanisms for phosphate ester hydrolysis were addressed from the perspective of the catalytic mechanism and the linear mechanism was confirmed. The pK_a shifts of His12 and His119 from 5.8 and 6.2^{92,102,103,104} to 6.2 and 8.10¹⁰⁵ were reported upon binding of the substrate. According to the kinetic studies, the turnover number of the reaction lies in the range of 2000-3000 s⁻¹, while the Michaelis constant is in the range of 1-5 mM⁷. The second order rate constant, k_{cat}/K_m for cleavage of UpA by RNase A is 2.3×10^6 M⁻¹s⁻¹. Therefore, the calculated dissociation constant for the reaction is 2×10^{-15} M. An approximately 10^4 fold decrement of enzyme affinity was observed upon the elimination of the imidazole of His12 and His119. Mutational effects were addressed for active site and sub-site residues in an ample number of studies (Figure 2.4)^{106,107,108,109,110,111,112,113}. One mutational effect study for Lys41 was believed to explain its role in the mechanism as a hydrogen bond donor in the transition state. Further, the crystal and NMR structures of native and inhibitor bound enzymes together with knowledge of the mechanism provides useful information about orientation and the role of the active site and sub-site residues. Molecular dynamics simulations have enabled researchers to observe key stages along the reaction coordinate and the dynamical behavior of the active site residues.

Although there have been plenty of studies conducted in the past, some debate with regard to the catalytic mechanism still exists. The proposed mechanisms by Mathias and Rabin, Witzel, Wang, Breslow and Wladkowski are still under discussion and criticism in the literature. The divergence of the pK_a values obtained for the His12 and His119 from imidazole have been questioned since large perturbations of pK_a values in the active site are not very well known⁹¹. The ambiguity regarding the role of Lys41 as the general base in catalysis is still under discussion. The details of individual proton transfer states, the protonation states of the active site residues along the reaction coordinate¹⁰⁰, and the role of some other residues around the active site¹¹⁴ are still not completely understood. In particular, the nature of the transition states and the intermediate, together with roles of active site residues are still of interest.

We answer some of these questions with the aid of multiscale computational modeling tools. First, the dynamics of the substrate bound enzyme is addressed with MD simulations and the conformational contribution of the active site residues are examined. The stability of each stage along the reaction path is also explored. Secondly, a detailed picture of the reaction mechanism is explored with QM/MM simulations. Thirdly, the effect of pH and titration curves, and pK_a s of the active site residues are validated with Constant-pH simulations. Finally, improvements are made for the constant-pH method to study ribozyme systems.

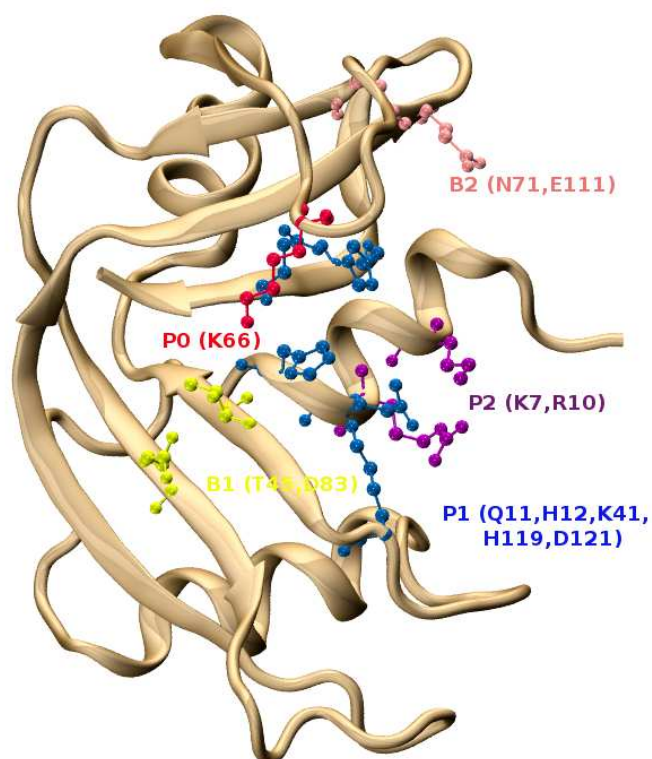


Figure 2.3: Active-site P1 and the sub-sites; B1,B2,B3,P0 and P2 of RNase A. The active site is made up of Gln11, His12, Lys41, His119 and Asp121. The P2 and P0 sub-sites are made up of Lys7, Arg10 and Lys66. The B1 and B2 sub-sites are defined by the Thr45, Asn83 residues and the Asn71, Glu111 residues, respectively¹.

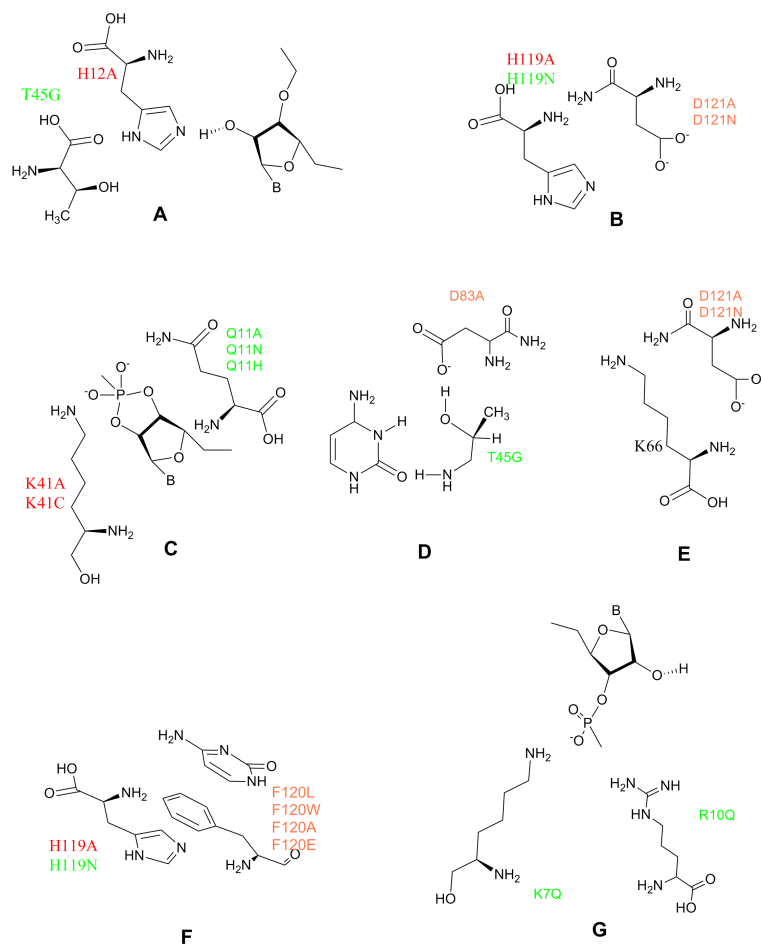


Figure 2.4: Mutational effects of key active site residues. The red color represents the strongest effect. The orange color represents a medium effect and the green color represents a weak effect on catalytic rate.

Chapter 3

Catalytic mechanism of RNase A, 1: Characterization of the structure and dynamics in solution at different stages along the reaction path

Bovine pancreatic ribonuclease A (RNase A) is an important, well-studied enzyme that catalyzes the degradation of RNA in two steps: transphosphorylation to form a 2',3'-cyclic phosphate, followed by hydrolysis to form a 3'-nucleotide. Despite the wealth of structural, biochemical and biophysical studies of this enzyme, many questions remain with regard to a detailed atomistic picture of mechanism. Computational models can aid in providing a unified interpretation of a broad range of experiments, as well as make experimentally testable predictions, that allow a consensus view of mechanisms to emerge. Toward that end, we use molecular dynamics and free energy simulations to study the structure and dynamics of RNase A in a crystalline environment and in solution at different stages along its catalytic reaction path. In this way, we identify a key conformational transition, not apparent from the crystallographic structure, involving a side chain flip of His12 into an active conformation that positions it within hydrogen bonding distance of the 2'OH nucleophile, and in a capacity that would allow it to act as a general base in transphosphorylation. Further, we are able to study the response of active site residues due to progression of the reaction coordinate, and identify catalytic protonation states of key active site residues. Results suggest that Lys41 is unlikely to play the role of the general base in transphosphorylation, but rather acts to increase the acidity of the 2'OH nucleophile in order to facilitate proton transfer in general base catalysis, and provides direct electrostatic stabilization of the anionic phosphorane transition state. This transition state model has been used to compute kinetic isotope effects that are in close agreement with experiment. His119 is positioned to act as the general acid in transphosphorylation to form the 2',3'-cyclic phosphate product, where it remains poised to abstract a proton from a nearby water molecule and thus play the reverse role of general base in the hydrolysis step. Analogously, His12, in fully protonated form, is poised to act as the general acid in the hydrolysis step. These simulations characterize the relevant conformational events and protonation states that lead to a catalytically active state, and thus provide new insights that establish a departure point and lay the groundwork for further investigation of the chemical mechanism using combined quantum mechanical/molecular

mechanical methods.

3.1 Introduction

Bovine pancreatic ribonuclease A (RNase A) is the best-characterized member of the family of pancreatic ribonucleases, and serves as an important model system for protein structure, folding and catalysis¹. RNase A is a small (124 residue, 13.7 kDa) enzyme that catalyzes the cleavage of single-stranded RNA via two steps: transphosphorylation to form a 2',3'-cyclic phosphodiester, followed by hydrolysis to form a 3'-nucleotide^{115,1} (Figure 3.1). Owing to its stability and ease of purification, combined with its impressive rate enhancement (over 11 orders of magnitude), RNase A has been the focus of enzymatic investigations starting from the first pioneering studies in the early 1960s^{92,93,94,95,96}.

There has been a major effort to elucidate the detailed reaction mechanism of RNase A, since becoming the first enzyme to be crystallized.^{116,117} The interpretation of subsequent experimental data led to several mechanistic proposals^{118,119,120,121,122}, from which the specific roles of key active site residues were implicated,^{123,121} and transition states were identified¹²². Witzel proposed His12 and the O2 of the pyrimidine nucleobase act as key players in the catalytic mechanism¹¹⁸. Wladkowski's studies suggest that Lys41 is involved in an intimate role, most likely acting as the general base^{100,2,121}. Usher's mechanism involves a possible pseudorotation of a pentacoordinate intermediate¹¹⁹, whereas Breslow's mechanism suggests a phosphorane-like intermediate¹²⁰. The most recent experimental study of kinetic isotope effects (KIE) concludes an involvement of negatively charged late-like transition state in the RNase A catalytic mechanism¹²². Among those proposed mechanisms, the ones, that have received the greatest attention, are the classical general acid-base mechanism^{92,93,94,95,96} and the triester-like mechanism^{120,124}. Several different variations of the above mechanisms and alternatives have, also, been postulated based on the experimental and theoretical studies^{122,125}. However, all of these studies have left behind an enormous amount of unanswered questions and debates concerning the detailed mechanism, restricting the goal of resolving a universally accepted mechanism. Largely, the questions have been focused on the roles, orientations and conformations of the active site residues and the nature of the transition states.

The first step towards understanding the mechanism is identification of the active site residues and their orientations which provide information on the proper active state for the reaction to proceed. Experimental studies such as pK_a measurements, kinetic and mutational effect studies bring a wealth of information in the identification of key residues. The crystal structures grant a fair explanation of orientations of the enzyme key residues and its ligands, useful in predicting mechanistic

pathways. Nevertheless, the crystal structure itself only partially describes the mechanistic details of enzyme catalysis¹²⁶: the crystallization conditions such as buffer pH, ionic strength, temperature and solvent medium aren't fully consistent with the biological conditions which facilitate the reaction. Inhibitor bound crystal structures represent the orientation of an active site suitable for the inhibitory action by the enzyme, misleading a detailed mechanistic picture. The conformational differences of the active site residues, governed by the crystal and solution medium, significantly influence the enzyme functionality. Further, the crystal structure is incapable of offering a complete picture of the protonation states of the active site residues that are important for proton transfer enzymatic reactions. The crystal structure of the inactive state, d[CpA] bound RNase A (PDB ID: 1RPG), used in this study, solved to 1.4 Å resolution at pH 5.2 in 20 mM phosphate, 20 mM acetate equilibrated with 50% of MPD ((4S)-2-Methyl-2,4-pentanediol), contradicts the physiological conditions. The protonation states of the key active site residues, His12 and His119, appear to be fully protonated in the crystal environment; His12 is hydrogen bonded to the Thr45 backbone and one of the non-bridging phosphoryl oxygens, His119 interacts with both O5' of the substrate and the Asp121 residue. Lys41 exists in two side-chain conformations and interrupt the inactive substrate stability. Therefore, a close examination of conformations, protonation states and orientations of the active site in the crystal environment is necessary before making direct conclusions of the mechanistic details from the crystal.

In this study, the information from the crystal structure of the inactive substrate (d[CpA]) bound RNase A has been used as the departure point to identify the detailed mechanism using novel computational tools. The theoretical methods, and in particular molecular simulations, are powerful tools to aid in the interpretation of mechanistic experiments by providing detailed dynamical information at the atomic level. The chemical catalytic steps of an enzyme reaction are often studied using fully quantum mechanical methods such as those based on density-functional or wave-function theory, and can be extremely computationally intensive such that it is only practical to apply them to highly truncated models of the active site, and exploration of "adiabatic" potential energy pathways that neglect conformational fluctuations and other important dynamical effects. Alternatively, combined quantum mechanical/molecular mechanical (QM/MM) simulations using fast, approximate quantum models can be used to more extensively explore the relevant free energy landscape for the reactions that can be more directly compared with experiments. In both cases, one must depart from a presumed catalytically relevant "active state", and ascertain the probability of this state being observed by characterizing the structure and dynamics in solution. Rarely can this information be deduced directly from X-ray crystallographic data alone, since crystal structures provide only a static picture of a deactivated enzyme state that is not in solution.

The purpose of the present work is to develop a model for the structure and dynamics of the active state of RNase A in solution that is consistent with available experiments, and that provides a quantitative departure point for further QM/MM investigation into the chemical steps of catalysis. In doing so, we perform simulations also of the crystal structure¹²⁷ itself in order to validate our methods and models, and better attribute predicted structural relaxation in going from the crystal into solution from possible artifacts of the computational models. The results of the simulations convincingly point out that the crystal structure is not in a catalytically relevant conformation or protonation state, and the enzyme must undergo key transitions in solution in order to be active. The active “reactant” state, along with the transition and product states further along the reaction coordinate, are characterized in terms of structure, protonation state of key residues, hydrogen bonding, and inline fitness, results are compared with available experiments, and their implications into catalytic mechanism are discussed.

3.2 Methods

Simulations were initiated from the crystallographic structure of RNase A, complexed with deoxycytidyl-3'-5'-deoxyadenosine [d(CpA)] (PDB ID:1RPG) and solved at 1.4 Å resolution¹²⁷. Doubly occupied atoms of the crystal amino acid residues that are lower in occupancy were removed in order to preserve the standard residues. The missing hydrogens of the inactive and active substrates, and the hydroxyl groups (OH) at C2' of the active substrate ribose were added using the LEaP module of AMBER14³⁷. A crystal simulation of the inactive substrate bound enzyme, and solution simulations of the active state and each stage along the reaction path were carried out with different combinations of protonation states and conformations of the His12 residue as listed in Table 3.1. A detailed description of each system (Table 3.1) can be found in the Supporting Information. Partial charges for the 2-methyl-2,4-pentadiol molecule (MPD) and the other non-standard substrate residues - the activated precursor (AP), late transition state mimic (TS) and the 2',3'-cyclic phosphate product (P) - were derived using the RESP fitting procedure in the RED program¹²⁸, and the bond and angle parameters were generated using high level *ab initio* quantum calculations. The crystal simulation system (Xtal) was prepared by adding 14 chloride ions, 485 TIP4P-Ew¹²⁹ water molecules with the associated monovalent parameters of Joung and Cheatham¹³⁰, and MPD to accurately reproduce the experimental crystal mother liquor conditions. One crystal unit cell, space group P1211, containing two RNase A monomers was extended to infinity via periodic boundary conditions. Production runs were performed in the NVT ensemble for 100 ns after initial energy minimization and system equilibration over a period of 12 ns during which heavy atom positional restraints were gradually

released (extended details can be found in the Supporting Information). The systems for the solution simulations were prepared by adding ~ 10958 of TIP4P-Ew¹²⁹ water molecules in a cubic box along with 140 mM NaCl (27 Na^+ , Cl^- ions) and neutralizing with extra Cl^- ions. The temperature and pressure conditions, 300 K and 1 atm, were regulated using the Langevin thermostat and Berendsen barostat, respectively, as implemented in AMBER14³⁷. Bonds to hydrogens were constrained using the SHAKE algorithm¹³¹. Periodic boundary conditions were applied and long range electrostatics were computed using the smooth particle-mesh Ewald method^{132,133} with a 9 Å cut off for non-bonded interactions. The systems were equilibrated for 10 ns, and production simulations were performed for 400 ns, with a 2 fs time steps.

3.2.1 Determination of the “active state”

The dynamics, protonation states, conformations and orientations of the active site residues and the substrate of the reactant state provide solid information in identification of a reaction path. Especially, in enzymatic acid-base proton transfer mechanisms, the in-line attack angle ($\text{O2}'\text{-P-O5}'$ angle), orientation of hydrogens, distance between the donor hydrogen and the acceptor (DH-A distance) and donor-hydrogen-acceptor angle (D-H-A angle) play a critical role in determination of a favorable reactant state, called the “active state”, which provides essential information of a detailed mechanism. The criteria for establishing the “active state” is represented by in-line attack angle (θ_{Inl}) $\geq 125^\circ$, the hydrogen-acceptor distance (d_{DH--A}) ≤ 2.4 Å and the hydrogen bond angle (θ_{D-H-A}) $\geq 120^\circ$, where D and A are the donor and acceptor atoms, respectively. The nucleophilic $\text{O2}'$ of the substrate is considered the donor atom at all times. The acceptor atom is either $\text{N}\zeta$ of Lys41 for the reactant states characterized by neutral Lys41, or $\text{N}\epsilon$ or δ for the reactant states characterized by neutral His12.

The percent occupancies and the average values of each property were calculated by obtaining the number of satisfactory steps and the satisfactory values of each criteria along the trajectories. The percentile values of co-occupancy of two or three selected criteria were calculated by extracting the number of steps that occupy all the satisfactory values simultaneously. The percent co-occupancy of all three properties $\geq 80\%$ is chosen as the cut off to determine the best candidate for the “active state”. The key average distances between the active site residues and the substrate were obtained in order to gain structural information about the reactant state.

Table 3.1: Summary of the simulated systems, including protonation states of the active site residues and conformations of His12

Model	Conf ^a	H12	H119	K41
Xtal	NF	+	+	+
R(H12 ^δ)	NF	δ	+	+
R ^F (H12 ^δ)	F	δ	+	+
R(H12 ^ε)	NF	ε	+	+
R(H12 ^δ , K41 ^ζ)	NF	δ	+	0 _ζ
R(K41 ^ζ)	NF	+	+	0 _ζ
R(H12 ^ε , K41 ^ζ)	NF	ε	+	0 _ζ
R	NF	+	+	+
AP	NF	+	+	+
AP ^F	F	+	+	+
TS	NF	+	+	+
TS ^F	F	+	+	+
P(H119 ^δ)	NF	+	δ	+
P ^F (H119 ^δ)	F	+	δ	+
P(H119 ^ε)	NF	+	ε	+
P ^F (H119 ^ε)	F	+	ε	+

Abbreviations: Xtal, Crystal structure; R, Reactant state; AP, Activated Precursor state; TS, Transition state; P, Product state. The protonation states of the active site residues: ‘+’, fully protonated; δ or ε, neutral and δ or ε site protonated; 0_ζ, neutral and ζ site deprotonated. His12 conformations: F, imidazole side chain is flipped; NF, imidazole side chain is non-flipped.

3.3 Results and Discussion

In this study we apply molecular dynamics simulations to explore the orientations and the protonation states of the active site residues in the crystal environment for direct comparison with the crystal structure, and the dynamical behavior of the active site residues and the substrate of each stages along the reaction coordinate of RNase A transphosphorylation.

The crystal structure of the deactivated RNase A is not catalytically relevant

The inactive substrate bound enzyme (Xtal) was simulated to explore the conformations and the protonation states of the active site residues in the crystal environment by introducing a fully protonated active site (protonated His12, Lys41 and His119) to the system. Then the key distances were directly compared with the crystal structure in order to confirm the protonation states to understand the relevance of the representation of inactive state bound enzyme to the active substrate bound state.

The monomer’s root mean square deviation (RMSD) compared to the experimental model converged quickly (Figure 3.14). The RMSD of the average simulation structure over 100 ns of simulation was 0.34 Å for all backbone atoms, 0.70 Å for all heavy atoms, 1.12 Å for all heavy atoms of the active site. Mean fluctuations, as represented by the B-factors ($B = 8/3 * \pi^2 * msf$, where msf is the mean squared fluctuation of an atom) are in good agreement with experiment: the Pearson correlation coefficient is 0.67 and the RMSD of the B-factors is 3.94 Å² (Figure 3.14). The overall

structural integrity was preserved in the simulation, thus supporting the starting crystal structure.

The His12 residue is stabilized in the crystal structure by the key interactions with the non-bridging phosphoryl oxygens and the backbone of the Thr45 having H12:NE2-OP1 and H12:ND1-T45:O distances 2.76 Å and 2.70 Å, respectively. A Fully protonated His12 in the crystal simulation also maintained the integrity of the active site by having the interactions of 3.11, 2.87 Å and 2.82, 2.82 Å for H12:NE2-OP1 and H12:ND1-T45:O, in first and second asymmetric unit cells (Table 3.8), respectively. Therefore, the results confirms a presence of a fully protonated His12 in the inactive state in a crystal environment. Further, the Lys41 residue possesses a higher flexibility in the crystal and occupies two conformations with a C2'–NZ distance of 4.14 or 9.31 Å. This trend is preserved in the crystal simulation for the fully protonated Lys41, with a C2'–NZ distance of 4.67 and 7.10 Å in first and second asymmetric unit cells. (Figure 3.10). Although, the second unit cell yields distance of 8.95 Å, the overall results support the presence of a fully protonated Lys41 in a crystal environment. Finally, His119 is found to be in the active 'A' conformation in the crystal structure interacting with the O5' of the inactive substrate and Asp121, with H119:ND1– O5' and D121:OD– H119:NE2 distances 2.66 Å and 2.62 Å. In both crystal unit cells, similar stabilization was observed with a fully protonated His119 having 3.13, 2.92 Å and 3.86, 4.52 Å of H119:ND1–DA3:126:O5' and H119:NE2–D121:OD1 distances, respectively. Therefore, the simulation results suggests His12, Lys41 and His119 are in their fully protonated states in the inactive crystal structure. A fully protonated His12 residue is less likely to act as a general base. Further, the conformations of Lys41 induced by the inactive substrate are less likely to exist in the active state. Therefore, the protonation states and conformations of the active site residues doesn't necessarily represent a catalytically active reactant state.

Table 3.2: Average key distances of the active site in the inactive state crystal simulation

Atomic pair	1RPG (Å)	Xtal (unit1) (Å)	Xtal (unit2) (Å)
H12:NE2 – DA126:OP1	2.76	3.11 (0.29)	2.87 (0.22)
H12:NE2 – DA126:OP2	4.31	3.04 (0.34)	3.83 (0.57)
H12:ND1 – T45:O	2.70	2.82 (0.11)	2.82 (0.11)
K41:NZ – DC125:C2' ^a	4.14	4.67 (0.28)	-
	9.31	7.10 (0.67)	8.95 (1.77)
H119:NE2 – DA126:O5'	4.85	5.20 (0.31)	5.01 (0.22)
H119:ND1 – DA126:O5'	2.66	3.13 (0.31)	2.92 (0.22)

Comparison of the average key active site contacts in asymmetric unit cells of the crystal simulations to the crystal structure. The standard deviations are listed in the parentheses. ^a the two conformations of Lys41 side chain in the crystal structure is represented by the K41:Nζ - DC125:C2' of deoxyribose.

His12, with its side chain flipped is positioned to act as the general base in transphosphorylation

The reactant state simulations were performed to elucidate the debate of general base between Lys41 and His12, and to identify the best reactant state for RNase A transphosphorylation. Accordingly, a comparison of two sets of data for the reactant states, represented by either neutral Lys41 or His12, is presented in this section.

Three possible reactant states include a neutral Lys41: $R(H12^\delta, K41^\zeta)$, $R(K41^\zeta)$ and $R(H12^\epsilon, K41^\zeta)$. The occupancies of each criteria for all of these reactant states except the θ_{Inl} of the $K41^\zeta$ state, yield less than 80% as shown in the Figure 3.6 and Table 3.3. Further analysis of co-occupancy for the θ_{Inl} and d_{DH--A} ($O2'H-NZ$) shows a few data points in the region of interest, suggesting a low probability for Lys41 to act as the general base (Figure 3.5 -left panel). In all the simulations, the side chain of neutral Lys41 has lost both its stability and the interaction with $O2'H$, and drifted away from $O2'$, so that less likely to accept the proton. Consequently, the substrate stability has disrupted causing fluctuations of the ribose conformations. The comparison of the sugar pucker conformations shows frequent transitions between $C3'$ -endo conformation to $C2'$ -endo conformations (Figure 3.8, Figure 3.9). Therefore, the neutral state of Lys41 is unstable and unlikely to act as the general base.

The role of His12 in the reactant state was explored with three simulations include a neutral His12: $R(H12^\delta)$, $R(H12^\epsilon)$ and $R^F(H12^\delta)$. The side chain conformation of His12 in the $R(H12^\delta)$ and $R(H12^\epsilon)$ states was derived from the crystal coordinates and was a non-flipped conformation, whereas the flipped conformation in $R^F(H12^\delta)$ state. Among all the reactant state models, $R(H12^\delta)$ is the best guess resembling the coordinates from the crystal structure, as the vacant $N\epsilon$ of the His12 locates in the closest vicinity of the $O2'H$. However, the results shows 4.00 Å of average distance between $N\epsilon$ site and $O2'H$, thus resulting 0% of the occupancy of $d_{DH--A} \leq 2.4$ Å. Although, the $\theta_{inl} \geq 125^\circ$ for 94% of the time, the co-occupancy of the d_{DH--A} vs. θ_{inl} is 0% as shown in the right panel of the Figure 3.5. The $R^F(H12^\delta)$ state, on the other hand, occupies 90.5% of the in the active region of the d_{DH--A} vs. θ_{inl} plot having 2.0 Å of $N\epsilon-OH2'$ average distance and the $\theta_{D-H-A} \geq 120^\circ$ also occupied for 98.4% of the time. Further, the $N\delta$ proton of His119 interact with $O5'$ of the substrate having lowest average distance, 2.12 Å, in comparison to the other states. The Lys41 also locates 3.24 Å away from the $O2'$ with fixed side-chain conformation, whereas the fluctuating side-chain of the other reactant models. Therefore, the orientation of the active site residues suggests $R^F(H12^\delta)$ is the best candidate for the active state.

The key structural features are further investigated to verify the substrate stability of the $R^F(H12^\delta)$ state. Especially, the stabilization by the Asp121, Thr45 and Asp83 residues and the structure of the substrate are addressed in detail. Asp121 is known to form a catalytic dyad with His119 to increase substrate stability, although there is no substantial effect on the catalysis¹¹⁴. The crystal structure contains a preserved dyad with 2.62 Å of H119:NE2-D121:OD1 distance. Among

all the reactant state models, the $R^F(H12^\delta)$ state exhibits a most stable dyad having 2.83 Å of the H119:NE2–D121:OD1 distance, resembling the crystal structure (Table 3.4). Asp83 and Thr45 residues also found to be important in stabilizing a pyrimidine ring of the substrate. Pyrimidine ring of cytidine is stabilized by the Thr45 residue in the $R^F(H12^\delta)$ state, but the stabilization by Asp83 residue is not suggested in the simulation. The glycosidic angle of both cytidine and adenosine closely represents the crystal conformation in the $R^F(H12^\delta)$ state. The sugar pucker of cytidine ribose strictly stays in C3'-endo conformation, suggesting an optimal A-form¹³⁴ as opposed to the C2'-endo conformation of the inactive state. The overall stability of the substrate supports $R^F(H12^\delta)$ state as the most reasonable reactant state.

Table 3.3: The percent occupancy for the in-line attack angle (θ_{Inl}), occupancies of DH–A distances (d_{DH--A}), D–H–A angles (θ_{D-H-A}) and co-occupancies of d_{DH--A} vs. θ_{Inl} for the reactant state models.

Reactant state	Donor	Acceptor	Occ. d_{DH--A} (Å)	Occ. θ_{D-H-A}	Occ. θ_{Inl}	Co-occ.
R(H12 $^\delta$)	O2'	H12:NE2	0.2	19.9	94.6	0.2
$R^F(H12^\delta)$	O2'	H12:NE2	93.7	98.4	96.3	93.6
R(H12 $^\epsilon$)	O2'	H12:ND1	0.0	26.0	31.2	0.0
R(H12 $^\delta$, K41 $^\zeta$)	O2'	H12:NZ	0.0	10.4	1.5	0.0
R(K41 $^\zeta$)	O2'	K41:NZ	0.13	1.0	81.1	0.1
R(H12 $^\epsilon$,K41 $^\zeta$)	O2'	K41:NZ	0.8	44.9	52.1	0.8

Abbreviations: Occ; the satisfactory occupancies of each properties discussed in the main text. Co-occ; the conditional occupancy of all three properties. The donor atom (D) is defined as O2' for all cases.

Lys41 facilitates the nucleophilic activation and stabilize the transition state through hydrogen bonding; His119 is positioned to act as the general acid in the transition state

The activated precursor state forms after extraction of O2'H by His12 side chain, leaving a negative charge on O2'. Subsequently, the transition state establishes upon a nucleophilic attack on phosphoryl group by O2', resulting a penta-coordinated structure. Stability of both of the states is maintained by the active site interactions, originated from fully protonated key residues. The stability of each states, AP^F and TS^F, are reflected in the simulations, having active site and substrate RMSD of 0.78 (0.14) , 1.88 (0.14) Å and 0.67 (0.10), 1.09 (0.12) Å, respectively, with respect to the crystal structure.

In AP^F state, both N ϵ of His12 and N ζ of Lys41 are interacting with the negatively charged O2'. The interactions are maintained in the simulation having H12:NE–CS125:O2' and K41:NZ–CS125:O2' distance of 2.79 and 2.80 Å, respectively (Table 3.5). The protonated His119 also locates near O5' of the substrate having H119:HD–A126:O5' distance of 2.50 Å, stabilizing and positioning the substrate for nucleophilic attack. In the substrate, O2' locates at closer proximity to the phosphoryl oxygens as a result of tight holding of the negatively charged O2' by positively charged

Lys41. Consequently, sugar pucker of cytosine ribose also strictly stay at C3'-endo conformation, structurally favoring the nucleophilic attack. Transition state mimic simulation also shows a well preserved integrity of the active site. Comparison of the key active site contacts to the substrate with the uridine vanadate complexed RNase A (UV-RNase A) transition state mimic structure¹³⁵ indicates fair agreement as listed in the Table 3.6. Especially, the distance between K41:NZ and O2' is in excellent agreement having 2.82 Å distance in comparison to the 2.73 Å distance observed in the crystal structure (Table 3.6). The distance indicates a stabilization of the transition state by Lys41 residue. Both O2' and O5' take apical position of the penta-coordinate structure and interacts with His12 and His119 residues, respectively. His12 shows His119 maintains 1.76 Å of hydrogen bond to O5'. The overall results conclude, Lys41 stabilizes the accumulating negative charge and His119 position to donate a proton to the O5' to facilitate the P-O5' bond cleavage.

His119 acts as the general base in hydrolysis of 2',3'-cyclic phosphate

Simulations of 2',3'-cyclic phosphate product state bound RNase A mimics were performed to identify the protonation states and the conformations of the participating active site residues in the hydrolysis step. Especially, the favorability of, having different conformations of protonated His12 and occupying a proton on either N δ or N ϵ sites of neutral His119, was explored. The simulations show protonated His12 flipping back into original conformation along the course of the simulation. Therefore, the results of His12 non-flipped states, P(H119 $^{\epsilon}$) and P(H119 $^{\delta}$), simulations are discussed in detail.

The key contacts of the active site residues derived from the simulations were compared with the uridine-vanadate transition state mimic bound RNase A¹³⁵(Table 3.7). A well preserved Asp121-His119 catalytic dyad has been observed in the crystal structure having H119:NE2-D121:OD1 and H119:NE2-D121:OD2 distances of 2.65 Å and 4.73 Å, respectively, resulting 4.73 Å of largest separation between His119 and Aps121. The P(H119 $^{\delta}$) state indicates 9.71 Å of the largest separation, while P(H119 $^{\epsilon}$) state shows a 5.56 Å. Each distances are occupied <3.5 Å for \sim 20-27% times in P(H119 $^{\epsilon}$) state. The transphosphorylation step also favor the P(H119 $^{\epsilon}$), having His119 N δ site as the proton donor for the leaving group. Further, in this state, 5.68 Å of distance between phosphoryl group and the vacant N δ site allows to accommodate a water molecule for hydrolysis to occur. Extended analysis of the radial distribution functions indicates a higher water accessibility in P(H119 $^{\epsilon}$) in comparison to the P(H119 $^{\delta}$) state (Figure 3.15).

The protonated His12 in the P(H119 $^{\epsilon}$) state interacts with the non-bridging phosphoryl oxygens of the substrate with 3.32 Å and 4.47 Å of average distances of the N ϵ of His12 to the O1P and O2P oxygens, respectively. The distance between Thr45 and His12 is also maintained having 2.81 Å distance of H12:ND1-T45:O, agreeing with 2.68 Å distance reflected in the UV-RNase A transition

state mimic. The distance to the O2' from N ϵ site is 3.46 Å, suggesting His12 likely to be the general acid. The overall results suggest, N δ site of the P(H119 $^{\epsilon}$) state positioned to abstract a proton from a water molecule to act as the general base and protonated His12 likely to act as the general acid.

3.4 Conclusions

Our goal in this work is to apply molecular dynamic simulations, a multi-scale computational method, to explore the dynamical behavior of the “active state” of RNase A catalysis the stages along the reaction path. We observe that a proper “active state” for the RNase A catalysis includes fully protonated Lys41, His119, and neutral His12 with flipped side chain, where the vacant N ϵ site is positioned to take the proton from O2'H. The results further confirmed the role of Lys41 is, to stabilize the accumulating negative charge on the activated precursor and the transition states, but not to act as the general base. Finally, N δ site of the neutral His119 involve in hydrolysis step to abstract a proton from a water molecule while His12 acts as the general acid. This study opens up a novel way of applying multi-scale computational tools to identify mechanistic detail in the enzyme catalysis.

3.5 Supporting Information

3.5.1 Extended details of the simulated systems

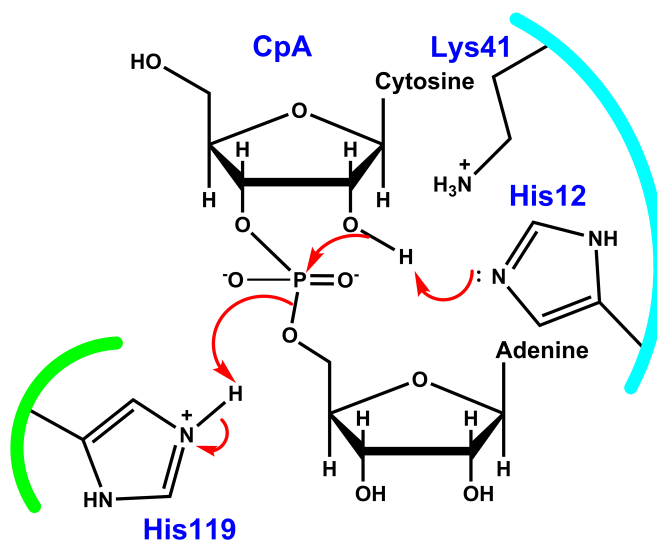
Table 3.1 in the main text summarizes all 16 systems simulated in this study, belonging to six different categories: crystal (Xtal), reactant state (R), activated precursor state (AP), late transition state (TS), and product state (P). In the crystal simulation (Xtal), the active site of the enzyme is occupied with the inactive substrate (d(CpA)), and fully protonated histidine residues, His12, His48, His105, and His119, as well as Lys41. The reactant state (R) contains a protonated O2' oxygen, a deprotonated general base (His12 or Lys41), and a fully protonated general acid (His119). Seven different reactant states models were simulated to explore the effect of protonation states and the active site residue conformations. The information of each reaction state can be found in the Table 3.1 of the main text. The activated precursor (AP) state bares a negative charge on O2' after donation of its hydrogen to the general base. The general base and the general acid are fully protonated at this stage. Only two different AP states, which differ in their conformations of the His12 side chain, were investigated in this study. The transition state mimic (TS), pentacoordinated and negatively charged, also contains a fully protonated active site. Two different simulations were performed for this state with flipped and non-flipped conformations of His12. Finally, four possible

2',3'-cyclic phosphate product state models, having fully protonated His12 with flipped or non-flipped conformations and neutral His119 in both HIE or HID protonation states, were simulated.

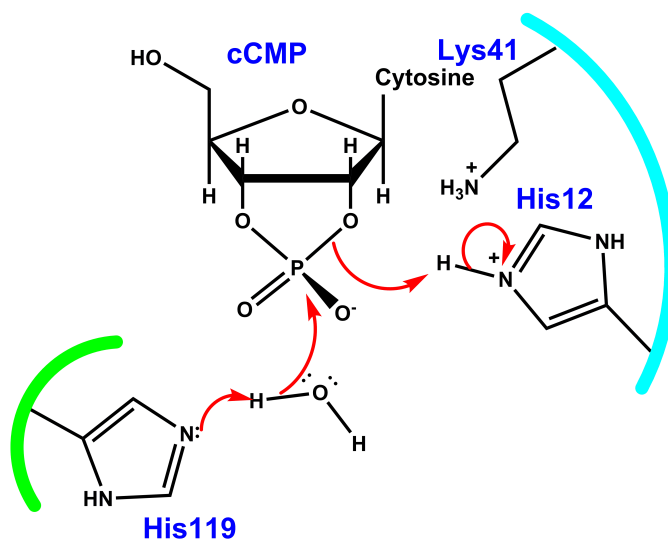
3.5.2 Preparation of the simulation supercell

Atomic coordinates were taken from the crystal structure of RNase A complexed with inactive DNA, d[CpA] (PDB ID: 1RPG)¹²⁷. The protein formed crystals in the $P2_1$ space group, with two asymmetric units per monoclinic unit cell of dimensions $a=30.00$ Å, $b=38.270$ Å, $c=53.170$ Å, $\alpha=90.00^\circ$, $\beta=106.00^\circ$, and $\gamma=90.00^\circ$. One asymmetric unit consisted of a RNase A molecule, a d(CpA) ligand, one MPD molecule, and 164 crystallographic water oxygens. A “supercell” of $1 \times 1 \times 1$ unit cells was created by using the PropPDB module of the AmberTools14³⁷ package, measuring 30.00 Å \times 38.270 Å \times 53.170 Å and comprising 2 copies of the asymmetric unit.

In previous all-atom crystal simulations^{136,137,138}, solvent that was unaccounted for in the X-ray data was added to the simulation supercell until the experimental volume of the crystal was accurately reproduced by molecular dynamics simulations at the temperature and pressure of the crystal growth conditions. The same procedure was followed in this study. Briefly, trial equilibrations and 10 ns of production runs were carried out in the NPT ensemble. All protonated crystallographic atoms were simulated. Additional solvent not derived from the diffraction density was added using the following scheme: first 14 Cl⁻ ions were added to neutralize the structure. Next water molecules and Na⁺ and Cl⁻ ions were added in proportions mimicking the ionic composition of the crystal liquor (20 mM sodium phosphate, 20 mM sodium acetate). The decision to use Na⁺ and Cl⁻ ions instead of the phosphate and acetate that was used for crystallization to replicate the ionic strength was made based on previous experience. The average volume of the crystal unit cell in the NPT ensemble for 50 ns showed 99.91% of the experimental value confirming that the correct amount of solvent had been added. More details on crystal simulation methodology used here can be found in the work published by Janowski *et al.*¹³⁹.



Transphosphorylation



Hydrolysis

Figure 3.1: The general interpretation of transphosphorylation and hydrolysis steps in RNase A catalytic mechanism. In the transphosphorylation (left), His12 acts as the general base to take the proton from O2' to promote a nucleophilic attack on the phosphate group, and His119 acts as the general acid to donate a proton to O5' to facilitate the cleavage of P-O5' bond. In the hydrolysis mechanism (right), neutral His119 acts as the general base to take the proton from a water molecule, promoting a nucleophilic attack on the phosphate group, and protonated His12 donates a proton back to O2' to facilitate the cleavage of P-O2' bond.

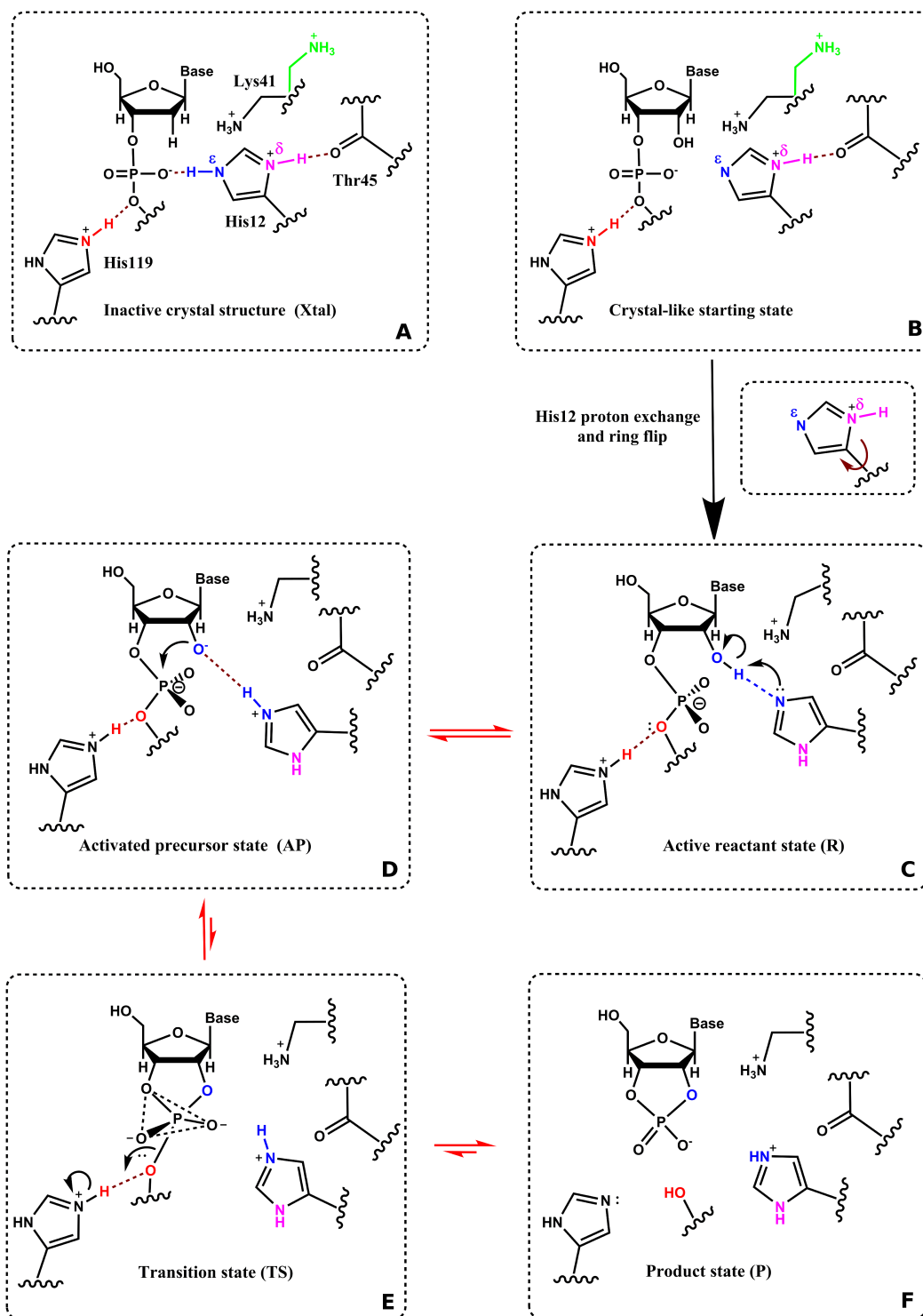


Figure 3.2: The schematic of the active site orientations: A- The crystal structure (with added protons). B- The reactant state with His12 non-flipped side chain ($R(H12^\delta)$). C- The reactant state with His12 flipped side chain ($R^F(H12^\delta)$). D- The activated precursor state with His12 flipped sidechain (AP^F). E- The late-transition state mimic with His12 flipped conformation (TS^F), F- 2', 3'- cyclic phosphate product state (P). Hydrogen bonds are indicated in dotted lines.

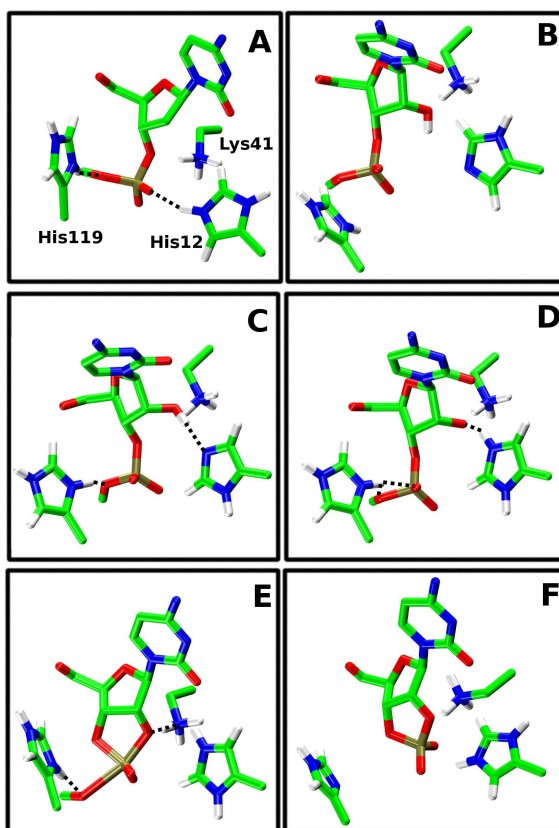


Figure 3.3: The average structures of active sites of each stage along the reaction path for, A- crystal structure, B- The reactant state model initiated from the coordinates of the crystal structure ($R(H12^\delta)$), C- “Active state” model, where His12 side chain is flipped and deprotonated at $N\epsilon$ site positioning to abstract the proton from $O2'$ ($R^F(H12^\delta)$). His119 and Lys41 are in fully protonated state, D- Activated precursor state, His12 has accepted the proton from $O2'$ (AP^F). The accumulating negative charge is stabilized by protonated Lys41. His119 is positioned towards the $O5'$ of the leaving group, E- Late transition state, His12, His119 and Lys41 are protonated in the active site. His119 is position to donate the proton to $O5'$ (TS^F), F- Product state where both His12 and Lys41 are fully protonated (P). His119 is deprotonated after the departure of the leaving group. The $N\delta$ site of His119 is likely to accept a proton from a water molecule in order to support the hydrolysis.

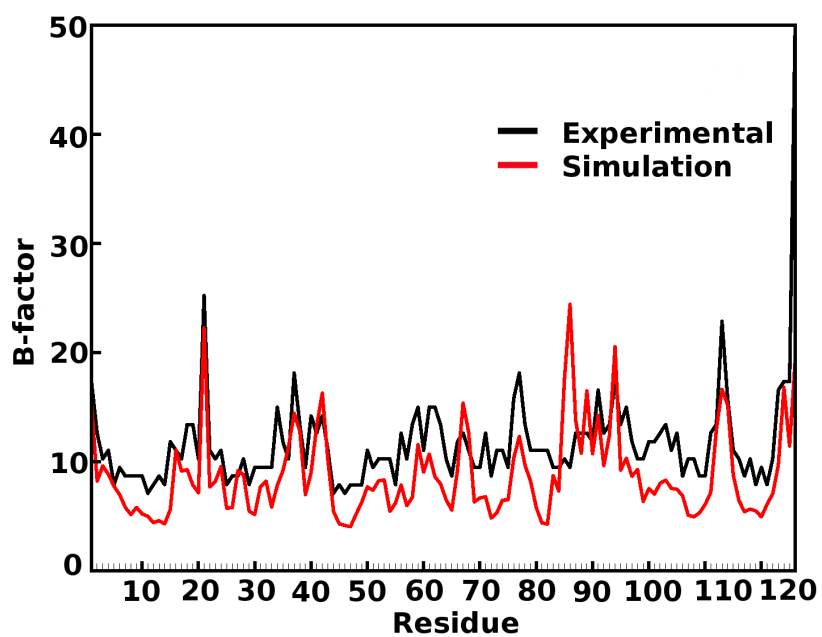
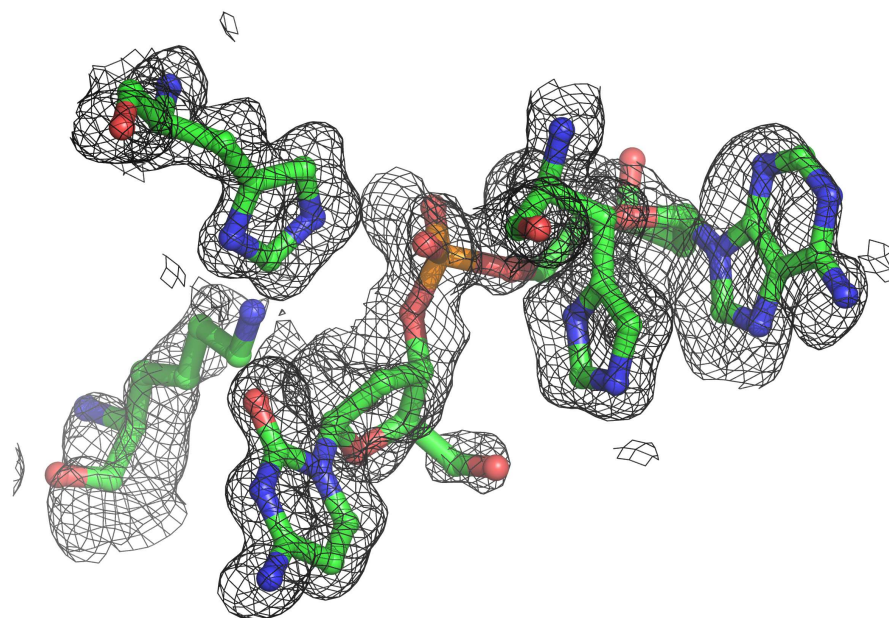


Figure 3.4: Comparison of experimental and crystal simulation characteristics. Top: overlay of the active site of the refined structure (PDB:1RPG) depicted as colored licorice and the simulation average electron density (black mesh, contoured at 1sigma). Bottom: Comparison of C-alpha B factor values from the crystal simulation (red) with the experimental values (black) for all residues in the enzyme and the inactive substrate. The Pearson correlation coefficient is 0.67 and RMSD of the B-factor is 3.94 Å

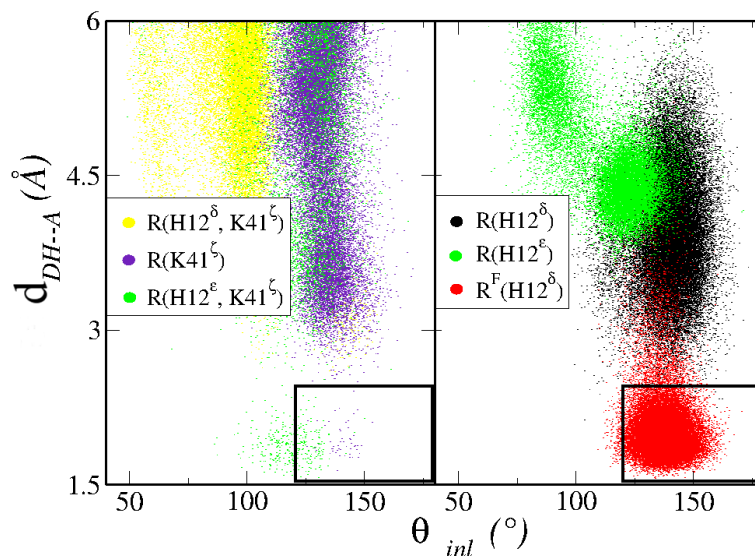


Figure 3.5: A distribution of the d_{DH--A} vs. θ_{Inl} plots for 6 different active states. Left: the distributions for the reactant states where Lys41 acting as the general base. Right: the distributions for the reactant states where His12 acting as the general base.

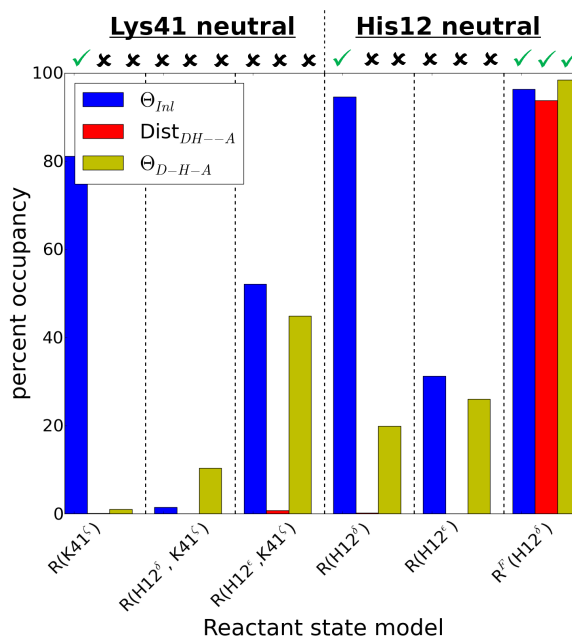


Figure 3.6: The percent occupancies of the inline attack angle (θ_{Inl}), donor acceptor distance (d_{DH--A}) and D-H-A angle (θ_{D-H-A}) for six different reactant state mimics. The first and second three sets of the plot represent the reactant states with neutral Lys41 and neutral His12, respectively. The reactant state $R^F(H12^\delta)$, meet the criteria to be the best candidate for the “active state”. The properties, that reached 80% or more of each criteria, is indicated in a green check mark; otherwise, indicated in a black ‘x’ mark.

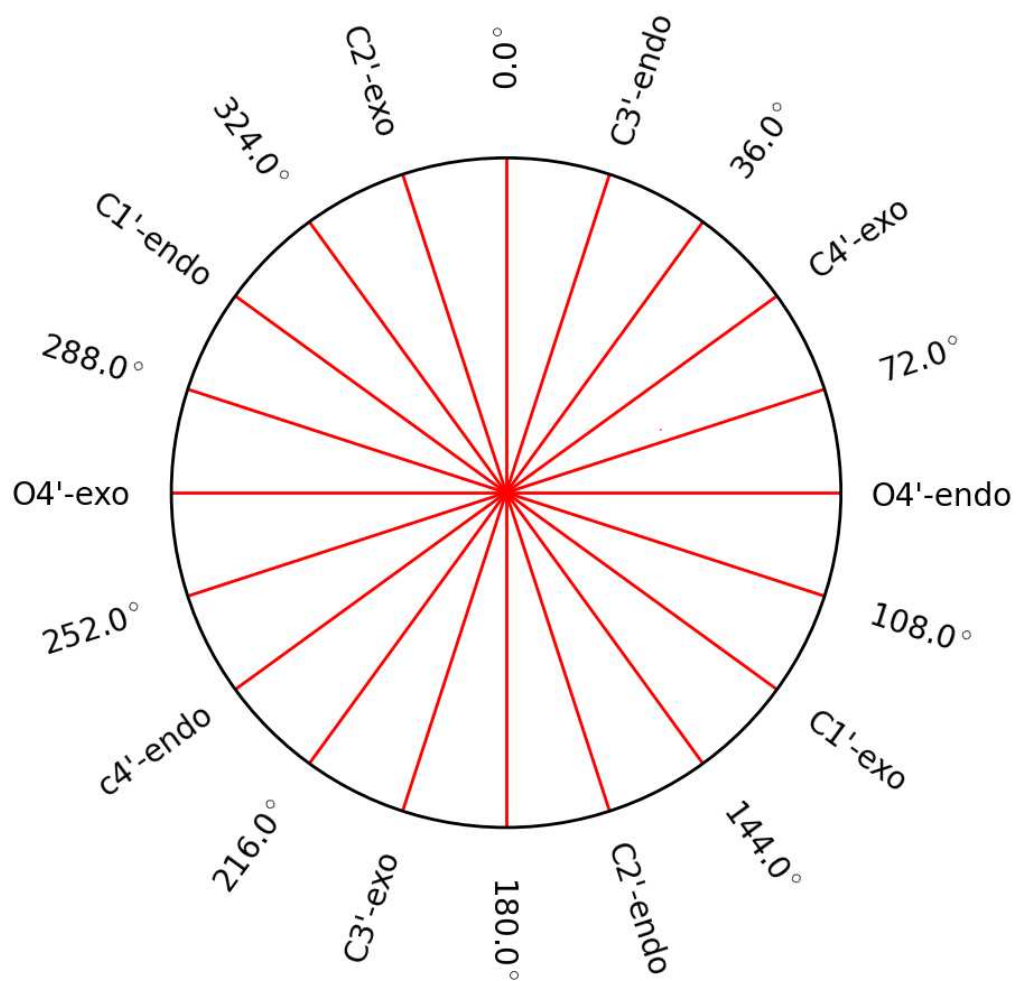


Figure 3.7: Pseudorotation cycle for furanose conformations. The values and labels along the circumference indicates the phase angle of pseudorotation and the conformation of the sugar.

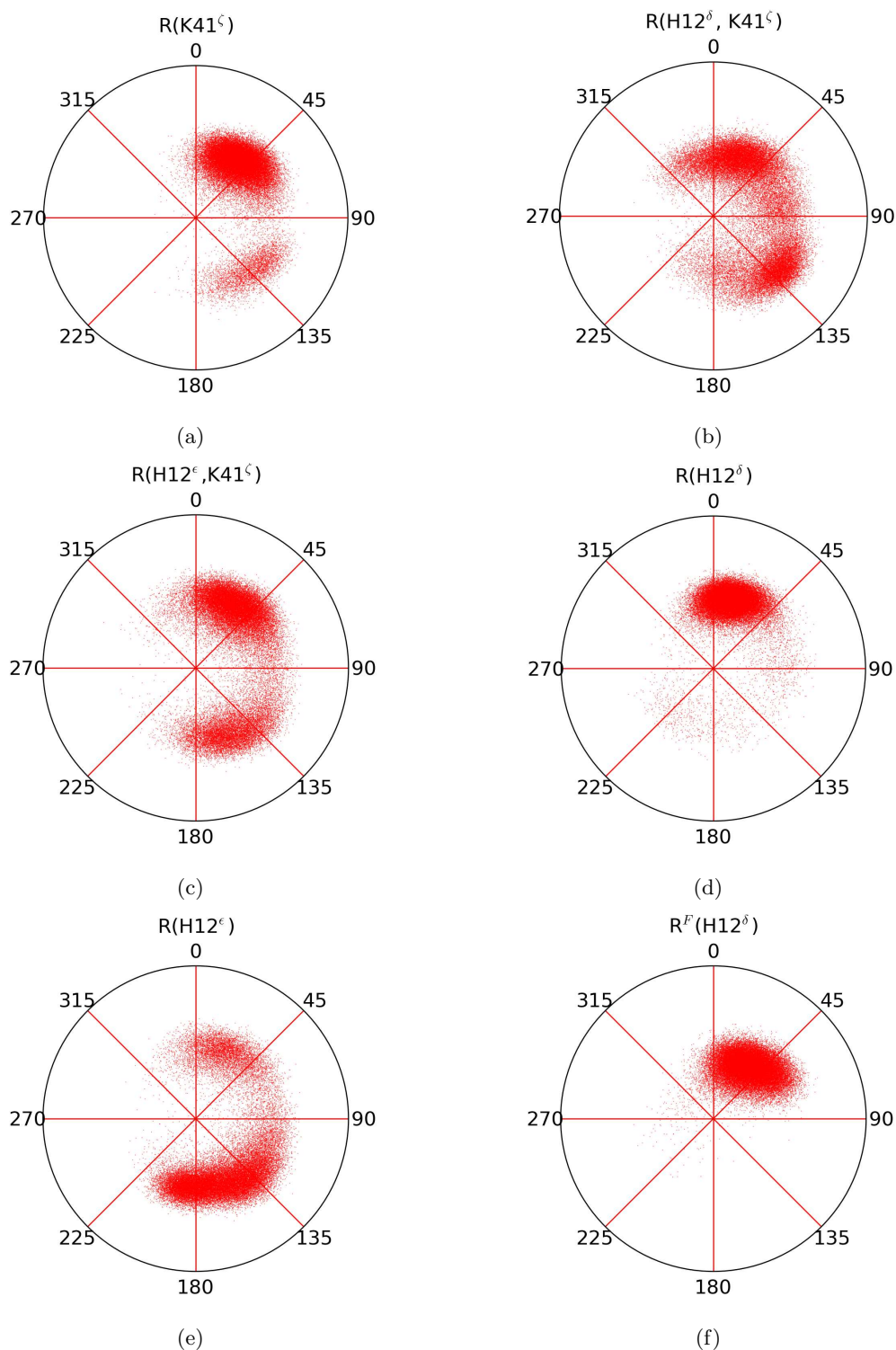


Figure 3.8: Polar plots of sugar pucker for cytosine ribose in the reactant state models: a) $R(K41^\zeta)$, b) $R(H12^\delta, K41^\zeta)$, c) $R(H12^\epsilon, K41^\zeta)$, d) $R(H12^\delta)$, e) $R(H12^\epsilon)$, f) $R^F(H12^\delta)$. Each conformations are shown with the analogous phase angle. The amplitude is shown along the coordinates of the radius. The sugar pucker of $R^F(H12^\delta)$ state stays in C3'-endo conformation.

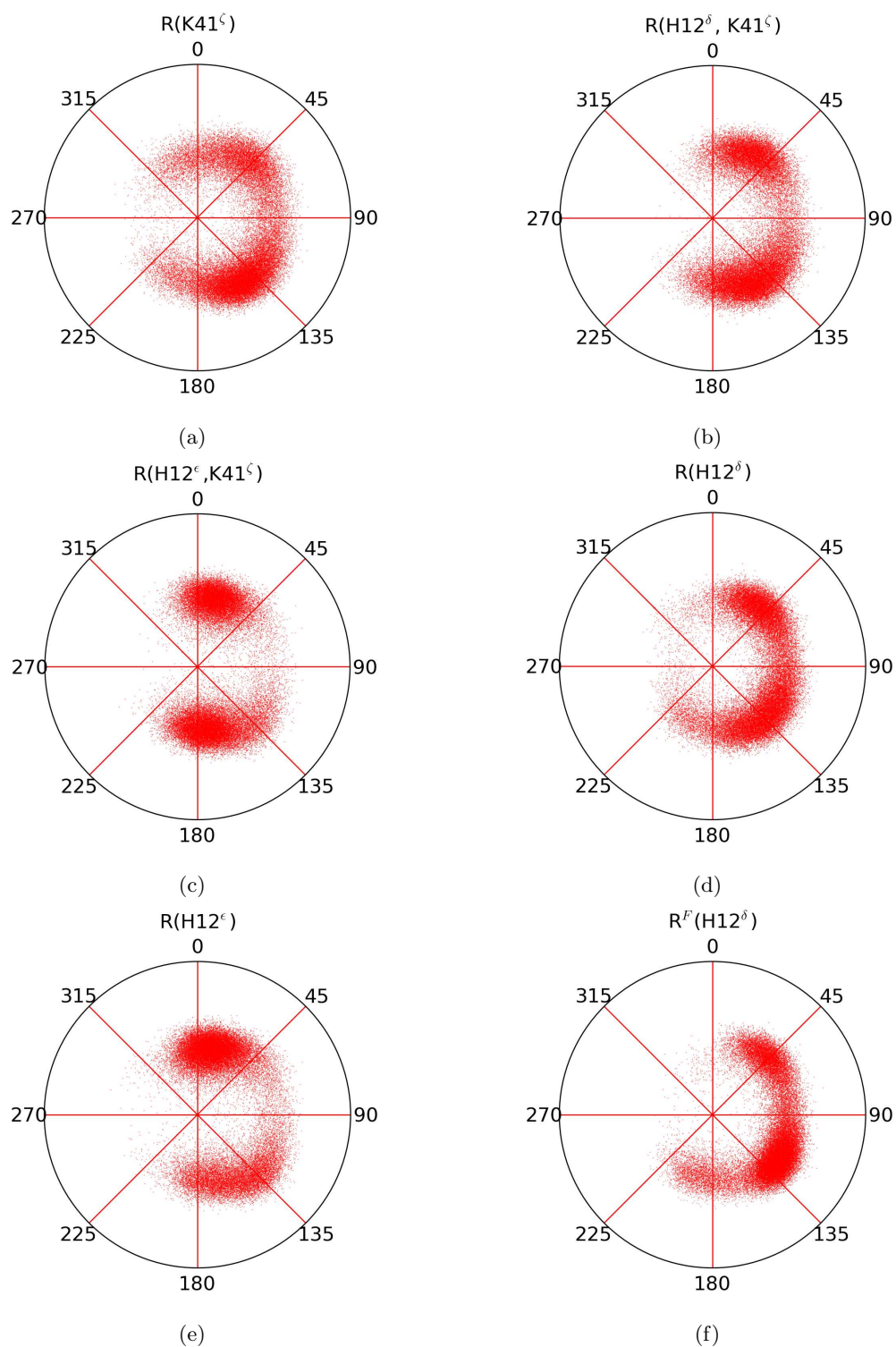


Figure 3.9: Polar plots of sugar pucker for adenosine ribose in the reactant state models: a) $R(K41^\zeta)$, b) $R(H12^\delta, K41^\zeta)$, c) $R(H12^\epsilon, K41^\zeta)$, d) $R(H12^\delta)$, e) $R(H12^\epsilon)$, f) $R^F(H12^\delta)$. Sugar pucker shows frequent transitions between C2'-endo and C3'-endo conformations.

Table 3.4: Summary of average distance of key active site interactions in the reactant state model simulations.

Atomic pair	1RPG	R(H12 ^δ)	R ^F (H12 ^δ)	R(H12 ^ε)	R(H12 ^δ , K41 ^ζ)	R(K41 ^ζ)	R(H12 ^ε ,K41 ^ζ)	R
H12:NE2–A126:OP1	2.76	4.34 (0.60)	3.78 (0.31)	5.80 (1.72)	6.94 (1.66)	5.58 (1.28)	7.76 (1.56)	5.20 (1.63)
H12:NE2–A126:OP2	4.31	5.19 (0.50)	4.26 (0.45)	6.76 (1.85)	7.50 (1.42)	4.98 (1.03)	6.61 (1.93)	5.02 (1.92)
H12:NE2–C125:O2′	3.05	4.27 (0.48)	2.92 (0.23)	5.82 (2.92)	4.59 (1.10)	3.30 (0.31)	9.57 (2.14)	3.48 (0.39)
H12:NE2–C125:HO2′	-	3.97 (0.56)	2.02 (0.28)	5.83 (2.70)	4.55 (0.93)	3.51 (0.27)	9.04 (2.08)	3.63 (0.44)
H12:ND1–N44:OD1	3.40	4.89 (0.46)	5.90 (0.46)	5.01 (1.03)	4.09 (0.39)	4.46 (0.59)	5.69 (0.69)	3.98 (0.54)
H12:ND1–T45:O	2.70	2.86 (0.12)	5.17 (0.28)	4.93 (0.55)	2.84 (0.11)	2.80 (0.10)	5.17 (0.62)	2.80 (0.10)
Q11:O–N44:ND2	2.78	2.91 (0.16)	2.96 (0.23)	4.06 (1.18)	2.96 (0.18)	2.84 (0.11)	3.00 (0.41)	2.89 (0.15)
Q11:OE1–A126:OP1	6.20	5.49 (0.86)	6.32 (0.80)	6.12 (2.03)	5.66 (1.48)	5.73 (1.34)	7.79 (1.90)	6.21 (1.25)
Q11:OE1–K41:NZ	4.39	3.82 (1.11)	4.55 (1.38)	6.21 (2.45)	5.10 (0.88)	5.54 (0.99)	5.97 (1.37)	6.19 (1.88)
Q11:NE2–K41:NZ	4.26	4.94 (0.80)	4.30 (1.31)	6.14 (1.89)	3.91 (0.89)	4.42 (0.95)	4.65 (1.43)	6.09 (1.42)
Q11:NE2–A3:126:OP1	4.18	4.06 (0.88)	5.61 (1.39)	5.74 (2.47)	5.14 (1.30)	6.13 (1.09)	7.42 (2.04)	6.96 (1.40)
Q11:NE2–A3:126:OP2	2.92	3.32 (0.77)	3.86 (1.13)	5.43 (2.65)	5.96 (1.81)	5.11 (1.26)	6.05 (2.01)	6.39 (1.74)
K41:NZ–C125:O2′	3.12	3.28 (0.97)	3.24 (0.80)	6.13 (2.30)	6.69 (1.63)	5.43 (1.56)	9.26 (2.91)	6.66 (2.48)
K41:NZ–C125:HO2′	-	3.66 (0.97)	3.68 (0.83)	6.20 (2.26)	6.82 (1.73)	5.78 (1.64)	8.97 (2.73)	6.85 (2.61)
K41:NZ–A126:OP1	4.94	6.69 (0.71)	5.98 (0.89)	5.93 (1.74)	6.16 (1.28)	7.87 (0.98)	9.27 (1.91)	7.60 (1.76)
K41:NZ–A126:OP2	4.75	6.17 (0.76)	4.98 (0.99)	5.86 (2.20)	6.91 (1.61)	7.71 (0.99)	7.96 (1.76)	7.58 (1.92)
K41:NZ–C125:O3′	3.67	4.93 (0.62)	3.96 (0.83)	5.55 (2.06)	6.15 (1.00)	5.89 (0.91)	8.53 (2.02)	6.59 (1.64)
K41:NZ–N44:OD1	2.59	3.10 (0.79)	3.02 (0.60)	3.87 (1.52)	4.27 (1.08)	3.68 (1.03)	4.30 (1.30)	5.75 (2.03)
H119:NE2–D121:OD1	2.62	5.11 (1.76)	2.83 (0.23)	8.85 (2.59)	4.49 (1.38)	4.37 (1.18)	4.16 (1.30)	3.78 (1.14)
H119:NE2–D121:OD2	4.63	5.17 (1.65)	4.80 (0.22)	8.58 (2.14)	4.37 (1.23)	4.46 (1.24)	4.93 (1.46)	4.25 (1.02)
H119:NE2–A126:O5′	4.85	4.58 (0.55)	4.96 (0.18)	4.61 (0.97)	4.78 (0.63)	5.11 (0.43)	5.18 (0.93)	5.12 (0.46)
H119:HE2–A126:O5′	-	4.97 (0.85)	5.86 (0.20)	4.89 (1.28)	5.28 (0.77)	5.91 (0.48)	5.65 (1.21)	5.96 (0.52)
H119:ND1–A126:O5′	2.66	3.74 (0.54)	2.94 (0.17)	4.21 (1.18)	3.71 (0.68)	3.24 (0.45)	4.32 (0.71)	3.21 (0.43)
H119:HD1–A126:O5′	-	3.42 (0.89)	2.12 (0.24)	4.14 (1.67)	3.31 (0.87)	2.40 (0.54)	4.05 (0.90)	2.40 (0.50)
H119:ND1–A126:OP1	3.80	3.43 (0.63)	2.87 (0.18)	4.49 (1.47)	5.09 (1.30)	3.17 (0.37)	4.75 (1.30)	3.69 (1.05)
H119:ND1–A126:OP2	5.00	4.47 (0.55)	4.79 (0.13)	5.35 (1.35)	4.96 (0.83)	3.83 (0.77)	4.71 (1.43)	4.22 (0.80)
H119:ND1–C125:O3′	4.43	5.44 (0.56)	4.24 (0.18)	5.52 (1.26)	4.00 (0.64)	4.85 (0.28)	5.83 (1.09)	4.16 (0.55)
F120:O–C125:O3′	5.13	6.44 (0.40)	5.01 (0.26)	7.36 (1.94)	4.83 (0.96)	5.41 (0.49)	8.05 (1.07)	4.86 (0.53)
F120:N–A126:OP1	3.07	4.93 (0.44)	2.90 (0.17)	6.74 (2.06)	7.52 (1.77)	4.21 (1.18)	8.06 (1.48)	4.83 (2.10)
F120:N–A126:OP2	5.49	6.52 (0.58)	5.19 (0.22)	8.19 (1.68)	7.61 (0.91)	4.34 (1.17)	7.67 (1.86)	5.40 (1.34)
C125:O2′–A126:OP1	3.03	3.78 (0.36)	3.88 (0.27)	3.62 (1.02)	4.77 (0.38)	3.80 (0.69)	3.87 (0.78)	4.03 (0.82)
C125:O2′–A126:OP2	4.41	4.14 (0.35)	4.30 (0.27)	4.54 (0.46)	4.85 (0.42)	3.73 (0.65)	4.38 (0.41)	4.01 (0.80)
A126:OP1–C125:HO2′	-	3.35 (0.51)	3.61 (0.34)	3.13 (1.55)	4.46 (0.74)	3.31 (0.85)	3.39 (1.00)	3.69 (1.17)
A126:OP2–C125:HO2′	-	3.43 (0.50)	4.07 (0.37)	4.16 (0.90)	4.69 (0.58)	3.04 (0.68)	3.84 (0.65)	3.52 (1.08)
T45:H–C125:O2	-	1.93 (0.14)	2.09 (0.43)	6.94 (5.20)	1.96 (0.17)	1.91 (0.13)	14.36 (2.42)	1.96 (0.17)
T45:OG1–C125:N3	2.79	2.86 (0.14)	3.12 (0.27)	8.53 (5.98)	2.86 (0.14)	2.93 (0.18)	16.89 (2.12)	2.91 (0.21)
T45:HG1–ASP:83:OD1	-	5.73 (0.77)	5.94 (0.90)	5.52 (1.33)	5.52 (0.95)	6.20 (0.87)	5.57 (1.32)	5.97 (0.96)
T45:HG1 – ASP:83:OD2	-	5.86 (0.77)	5.77 (0.93)	5.53 (1.25)	5.92 (0.94)	6.25 (0.86)	5.58 (1.43)	5.90 (0.99)
C125:N4 – ASP:83:OD1	6.07	6.22 (1.05)	7.22 (1.12)	11.75 (6.27)	5.72 (1.13)	6.70 (1.14)	21.84 (2.74)	6.29 (1.10)
C125:N4 – ASP:83:OD2	4.16	6.40 (1.06)	7.02 (1.19)	11.71 (6.46)	6.11 (1.07)	6.79 (1.13)	21.86 (2.67)	6.22 (1.21)
ASN:71:ND2 – A126:N1	2.94	6.71 (2.16)	3.45 (1.03)	5.00 (2.05)	3.71 (1.23)	5.71 (2.80)	3.41 (1.30)	3.16 (0.58)
ASN:71:OD1 – A126:N6	2.84	6.51 (2.30)	3.13 (0.95)	5.29 (2.65)	3.56 (1.29)	5.90 (2.96)	3.33 (1.22)	3.06 (0.50)

The table compares the average key distances in the active site for the set of reactant state models, described in the Table 3.1, with the distances from inactive state bound crystal structures. C125 and A126 are the residues of the dinucleotide substrate CpA.

Table 3.5: The average key distances in the active site of the activated precursor state models.

Atomic pair	1RPG	AP	AP ^F
H12:NE2 – A126:OP1	2.76	7.23 (0.41)	4.06 (0.38)
H12:NE2 – A126:OP2	4.31	7.75 (0.33)	4.89 (0.46)
H12:NE2 – CS125:O2'	-	3.11 (0.36)	2.79 (0.15)
H12:HE2 – CS125:O2'	-	2.46 (0.44)	1.91 (0.21)
H12:ND1 – N44 :OD1	3.40	3.79 (0.36)	5.93 (0.47)
H12:ND1 – T45:O	2.70	2.81 (0.11)	5.28 (0.23)
Q11:O – N44 :ND2	2.78	2.90 (0.17)	2.89 (0.14)
Q11:OE1 – A126:OP1	6.20	6.53 (0.90)	5.62 (0.71)
Q11:OE1 – K41:NZ	4.39	5.09 (1.66)	7.12 (0.97)
Q11:NE2 – K41:NZ	4.26	5.64 (1.26)	6.52 (1.39)
Q11:NE2 – A126:OP1	4.18	7.18 (0.91)	6.08 (1.19)
Q11:NE2 – A126:OP2	2.92	9.23 (0.89)	4.22 (0.97)
K41:NZ – CS125:O2'	-	4.30 (1.93)	2.80 (0.30)
K41:NZ – A126:OP1	4.94	5.46 (0.78)	6.38 (0.37)
K41:NZ – A126:OP2	4.75	7.07 (0.75)	5.87 (0.46)
K41:NZ – CS125:O3'	3.67	5.47 (1.09)	4.57 (0.34)
K41:NZ – N44 :OD1	2.59	4.22 (1.63)	3.00 (0.41)
H119:NE2 – D121:OD1	2.62	4.91 (0.99)	3.39 (0.89)
H119:NE2 – D121:OD2	4.63	3.72 (1.13)	4.38 (0.86)
H119:NE2 – A126:O5'	4.85	4.73 (0.31)	4.90 (0.24)
H119:HE2 – A126:O5'	-	5.46 (0.39)	5.76 (0.27)
H119:ND1 – A126:O5'	2.66	3.24 (0.27)	3.10 (0.27)
H119:HD1 – A126:O5'	-	2.76 (0.42)	2.50 (0.34)
H119:ND1 – A126:OP1	3.80	5.05 (0.20)	2.80 (0.23)
H119:ND1 – A126:OP2	5.00	3.73 (0.40)	4.80 (0.16)
H119:ND1 – CS125:O3'	4.43	3.24 (0.36)	4.22 (0.32)
F120:O – CS125:O3'	5.13	4.74 (0.38)	5.20 (0.39)
F120:N – A126:OP1	3.07	7.68 (0.36)	3.18 (0.46)
F120:N – A126:OP2	5.49	6.84 (0.43)	5.45 (0.43)
CS125:O2' – A126:P	-	4.26 (0.12)	3.77 (0.17)
CS125:O2' – A126:OP1	3.03	4.61 (0.28)	3.88 (0.24)
CS125:O2' – A126:OP2	4.41	5.18 (0.12)	4.20 (0.24)
A126:OP1 – H12:HE2	-	6.31 (0.44)	3.98 (0.43)
A126:OP2 – H12:HE2	-	6.91 (0.38)	4.78 (0.49)
T45:H – CS125:O2	-	1.92 (0.14)	1.93 (0.16)
T45:OG1 – CS125:N3	2.79	2.93 (0.18)	2.88 (0.16)
T45:HG1 – D83:OD1	-	5.84 (0.96)	5.73 (0.81)
T45:HG1 – D83:OD2	-	5.69 (0.98)	6.02 (0.76)
CS125:N4 – D83:OD1	6.07	6.28 (1.02)	6.42 (1.1)
CS125:N4 – D83:OD2	4.16	6.15 (1.03)	6.88 (1.01)
N71:ND2 – A126:N1	2.94	3.03 (0.20)	4.82 (2.06)
N71:OD1 – A126:N6	2.84	3.02 (0.32)	5.48 (1.45)

A comparison of the average distance in the active site of the activated precursor simulations represented for His12 non flipped (AP) and flipped states (AP^F). CS125 and A126 are the substrate residues in the activated precursor state. CS125 is deprotonated at O2'.

Table 3.6: The comparison of the key average distances in the transition state consist of His12 flipped conformation (TS^F), to the distances in the ribonuclease A-uridine vanadate complex.

Atomic pair	1RUV	TS^F
H12:NE2 – A3:126:OP1	2.72	4.51 (0.47)
H12:NE2 – A3:126:OP2	2.87	3.54 (0.32)
H12:NE2 – C5:125:O2'	3.39	3.16 (0.31)
H12:ND1 – N44:OD1	3.47	4.79 (0.39)
H12:ND1 – T45:O	2.69	5.13 (0.23)
Q11:O – N44:ND2	2.80	2.86 (0.13)
Q11:OE1 – A3:126:OP1	5.47	4.01 (0.62)
Q11:OE1 – K41:NZ	2.87	4.25 (1.13)
Q11:NE2 – K41:NZ	3.85	5.72 (1.00)
Q11:NE2 – A3:126:OP1	3.85	5.58 (0.92)
Q11:NE2 – A3:126:OP2	2.87	7.61 (0.90)
K41:NZ – C5:125:O2'	2.73	2.82 (0.14)
K41:NZ – A3:126:OP1	4.83	3.46 (0.35)
K41:NZ – A3:126:OP2	3.34	5.04 (0.21)
K41:NZ – C5:125:O3'	4.86	4.51 (0.22)
K41:NZ – N44:OD1	2.81	2.86 (0.19)
H119:NE2 – D121:OD1	2.66	5.36 (0.90)
H119:NE2 – D121:OD2	4.73	5.38 (1.22)
H119:NE2 – A3:126:O5'	6.13	4.87 (0.10)
H119:HE2 – A3:126:O5'	6.06	5.79 (0.11)
H119:ND1 – A3:126:O5'	6.13	2.74 (0.08)
H119:HD1 – A3:126:O5'	6.04	1.76 (0.09)
H119:ND1 – A3:126:OP1	3.69	4.91 (0.11)
H119:ND1 – A3:126:OP2	4.81	3.61 (0.20)
H119:ND1 – C5:125:O3'	3.16	3.09 (0.17)
F120:O – C5:125:O3'	3.65	4.2 (0.29)
F120:N – A3:126:OP1	2.86	5.65 (0.16)
F120:N – A3:126:OP2	5.24	2.98 (0.17)
C5:125:O2' – A3:126:OP1	2.99	2.49 (0.06)
C5:125:O2' – A3:126:OP2	2.52	2.56 (0.06)
T45:H – C5:125:O2	2.11	2.13 (0.28)
T45:OG1 – C5:125:N3	2.79	2.98 (0.30)
T45:HG1 – D83:OD1	2.50	5.53 (1.23)
T45:HG1 – D83:OD2	4.38	5.59 (1.11)
C5:125:N4 – D83:OD1	-	5.98 (0.99)
C5:125:N4 – D83:OD2	-	5.99 (0.99)
N71:ND2 – A3:126:N1	-	7.63 (1.84)
N71:OD1 – A3:126:N6	-	8.16 (2.04)

The average distances from the TS^F state is compared with the distances from ribonuclease A-uridine vanadate complex ¹³⁵ (PDB ID: 1RUV)

Table 3.7: Key average distance in the active site of 2',3'-cyclic phosphate product state bound RNase A.

Atomic pair	1RUV	$P^F(\text{H119}^\epsilon)$	$P(\text{H119}^\epsilon)$	$P^F(\text{H119}^\delta)$	$P(\text{H119}^\delta)$
H12:NE2 – CP:125:O31	2.73	3.43 (0.92)	3.32 (1.07)	2.88 (0.23)	2.91 (0.38)
H12:NE2 – CP:125:O32	3.92	4.74 (0.54)	4.47 (0.65)	4.23 (0.41)	4.25 (0.43)
H12:NE2 – CP:125:O2'	3.39	3.34 (0.46)	3.46 (0.50)	3.43 (0.25)	3.38 (0.27)
H12:ND1 – N44:OD1	3.47	4.33 (0.74)	3.98 (0.48)	3.76 (0.38)	3.81 (0.42)
H12:ND1 – T45:O	2.68	3.48 (1.09)	2.81 (0.11)	2.80 (0.10)	2.79 (0.10)
Q11:O – N44:ND2	2.80	2.87 (0.13)	2.88 (0.13)	2.88 (0.14)	2.87 (0.13)
Q11:OE1 – CP:125:O31	5.48	4.32 (0.95)	4.78 (1.22)	4.99 (0.87)	5.04 (0.90)
Q11:OE1 – K41:NZ	2.87	6.69 (1.54)	6.62 (1.87)	4.06 (1.53)	4.91 (1.88)
Q11:NE2 – K41:NZ	3.86	6.13 (1.31)	6.13 (1.45)	4.84 (1.06)	5.12 (1.24)
Q11:NE2 – CP:125:O31	3.85	4.57 (1.14)	4.89 (1.27)	5.71 (1.19)	4.86 (1.15)
Q11:NE2 – CP:125:O32	2.87	4.67 (1.40)	4.10 (1.03)	4.54 (0.99)	3.80 (0.86)
K41:NZ – CP:125:O2'	2.73	5.29 (1.66)	5.98 (1.92)	4.01 (1.39)	5.14 (1.89)
K41:NZ – CP:125:O31	4.83	5.96 (1.73)	7.22 (1.93)	5.30 (1.12)	6.52 (1.83)
K41:NZ – CP:125:O32	3.34	4.03 (1.65)	5.23 (2.04)	3.28 (1.03)	4.44 (1.76)
K41:NZ – CP:125:O3'	4.86	5.62 (1.55)	6.36 (1.77)	4.81 (1.08)	5.86 (1.68)
K41:NZ – N44:OD1	2.81	5.00 (1.48)	5.34 (1.68)	4.12 (1.48)	4.91 (1.80)
H119:NE2 – D121:OD1	2.65	5.39 (2.15)	5.34 (2.44)	10.01 (2.54)	9.71 (2.42)
H119:NE2 – D121:OD2	4.73	5.20 (2.27)	5.56 (2.34)	9.37 (2.37)	9.38 (2.06)
H119:NE2 – CP:125:O5'	6.13	6.96 (2.25)	7.74 (2.04)	7.86 (1.53)	8.09 (1.23)
H119:ND1 – CP:125:O5'	6.13	7.16 (1.81)	7.74 (1.77)	7.52 (1.39)	7.85 (1.38)
H119:ND1 – CP:125:O31	3.68	5.22 (1.00)	5.01 (0.93)	4.53 (1.13)	4.77 (1.08)
H119:ND1 – CP:125:O32	4.81	6.88 (1.21)	6.59 (1.27)	5.59 (1.12)	5.85 (1.03)
H119:ND1 – CP:125:O3'	3.15	5.37 (1.22)	5.39 (1.25)	4.75 (1.13)	5.00 (1.20)
F120:O – CP:125:O3'	3.65	5.22 (1.13)	5.28 (1.11)	3.91 (0.47)	3.90 (0.57)
F120:N – CP:125:O31	2.86	5.61 (1.18)	4.98 (1.25)	3.26 (0.45)	3.29 (0.54)
F120:N – CP:125:O32	5.24	7.63 (1.06)	7.04 (1.25)	5.61 (0.34)	5.60 (0.43)
CP:125:O2' – CP:125:O31	3.00	2.49 (0.06)	2.49 (0.06)	2.48 (0.06)	2.47 (0.06)
CP:125:O2' – CP:125:O32	2.52	2.50 (0.06)	2.50 (0.06)	2.48 (0.06)	2.49 (0.06)
T45:H – CP:125:O2	2.11	2.02 (0.30)	1.95 (0.15)	1.97 (0.17)	1.95 (0.16)
T45:OG1 – CP:125:N3	2.79	3.19 (0.73)	2.84 (0.15)	2.90 (0.19)	2.90 (0.17)
T45:HG1 – D83:OD1	-	5.36 (1.47)	5.83 (0.77)	5.60 (1.06)	5.64 (0.99)
T45:HG1 – D83:OD2	-	5.43 (1.55)	5.74 (0.78)	5.57 (1.08)	5.85 (1.03)
CP:125:N4 – D83:OD1	-	5.91 (1.25)	6.09 (1.00)	5.80 (1.05)	5.80 (1.02)
CP:125:N4 – D83:OD2	-	6.03 (1.30)	6.01 (1.01)	5.75 (1.08)	6.07 (1.14)
N71:ND2 – CP:125:N1	15.15	17.00 (1.56)	17.03 (1.68)	15.54 (1.38)	15.58 (1.37)

The average distance in the active site residues of the 2',3'-cyclic phosphate product state are compared with the active site of the uridine-vanadate transition state mimic bound RNase A¹³⁵. Results were reported for four different systems, $P(\text{H119}^\delta)$, $P^F(\text{H119}^\delta)$, $P(\text{H119}^\epsilon)$, $P^F(\text{H119}^\epsilon)$ where His119 is deprotonated at the N ϵ and N δ sites while protonated His12 is in its flipped or not-flipped state.

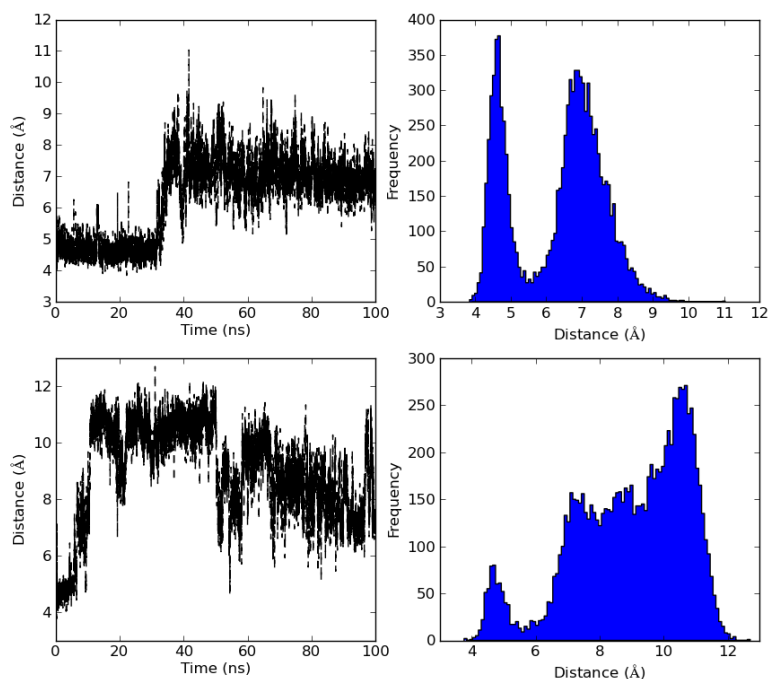


Figure 3.10: K41:NZ–C125:C2' distance vs. simulation time plots and the distribution of K41:NZ–C125:C2' distance for both crystal unit cells. The results indicates two conformations of Lys41 side chain in the first unit cell. The conformation centered at distance 4.67 Å is lower in occupancy comparative to the conformation at distance 7.10 Å in the first unit cell. A broad conformational fluctuation is observed in the second unit cell and the average value is located at the 8.95 Å.

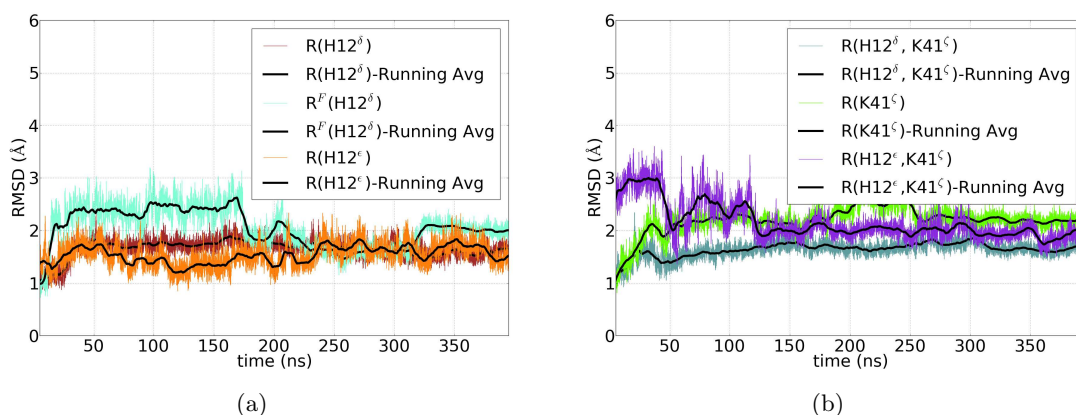


Figure 3.11: The plots of backbone RMSD of the reactant state models with respect to the crystal structure. a) The RMSD plots for the reactant state model when His12 act as the general base where His12 is in neutral state. b) The RMSD plots for the reactant state model when Lys41 is in deprotonated state.

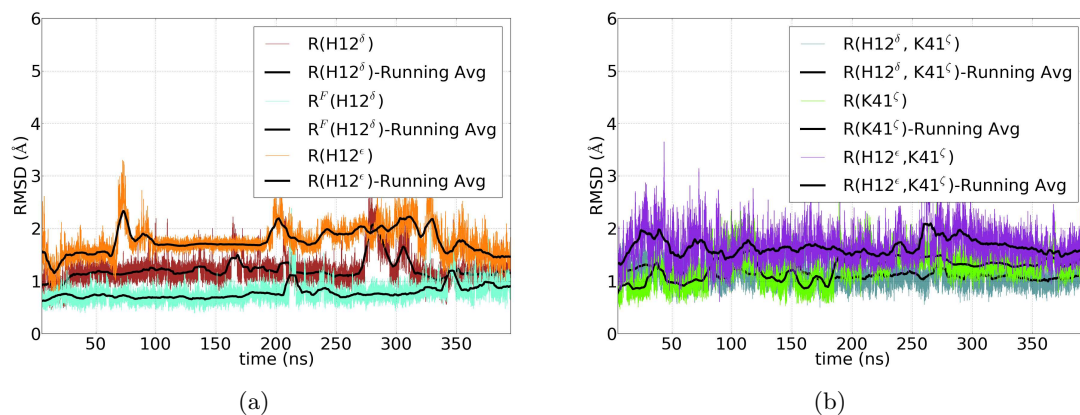


Figure 3.12: The RMSD of the active site residues (His12, His119 and Lys41) in the reactant states models with respect to the crystal structure. a) The RMSD plots for the reactant state model when His12 act as the general base where His12 is in neutral state. b) The RMSD plots for the reactant state model when Lys41 is in deprotonated state.

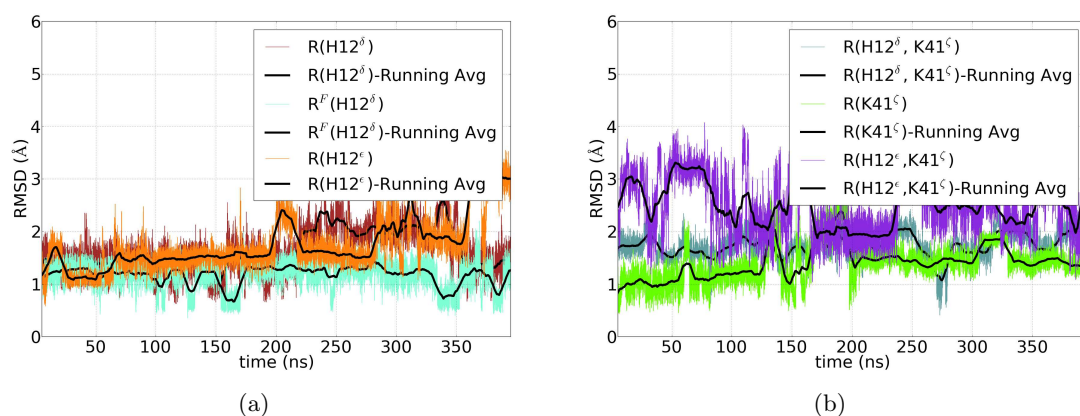


Figure 3.13: RMSD of the substrate in comparison to the inactive substrate in the crystal structure is shown in the figure. a) The set of RMSD plots of the substrate, which represent the His12 deprotonated state in the reactant state model. b) RMSD plots of the substrate when Lys41 is in its deprotonated state.

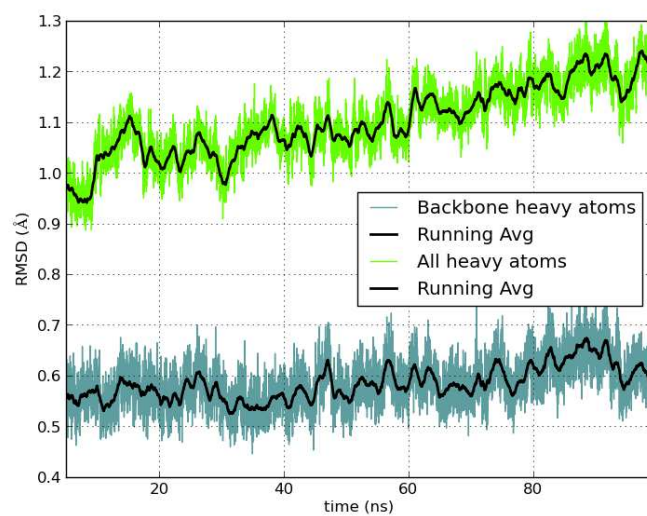


Figure 3.14: RMSD of average structure for 100 ns of the crystal simulation (Xtal) is 0.34 Å for all backbone atoms and 0.70 Å for all heavy atoms.

Table 3.8: The average distance in the active site from the crystal simulation

Atomic Pair	IRPG	crystal unit1	crystal unit2
H12:NE2 – DA3:126:OP1	2.76	3.11 (0.30)	2.87 (0.22)
H12:NE2 – DA3:126:OP2	4.31	3.04 (0.34)	3.83 (0.57)
H12:ND1 – N44:OD1	3.40	3.62 (0.29)	3.45 (0.33)
H12:ND1 – T45:O	2.70	2.82 (0.11)	2.82 (0.11)
Q11:O – N44:ND2	2.78	2.84 (0.12)	2.85 (0.12)
Q11:OE1 – DA3:126:OP1	6.20	6.76 (0.80)	5.72 (0.98)
Q11:OE1 – K41:NZ	4.39/4.41	3.19 (0.85)	5.97 (1.93)
Q11:NE2 – K41:NZ	4.26/6.33	4.06 (0.68)	6.54 (1.47)
Q11:NE2 – DA3:126:OP1	4.18	5.89 (0.86)	4.71 (0.98)
Q11:NE2 – DA3:126:OP2	2.92	3.54 (0.79)	3.25 (0.53)
K41:NZ – DA3:126:OP1	4.94/9.16	6.67 (1.31)	9.53 (1.95)
K41:NZ – DA3:126:OP2	4.72/7.12	4.98 (1.23)	7.27 (1.86)
K41:NZ – DC5:125:O3'	3.67/7.53	4.59 (1.22)	8.43 (2.05)
K41:NZ – N44:OD1	2.59/7.85	5.07 (1.33)	8.45 (1.87)
H119:NE2 – D121:OD1	2.62	4.30 (1.02)	3.65 (1.03)
H119:NE2 – D121:OD2	4.63	4.10 (1.29)	4.27 (1.02)
H119:NE2 – DA3:126:O5'	4.85	5.20 (0.31)	5.01 (0.22)
H119:HE2 – DA3:126:O5'	5.77	6.10 (0.32)	5.91 (0.24)
H119:ND1 – DA3:126:O5'	2.66	3.14 (0.31)	2.92 (0.22)
H119:HD1 – DA3:126:O5'	1.65	2.27 (0.38)	1.98 (0.27)
H119:ND1 – DA3:126:OP1	3.80	2.93 (0.27)	3.58 (0.45)
H119:ND1 – DA3:126:OP2	5.00	4.92 (0.15)	4.92 (0.14)
H119:ND1 – DC5:125:O3'	4.43	4.26 (0.19)	3.55 (0.43)
F120:O – DC5:125:O3'	4.70	4.94 (0.27)	3.94 (0.56)
F120:N – DA3:126:OP1	3.07	2.86 (0.14)	3.17 (0.39)
F120:N – DA3:126:OP2	5.49	5.02 (0.27)	5.51 (0.41)
T45:OG1 – DC5:125:N3	2.79	2.87 (0.16)	2.87 (0.15)
T45:HG1 – D83:OD1	5.95	5.73 (0.91)	5.37 (1.04)
T45:HG1 – D83:OD2	4.05	5.68 (0.92)	5.47 (1.01)
DC5:125:N4 – D83:OD1	6.07	5.87 (0.97)	5.31 (1.14)
DC5:125:N4 – D83:OD2	4.16	5.76 (1.00)	5.38 (1.17)
N71:ND2 – DA3:126:N1	2.94	3.18 (0.25)	3.09 (0.19)
N71:OD1 – DA3:126:N6	2.84	2.91 (0.14)	2.90 (0.13)

The average distance in the active site residues are compared with the distance in the crystal structure.

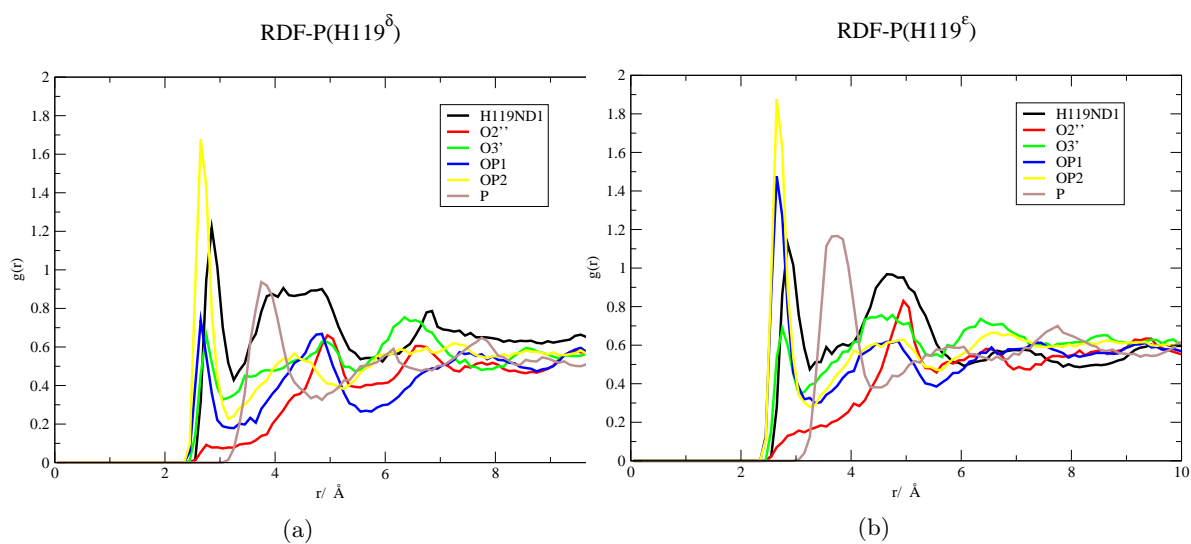


Figure 3.15: The radial distribution functions ($g(r)$) of water molecules around N^δ of His119, $\text{O2}'$, $\text{O3}'$, phosphoryl oxygens (OP1, OP2) and phosphate (P). a) The plots for the product state when neutral His119 is having a proton on N^δ site b). The plots for the product state when neutral His119 is having a proton on N^ϵ site

Chapter 4

Interpretation of pH–Activity Profiles for Acid–Base Catalysis from Molecular Simulations

The measurement of reaction rate as a function of pH provides essential information about mechanism. These rates are sensitive to the pK_a values of amino acids directly involved in catalysis that are often shifted by the enzyme active site environment. Experimentally observed pH-rate profiles are usually interpreted using simple kinetic models that allow estimation of “apparent pK_a ” values of presumed general acid and base catalysts. One of the underlying assumptions in these models is that the protonation states are uncorrelated. In the present work, we introduce the use of constant pH molecular dynamics simulations in explicit solvent (CpHMD) with replica exchange in the pH-dimension (pH-REMD) as a tool to aid in the interpretation of pH-activity data of enzymes, and test the validity of different kinetic models. We apply the methods to RNase A, a prototype acid/base catalyst, to predict the macroscopic and microscopic pK_a values, as well as the shape of the pH-rate profile. Results for apo and cCMP-bound RNase A agree well with available experimental data, and suggest that deprotonation of the general acid and protonation of the general base are not strongly coupled in transphosphorylation and hydrolysis steps. Stronger coupling, however, is predicted for the Lys41 and His119 protonation states in apo RNase A, leading to the requirement for a microscopic kinetic model. This type of analysis may be important for other catalytic systems where the active forms of implicated general acid and base are oppositely charged and more highly correlated. These results suggest a new way for CpHMD/pH-REMD simulations to bridge the gap with experiments to provide a molecular-level interpretation of pH-activity data in studies of enzyme mechanisms.

Acid-base catalysis is a common catalytic strategy in protein and RNA enzymes¹⁴⁰, and is employed in the cleavage of the RNA phosphodiester backbone by RNase A^{1,91} as well as small nucleolytic RNA enzymes¹⁴¹. General base and acid catalysts facilitate nucleophile activation through proton abstraction, and promote leaving group departure through proton donation, respectively. The observed reaction rate is assumed to be proportional to the probability of finding the enzyme in a “catalytically active” state with the acid protonated (*i.e.*, able to donate a proton), and the base deprotonated (*i.e.*, able to receive a proton). This conditional probability will thus be a function of the pH, and the pH-rate curves will be sensitive to the pK_a values of the general acid and base.^{49,39}

The measurement of reaction kinetics as a function of pH provides vital information about mechanism; however, the interpretation of this data is not always straightforward.³⁹ Experimentally determined pH-rate curves are commonly fit to a simple equilibrium model where the apparent pK_a values of the general acid and base appear as independent parameters. When protonation states are strongly coupled as they often are in enzyme active sites, irregular titration behavior occurs, requiring a more detailed theoretical analysis^{49,40,41}. Recently, computational methods have emerged that allow molecular simulations in explicit solvent to be performed under constant pH conditions (CpHMD), and the conditional probabilities of correlated protonation events to be directly determined^{4,142,143,144,145,43,146,147}. CpHMD can be used in conjunction with replica exchange molecular dynamics in the pH dimension (pH-REMD) in order to enhance sampling of important states while at the same time providing information over a range of pH values that can be used to predict complex titration curves^{4,50}. The present work reports the first application of the CpHMD/pH-REMD method to the prediction of the pH-rate curves for the apo and the 2',3'-cyclic phosphate (2'O-transphosphorylation product) bound RNase A, a prototype acid-base catalyst.

RNase A catalyzes a 2'O-transphosphorylation of a bound RNA substrate that involves cleavage of the phosphodiester backbone to form a 2',3'-cyclic phosphate and 5'-hydroxyl termini^{1,91}. In a subsequent reaction RNase A catalyzes the hydrolysis of the cyclic phosphate to form a 3' phosphate. Both transphosphorylation and hydrolysis involve general acid-base catalysis, and thus are strongly pH-dependent. The kinetics of RNase A have been extensively studied^{93,148}, including analysis of the roles of His12 and His119¹⁰⁶ and the pH-dependence of substrate association¹⁴⁹. In the present work, we examine the effect of pH on the acid-base catalytic step in RNase A 2'O-transphosphorylation and hydrolysis. We do not consider here the effect of pH on substrate association and binding, which is known to be important¹⁴⁹. Extension of the theoretical framework to take into account the added dimension of substrate binding is possible, but this requires technical details that are beyond the scope of this first application. Nonetheless, we note that very recent progress in this area has been reported¹⁵⁰.

Figure 4.1 illustrates the putative mechanism of RNA cleavage via transphosphorylation and hydrolysis of cytidyl-3'-5'adenosine (CpA) and 2',3'-cyclic phosphate by RNase A^{1,91}. The His119/His12 pair is generally accepted general acid/base pair in transphosphorylation (although other mechanisms have been proposed and discussed⁹¹). His12 abstracts the proton from O2' to facilitate the nucleophilic attack on the adjacent phosphorus atom. His119 act as the general acid to donate a proton to the O5' leaving group, resulting in a 2',3'-cyclic phosphate. Subsequent hydrolysis occurs by the action of His119 as the general base to abstract the proton from a water molecule in order to facilitate nucleophilic attack. The His12 residue acts as the general acid to donate a proton to O2' and leads to the 3'-phosphomonoester final product. Note that some reports in the literature^{100,3} have suggested that Lys41, rather than His12, may play the role of the general base in transphosphorylation. Although this mechanism is not widely accepted, we nonetheless consider it for comparison.

Putative Mechanism of Transphosphorylation (left) and Hydrolysis (right) by RNase A

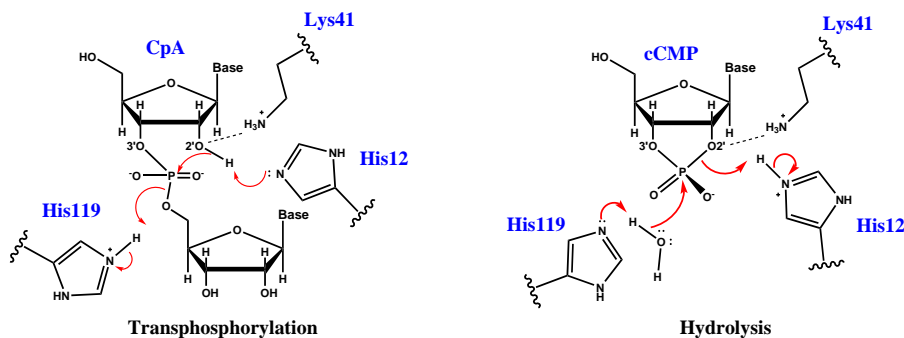


Figure 4.1: The generally accepted view is that H119/H12 acts as the general acid/base pair in transphosphorylation, and their roles are reversed in hydrolysis. However, there has been some debate in the literature that alternatively, K41 might act as general base in transphosphorylation, although this is less widely accepted.

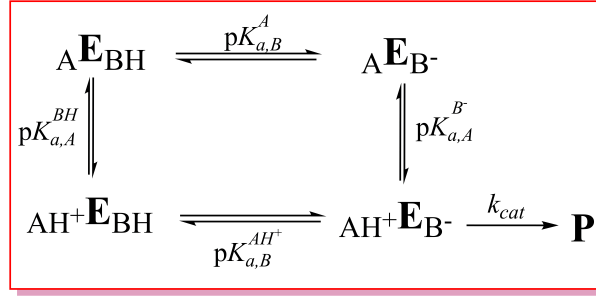
4.1 Methods

Interpretation of pH-rate data for general acid-base catalysis

Figure 4.1.1 illustrates a microscopic kinetic model for general acid-base catalysis used to interpret pH-activity data. The underlying kinetic assumption is that the catalytic rate is proportional to the active species, $AH^+E_{B^-}$, which has the general base deprotonated (B^-) so as to be able to accept a proton and activate the nucleophile, while the general acid (AH^+) is protonated so as to be able to donate a proton to the leaving group.

The pH-dependent probabilities for each of the four micro-states illustrated in Figure 4.1.1 can

Microscopic general acid-base protonation state model used to interpret pH-activity data



Scheme 4.1.1: $AH^+E_{B^-}$, AH^+E_{BH} , $A E_{BH}$, $A E_{B^-}$ are four microstates. E stands for “enzyme” and subscripts A and B indicate the acid and base, respectively. The pK_a shifts discussed in the text are defined as $\Delta pK_{a,A} = pK_{a,A}^{BH} - pK_{a,A}^{B^-}$, and $\Delta pK_{a,B} = pK_{a,B}^A - pK_{a,B}^{AH^+}$. Note the constraint of the thermodynamic cycle ensures $\Delta pK_{a,A} + \Delta pK_{a,B} = 0$, and a positive value for $\Delta pK_{a,B}$ indicates anticooperative coupling of protonation states (i.e., protonation of the acid site disfavors protonation of the base), whereas a negative value of $\Delta pK_{a,B}$ indicates cooperative coupling (i.e., protonation of the acid site *favours* protonation of the base). The “apparent pK_a ” model discussed in the text involves fitting of pH-rate data under the constraint that $\Delta pK_{a,A} = -\Delta pK_{a,B} = 0$ (see text). This scheme does not consider the pH-dependence of substrate binding, which is also important for a complete kinetic characterization.

be described by the partition function Q

$$Q = 1 + 10^{pK_{a,B}^{AH^+} - pH} + 10^{pK_{a,B}^{AH^+} - pK_{a,A}^{BH}} + 10^{pH - pK_{a,A}^{B^-}} \quad (4.1)$$

from which the probabilities (fractions) for each state can be determined as:

$$f_{(AH^+/B^-)} = 1/Q \quad (4.2)$$

$$f_{(AH^+/BH)} = 10^{pK_{a,B}^{AH^+} - pH}/Q \quad (4.3)$$

$$f_{(A/BH)} = 10^{pK_{a,B}^{AH^+} - pK_{a,A}^{BH}}/Q \quad (4.4)$$

$$f_{(A/B^-)} = 10^{pH - pK_{a,A}^{B^-}}/Q \quad (4.5)$$

Note that the reference energy in the partition function is taken as that of the active state $AH^+E_{B^-}$ (set to zero energy). The microscopic pK_a values can be determined from fitting to these four fractions simultaneously under the constraint that their related free energy values sum to zero in accord with the thermodynamic cycle shown in Figure 4.1.1. The plots of the data used in deriving microscopic pK_a values can be found in supporting information.

A common assumption in the interpretation of pH-rate profiles is that the protonation/deprotonation events are uncorrelated; *i.e.*, an equivalence is assumed between microscopic pK_a values $pK_{a,A}^{BH} = pK_{a,A}^{B^-}$ and $pK_{a,B}^{AH^+} = pK_{a,B}^A$, or equivalently that $\Delta pK_{a,A} = -\Delta pK_{a,B} = 0$, where $\Delta pK_{a,A} = pK_{a,A}^{BH} - pK_{a,A}^{B^-}$ and

$\Delta pK_{a,B} = pK_{a,B}^A - pK_{a,B}^{AH+}$ (for more complete discussion, see reference³⁹). Fitting of the pH-rate data under these constraints leads to apparent pK_a values for the general acid and base, and we will henceforth refer to this model as the “apparent pK_a model” to distinguish it from the “microscopic pK_a model” which allows this coupling parameter to be optimized. Alternatively, direct determination of the macroscopic pK_a values of the individual acid and base sites involves fitting to each of the corresponding independent acid and base fractions

$$f_{(AH^+)} = f_{(AH^+/BH)} + f_{(AH^+/B^-)} \quad (4.6)$$

$$f_{(B^-)} = f_{(A/B^-)} + f_{(AH^+/B^-)} \quad (4.7)$$

using the Hill equation

$$f_{(d)} = 1 / (1 + (10^{n(pK_a - pH)})) \quad (4.8)$$

where $f_{(d)}$ is the deprotonated fraction of the residue of interest (either $f_{(AH)}$ or $f_{(B^-)}$) and n is the Hill coefficient.

Molecular dynamics simulations at constant pH

The starting structures for the simulations were prepared by modifying the crystallographic structure of apo RNase A (PDB ID:1KF5)¹⁵¹ and RNase A complexed with deoxycytidyl-3',5'-deoxyadenosine [(d(CpA))] (PDB ID:1RPG)¹²⁷ solved at 1.15 Å and 1.4 Å resolution, respectively. Two separate sets of explicit solvent constant pH replica exchange molecular dynamics (CpHMD/pH-REMD)⁴ simulations were performed using developmental version of Amber 12¹⁵² molecular dynamics package on apo RNase A and the product state, 2',3'-cyclic phosphate (cCMP) bound RNase A. The apo and complexed RNase A structures were each placed in a cubic box of TIP4P-Ew¹²⁹ water molecules having a buffer of at least 10 Å on each side, and neutralized by chloride counterions for the expected charge of the systems (standard amino acid protonation states) at pH 7 with the exception of His12 and His119 treated as fully protonated. Note that the net number of explicit ions are constant in all the simulations, so the net charge of the system does vary as titratable residues change protonation state. This point was addressed specifically with regards to the titration process in Swails et al.⁴ and shown not to affect the titration since the sampling of protonation states occurs using an implicit solvent model based on the conformations generated with explicit solvent simulation. After initial minimization and equilibration with molecular dynamics, pH-REMD simulations in explicit solvent⁴ were performed at 300K and 1 bar using a total of 24 replicas corresponding to different pH values between 2 and 13. During these simulations, His12, Lys41 and His119 were treated as titratable. Long range electrostatics were computed using the smooth particle-mesh Ewald

(PME) method^{132,133}. Note that the PME method implicitly introduces a background term that neutralizes the system (sometimes referred to as a "neutralizing plasma") that is further corrected for finite size effects¹⁵³ with a volume-dependent term. Other corrections can be made to remove pressure and free energy artifacts for charged periodic systems¹⁵⁴, but in previous studies⁴ these were not found to be necessary and were not used here. Simulations were performed using a 2 fs integration step, with exchanges between adjacent replicas attempted every 200 fs, and carried out to 74 and 104 ns for each replica of apo and cCMP-bound RNase A. The 60 ns of production simulation was analyzed and reported for both systems. A full description of the simulation protocol, pH-activity curves, convergence tests on the sampled protonation states, and details about the side-chain conformations, hydrogen bonding networks in the active site and binding of cCMP at different pH values can be found in the supporting information.

4.2 Results and Discussion

Simulations accurately predict the macroscopic pK_a s for apo and cCMP-bound RNase A

Analysis of the pH-REMD simulations provides values for the fractions $f_{(AH^+/B^-)}$, $f_{(AH^+/BH)}$, $f_{(A/BH)}$, and $f_{(A/B^-)}$ described in eqs 4.2-4.5, and hence via eqs 4.6 and 4.7 the overall acid and base fractions, $f_{(AH^+)}$ and $f_{(B^-)}$. The titration curves for His12, His119 and Lys41 for both apo and cCMP-complexed RNase A can be found in Figure 4.2. Each fraction is observed to exhibit near-ideal Henderson-Hasselbalch behavior. Fitting the acid and base fractions to the Hill equation (eq 4.8), allows direct evaluation of the simulated macroscopic pK_a s. Comparison of the simulated and experimental pK_a values are listed in Table 5.3 for both apo and cCMP-complexed RNase A. In the case of apo RNase A, the simulated pK_a values for His12, His119 and Lys41 (5.95, 6.23 and 9.26) are within approximately 0.3 units of the corresponding experimental values^{155,156}. The simulations of RNase A complexed with cCMP predict shifted pK_a values for His12, His119 and Lys41 to 7.95, 7.17 and 9.65, respectively and are in reasonable agreement with the experimentally estimated values 8.0, 7.4 and 9.11, respectively^{157,156}. In the course of the simulations over the pH range 2-8, the integrity of the active site is maintained. Beyond pH 8, the interaction between Asp121 and His119 that form the His-Asp catalytic dyad begins to become displaced.¹⁵⁸ At high pH values, the cCMP becomes more loosely bound¹⁵⁹ in the binding pocket, leading to larger fluctuations and greater difficulty sampling. Details about the key structural features of the simulations at each pH are provided in the supporting information.

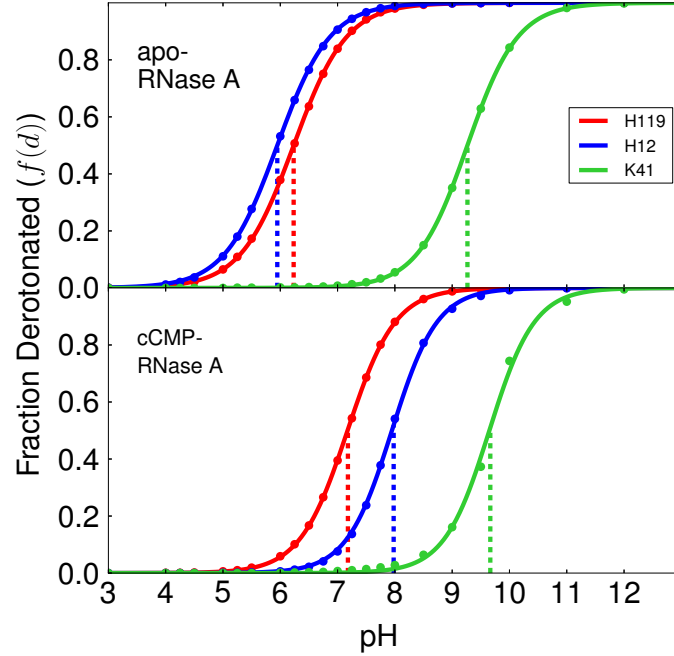


Figure 4.2: Titration curves (lines) fitted to simulation data (points) with the Hill equation [eq 4.8]. The upper panel represents the apo enzyme and the lower panel represents the cCMP-bound enzyme. The Hill coefficients for His119, His12 and Lys41 are 0.94, 0.94, 0.98 for the apo enzyme and 1.04, 1.09, 1.10 for the cCMP-bound enzyme, respectively.

Table 4.1: Comparison of experimental and simulated macroscopic pK_a s for apo and cCMP-RNase A using the Hill equation and the “apparent pK_a ” model.

	His12	His119	Lys41
<i>Apo RNase A</i>			
Expt.	5.8 ^a	6.2 ^a	9.0 ^b
Hill Eq.	5.95 (0.14)	6.23 (0.07)	9.27 (0.07)
Expt. (“Apparent” pK_a estimate) ^c	4.9-5.2	6.3-6.9	–
“Apparent” pK_a model with H12(B)/H119(A)	5.88 (0.14)	6.27 (0.08)	–
“Apparent” pK_a model with K41(B)/H119(A)	–	6.17 (0.21)	8.81 (0.21)
<i>cCMP-RNase A</i>			
Expt. (3'-CMP)	8.0 ^d	7.4 ^d	9.11 ^e
Hill Eq.	7.95 (0.15)	7.17 (0.12)	9.65 (0.16)
Expt. (cCMP “Apparent” pK_a) ^f	8.10	6.30	–
Expt. (cCMP “Apparent” pK_a) ^g	9.0	6.25	–
“Apparent” pK_a model with H119(B)/H12(A)	7.92 (0.16)	7.16 (0.12)	–
“Apparent” pK_a model with H119(B)/K41(A)	–	7.18 (0.13)	9.64 (0.16)

The calculated macroscopic pK_a values using the Hill equation [eq 4.8] and the “apparent pK_a ” model for the apo and cCMP-bound RNase A were compared with the experimental macroscopic and kinetic pK_a (Expt.) values from the literature. Statistical error estimates, shown in parentheses, were obtained as standard deviations derived from values computed from data in 5 ns intervals over the 60 ns of production.

^aTaken from ¹⁵⁵; ^bTaken from ¹⁵⁶; ^cRange of values estimated from values reported in ^{149,122};

^dTaken from ¹⁵⁷; ^eTaken from ¹⁵⁶; ^fTaken from ⁹²; ^gTaken from ¹⁴⁸;

3'-CMP, cCMP are the substrates: 3'-cytidine monophosphate, 2',3'-cyclic phosphate

Simulated protonation states can be interpreted using a microscopic pK_a model

The microscopic model illustrated in Figure 4.1.1 was fit to the simulated fractions of His12 as the general base/acid and His119 as the general acid/base for apo/cCMP-RNase A (Figure 4.3). The model, which has three independent parameters, fits the simulation data extremely well. This general trend is that the curves for the cCMP-complex shift to higher pH values relative to those for the apo enzyme. This trend is primarily due to the existence of the negatively charged cyclic phosphate on cCMP that is in fairly close proximity to the titratable residues in the active site. Further, the microscopic pK_a values derived from the fitting are in very good agreement with those measured from NMR experiments¹⁵⁸ for the apo enzyme (Table 4.2). For the cCMP-bound complex, to our knowledge, there currently does not exist experimental microscopic pK_a values that would allow a direct comparison. Nonetheless, there are microscopic pK_a values reported for a 3'-UMP inhibitor¹⁵⁸. The microscopic pK_a values for His12 and His119 in the 3'-UMP bound RNase A indicate a different trend than the experimental pK_a values for 3'-CMP and cCMP-bound RNase A complexes^{157,92,148} and the simulated values reported here (Table 5.3); nonetheless, we can compare the magnitude of the coupling (ΔpK_a values in Table 4.2) between protonation states.

Table 4.2: Comparison of experimental and simulated microscopic pK_a s using the microscopic model illustrated in Figure 4.1.1 for apo and cCMP-bound RNase A.

	$pK_{a,B}^{AH+}$	$pK_{a,B}^A$	$\Delta pK_{a,B}$	$pK_{a,A}^{BH}$	$pK_{a,A}^{B-}$	$\Delta pK_{a,A}$
<i>Apo-RNase A with His12(B)/His119(A)</i>						
Expt. ^a	5.87	6.18	0.31	6.03	6.34	-0.31
Microscopic model	5.88 (0.14)	6.05 (0.15)	0.17	6.13 (0.05)	6.30 (0.10)	-0.17
<i>Apo-RNase A with Lys41(B)/His119(A)</i>						
Microscopic model	8.29 (0.14)	9.26 (0.07)	0.97	6.23 (0.07)	7.21 (0.21)	-0.97
<i>cCMP-RNase A with His119(B)/His12(A)</i>						
Expt. (3'-UMP) ^a	7.95	7.85	-0.1	6.45	6.35	0.1
Microscopic model	7.20 (0.12)	6.99 (0.20)	-0.21	8.12 (0.17)	7.92 (0.16)	0.21
<i>cCMP-RNase A with His119(B)/Lys41(A)</i>						
Microscopic model	7.17 (0.12)	7.41 (0.40)	0.24	9.40 (0.18)	9.64 (0.16)	-0.24

The calculated microscopic pK_a values are derived from the thermodynamic cycle illustrated in Figure 4.1.1, which contains three free parameters that were fit to the acid/base fractions $f_{(AH^+/B^-)}$, $f_{(AH^+/BH)}$, $f_{(A/BH)}$, and $f_{(A/B^-)}$. The pK_a shifts are defined as $\Delta pK_{a,A} = pK_{a,A}^{BH} - pK_{a,A}^{B-}$, and $\Delta pK_{a,B} = pK_{a,B}^A - pK_{a,B}^{AH+}$. Note the constraint of the thermodynamic cycle in Figure 4.1.1 ensures $\Delta pK_{a,A} + \Delta pK_{a,B} = 0$, and a positive value for $\Delta pK_{a,B}$ indicates anticooperative coupling of protonation states (i.e., protonation of the acid site disfavors protonation of the base), whereas a negative value of $\Delta pK_{a,B}$ indicates cooperative coupling (i.e., protonation of the acid site *favours* protonation of the base). Statistical error estimates, shown in parentheses, were obtained as standard deviations derived from values computed from data in 5 ns intervals over the 60 ns of production. Results were compared with the experimental (Expt.) values obtained from NMR ^aTaken from Ref. 158 in order to validate the model. In order to explore the degree to which this weak coupling persists with other plausible general base and acid pairs, the general base was replaced by Lys41 for each case and the resulting microscopic pK_a values were determined.

The experimental microscopic pK_a values suggest that the His12 and His119 protonation states are weakly coupled in both the apo enzyme and the 3'-UMP bound RNase A ($\Delta pK_{a,B}$ values of 0.31 and -0.1 pK_a units, respectively). The calculated microscopic pK_a values are in reasonable agreement, and predict the apo and cCMP-bound RNase A have $\Delta pK_{a,B}$ values of 0.17 and -0.21

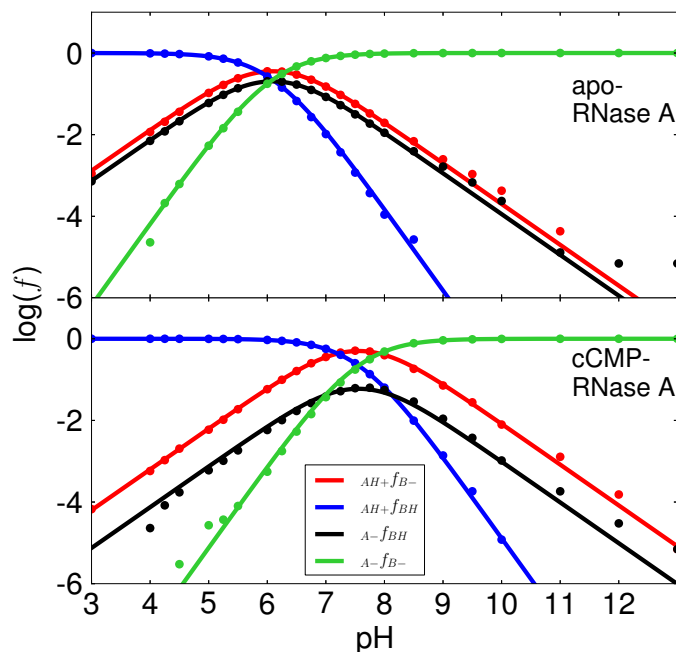


Figure 4.3: *Microscopic pK_a model results with H12/H119 acting as the general acid and base.* The plots of the logarithm of protonated fractions, $\log(f)$, versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data for all fractions to the equations derived from the microscopic model (Figure 4.1.1) with His12/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively, as depicted in Figure 4.1. The simulation data fits well with RMS errors of 0.22 (apo) and 0.17 (cCMP) for the $\log(f)$ values. The $\log(f)$ maximums (-0.45 and -0.30) for the curve of the active fraction, $f_{(AH^+/B^-)}$, are at 6.1 (apo) and at 7.5 (cCMP), respectively.

pK_a units, respectively. Overall, the experimental and calculated coupling between His12 and His119 protonation states is fairly weak. The observed weak coupling can be explained by the fact that the acid and base protonation sites are fairly far apart in the active site, and in the catalytically active state, involve interactions between a neutral and a positively charged residue (as opposed to oppositely charged residues). A fairly striking result is that the experimental $\Delta pK_{a,B}$ value for the 3'-UMP bound RNase A is negative, suggesting that protonation of one of the active-site histidine residues *favors* protonation of the other. This observation has been made experimentally¹⁵⁸ and has been explained as a result of enhanced interactions with the phosphate when both sites are protonated. The simulation results reported here are completely consistent with this interpretation (see SI for additional details), and as indicated in Table 4.2, suggest that only in the phosphate-bound systems does such cooperative coupling occur. The fact that the simulation results (both for

the apo and cCMP-complex enzymes) can be precisely fit to all four fractions by the 3-parameter microscopic kinetic model lends credence to its validity.

“Apparent” (uncorrelated) pK_a model for His12/His119 is justified for the catalytic steps in RNase A

The experimental analysis of pH-rate data involves fitting the observed rate curve to a simple kinetic model with apparent pK_a values for the presumed general acid and base. The correspondence of directly measured macroscopic pK_a values of presumed general acid and base residues with the apparent pK_a values is often interpreted as indirect evidence supportive of their catalytic roles. The computational analog of this procedure would be to fit only the active fraction, $f_{(AH^+/B^-)}$, to the thermodynamic cycle shown in Figure 4.1.1, but with only two free parameters, $pK_{a,A}$ and $pK_{a,B}$ (or alternatively, under constraints that $pK_{a,A}^{B-} = pK_{a,A}^{BH}$ and $pK_{a,B}^{AH+} = pK_{a,B}^A$). The fitted apparent pK_a curve for His12 and His119 as the general acid and base, respectively, is shown in Figure 4.4, and the values for the apparent pK_a s are shown in Table 5.3. Note that although the apparent pK_a values only consider the active fraction in the fitting, as would be the procedure used to fit experimental curves, these parameters also determine the fractions of the other protonation states in accord with the partition function in eq 4.1. Unlike experiment, these non-active fractions are available from the simulations, and thus available for comparison. The fit to the active fraction (red line in Figure 4.4) is excellent; however, model curves for the other fractions, particularly $f_{(A-/BH)}$, are somewhat worse than for the curves shown in Figure 4.3 that involved fitting the microscopic model to all of the fractions simultaneously. Nonetheless, the apparent pK_a values fitted to the active fraction are in remarkably close agreement (within 0.07 pK_a units) to the macroscopic pK_a values that were derived from fitting the protonation fractions to the Hill equation (Table 5.3). This is a direct consequence of the very weak coupling between the His12 and His119 protonation states. Under these conditions, the assumptions of independent protonation events and the interpretation of the apparent pK_a values used to gain insight into general acid-base mechanism is a valid one. Note, however, the present discussion assumes that the only protonation events that are affecting the catalytic activity are those of the general acid and base, and in many cases this may not be true.

Coupling of protonation states for His119 and Lys41 is significant

The above example of general acid-base catalysis in apo RNase A, with His12 and His119 as the presumed general acid and base, respectively, indicates that treatment of the protonation states as uncorrelated within the framework of the “apparent pK_a ” model is justified. However, it should not be assumed that this is a general phenomena. In order to demonstrate this point, we consider

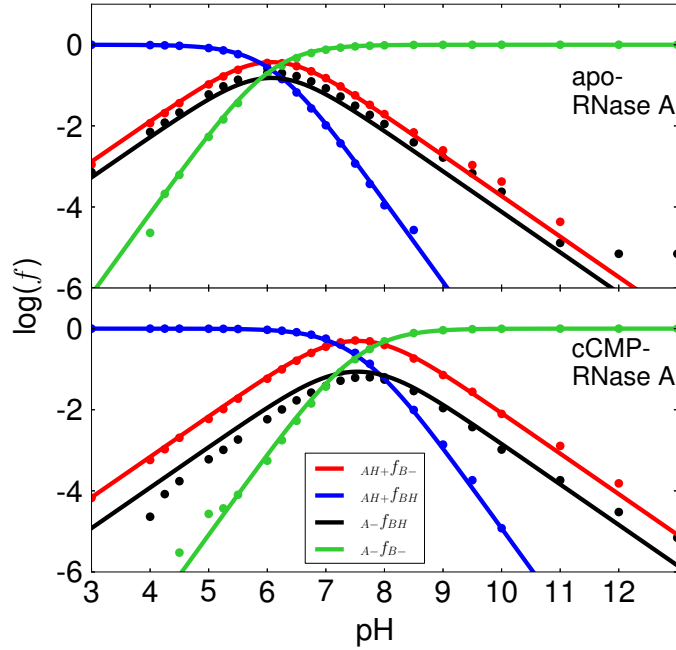


Figure 4.4: *Apparent pK_a model results with H12/H119 acting as the general acid and base.* The plots of the protonated fractions, $\log(f)$, versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data to ONLY the ACTIVE FRACTION ($f_{(AH^+/B^-)}$, shown in red) with His12/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively, as depicted in Figure 4.1. The model assumes no coupling of the microstates, and although only the active fraction is considered in the fitting (as would be the case experimentally), the parameters nonetheless are able to determine the other fractions which can be compared with the simulation data. The upper panel represents the apo RNase A and the lower panel represents the cCMP bound RNase A. The RMS errors for the $\log(f)$ values are 0.25 (apo) and 0.18 (cCMP) respectively. The $\log(f)$ maximums (-0.43 and -0.30) for the curve of the active fraction, $f_{(AH^+/B^-)}$, are at 6.1 (apo) and at 7.5 (cCMP), respectively.

the scenario whereby the role of general base is replaced by Lys41 in the kinetic model for the apo RNase A enzyme. Evidence suggests this is likely not the biological role of Lys41 in catalysis by RNase A, but for the purposes of demonstration, it is still instructive to examine.

The microscopic model is able to fit the simulation data very well for all fractions (Figure 4.5). However, considerably stronger protonation state coupling ($\Delta pK_{a,A} = \Delta pK_{a,B} = 0.97$) is observed between Lys41 and His119 relative to that of His12 and His119 (Table 4.2). The apparent pK_a values for Lys41 and His119 are 8.81 and 6.17, which differ somewhat from the macroscopic pK_a values obtained from the Hill equation (9.27 and 6.23, respectively). Moreover, the “apparent pK_a model” does not closely reproduce the protonation state fractions (Figure 4.6), including the active

fraction that was used in the fitting. Although the slopes of the higher and lower pH-regimes of the pH-rate curve is not influenced by the interactions between the residues, the maximum probability is observed to be approximately 3 times larger. However, the predicted pK_a from the microscopic model illustrates that the “apparent pK_a model” has limitations with regard to mechanistic interpretation in the regime where key protonation states are more strongly coupled. This may be particularly relevant for some RNA enzymes where the active form of the general acid and base are oppositely charged species that can be expected to exhibit stronger electrostatic interactions.

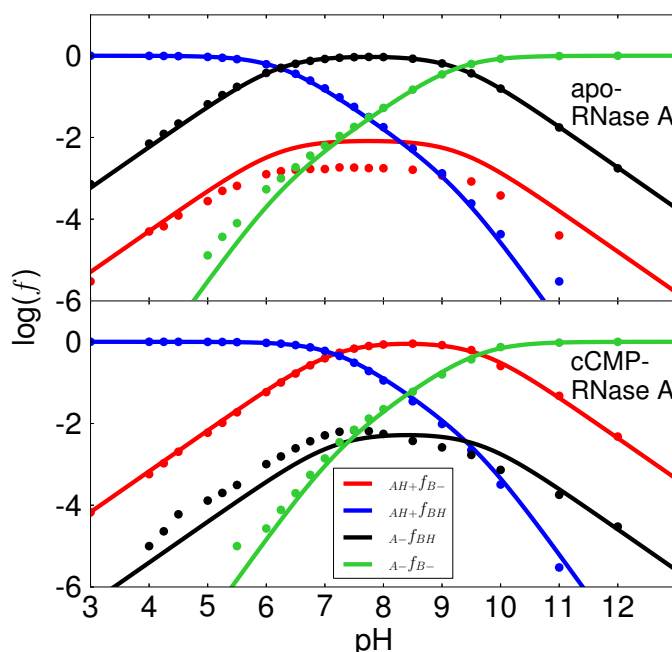


Figure 4.5: *Microscopic pK_a model results with K41/H119 acting as the general acid and base.* The plots of the logarithm of protonated fractions, $\log(f)$, versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data for all fractions to the equations derived from the microscopic model (Figure 4.1.1) with Lys41/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively. This is not the generally accepted mechanism depicted in Figure 4.1, but has nonetheless received some support in the literature^{2,3} and so is considered here for comparison. The RMS errors for the $\log(f)$ values are 0.31 (apo) and 0.44 (cCMP) respectively. The $\log(f)$ maximums (-2.09 and -0.05) for the curve of the active fraction, $f_{(AH^+/B^-)}$, are at 7.7 (apo) and at 8.4 (cCMP), respectively.

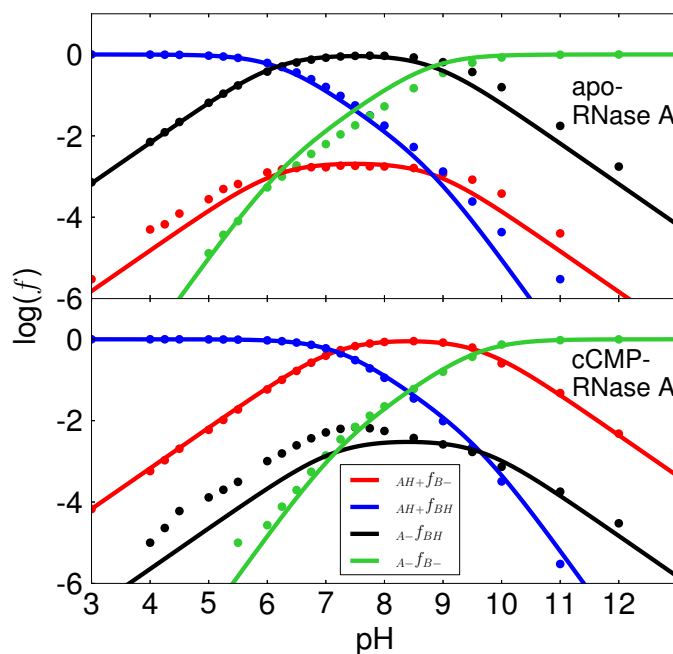


Figure 4.6: *Apparent pK_a model results with K41/H119 acting as the general acid and base.* The plots of logarithm of protonated fractions, $\log(f)$, versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data to ONLY the ACTIVE FRACTION ($f_{(AH^+/B^-)}$, shown in red) with Lys41/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively. This is not the generally accepted mechanism depicted in Figure 4.1, but has nonetheless received some support in the literature^{2,3} and so is considered here for comparison. The model assumes no coupling of the microstates, and although only the active fraction is considered in the fitting (as would be the case experimentally), the parameters nonetheless are able to determine the other fractions which can be compared with the simulation data. The upper panel represents the apo RNase A and the lower panel represents the cCMP bound RNase A. The RMS errors for the $\log(f)$ values are 0.33 (apo) and 0.48 (cCMP) respectively. The $\log(f)$ maximums (-2.69 and -0.11) for the curve of the active fraction, $f_{(AH^+/B^-)}$, are at 7.5 (apo) and at 8.4 (cCMP), respectively.

Summary and Perspective

Recently, advances in computational methods have allowed simulations of biological molecules to be performed in explicit solvent under constant pH conditions. These simulations allow conformations and protonation states to be sampled together across a range of pH conditions. This enables the prediction of pH-rate curves from molecular simulation, as well as tools to provide atomic-level interpretation of pH-activity data. More importantly, this method allows one to quantify the probability of finding an enzyme system in a catalytically active protonation state from which it is capable to go on to react with a pseudo first-order rate constant in the catalytic chemical step. Together with other methods, such as combined quantum mechanical/molecular mechanical simulations that can be used to map the free energy landscape for the catalytic chemical steps of the reaction, a complete description of catalysis can be obtained and compared directly with experimental data. The results described here demonstrate promise for theory and experiment to work together to understand enzyme mechanisms.

4.3 Supporting Information

4.3.1 Molecular Dynamics

System preparation

The crystal structure of apo RNase A, resolved at 1.15 Å, was obtained from the Protein Data Bank (PDB ID:1KF5)¹⁵¹. For atoms with double occupancy (including those in residues Glu9, Gln11, Ser15, Ser16, Thr17, Ser18, Ser23, Cys65, Cys72, Ser77, Arg85, Ser89, Gln101, Asn103, Lys104 and Ser123), only the highest occupancy was kept. The crystal structure of deoxycytidyl-3',5'-deoxyadenosine [(d(CpA))] complexed RNase A (PDB ID:1RPG)¹²⁷, resolved at 1.4 Å, was used to prepare the enzyme substrate bound complex of RNase A. The doubly occupied residues Leu35, Lys41, Asn67 and Gln69 were modified in the same way as discussed above. In the crystal structure, which was at low pH, His12 is clearly fully protonated, and forms strong hydrogen bonds with Thr45 and the non-bridge phosphoryl oxygen. Preliminary constant pH MD simulations revealed that this initial conformation, which is not catalytically relevant in transphosphorylation and supported only at low pH, had a deep minima that required very long equilibration times to overcome at higher pH. To create an initial starting point that would facilitate sampling of relevant high and low pH conformational states, His12 was rotated around the $C\beta-C\gamma$ bond to avoid the conformational trap triggered by the solution pH used in crystallization. The time scale of the simulations was sufficient to see transitions from the flipped to the non-flipped states.

The product state, 2',3'-cyclic phosphate (cCMP), was created by removing the leaving group

from the CpA structure and modeling the transphosphorylation reaction product. The charge parameters for cCMP were generated with RESP¹⁶⁰ in Antechamber¹⁶¹.

4.3.2 pH-REMD simulations

Representative snapshots from 100 ns long NVT MD simulations of apo and cCMP-bound RNase A in a cubic box of 20 Å clearance were chosen as starting structures. The solute was resolvated in a smaller cubic box of 10 Å clearance for both systems. In cCMP bound structure, the GB radii of the oxygens O3', O2', O1P, O2P and O2 (AMBER atom names) on the substrate of the product state were adjusted to 1.4 Å to be accessible to solvent. The systems were equilibrated in a NPT ensemble for 18 ns with a 2 fs time step by maintaining the temperature at 300K and pressure at 1 bar using the Langevin¹⁶² thermostat and the Berendsen¹⁶³ barostat, respectively.

24 replicas were chosen in the pH range of 2 to 13 with 1 pH-unit spacing except in the range of pH 4 to 9.5 with 0.25 pH-unit intervals (excluding pH=4.75 and pH=5.75 in order to match the available number of nodes to a multiple of number of replicas). The proton exchanges were attempted every 50 fs in all the simulations using a modified GB model for biomolecules as implemented in *sander*¹⁶⁴. The relaxation of the system after proton exchanges was performed in the NVT ensemble with all bonds containing hydrogens constrained using SHAKE^{165,131}. The constant temperature (300K) of the system was maintained by applying Langevin dynamics with a collision frequency of 2 ps⁻¹. A direct space cutoff for electrostatic interactions was set to 9 Å. Exchanges between adjacent replicas were attempted every 200 fs. simulations was performed for 74 ns for apo RNase A and 106 ns for cCMP-bound RNase A. The production simulations was chosen as last 60 ns to analyze the results.

4.3.3 pH-activity curves

Figures S1-S4 illustrate the plots of pH vs. probability generated using the simulated fractions of microstates $_{AH^+}E_{B^-}$, $_{AH^+}E_{BH}$, $_{AE_{BH}}$ and $_{AE_{B^-}}$ by fitting to Eq 1-7 in the main text which used in derivation of the parameters in the main text (Table 5.3, Table 4.2). Figure S1 shows the plots obtained for the acid/base pairs His119/His12 for apo RNase A and His12/His119 for cCMP-bound RNase A, using the microscopic model (Figure 4.1.1). The comparison of the same data is shown in Figure S2 for "apparent pK_a " model, where only two parameters were obtained from the red curve, which represents the active fraction, $f_{(AH^+/B^-)}$ and then the resulting two parameters were plugged back into the other fraction equations in order to compare the fitting of the curves with the microscopic model.

The plots shown in Figure S3 and Figure S4 are obtained for the acid/base pair His12/His119 for apo RNase A and Lys41/His119 for the cCMP-bound enzyme.

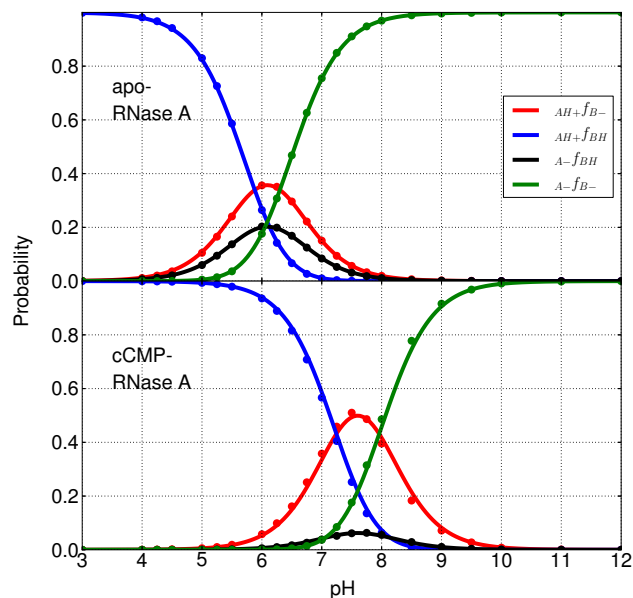


Figure 4.7: *Microscopic pK_a model results with H12/H119 acting as the general acid and base.* The plots of the probability versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data to the equations derived from the microscopic model with His12/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively. These plots were used in deriving microscopic pK_a parameters discussed in the main text (Table 5.3, Table 4.2).

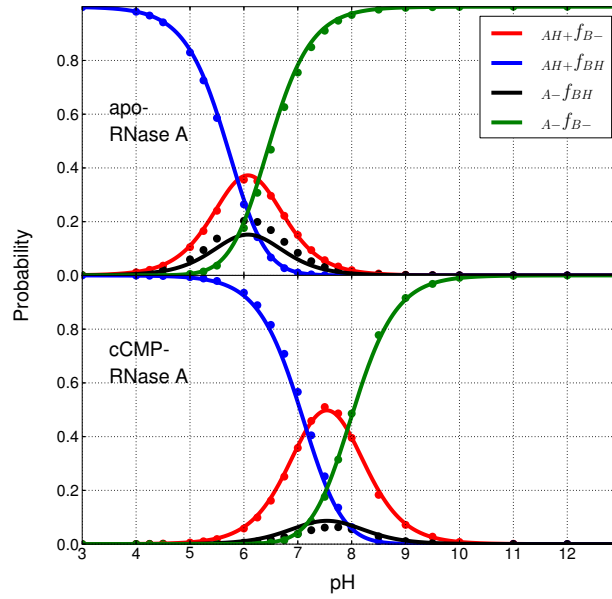


Figure 4.8: *Apparent pK_a model results with H12/H119 acting as the general acid and base.* The plots of the probability versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data to ONLY the ACTIVE FRACTION ($f_{(AH^+/B^-)}$, shown in red) with His12/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively. These plots were used in deriving apparent pK_a parameters discussed in the main text (Table 5.3, Table 4.2).

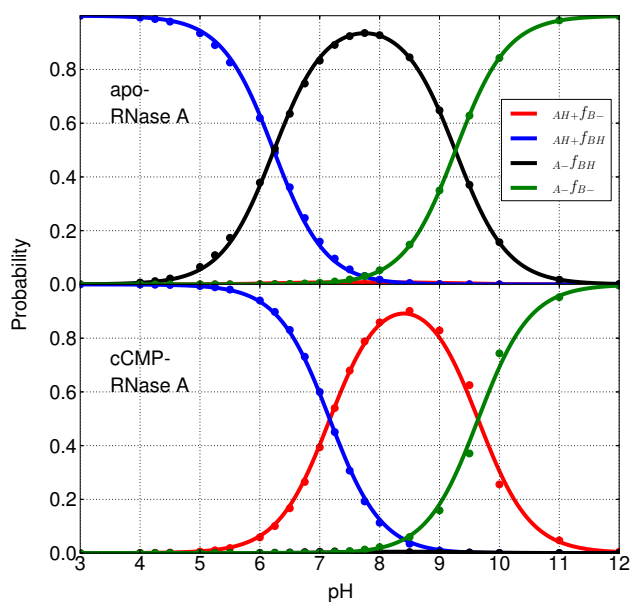


Figure 4.9: *Microscopic pK_a model results with K41/H119 acting as the general acid and base.* The plots of the probability versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data to the equations derived from the microscopic model with Lys41/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively. This is not the generally accepted mechanism. These plots were used in deriving microscopic pK_a parameters discussed in the main text (Table 5.3, Table 4.2).

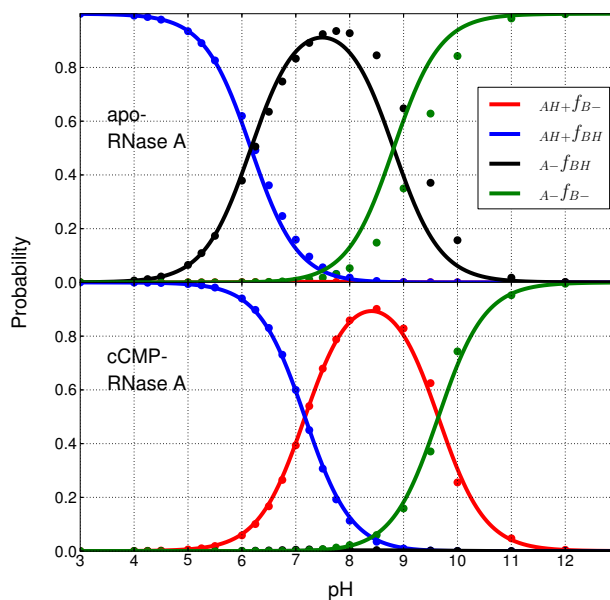


Figure 4.10: *Apparent pK_a model results with K41/H119 acting as the general acid and base.* The plots of logarithm of probability versus pH of each microstate for apo (top) and cCMP-bound RNase A (bottom) were obtained by fitting the simulation data to ONLY the ACTIVE FRACTION ($f_{(AH^+/B^-)}$, shown in red) with Lys41/His119 acting as the general base/acid for apo RNase A (transphosphorylation model) and the general acid/base for cCMP-bound RNase A (hydrolysis model), respectively. This is not the generally accepted mechanism. These plots were used in deriving apparent pK_a parameters discussed in the main text (Table 5.3, Table 4.2).

4.3.4 Stability of cCMP in the active site

The stability of the cCMP substrate in the RNase A active site is investigated by examining the average distance between His12/His119 and cCMP. Table 4.3 shows the occupancy of the H12:NE2-cCMP:O2' and H119:ND1-cCMP:O3' bonds and their average distances at the 5 Å cutoff. The results show a loosely bound substrate at pH > 10.

Table 4.3: The occupancy of the key contacts to the substrate (cCMP) at 5 Å cutoff along the pH ladder and the average distances. The numbers in parenthesis are standard deviations.

pH	H12:NE2-cCMP:O2'		H119:ND1-cCMP:O3'	
	Occupancy (%)	Avg. (Å)	Occupancy (%)	Avg. (Å)
2.00	98.1	3.4 (0.4)	88.4	3.6 (0.5)
3.00	98.2	3.4 (0.4)	88.1	3.6 (0.5)
4.00	98.0	3.4 (0.4)	87.7	3.6 (0.5)
4.25	98.0	3.4 (0.4)	87.9	3.6 (0.5)
4.50	98.2	3.4 (0.4)	88.2	3.6 (0.5)
5.00	98.2	3.4 (0.4)	87.8	3.6 (0.5)
5.25	97.9	3.4 (0.4)	87.6	3.6 (0.5)
5.50	97.7	3.4 (0.4)	86.7	3.6 (0.5)
6.00	97.4	3.4 (0.4)	84.6	3.6 (0.5)
6.25	96.5	3.4 (0.4)	82.1	3.6 (0.6)
6.50	95.7	3.4 (0.4)	78.3	3.6 (0.6)
6.75	93.6	3.4 (0.4)	71.4	3.6 (0.6)
7.00	90.0	3.4 (0.4)	63.3	3.6 (0.6)
7.25	83.6	3.4 (0.3)	53.5	3.7 (0.6)
7.50	74.6	3.4 (0.3)	43.6	3.8 (0.6)
7.75	61.4	3.4 (0.4)	32.8	3.8 (0.6)
8.00	46.7	3.4 (0.4)	24.4	3.9 (0.6)
8.50	23.5	3.4 (0.4)	14.2	4.1 (0.6)
9.00	13.1	3.5 (0.5)	10.0	4.2 (0.6)
9.50	9.1	3.7 (0.6)	8.1	4.2 (0.5)
10.00	6.8	3.9 (0.6)	6.6	4.3 (0.5)
11.00	4.8	4.1 (0.6)	5.1	4.3 (0.5)
12.00	5.0	4.1 (0.6)	5.1	4.3 (0.5)
13.00	5.0	4.1 (0.5)	5.3	4.3 (0.5)

4.3.5 Convergence of fractions of protonations

Figure S5 and S6 show the cumulative running average of the fraction of protonated residues as a function of time for apo and cCMP-bound RNase A, respectively. Figures S7 and S8 show the running averages (1 ns window size) of the fraction of protonated residues in apo and cCMP-bound RNase A, respectively.

4.3.6 Structural interpretation of the pK_a shifts of the active-site residues in cCMP-bound RNase A

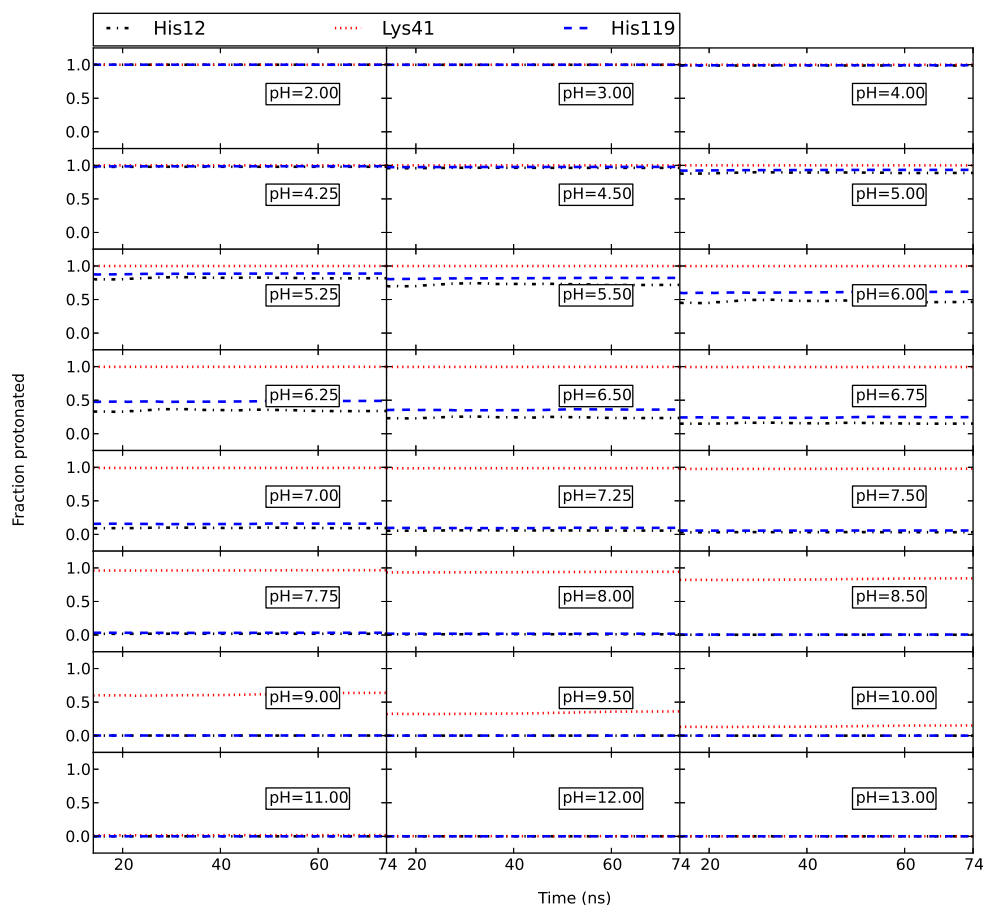


Figure 4.11: The cumulative averaged protonated fractions of His12, Lys41 and His119 at each pH for apo-RNase A

Table 4.4: The occupancy of the key contacts to the substrate at 5 Å cutoff along the pH ladder and the average distance

Bond	pH=7		pH=7.5		pH=8	
	Avg. (Å)	Occupancy (%)	Avg. (Å)	Occupancy (%)	Avg. (Å)	Occupancy (%)
cCMP:P-H12:NE2	3.8	89.6	3.8	74.0	3.8	45.7
cCMP:P-H119:ND1	3.9	61.7	3.9	41.2	4.0	21.1
cCMP:O1-H119:ND1	3.3	68.4	3.4	48.4	3.6	27.1
cCMP:O1-H119:NE2	4.7	44.1	4.7	35.5	4.6	22.8
cCMP:O2-H119:ND1	4.5	40.3	4.4	27.6	4.3	14.1
cCMP:O1-H12:ND1	4.5	68.2	4.5	57.1	4.5	35.2
cCMP:O1-H12:NE2	3.5	86.5	3.5	71.8	3.5	44.6
cCMP:O2-H12:NE2	4.0	83.3	4.1	67.7	4.2	40.9

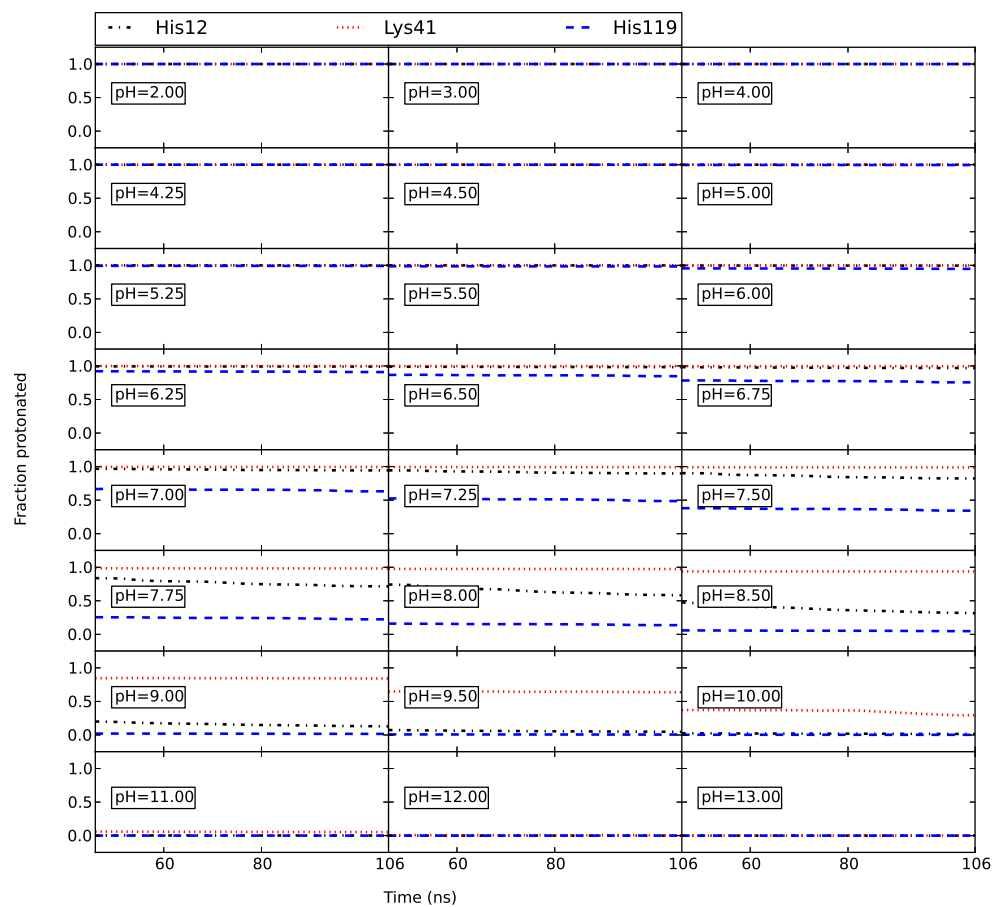


Figure 4.12: The cumulative averaged protonated fractions of His12, Lys41 and His119 at each pH for cCMP bound RNase A

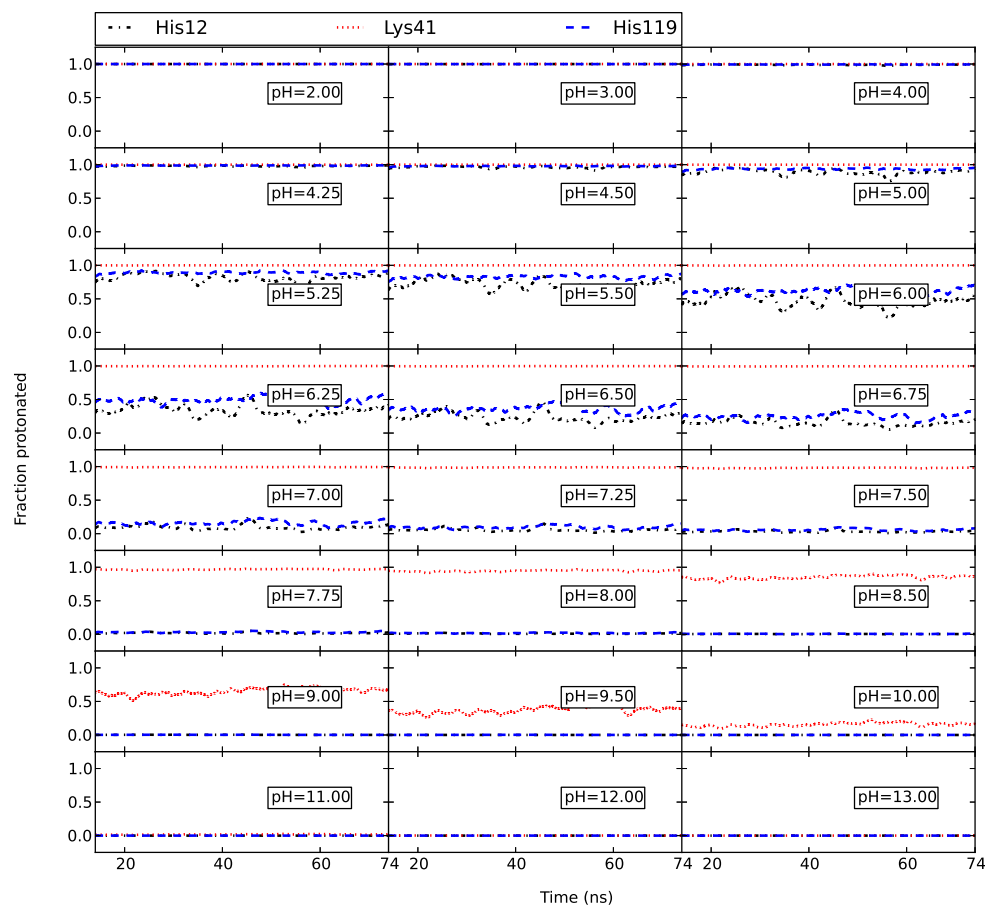


Figure 4.13: The running averages (1 ns window) of the fraction of protonated His12, Lys41 and His119 at each pH for apo RNase A

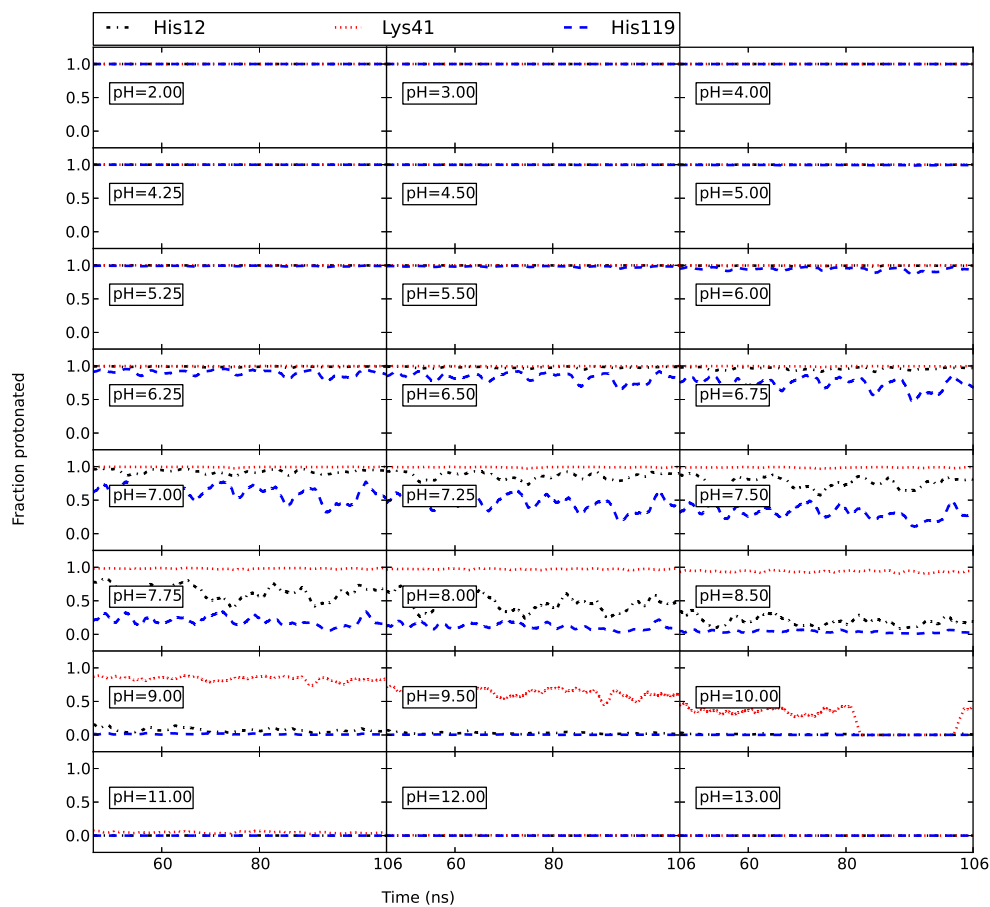


Figure 4.14: The running averages (1 ns window) of the fraction of protonated His12, Lys41 and His119 at each pH for cCMP-bound RNase A

4.3.7 The structural information of His119-Asp121 dyad

Table 4.5: The occupancy of H119:N(NE2/ND1)-D121:O(OD1/OD2) distance along the pH-ladder for apo RNase A at cutoff 5 Å

pH	D121:OD1-H119:NE2 occupancy (%)	D121:OD1-H119:ND1 occupancy (%)	D121:OD2-H119:NE2 occupancy (%)	D121:OD2-H119:ND1 occupancy (%)
2.00	61.0	23.8	57.5	18.0
3.00	61.6	24.2	58.1	18.5
4.00	60.7	23.9	57.3	18.5
4.25	60.4	23.8	57.1	18.4
4.50	60.2	24.2	56.9	18.5
5.00	57.7	22.8	54.3	17.7
5.25	55.9	22.4	53.3	18.2
5.50	52.4	21.4	49.9	17.2
6.00	43.0	18.4	41.4	16.1
6.25	35.8	15.4	34.6	14.3
6.50	30.0	14.0	29.4	13.0
6.75	23.6	11.0	22.9	11.3
7.00	19.4	9.9	19.1	9.9
7.25	15.7	8.7	15.5	9.1
7.50	14.7	8.4	14.1	8.6
7.75	12.9	7.5	12.7	8.1
8.00	12.9	7.6	12.7	8.6
8.50	12.1	7.3	11.9	7.8
9.00	12.2	7.2	12.2	7.8
9.50	12.3	7.1	12.7	7.7
10.00	12.9	7.3	13.1	7.5
11.00	12.5	7.1	13.0	7.1
12.00	12.7	7.2	13.3	7.2
13.00	12.6	7.2	13.0	7.3

Table 4.6: The occupancy of H119:N(NE2/ND1)-D121:O(OD1/OD2) distance along the pH-ladder for cCMP-bound RNase A at cutoff 5 Å

pH	D121:OD1-H119:NE2 occupancy (%)	D121:OD1-H119:ND1 occupancy (%)	D121:OD2-H119:NE2 occupancy (%)	D121:OD2-H119:ND1 occupancy (%)
2.00	30.9	12.3	29.2	7.6
3.00	30.7	12.2	28.9	7.5
4.00	30.6	12.3	29.4	7.7
4.25	31.4	12.5	29.4	7.6
4.50	30.9	12.6	29.4	7.6
5.00	30.7	12.3	29.3	7.8
5.25	31.2	12.4	29.4	7.9
5.50	30.1	11.9	28.8	7.7
6.00	29.5	11.9	27.8	7.3
6.25	28.6	11.4	27.1	7.0
6.50	27.3	11.0	26.0	6.9
6.75	24.8	9.7	23.5	6.4
7.00	22.2	8.4	21.7	6.2
7.25	20.0	7.3	19.7	6.0
7.50	17.5	5.6	18.0	6.1
7.75	15.6	4.6	17.0	6.5
8.00	15.7	4.7	17.9	7.2
8.50	16.0	4.8	19.5	8.9
9.00	16.4	5.3	20.0	9.2
9.50	16.4	5.5	20.3	9.5
10.00	16.9	5.7	20.2	9.0
11.00	17.0	6.0	20.2	9.2
12.00	17.2	5.9	20.6	9.2
13.00	16.8	6.1	20.1	8.9

Chapter 5

Catalytic mechanism of RNase A, 2: Prediction of reaction path of Ribonuclease A transphosphorylation from QM/MM simulations

Bovine Pancreatic Ribonuclease A (RNase A) catalyzes the O2'-transphosphorylation and hydrolysis of nucleotides, and is an important benchmark system to study phosphoryl transfer reactions. Several mechanisms for RNase A has been proposed and debated in the literature, with support coming from both theory and experiment. The most commonly accepted mechanism involves a participation of His12 and His119 as the general base and the general acid. Nonetheless, a detailed description of the mechanism and transition state that is consistent with a wide range of experimental measurements has yet to emerge. Here, we explore the energetics of the reaction path using combined quantum mechanical/molecular mechanical (QM/MM) free energy simulations, in order to elucidate the rate enhance of the catalysis and understand the involvement of active site residues and explore the different mechanistic pathways. The two identified mechanistic pathways are the general acid base mechanism with dianionic transition state and triester-like transition state. In this work, a dianionic transition state is observed when the flipped His12 or Lys41 acting as the general base, for a QM-region defined with all three active site residues and the dinucleotide, CpA. The calculated barrier for the reaction, 16.25 kcal/mol, reproduces a value within the range of barriers 13-19 kcal/mol, predicted from the experimental studies for different substrates. A QM region with a partially occupied substrate and the three active site residues predict a triester-like transition state of 26.47 kcal/mol barrier, when His12 is in non-flipped conformation. Therefore, a triester-like mechanism is un-likely. The preliminary results concludes the flipped His12 act as the general base while proceeding through a dianionic transition state. Further, the role of Lys41 is found to assists the activation of the nucleophile and provides electrostatic stabilization of the transition state. Results are consistent with available pH-rate profiles, kinetic isotope effects studies.

5.1 Introduction

Ribonuclease A has been subjected to extensive studies in the literature for fast few decades^{166,1,115,38,1,167,91}. The catalytic efficiency of the enzyme is identified with 10^{12} of rate enhancement comparative to the uncatalyzed *in vitro* reaction^{168,1,169}. Numerous amount of studies have been performed to identify the efficiency of the enzyme and the mechanistic details of the catalysis. After initial discovery of the mechanism by Mathias and Rabin^{105,92,93,94,95,96}, several mechanisms were proposed using both experimental and theoretical studies^{118,98,99,2,170,122,125}. Nonetheless, the rate enhancement and the details of the mechanism of ribonuclease A remains under active discussions. RNA degradation by Ribonuclease A occurs in two steps: (1) Transphosphorylation, where the general base of the enzyme accepts the proton from O2' to facilitate the nucleophilic attack on the phosphate group to form P-O2' bond, followed by donation of a proton from the general acid to O5' to complete the cleavage of P-O5' bond to form 2',3'-cyclic phosphate product. His12 and His119 are widely proposed for the role of the general base and the general acid, respectively. (2) Hydrolysis step, where ring opening of 2',3'-cyclic phosphate to form a nucleotide, with the support of His12 and His119 residues acting as the general acid and the general base. The general mechanism is depicted in the Figure 3.1.

There are two types of proposed mechanisms: concerted mechanism and step-wise mechanism. Both general acid and base steps occur simultaneously in the concerted type mechanism, whereas, the general acid step of the stepwise mechanism occur sequentially after the general base step. Variations the stepwise mechanism is proposed experimentally and theoretically^{99,170,122}. One of recent theoretical study shows a stepwise mechanism with a triester-like intermediate is most likely, in comparison to the concerted mechanism¹²⁵. Another type of step-wise mechanism was proposed by Gu et al. which involve dianionic transition state¹²².

In this work, QM/MM simulations, a multiscale computational method, is applied to identify the detailed mechanism of Ribonuclease A catalysis. As preliminary study, 2D-free energy surfaces for general acid vs phosphoryl transfer and the general base vs. phosphoryl transfer reaction coordinates were obtained by applying umbrella sampling technique. The reaction paths and the reaction barriers were derived from 2D free energy surfaces. The conformational effect of His12 side chain and the size effect of the QM region was explored.

5.2 Materials and Method

5.2.1 Molecular dynamics simulations

The simulations were initiated from two transition state mimics bound Ribonuclease A, where the side chain conformation of His12 is either flipped or non-flipped relative to the starting geometry of the inactive state bound crystal structure (PDB ID: 1RPG)¹²⁷. The non-flipped state of His12 side chain represents the conformation from the crystal structure, while the flipped state represents a 180° of rotation of the His12 side chain around the C β -C γ bond. The active site residues: His12, His119 and Lys41, are fully protonated and positively charged, while the total charge of the transition state mimic substrate is -2. The systems were neutralized by adding 5 of Cl⁻ ions, along with 22 of Na⁺ and 22 of Cl⁻ ions to achieve 140 mM of physiological concentration of NaCl. The systems were pre-equilibrated in a cubic box of TIP4P-Ew solvent environment of \sim 8500 water molecules for 4 ns in NPT conditions with periodic boundaries using AMBER ff14SB³⁷ force field. The temperature and pressure were regulated using Langevin thermostat¹⁷¹ and Berendsen barostat¹⁶³. Long range electrostatics were treated with Particle Mesh Ewald (PME) method¹³².

5.2.2 QM/MM simulations

The QM/MM simulations were initiated from the pre-equilibrated systems described above. The AM1/d-PhoT semi-empirical Hamiltonian⁶³ was used for all the simulations. Three QM regions were simulated in RNase A system: (1) A full QM region, with 102 atoms, include His12, His119, Lys41 and the full substrate dinucleotide (CpA) (DNU) to determine the detail mechanism of RNase A and conformational dependence of His12 residue. (2) A small QM region with 53 QM atoms include His12, His119, Lys41 and truncated substrate (RDNU) to investigate the effect of the number of atoms in the QM region. (3) A QM region with 88 QM atoms include His12, His119 and full substrate to explore the role of Lys41 residue (NoK41). Finally, the uncatalyzed reaction (Uncat) was explored using a system of solvated dinucleotide. The systems were minimized and equilibrated with NPT conditions for 50 ps. The production simulations were performed in NVT conditions for 100 ps with 1fs time step, per each of 528 windows. The summary of simulations, corresponding abbreviations and details of QM regions are listed in the Table 5.1 and QM regions are illustrated in the Figure 5.1.

5.2.3 Umbrella sampling

Umbrella sampling technique is used to compute the potential of mean force in order to predict the free energy profiles of the reaction paths, by applying a harmonic biasing potential with a 100

Table 5.1: The summary of QM/MM simulations and the QM regions

Model/Abbreviation	His12 side chain conformation	Description of the QM region
DNU		
DNU-TS-F _{Acid}	F	CpA, H12/F, K41, H119
DNU-TS-F _{Base}	F	CpA, H12/F, K41, H119
DNU-TS-NF _{Acid}	NF	CpA, H12/NF, K41, H119
DNU-TS-NF _{Base}	NF	CpA, H12/NF, K41, H119
RDNU		
RDNU-TS-NF _{Acid}	NF	Truncated CpA, H12/NF, K41, H119
RDNU-TS-NF _{Base}	NF	Truncated CpA, H12/NF, K41, H119
NoK41		
NoK41-TS-F _{Acid}	F	CpA, H12/F, H119
NoK41-TS-F _{Base}	F	CpA, H12/F, H119
Uncat		
Uncat	NA	CpA

Abbreviations: CpA- dinucleotide substrate, DNU- fully occupied dinucleotide and three active site residues in the QM region. RDNU- partially occupied substrate, the region below C1' and C3' of cytosine and above the C5' of adenosine, with three active site residues in the QM region. NoK41- CpA is occupied while Lys41 is unoccupied in the QM region. Uncat- dinucleotide in non-enzymatic environment, only dinucleotide presents in the QM region.

kcal mol⁻¹ Å² of force constant along the reaction path. The difference between the distances of breaking bond and the forming bond in the phosphoryl transfer and the proton transfer reactions were considered as the generalized coordinates as shown in the Figure 5.2. The figure indicates: (1) Phosphoryl transfer vs. general base, where phosphoryl transfer reaction coordinates is defined by R4-R3 and general base reaction coordinate is defined by R1-R2. (2) Phosphoryl transfer vs. general acid, where the general acid reaction coordinate is defined by R6-R5. Here, R1 = O2'-OH2', R2= OH2'-H12:ND1, R3 = O2'-P, R4= P-O5', R5 = H119:HD1-O5', R6= H119:ND1-H119:HD1. The biased data obtained from umbrella sampling was unbiased using the vFEP¹⁷² method to determine the free energies.

5.2.4 Kinetic isotope effects calculations

For both mechanisms, the reactant (RT) and rate-limiting transition state (TS) structures were obtained from the 2D QM/MM free energy profiles followed by 120 ps simulated annealing which finally cooled the system to 0 K. Generalized distance restraints on the two reaction coordinates were present in the simulated annealing of TS but not RS. Finite difference calculations of all atoms in the QM region were then performed based on the annealed RT and TS structures to generate the subsystem Hessian matrices of the QM-region. Kinetic isotope effects were then computed via Bigeleisen equation¹⁷³ using the vibrational frequencies generated from the diagonalization of mass-weighted Hessian matrices.

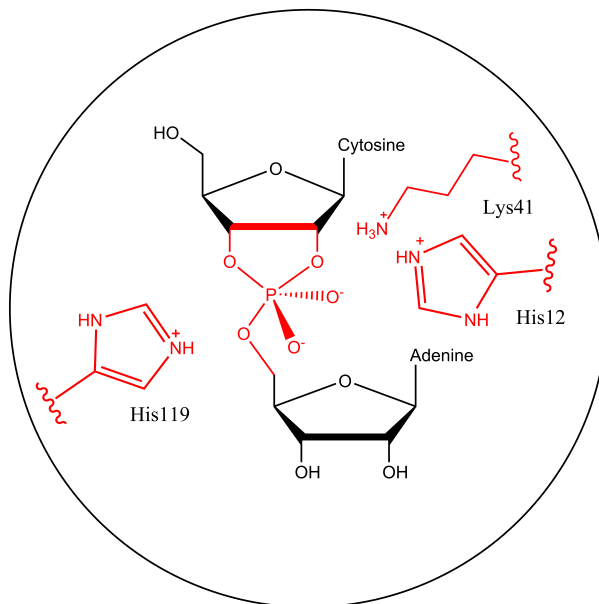


Figure 5.1: The schematic of QM regions indicates both DNU and RDNU regions. The black circle represent the full QM region (DNU), where the whole dinucleotide and the active site residues His12, Lys41 His19 are included in the QM region. The red color represent the small QM region (RDNU) where the truncated form of the dinucleotide and the active site residues are included in the QM region.

5.3 Results and Discussion

The preliminary results obtained from QM/MM simulations is presented in this section. A detailed mechanism of RNase A catalysis, the roles of active site residues and the contribution of His12 side chain conformations are explored. The structural effect of the reaction barrier and the mechanism is addressed by changing the sizes of the QM region, and exploring the uncatalyzed reaction.

5.3.1 Determination of initial structure for QMMM simulations

QM/MM simulations are time consuming and computationally costly, therefore, a reasonable starting structure is necessary for initiation of a QM/MM simulation. The crystal structures can be only found in the form of an inhibitor bound complex. An inactive substrate (d[CpA]) bound crystal structure¹²⁷ is available for RNase A, however, it is only a partially representative of an “active state”. Therefore, the effort we put forward to identify a reasonable “active state” in the chapter 3, suggested a flipped His12 acting as the general base in the “active state”. In this state (Figure 5.3), N ϵ of His12 locates at 2.02 Å away from the OH2' with 159° angle of O2'-OH2'-NE:His12 and well poised to abstract the proton, as oppose to the orientation of non-flipped His12 found in the crystal

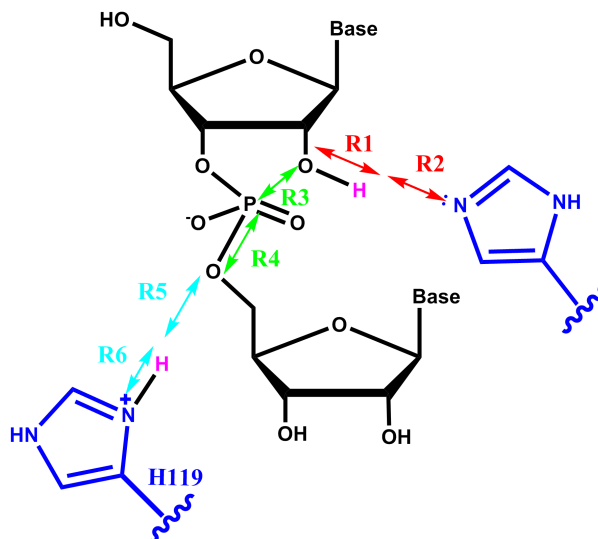


Figure 5.2: The reaction coordinates used in this study is represented in the figure. The protein residues (His12, His119) are colored in blue and the substrate is colored in black. His12 and His119 residues considered as the general base and the acid, respectively. The generalized coordinates are defined as the difference between the breaking bond and the forming bond; R1-R2, R4-R3 and R6-R5 where, R1 = O2'-OH2', R2= OH2'-H12:ND1, R3 = O2'-P, R4= P-O5', R5 = H119:HD1-O5', R6= H119:ND1-H119:HD1.

structure. However, we modeled transition state mimic for both flipped and non-flipped conformations of His12 for the comparison. The QM/MM simulations were initiated from the starting structures shown in the Figure 5.4. The transition state mimic at left, represents the His12 flipped state. The distances, O2'-H12:HE2, K41:NZ-O2' and H119:HD1-O5' are 2.45 Å, 2.62 Å and 1.85 Å, respectively. The transition state mimic at right, represents the His12 non-flipped state and the distances are, 3.46 Å, 2.74 Å, 1.76 Å respectively.

5.3.2 Obtaining the reaction paths

Departing from the transition state mimics (Figure 5.4), the QMMM simulations were performed along 2D generalized reaction coordinates. Biased umbrella potentials were applied in the simulation and later unbiased with the vFEP method to obtain 2D-free energy profiles for the systems listed in the Table 5.1. The minimum paths for each free energy profiles were obtained using Nudge Elastic Band (NEB) method.

5.3.3 Impact of the selection of the QM region

The size effect of the QM region was explored by including or excluding either active site residues or sections of the substrate as described in the Figure 5.1 and the Table 5.1. The reaction barriers,

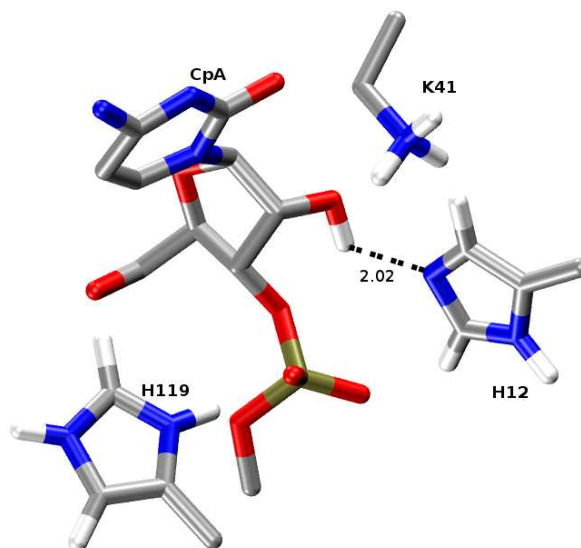


Figure 5.3: The average structure of the reactant state model from 400 ns long simulation (“active state”), where His12 has its flipped conformation. Ne of His12 is located at 2.02 Å away from the OH2’ and poised to abstract the proton with 159° of average hydrogen bond angle of O2’-OH2’-NE:His12

structure and the reaction path is compared with the results of full QM region.

5.3.4 Exploring the uncatalyzed reaction

Simulation of the uncatalyzed system was performed on the CpA substrate in a solvent environment. Assuming a major contribution of phosphoryl transfer step in the reaction mechanism, the reaction barrier was calculated from the 1D free energy profile for R4-R3 ($R_{P-O5'}-R_{P-O2'}$) using AM1d-PhoT Hamiltonian. The results provide a validation of application of the AM1d-PhoT Hamiltonian in phosphoryl transfer reactions and the associated error of the barrier, in comparison to the experiments.

5.3.5 Modeling uncatalyzed reaction

The predicted free energy for uncatalyzed reaction is slightly lower than the experiment

This section focuses on the preliminary results of the QM/MM simulation of the uncatalyzed reaction. The reaction barrier and the structures along the reaction path are discussed for 800 ps of production simulation. The experimental rate constant of substrate UpA for the uncatalyzed reaction, $k_{uncat} = 0.06 \text{ s}^{-1}$, is converted to 19.9 kcal/mol¹⁷⁴. In this work, the QM/MM simulations were performed on the CpA substrate, along the phosphoryl transfer reaction coordinate ($R_{P-O5'}-R_{P-O2'}$) to obtain the 1D free energy profile as shown in Figure 5.5. The reaction path indicates

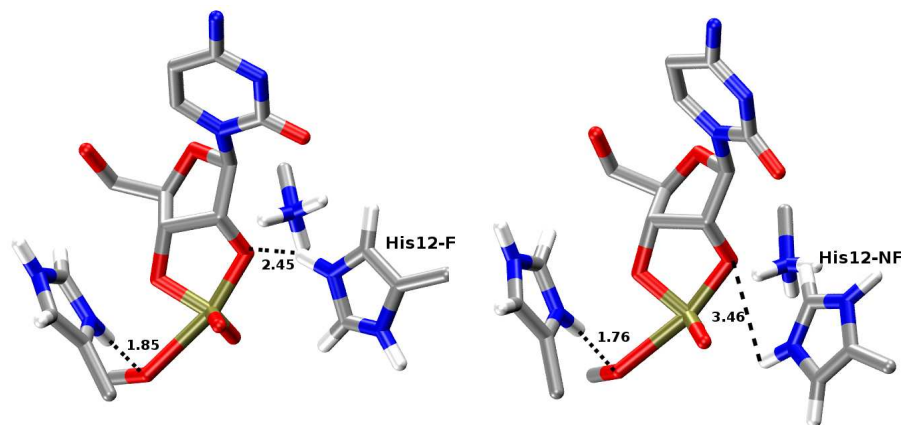


Figure 5.4: The departing structures for the QM/MM simulations. Left: the transition state mimic where His12 is in its flipped state, right: the transition state where His12 is in its non-flipped state.

a minimum and a transition state, with 20.80 kcal/mol of reaction barrier. The geometry of the minimum imply an activated precursor state with a negatively charged O2', as shown in the Figure 5.6. The equilibrium distance of 3.14 Å of P-O2' is smaller in comparison to the 3.89 Å of distance found in the crystal structure of the dinucleotide¹⁷⁵, favoring a nucleophilic attack on the phosphate group. Furthermore, slightly longer P-O5' bond (1.64 Å), in comparison to the crystal structure (1.58 Å) also supports a cleavage of the PO5'. In the transition state, a significantly shorter P-O2' bond (1.76 Å) and an elongated P-O5' (2.08 Å) reveal a late-like nature of the transition state. Our calculated barrier between the activated precursor state and the transition state, 20.80 kcal/mol, agree with the experimentally observed value between the reactant state and the transition state.

5.3.6 Reaction paths starting from His12 non-flipped conformation

The reaction paths obtained from the transition states containing non-flipped conformation of His12 are presented in this section. Both free energy profiles for phosphoryl transfer vs. general acid and general base coordinates are discussed for the full and small QM regions.

Phosphoryl transfer vs. general acid free energy profile indicates a minima with Lys41 deprotonated and a late-like transition state.

The free energy information for phosphoryl transfer vs. general acid was obtained from a 100 ps of production simulations. The profiles indicates two minima and a transition state as shown in the Figure 5.7. The protonated state of O2'H in the first minima suggests a nature of a reactant state with a deprotonated Lys41 at 1.01 Å distance from O2' to act as the general base. The protonated His12 is less likely to act as the general base. Lys41 could act as the general base upon a pK_a

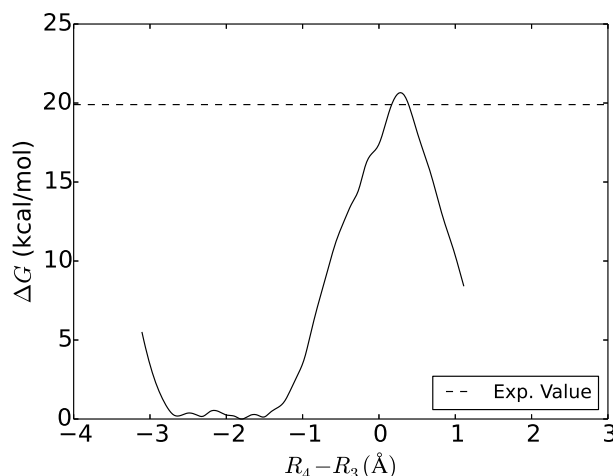


Figure 5.5: 1D PMF of uncatalyzed transphosphorylation reaction of CpA. A minimum and a transition state is observed along the reaction path with a 20.80 kcal/mol of reaction barrier. $R_4=R_{P-O5'}$ and $R_3=R_{P-O2'}$.

downshift toward the neutrality, costing approximately 3 kcal/mol in energy to lower two pK_a units from initial pK_a , 9^{156} to 7. According to our knowledge, pK_a of Lys41 in wild type or substrate bound RNase A has never been reported closer to the neutrality. Therefore, Lys41 is less likely to be the general base. A late-like transition nature of the transition state is found with 2.18 Å of elongated P-O5' bond and 1.77 Å of P-O2' bond. Although the barrier calculated from the first minimum is 15.13 kcal/mol, the value is not comparable with the experimentally observed reaction barrier from a ground state.

Phosphoryl transfer vs. general base free energy profile indicates a concerted like transition state

The free energy profile for phosphoryl transfer vs. general base of His12 non-flipped state, shown in the Figure 5.8, gives three minima and two transition states. The first minimum includes a deprotonated His12 at Nε nitrogen and the protonated O2', displaying a nature of a reactant state. The second minimum shows similar characteristics to the first minimum structure observed in the phosphoryl transfer vs. general acid step. The second transition state locates along the phosphoryl transfer reaction coordinates with 21.68 kcal/mol of overall reaction barrier indicating a concerted type transition state having 1.85 and 1.85 Å of P-O2' and P-O5' bond distances, respectively. The results predict a higher reaction barrier in comparison to the experiments.

The free energy profile from small QM region yields a triester-like transition state

In triester-like mechanism, a proton from His12 transfer to the non-bridging phosphoryl oxygen prior to the nucleophilic attack. The mechanism has been frequently suggested in experimental and

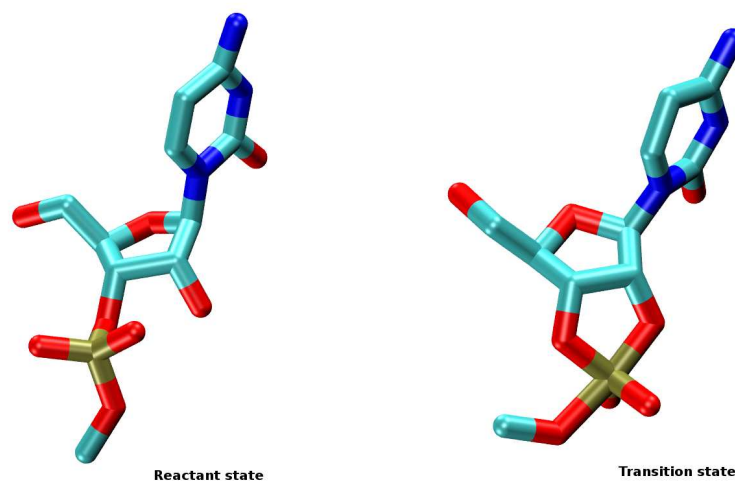


Figure 5.6: The structure of the minimum and the transition state for the uncatalyzed reaction. The minimum indicates a structure of a negatively charged activated precursor state and the transition state shows a negatively charged, late-like transition state.

theoretical studies for RNase A catalysis^{99,176,2,177,178,179}. The Figure 5.9 and Figure 5.10 represent the free energy profiles for both phosphoryl transfer vs. general acid and the base with 15.81 kcal/mol and 26.47 kcal/mol of reaction barriers, respectively. A lower barrier is monitored along the phosphoryl transfer vs. general acid step between a reactant state where Lys41 is deprotonated, and the triester-like transition state. Although, both free energy profiles predict a unique transition state, the nature of the predicted reactant states are different. Therefore, assuming a reactant state with His12 deprotonated is relevant in catalysis, the reaction path predicted from phosphoryl transfer vs. general base is most likely. The results shows the path is higher in energy, hence tri-ester like mechanism is unlikely.

5.3.7 Generating reaction paths for RNase A with His12 flipped conformation

A flipped conformation of His12 in the “active state” was proposed in our molecular dynamics simulation studies (chapter 3). This section further investigate the reaction path, barrier and the nature of the transition state, when His12 is flipped in the reactant state.

2D free energy profile of phosphoryl transfer vs. general base indicates an early transition state when flipped His12 is occupied in the reactant state

Further investigation of the phosphoryl transfer vs. general base was studied when His12 is in flipped conformation. The free energy profile for 25 ps long simulation results is depicted in the Figure 5.11. The profile indicates three minima and two transition states. The first minimum shows

Table 5.2: The energies of each state along the reaction path for the simulated systems

Simulation	MIN1 (kcal/mol)	TS1 (kcal/mol)	MIN2 (kcal/mol)	TS2 (kcal/mol)	MIN3 (kcal/mol)	Barrier (kcal/mol)
His12 Not-flipped						
DNU-TS-NF _{Acid}	0.00	15.13	-11.40	na	na	15.13
DNU-TS-NF _{Base}	0.00	15.52	8.82	21.68	5.46	21.68
RDNU-TS-NF _{Acid}	0.00	15.81	-19.54	na	na	15.81
RDNU-TS-NF _{Base}	0.00	26.47	25.52	na	na	26.47
His12 flipped						
DNU-TS-F _{Acid}	0.00	16.25	-12.24	na	na	16.25
DNU-TS-F _{Base}	0.00	12.48	9.44	21.91	1.84	21.91
NoK41						
NoK41-TS-F _{Acid}	0.00	21.47	-6.23			21.47
NoK41-TS-F _{Base}	0.00	24.66	4.47			24.66

The calculated free energy values for each stages along the reaction path for His12 flipped and Not-flipped systems are listed in the table. All the values were obtained from 100 ps long production simulation. The abbreviation DNU stands for the full QM region of all active site residues His12, Lys41 and His119 and CpA dinucleotide. RDNU represents reduced number of atoms of the CpA dinucleotide together with the active site residues.

a nature of a reactant with a protonated O2'. The P and O2' atoms are apart with 3.50 Å and His119 interacts with O5' having 2.06 Å distance. The second high energy minimum indicates a proton transfer from Lys41 to O2' suggesting a less likely state along the reaction path. The second transition state is recognized as an early-like transition state with 21.91 kcal/mol of overall reaction barrier. This result also suggest, the reaction is unlikely to proceed through a high energy early-like transition state initiated from a deprotonated Lys41. state.

2D free energy profile of phosphoryl transfer vs. general acid indicates a late transition state when His12 is in its flipped state with fully occupied CpA substrate in the QM region

The free energy profile indicates (Figure 5.12) two minima and one transition state. The first minimum shows a nature of a reactant state with a proton on O2' and well-position towards His12. The location of O2' also reasonably closer for nucleophilic attack with 3.28 Å of distance to the phosphate group. His119 residues also positioned towards O5' in this state. The transition state with 16.25 kcal/mol barrier shows a nature of a late-like transition state having 1.80 Å of P-O2' and 1.99 Å of P-O5' bond distances, respectively. The free energy profiles indicate solely occurrence of phosphoryl transfer with extremely subtle change along the general acid coordinate until reach the point of the transition state. His119 is well positioned towards O5' at this stage and ready for the proton transfer process having 1.85 Å of H119:HD1-O5' distance. Along the general acid coordinate, P-O2', H119:ND1-H bonds are formed and P-O5' bond is broken to form the 2',3'-cyclic phosphate

Table 5.3: The average distances at each stages for each system of simulation.

Simulation	O2'-H (Å)	P-O2' (Å)	P-O5' (Å)	H-O5' (Å)	H-OP1 (Å)	H-OP2 (Å)
Uncatalyzed						
MIN1	-	3.14	1.64	-	-	-
TS1	-	1.76	2.08	-	-	-
DNU-TS-NF_{Acid}						
MIN1	1.01	3.25	1.65	2.05	3.57	4.71
TS1	-	1.77	2.18	1.68	1.88	3.33
MIN2	-	1.68	3.29	1.00	1.89	3.72
DNU-TS-NF_{Base}						
MIN1	1.00	3.26	1.65	2.22	3.43	4.20
TS1	1.01	3.70	1.65	2.49	1.01	3.45
MIN2	1.01	3.07	1.66	1.93	2.52	3.32
TS2	-	1.87	1.85	1.85	1.83	3.30
MIN3	-	1.68	3.28	1.00	1.95	3.66
RDNU-TS-NF_{Acid}						
MIN1	0.99	2.86	1.66	2.02	1.90	3.19
TS1	-	1.87	1.73	1.86	1.15	3.08
MIN2	-	1.67	3.47	1.00	1.85	3.48
RDNU-TS-NF_{Base}						
MIN1	0.99	3.06	1.64	2.25	2.11	3.56
TS1	-	1.87	1.72	2.04	1.01	3.20
MIN2	-	1.77	1.77	1.92	1.00	3.13
DNU-TS-F_{Acid}						
MIN1	1.00	3.28	1.67	2.03	3.58	4.49
TS1	2.23	1.80	1.99	1.85	3.69	3.97
MIN2	2.16	1.68	3.28	1.00	3.74	4.14
DNU-TS-F_{Base}						
MIN1	1.00	3.50	1.67	2.06	4.26	4.87
TS1	1.39	3.64	1.65	2.19	4.10	4.22
MIN2	1.00	3.44	1.65	2.03	4.78	5.71
TS2	2.00	1.91	1.88	1.88	3.46	3.80
MIN3	2.05	1.69	3.28	1.00	3.82	3.76
NoK41-TS-F_{Acid}						
MIN1	1.85	3.85	1.66	2.04	4.35	4.95
TS1	2.32	1.80	2.22	1.69	3.53	3.87
MIN2	3.51	1.68	3.28	1.00	4.40	4.17
NoK41-TS-F_{Base}						
MIN1	1.87	3.86	1.66	2.18	4.13	4.63
TS1	1.89	1.89	1.86	1.89	3.15	3.82
MIN2	3.05	1.68	3.29	1.00	3.66	3.95

product state.

The role of Lys41 is identified as stabilization of the negative charge and reduction of the reaction barrier

The free energy profiles were obtained by removing Lys41 and including full dinucleotide, CpA in the QM region while His12 is left with flipped conformation in the active site. The barrier was predicted for both phosphoryl transfer vs. general acid and general base steps. The phosphoryl transfer vs. general acid step predicts 21.47 kcal/mol of reaction barrier for a late-like transition state while phosphoryl transfer vs. general base predicts a transition state with 24.66 kcal/mol of reaction barrier. Both barriers are elevated in the absence of Lys41 in the QM region, suggesting, Lys41 is essential for barrier reduction by stabilizing the negative charge on the transition state

Table 5.4: Comparison of calculated Kinetic Isotope Effect values with experimental measurements

His12	QM Region	QM Hamiltonian	$^{18}k_{LG}$	$^{18}k_{NUC}$	$^{18}k_{NPO}$
F	RDNU	AM1/d-PhoT	1.0179	0.9972	1.0017
F	RDNU	B3LYP	1.0122	0.9984	1.0008
F	DNU	AM1/d-PhoT	1.0282	0.9990	1.0057
NF	RDNU	AM1/d-PhoT	1.0083	1.0146	0.9991
NF	DNU	AM1/d-PhoT	1.0393	1.0004	1.0001
Expt. ¹²²			1.014(3)	0.994(2)	1.001(1)

Kinetic isotope effect (KIE) studies suggest a late-like transition state

The experimental KIE studies has identified a late-like transition state with 1.014, 0.994 and 1.001 values of $^{18}k_{LG}$, $^{18}k_{NUC}$ and $^{18}k_{NPO}$, respectively. The calculated values for the His12 flipped state closely agree with the experimental values as oppose to the values predicted for His12 non-flipped conformation (Table 5.4)

5.4 Conclusions

Hybrid QM/MM simulations are successfully applied to identify the details of catalyzed mechanism by RNase A. Our results imply a step-wise mechanism, which involve a late transition state with 16.25 kcal/mol energy barrier for “active state” with flipped His12 conformation. The nature of the transition state is consistent with the experimentally identified transition state using kinetic isotope effect studies. A triester-like transition state is observed with a non-flipped His12 in the “active state”, having high energy barrier and concluding the mechanism is unlikely. Moreover, the role of the Lys41 proposed to stabilize the negative charged transition state and reduce the reaction barrier.

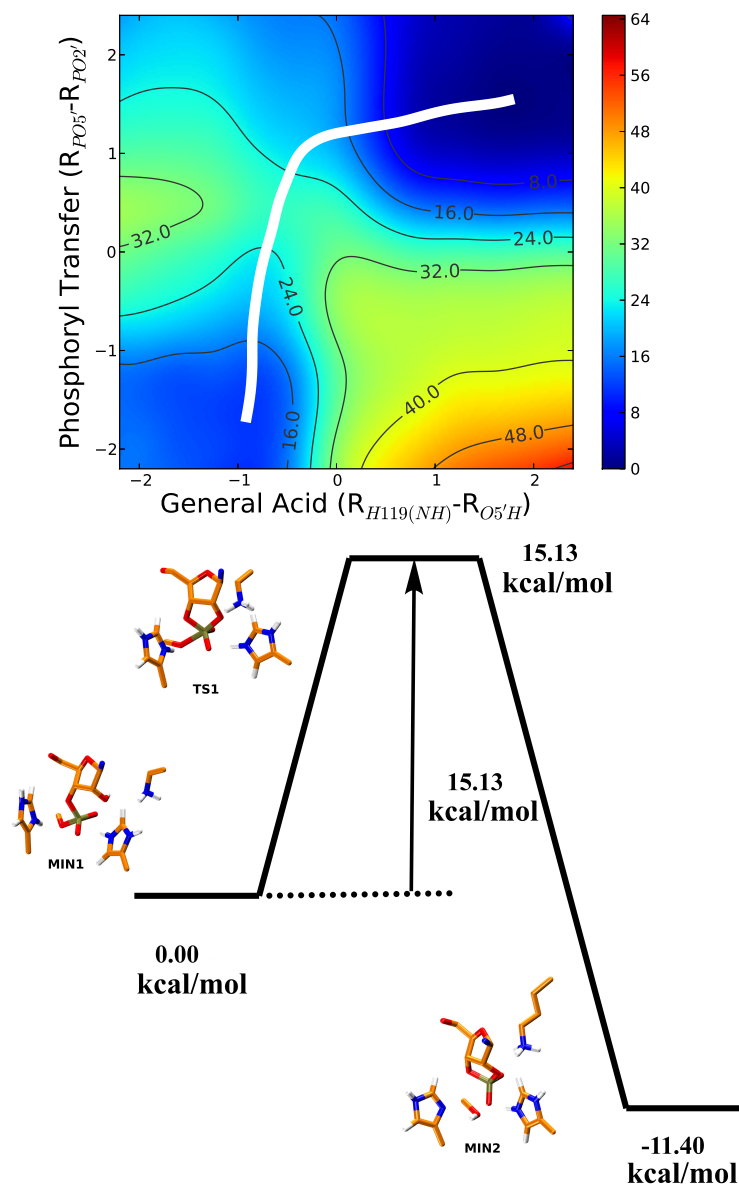


Figure 5.7: 2D free energy profile of phosphoryl transfer vs. general acid in DNU-TS-NF_{Acid} state where QM region is fully occupied with the substrate CpA and active site residues His12, Lys41 and His119. Two minimum and one dianionic transition state is observed. The barrier of the reaction is 15.13 kcal/mol. The first minima indicates deprotonated Lys41 acting as the general base.

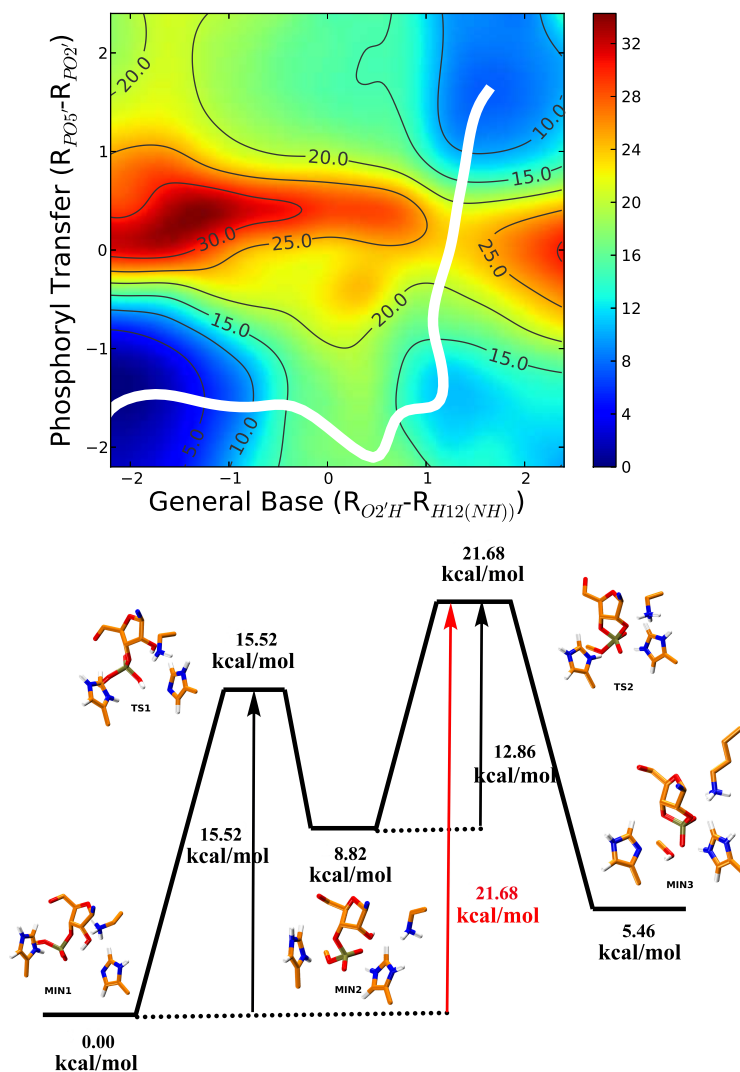


Figure 5.8: 2D free energy profile of phosphoryl transfer vs. general base for DNU-TS-NF_{Base} state where CpA substrate, His12, Lys41 and His119 are fully occupied in the QM region. Three minimum and two transition states are observed with 21.68 kcal/mol of overall reaction barrier from a reactant state of His12 acting as the general base.

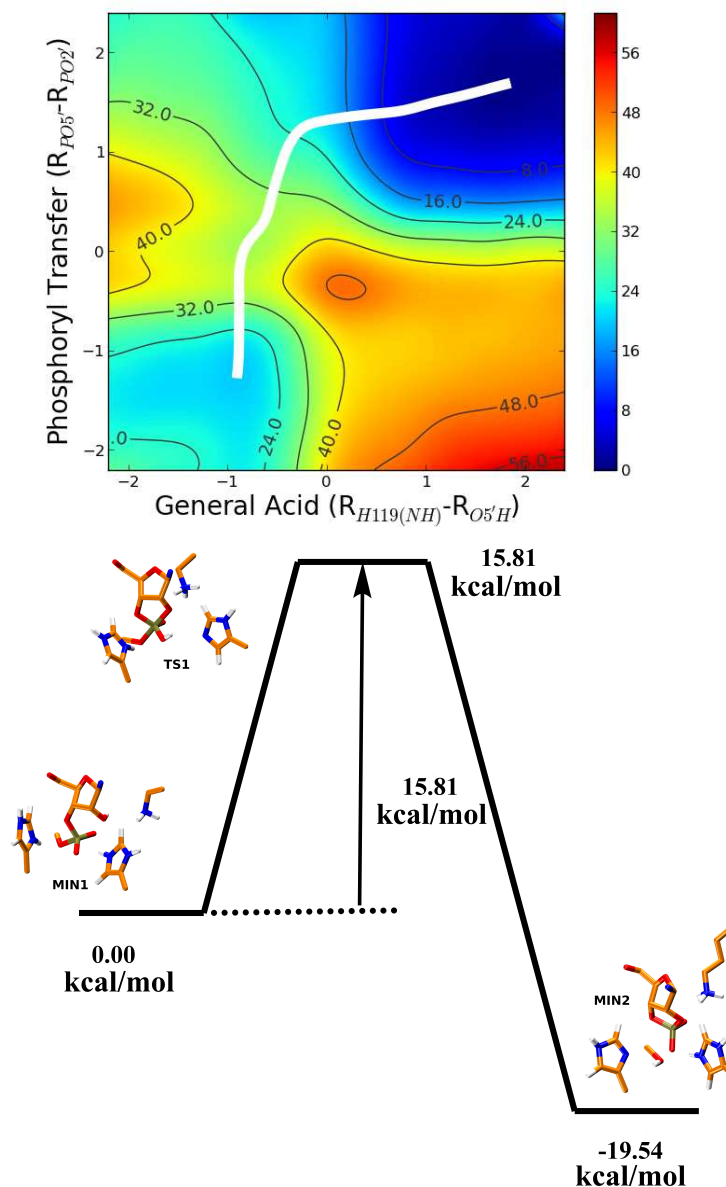


Figure 5.9: The 2D free energy profile for phosphoryl transfer vs. general acid in RDNU-TS-NF_{Acid} state where the QM region is fully occupied with His12, Lys41 and His119 and partially occupied with CpA substrate. Two minimums and a transition state are observed. The reactant state indicates Lys41 acting as the general base with 15.81 kcal/mol reaction barrier to the triester-like transition state.

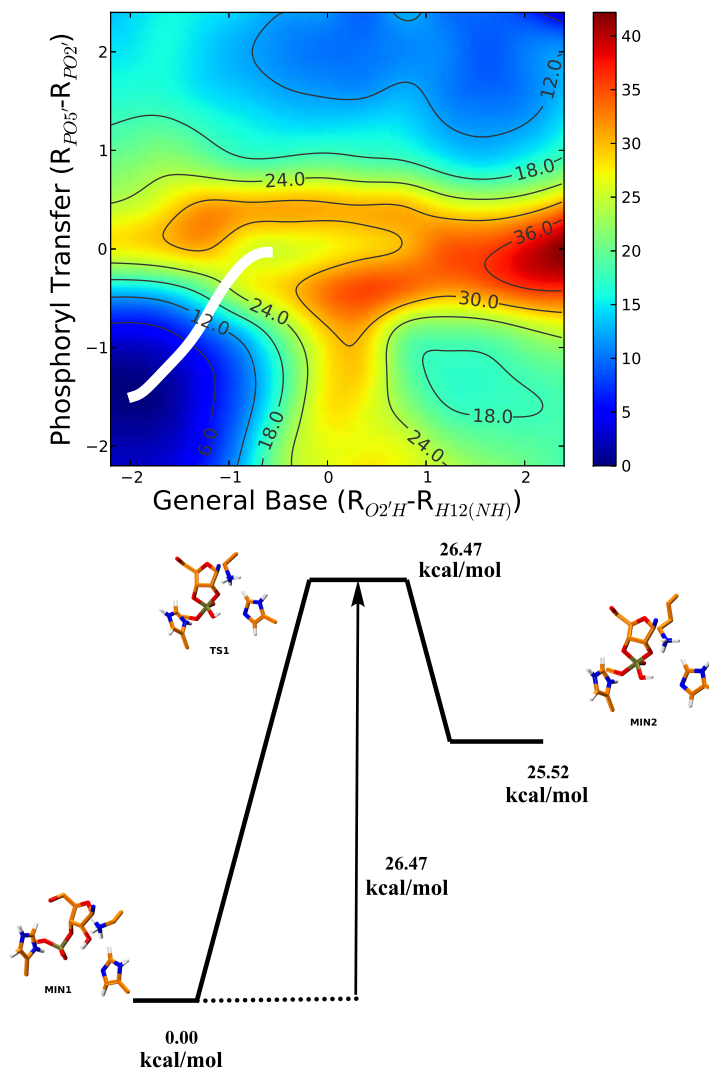


Figure 5.10: The 2D free energy profile for the phosphoryl transfer vs. general base step in RDNU-TS-NF_{Base} state with the truncate QM region. A triester-like transition state is observed along the reaction path with 26.47 kcal/mol reaction barrier.

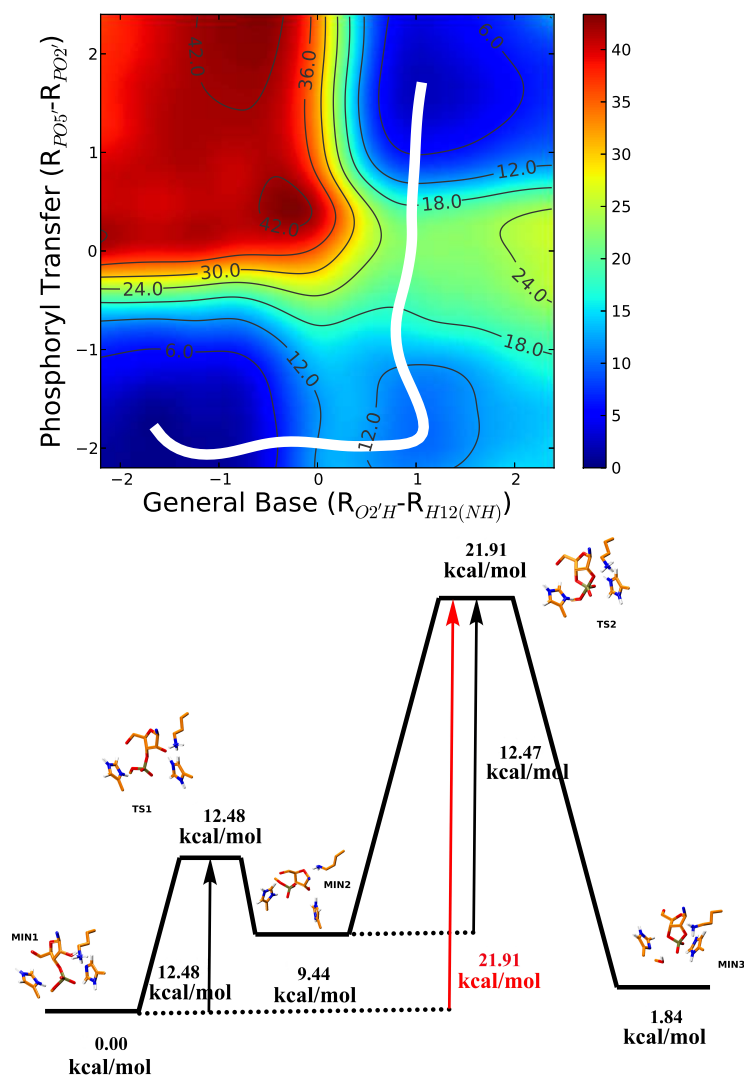


Figure 5.11: 2D free energy profile for the phosphoryl transfer vs. general base step when flipped His12, Lys41, His119 and CpA are occupied in the QM region (DNU-TS- F_{Base}). 21.91 kcal/mol of high barrier early transition state is observed.

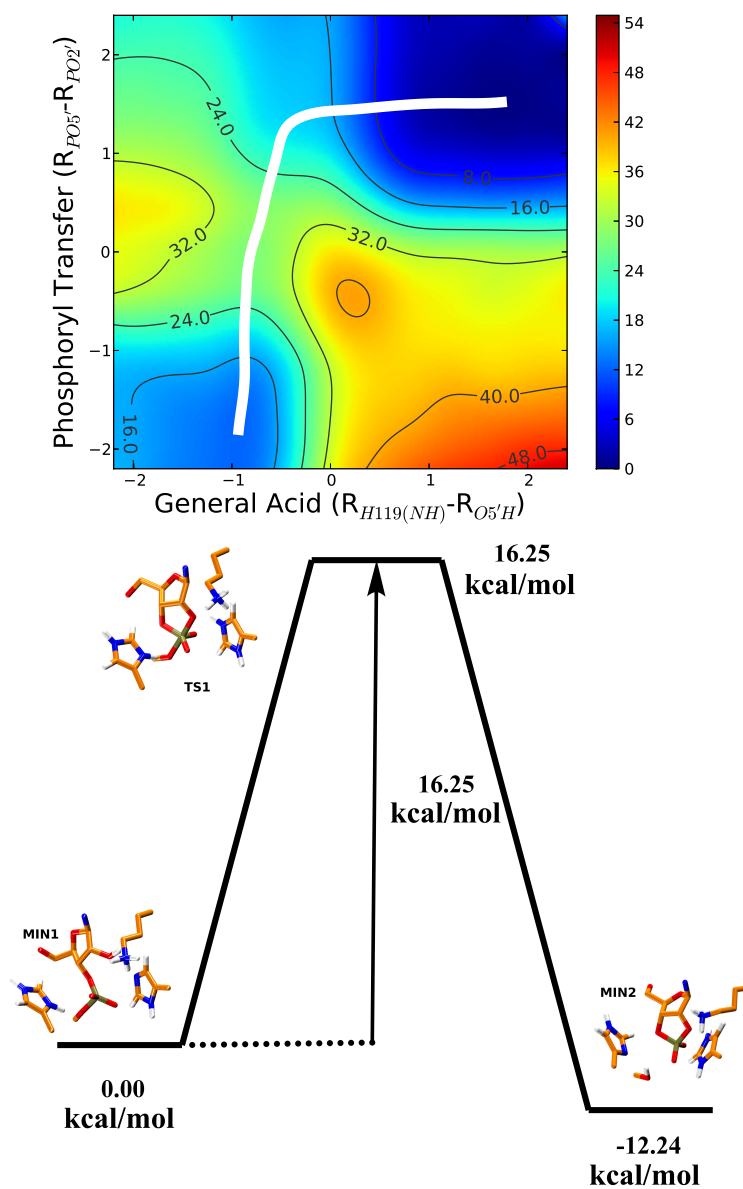


Figure 5.12: The 2D free energy profile for phosphoryl transfer vs. general acid step when flipped His12 and CpA substrate fully occupied in the QM region (DNU-TS- F_{Acid}). The reaction barrier is 16.25 kcal/mol from the reactant state to the dianionic transition state.

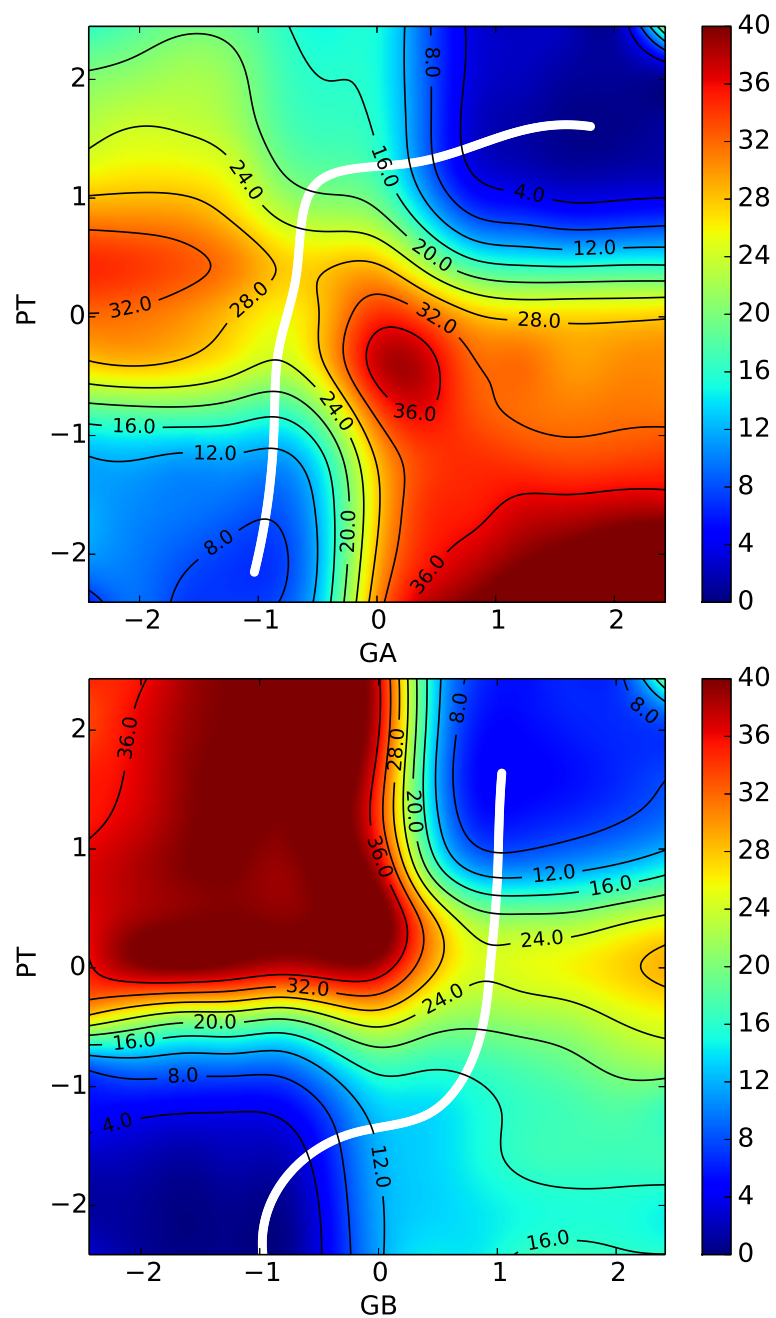


Figure 5.13: 2D free energy profile for the phosphoryl transfer vs. acid step (upper panel) and the phosphoryl transfer vs. base step (lower panel) when His12 is in its flipped state and Lys41 is absent in the QM region.

Chapter 6

Assessment of ionization of nucleic acids using constant-pH explicit solvent simulations

6.1 Introduction

Increasing evidence suggests, the ionization of key building blocks of RNA is critical for efficient RNA catalysis. These residues shift their pK_a s towards the neutrality in efficient RNA catalysis. The ionization of N1 site of adenosine and guanosine occur at pH 3.5 and 9.2, respectively, and N3 site of cytosine and uridine occur at pH 4.2 and 9.2, respectively^{180,181,182} as shown in the Figure 6.1. The pK_a values of these residues are located far from the neutrality, suggesting adenosine and cytosine are fully protonated, guanosine and uridine are deprotonated at physiological pH. However, the structural variations of RNA allow ionization equilibria of the nucleobases to shift towards a desired value in order to play acid or base role in enzyme catalysis^{183,184}. Therefore, Adenosine and cytidine show pK_a upshift while guanosine and uridine show pK_a s downshift. For example, pK_a downshift of G40 is implicated in *glms* ribozyme catalysis¹⁸⁵, pK_a upshift is implicated in C75 of hepatitis delta virus (HDVR)^{186,187}. Further, pK_a shifts of adenine also found in hairpin ribozyme¹⁸⁸. The effect of the shifted pK_a s on the dynamics of RNA structure is important to address in exploring the ribozyme catalysis. In particular, conformational changes upon protonation states take a lead in this topic. The pK_a shifts also indicated in smaller RNA systems consist of canonical and non-canonical base pairs. Non-canonical base pairs are good example to explore structural changes upon the shifted pK_a s^{189,190,191} including a study of A+.C base pair for the pH dependent conformational changes and pK_a shifts¹⁹². Here in this work, we obtained a set of parameters for nucleotides, the smallest units of the RNA, as reference compounds. The results were applied on the simple set of dinucleotides in order to assess the parameters. The parameters were further tested by applying on the systems consist of A+.C wobble pairs which are known to shifts their pK_a upon charged nucleotides.

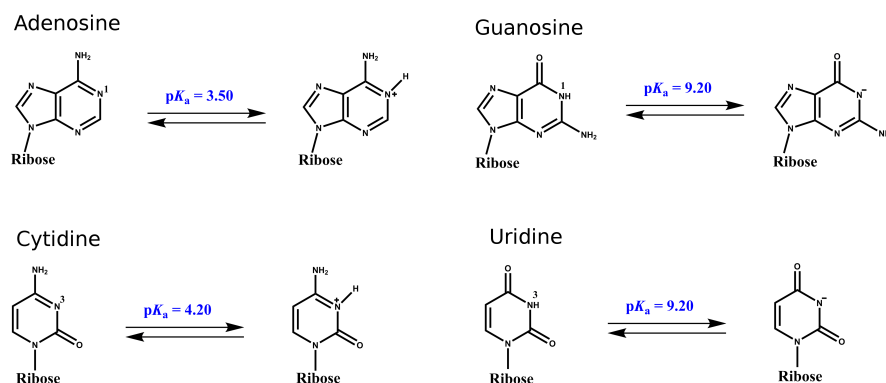


Figure 6.1: The figure illustrates the ionization of A,C,G,U nucleobases. The pK_a of N1 site of adenosine and guanosine is 3.50 and 9.20, respectively. The pK_a of N3 site of cytidine and uridine is 4.20 and 9.20, respectively.

6.2 Methods

6.2.1 Reference compounds

The reference compounds, adenosine monophosphate (AMP), cytidine monophosphate (CMP), guanosine monophosphate (GMP) and uridine monophosphate (UMP) was prepared using LeaP module in AMBER14 software package³⁷. The initial structures of non-terminal ends of the reference compounds were capped with Methoxy group (OCH3) at 5' terminal and methyl group (CH3) at 3' terminal. The parameters for the residues were described by the constant pH force field in AMBER14³⁷. Thermodynamic integration calculations were performed for, deprotonation of N1 site of adenosine and guanosine and, N3 site of cytidine and uridine to calculates the free energy difference (ΔG_{TI}) between protonated and deprotonated state of each residue and, applied to perform constant-pH simulations to calculate the pK_a values. The ΔG_{TI} was adjusted in order to match the correct experimental pK_a s of the reference compounds. The corrected reference energies were tested by repeating the calculation for each reference compounds.

6.2.2 RNA systems

The parameters obtained for the reference compounds were tested by applying on RNA systems to explore the pH dependence of the residues. Three main systems were subjected to study, dinucleotides: ApA, ApC, GpG, UpU, branch-point helix (BPH), exon splicing silencer 3 (ESS3) from HIV-1¹⁹³. All the systems were solvated with 10 Å of cubic box of TIP4PEW water molecules. The dinucleotides systems were solvated with 2250 of water molecules. The ESS3 and BPH took

approximately 10300 and 7460 of TIP4Pew water molecules to fully solvate the systems. Both ESS3 and BPH were neutralized by adding Na⁺ ions. The systems were equilibrated for 4 ns with 2 fs time step in a NPT ensemble. Then, the systems were subjected to replica exchange simulations along the pH ladder from pH =2 to 13 using 1 pH unit apart replicas. The replica exchange and proton exchange were performed for every 100 steps and 50 ps, respectively. Both residues of the dinucleotide systems were titrated at the same time. Only A7 was titrated in ESS3 system assuming the cytosine (C21) residue is most likely to be in fully protonated state. BPH was divided in to four systems where, (1) s1: A10, U11, U12 and U13 are titrating at the same time (2) s2: G4,U5,A17 and U18 titrating at the same time (3) s3: A6, A7 and U16 are titrating at the same time.

6.3 Results and discussion

The reference compounds

The free energy of deprotonation (ΔG_{TI}) of nucleobases were calculated from 4 ns long simulation. The calculated pK_a using ΔG_{TI} are 3.42, 4.21 , 9.14 and 9.13 for A, C, G and respectively. Therefore, the ΔG_{TI} values were corrected in order to match the experimental pK_a values. After the adjustment, the repeated calculations predicted closer pK_a values to the experimental values as shown in the Table 6.1. The running average of each reference compound is shown in the Figure 6.2-Figure 6.5, and indicates a convergence of fractions of protonation states of each residues. The Figure 6.6 illustrates the titration curves for each residues after applying the corrected free energy for each residue.

Table 6.1: The reference pK_a values, TI energies and adjusted reference energy for reference compounds.

Residue	Site	Ref pK_a	TI energy kcal/mol	Calc pK_a	Corrected Energy kcal/mol	Predicted pK_a
AMP	N1	3.50	14.84	3.42	14.96	3.51
CMP	N3	4.20	37.48	4.21	37.47	4.22
GMP	N1	9.20	-97.29	9.14	-97.37	9.18
UMP	N3	9.20	-136.33	9.13	-136.42	9.19

Application on dinucleotide systems

The parameters were applied on the dinucleotide systems to perform the constant-pH simulations on 5'-ApAp-3', 5'-ApCp-3',5'-GpGp-3' and 5'-UpUp-3' systems. Both residues of each systems were titrated at the same time. The calculated pK_a values agree with the experimentally observed values of the single nucleobases, as listed in the Table 6.2. The pK_a values of adenine residues of 5'-ApAp-3' are 3.44 and 3.70 respectively. These values closely reproduces the pK_a 3.50 of the adenine. The

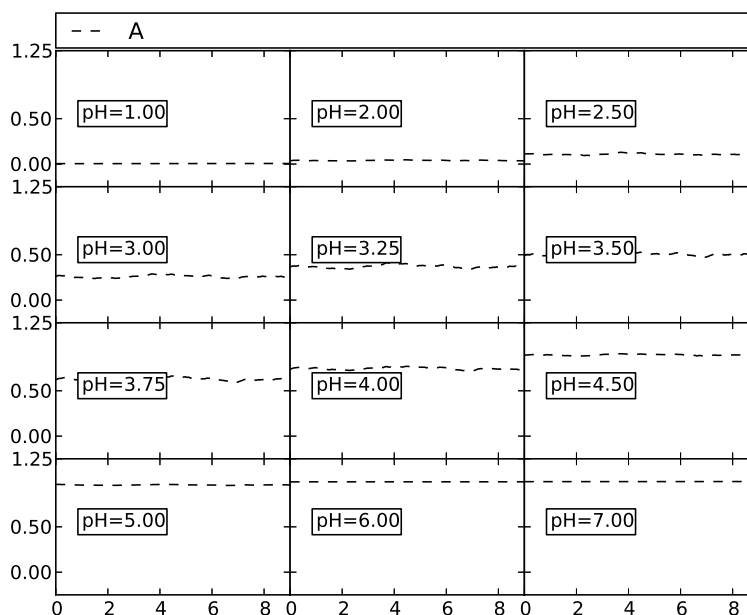


Figure 6.2: The running average, calculated for 1 ns time intervals, for 9 ns of the simulations for adenine nucleobase (A).

calculated Significant pK_a of guanine and uracil are upshifter and significantly different from the experimental values. The hill coefficient (n), an indication of an independently titrating residue (when $n=1$), for all the residues are >1 or <1 indicating a dependence on the nearby titrating residue.

Due to lack of experimental pK_a for the dinucleotides, the structural properties of each systems were compared with the experimentally available crystal structures in order to express the quality of our simulations. The stacking interactions are compared for each system with the criteria of the distance between neighboring glycosidic nitrogen $R(NN) < 5.0 \text{ \AA}$. The results are shown in the Figure 6.7, which explains the percent occupancy of stacking interactions for each system for five randomly selected pH values, pH=1, 4, 7, 9, 13 (Table 6.5 gives detail information). Experimental studies has suggested a higher stacking propensity between purine-purine bases, for which ApA and GpG belong in this category. In this study, lower stacking interactions has been observed for pyrimidine-pyrimidine bases while intermediate interactions for purine-pyrimidine and pyrimidine-purine bases. However, overall stacking interactions for ApC is higher in comparison to ApA interactions for lower solution pH values. UpU interactions are much higher in comparison to the GpG interactions at pH = 7 and 9. However, slight increment for GpG interactions is shown in pH=13.

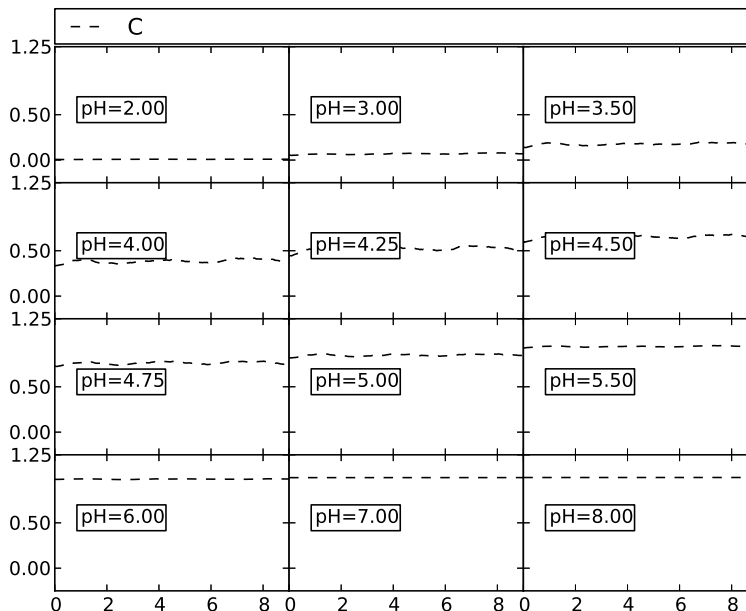


Figure 6.3: The running average, calculated for 1 ns time intervals, for 9 ns of the simulations for cytosine nucleobase (C).

Further validation of the structures were done by comparing the structural properties such as dihedral angle, sugar pucker to the available crystal structures of the dinucleotides, ApA 3'-5'-phosphodiester linkage and ApC 2'-5' link as indicated in the Table 6.3. A comparison of average dihedral angles is listed in the Table 6.3 for pH=2, 7 and 13. ApA closely represents the crystal structure at lower pH values. Overall structure of the dinucleotides agrees well with the experiment.

Table 6.2: The calculated pK_a for dinucleotides

System	Residue	Calc pK_a	Hill coefficient	Exp pK_a
AA	A1	3.44	0.67	3.50
	A2	3.70	-0.65	3.50
AC	A1	3.80	0.61	3.50
	C2	4.86	0.59	4.20
GG	G1	10.02	0.90	9.20
	G2	10.58	0.87	9.20
UU	U1	9.85	1.06	9.20
	U2	10.02	1.05	9.20

The calculated pK_a values (Calc pK_a) for the dinucleotides are compared with the experimentally observed pK_a of each nucleobase (Exp pK_a)

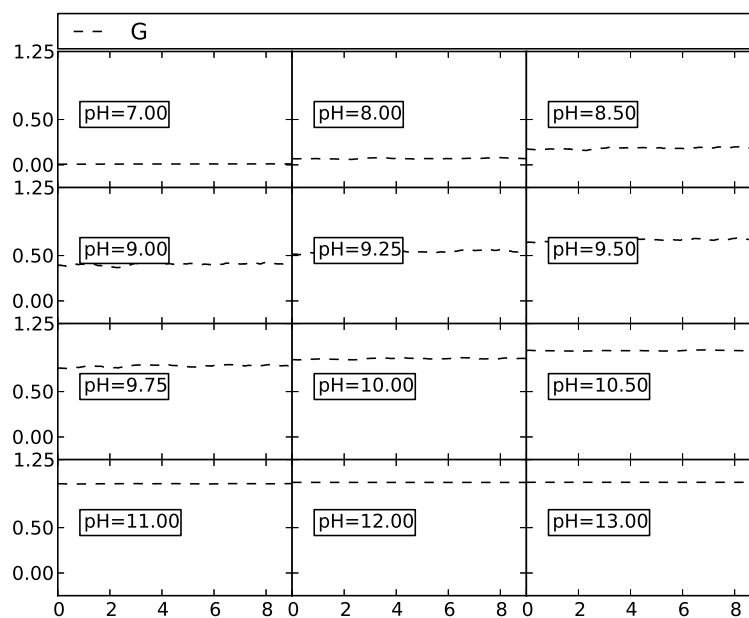


Figure 6.4: The running average, calculated for 1 ns time intervals, for 9 ns of the simulations for guanine nucleobase (G).

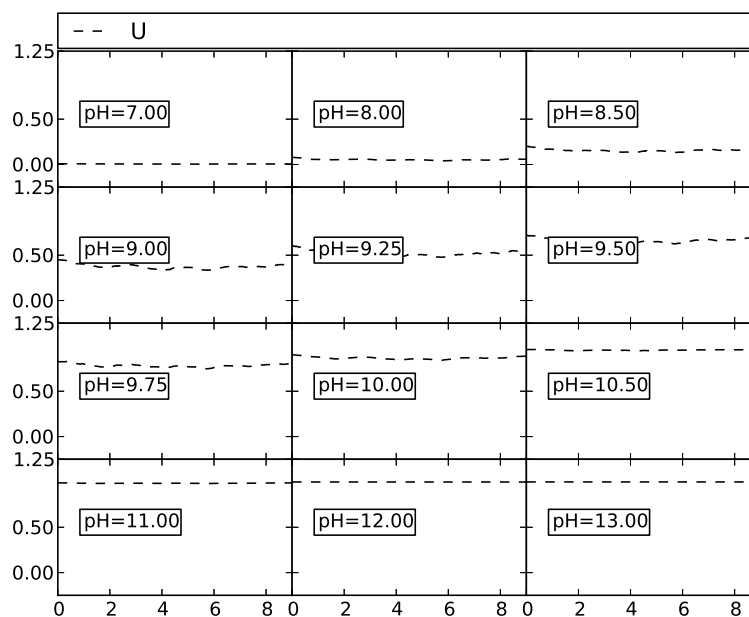


Figure 6.5: The running average, calculated for 1 ns time intervals, for 9 ns of the simulations for uracil nucleobase (U).

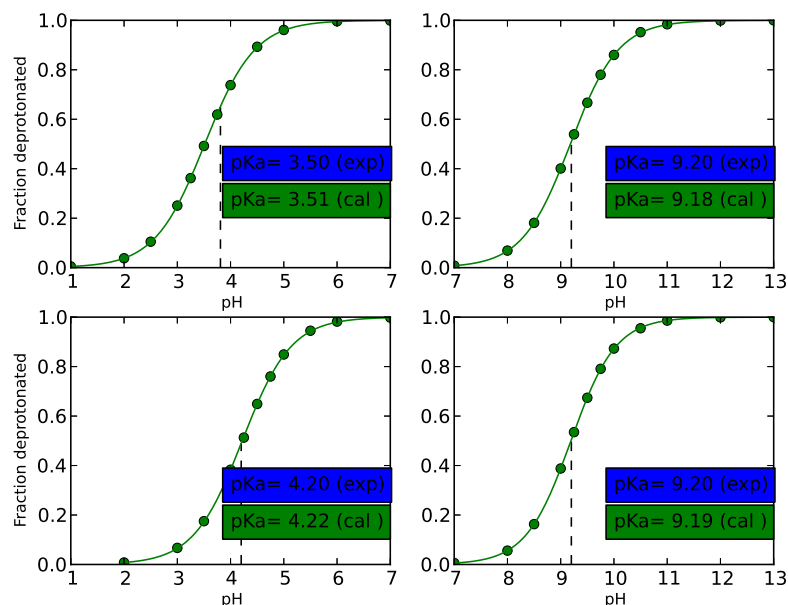


Figure 6.6: The figure shows the titration curves for the reference compounds after applying the corrected reference energy. The upper left shows the curve for adenine protonation in AMP. Upper right shows the guanine titration curve in GMP. Lower left shows the cytosine titration curve in CMP. Lower right shows the uracil titration curve in UMP. All the computed values are closely reproduce the experimental pK_a values.

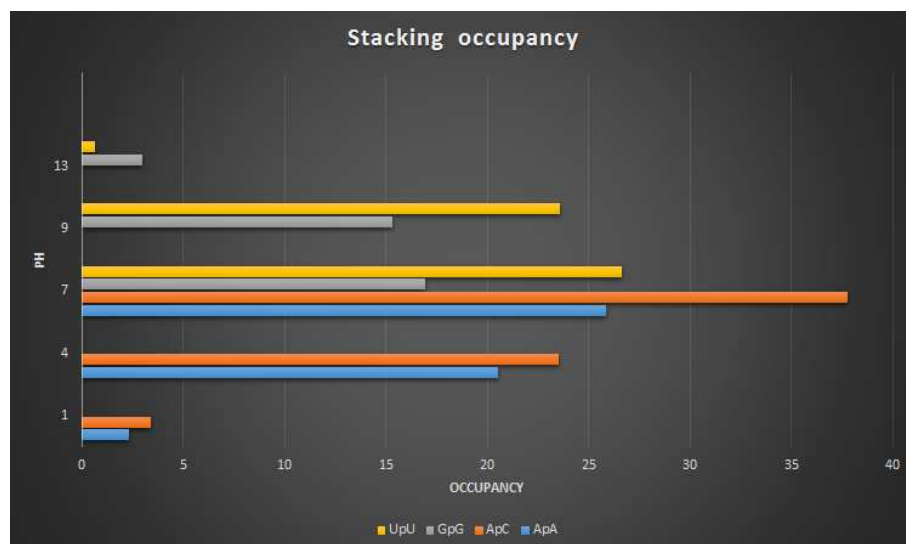


Figure 6.7: Occupancies of the stacking interactions between neighboring N atoms of each nucleobase is presented in the plot. The criteria for interaction is chosen as the distance between N atoms ($R(NN) < 5 \text{ \AA}$). ApC shows highest stacking interactions at pH=7 and the lowest interactions are observed in GpG.

Table 6.3: Comparison of calculated dihedral angles of dinucleotides with the experimental values

ApA	α	β		γ		δ		ζ	
	A3	A2	A3	A2	A3	A2	A3	A2	A3
pH =1	69.6 (54.4)	179 (91.9)	179.2 (94.2)	57.0 (95.2)	56.7 (96.9)	120.2 (71.6)	119.3 (68.5)	71.5	-
pH =4	-	179.2 (90.3)	179.0 (96.2)	58.0 (90.8)	59.0 (98.0)	-	113.5 (66.1)	-	-
pH =7	-	179.8 (86.3)	179.1 (96.0)	57.5 (85.5)	59.9 (98.8)	79.4 (57.3)	109.5 (63.0)	-72.2 (50.4)	-
ApC									
pH =1		179.3 (90.8)	178.3 (94.3)	56.2 (92.0)	57.3 (98.3)	124.8 (71.6)	123.8 (67.2)	-	-
pH= 4	-73.2 (45.8)	180.2 (94.4)	178.5 (95.6)	57.5 (92.4)	57.9 (97.8)	117.2 (64.3)	121.9 (59.0)	-64.6 (51.5)	-
pH=7	-74.3 (58.0)	181.2 (93.8)	177.9 (96.9)	57.6 (93.5)	58.3 (97.2)	112.3 (49.7)	119.7 (48.6)	-64.1 (65.8)	-

The average values of dihedral angles for the dinucleotide backbone was compared with the measured values from the crystal structures of ApA 3'-5'-phosphodiester linkage and ApC 2'-5' link^{194,195}. The values from the crystal structures are listed in the parenthesis.

Branch-point helix (BPH)

The parameters were validated by applying on a larger system, the Branch-point helix (BPH). The available pK_a values, determined with NMR spectroscopy, are directly compared with the calculated values. As shown in the Table 6.4, the results is in good agreement with the experimental values as well as another theoretical study.

Table 6.4: Comparison of calculated pK_a values of Branch-point helix (BPH) with experiment

Residue	System	Calc pK_a	Experimental pK_a	Theoretical pK_a ¹⁹⁶
G4	s2	14.5 (s2)		
U5	s2	14.0 (s2)		
A6	s3	2.8 (s3)	<5.0	2.5
A7	s3	3.1 (s3)	6.1	6.8
A10	s1	3.8 (s1)	<5.0	1.7
U11	s1	13.2 (s1)		
U12	s1	11.1 (s1)		
A13	s1	5.4 (s1)	<5.5	5.3
U16	s3	14.1 (s3)		
A17	s2	4.00 (s2)	<5.0	2.7
U18	s2	11.9 (s2)		

The calculated pK_a values for residues in BPH were compared with the experimentally observed values and a theoretically measured values.

The structure and dynamics of ESS3

The A+ of the A+·C wobble pair of ESS3 is subjected to titrate in the simulation. Here we report the structure and dynamics of the system along the pH ladder of 2 to 13. The RMSD of the backbone in comparison to the solution structure at each pH values is approximately 3.5 Å. The titration of A7 residue has not shown any significant effect on the backbone of the ESS3. However, the side chain of RMSD of the A+C wobble pair indicates higher fluctuations along the pH ladder with titrating adenine residue. An experimental study shows that the adenosine of highly conserved A+ C wobble pair titrates at $pH = 6.7$ ¹⁹³. Our calculated pK_a , 6.5, closely predicts the experimental value. The pK_a is upshifted in comparison to the reference pK_a value, 3.5. The calculated hill coefficient is 0.82 indicate slight dependence on the neighboring residues. However, significant structural changes are seen only at higher pH values. Adenine make base pairs with cytosine by interacting A7:N1H with C21:O and A7:H61 with C21:N3. These are strong at lower pH values where adenine is protonated at N1. However, at higher pH values the base pairs are dissociated as N1 of adenine gets deprotonated.

6.4 Conclusions

New parameters were obtained for application of the constant-pH simulations on nucleobases. The pK_a predictions for the test cases, dinucleotides, BPH and ESS3, are in fairly good agreement with

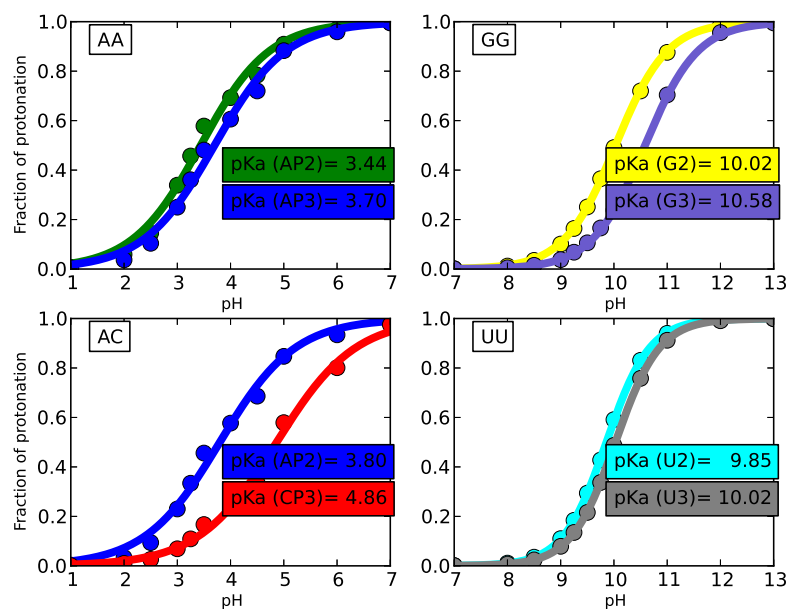


Figure 6.8: The calculated pK_a values and the titration curves for the dinucleotide systems.

the experimental values. The results will be applied to predicted the macroscopic and microscopic pK_a s and the pH-rate profiles for the ribozyme systems in the near future.

Table 6.5: The distance between neighboring glycosidic nitrogen of dinucleotides

pH	ApA:A2@N1-A3@N1 (Å)	ApC:A2@N1-C3@N3 (Å)	pH	GpG:G2@N1-G3@N1 (Å)	UpU:U2@N3-U3@N3 (Å)
1.00	10.53	11.01	7.00	6.66	6.02
2.00	10.23	10.85	8.00	6.55	6.09
2.50	10.05	10.70	8.50	6.76	6.14
3.00	9.15	10.08	9.00	6.67	6.24
3.25	8.86	9.88	9.25	6.78	6.48
3.50	8.35	9.44	9.50	6.92	6.57
3.75	8.26	8.97	9.75	7.21	6.89
4.00	8.00	8.81	10.00	7.52	7.37
4.50	7.16	8.19	10.50	8.73	8.20
5.00	6.98	7.77	11.00	10.12	8.70
6.00	6.79	7.17	12.00	11.84	8.93
7.00	6.93	7.38	13.00	11.93	9.06

The distance between neighboring glycosidic nitrogen of dinucleotides (R(NN)), 5'-ApAp-3', 5'-ApCp-3', 5'-GpGp-3' and 5'-UpUp-3', is calculated at each pH to observe the structural behavior along the pH ladder.

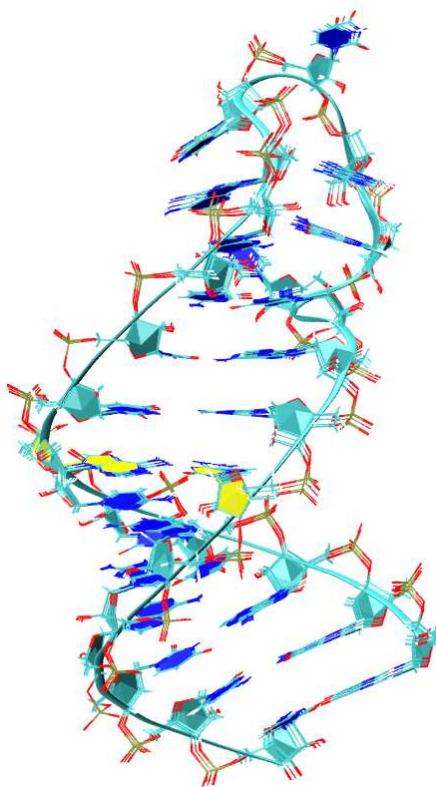


Figure 6.9: The structural overlay of the average structures of ESS3 along the pH ladder of pH=2 to 13. The nucleobases of A7 and C21 residues show slight structural fluctuation due to titration of A7 residue. The deviation of the pH has no significant effect on the backbone of the structure.

Bibliography

- [1] Ronald T. Raines. Ribonuclease A. *Chem. Rev.*, 98:1045–1065, 1998.
- [2] Brian D. Wladkowski, Morris Krauss, and Walter J. Stevens. Transphosphorylation catalyzed by ribonuclease A: Computational study using *ab Initio* effective fragment potentials. *J. Am. Chem. Soc.*, 117:10537–10545, 1995.
- [3] Brian D. Wladkowski, Paul Ostazeski, Sarah Chenoweth, Steven J. Broadwater, and Morris Krauss. Hydrolysis of cyclic phosphates by ribonuclease A: A computational study using a simplified *ab initio* quantum model. *J. Comput. Chem.*, 24:1803–1811, 2003. doi: 10.1002/jcc.10331.
- [4] 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.
- [5] 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.
- [6] Geoffrey M Cooper, editor. *The Cell*. Sinauer Associates, Boston, 2000.
- [7] Gordon G. Hammes. How Do Enzymes Really Work? *J. Biol. Chem.*, 283:22337–22346, 2008.
- [8] Mireia Garcia-Viloca, Jiali Gao, Martin Karplus, and Donald G. Truhlar. How enzymes work: Analysis by modern rate theory and computer simulations. *Science*, 303:186–195, 2004.
- [9] Alan Fersht. *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding*. W. H. Freeman, New York, 1998.
- [10] Sergio Martí, Maite Roca, Juan Andrés, Vicent Moliner, Estanislao Silla, Iñaki Tuñón, and Juan Bertrán. Theoretical insights in enzyme catalysis. *Chem. Soc. Rev.*, 33:98–107, 2004.
- [11] Stephen J. Benkovic and Sharon Hammes-Schiffer. A perspective on enzyme catalysis. *Science*, 301:1196–1202, 2003.

- [12] Maria T. Panteva, Thakshila Dissanayake, Haoyuan Chen, Brian K. Radak, Erich R. Kuechler, George M. Giambasu, Tai-Sung Lee, and Darrin M. York. *Multiscale Methods for Computational RNA Enzymology*, chapter 14. Elsevier, 2015.
- [13] Joseph Kraur. How Do Enzymes Work? *Science*, 242:533–540, 1988.
- [14] James F. Dama, Anton V. Sinitskiy, Martin McCullagh, Jonathan Weare, Benoît Roux, Aaron R. Dinner, and Gregory A. Voth. The theory of ultra-coarse-graining. 1. general principles. *J. Chem. Theory Comput.*, 9:2466–2480, 2013.
- [15] Michael Feig, John Karanicolas, and Charles L. Brooks, III. MMTSB tool set: Enhanced sampling and multiscale modeling methods for applications in structure biology. *J. Mol. Graph. Model.*, 22:377–395, 2004.
- [16] Alessio Lodola and Adrian J. Mulholland. *Computational enzymology*, volume 924 of *Methods Mol. Biol.*, chapter 4, pages 67–89. Humana Press, 2013.
- [17] Martin Meier-Schellersheim, Iain D. C. Fraser, and Frederick Klauschen. Multiscale modeling for biologists. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, 1(1):4–14, 2009.
- [18] Paul Sherwood, Bernard R. Brooks, and Mark S. P. Sansom. Multiscale methods for macromolecular simulations. *Curr Opin Struct Biol*, 18(5):630–640, 2008.
- [19] Marc W. van der Kamp and Adrian J. Mulholland. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry*, 52:2708–2728, 2013.
- [20] Darrin M. York and Tai-Sung Lee, editors. *Multiscale quantum models for biocatalysis: Modern techniques and applications*. Challenges and Advances in Computational Chemistry and Physics. Springer, New York, 2009.
- [21] Guanhua Hou and Qiang Cui. Stabilization of different types of transition states in a single enzyme active site: QM/MM analysis of enzymes in the alkaline phosphatase superfamily. *J. Am. Chem. Soc.*, 135:10457–10469, 2013.
- [22] Wangshen Xie, Modesto Orozco, Donald G. Truhlar, and Jiali Gao. X-Pol potential: An electronic structure-based force field for molecular dynamics simulation of a solvated protein in water. *J. Chem. Theory Comput.*, 5:459–467, 2009.
- [23] Mark Gerstein and Werner Krebs. A database of macromolecular motions. *Nucleic Acids Res.*, 26:4280–4290, 1998.

- [24] Chern-Sing Goh, Duncan Milburn, and Mark Gerstein. Conformational changes associated with protein-protein interactions. *Curr. Opin. Struct. Biol.*, 14:104–109, 2004.
- [25] Alex Gutteridge and Janet Thornton. Conformational change in substrate binding, catalysis and product release: an open and shut case? *FEBS Lett.*, 567:67–73, 2004.
- [26] Heather Beach, Roger Cole, Michelle L. Gill, and J. Patrick Loria. Conservation of us-ms Enzyme Motions in the Apo- and Substrate-Mimicked State. *J. Am. Chem. Soc.*, 127:9167–9176, 2005.
- [27] Oliver F. Lange, Nils-Alexander Lakomek, Christophe Farès, Gunnar F. Schröder, Korvin F A. Walter, Stefan Becker, Jens Meiler, Helmut Grubmüller, Christian Griesinger, and Bert L. de Groot. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science*, 320(5882):1471–1475, 2008.
- [28] Tomasz Wlodarski and Bojan Zagrovic. Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc. Natl. Acad. Sci. USA*, 106:19346–19351, 2009.
- [29] Larry R Masterson, Cecilia Cheng, Tao Yu, Marco Tonelli, Alexandr Kornev, Susan S. Taylor, and Gianluigi Veglia. Dynamics connect substrate recognition to catalysis in protein kinase A. *Nat. Chem. Biol.*, 6:821–828, 2010.
- [30] Sarah M. Sullivan and Todd Holyoak. Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. *Proc. Natl. Acad. Sci. USA*, 105:13829–13834, 2008.
- [31] Sonia Fieulaine, Adrien Boularot, Isabelle Artaud, Michel Desmadril, Frédéric Dardel, Thierry Meinnel, and Carmela Giglione. Trapping Conformational States Along Ligand-Binding Dynamics of Peptide Deformylase: The Impact of Induced Fit on Enzyme Catalysis. *PLoS Biol.*, 9:1001066, 2011.
- [32] Marcus P. D. Hatfield and Sándor Lovas. Conformational Sampling Techniques. *Curr. Pharm. Design*, 20:3303–3313, 2014.
- [33] Adam Liwo, Cezary Czaplewski, Stanisław Ołdziej, and Harold A Scheraga.
- [34] M. von Bergen, P. Friedhoff, J. Biernat, E.-M. Mandelkow J. Heberle, and E. Mandelkow. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA*, 44:98–104, 1958.

- [35] B. J. Alder and T. E. Wainwright. Phase Transition for a Hard Sphere System. *J. Chem. Phys.*, 27:1208–1209, 1957.
- [36] B. J. Alder and T. E. Wainwright. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.*, 31(2):459–466, 1959.
- [37] 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. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, 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.
- [38] 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.
- [39] Philip C. Bevilacqua. Mechanistic considerations for general acid-base catalysis by RNA: Revisiting the mechanism of the hairpin ribozyme. *Biochemistry*, 42:2259–2265, 2003.
- [40] G. Matthias Ullmann. Relations between protonation constants and titration curves in polyprotic acids: A critical view. *J. Phys. Chem. B*, 107:1263–1271, 2003.
- [41] Astrid R. Klingen, Elisa Bombarda, and G. Matthias Ullmann. Theoretical investigation of the behavior of titratable groups in proteins. *Photochem. Photobiol. Sci.*, 5:588–596, 2006.
- [42] António M. Baptista, Paulo J. Martel, and Steffen B. Petersen. Simulation of protein conformational freedom as a function of pH: Constant-pH molecular dynamics using implicit titration. *Proteins*, 27:523–544, 1997.
- [43] António M. Baptista, Vitor H. Teixeira, and Cláudio M. Soares. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.*, 117:4184–4200, 2002.
- [44] Jana Khandogin and Charles L. Brooks III. Constant pH molecular dynamics with proton tautomerism. *Biophys. J.*, 89:141–157, 2005.
- [45] Michael S. Lee, Freddie R. Salsbury, Jr., and Charles L. Brooks III. Constant-pH molecular dynamics using continuous titration coordinates. *Proteins*, 56:738–752, 2004.
- [46] John Mongan, David A. Case, and J. Andrew McCammon. Constant pH molecular dynamics in Generalized Born implicit solvent. *J. Comput. Chem.*, 25:2038–2048, 2004.

- [47] Douglas Poland. *Cooperative Equilibria In Physical Biochemistry*. Oxford University Press, 1978.
- [48] Yilin Meng and Adrian E. Roitberg. Constant pH Replica Exchange Molecular Dynamics in Biomolecules Using a Discrete Protonation Model. *J. Chem. Theory Comput.*, 6:1401–1412, 2010.
- [49] Alexey Onufriev, David A. Case, and G. Matthias Ullmann. A novel view of pH titration in biomolecules. *Biochemistry*, 40:3413–3419, 2001.
- [50] Satoru G. Itoh, Ana Damjanović, and Bernard R. Brooks. pH replica-exchange method based on discrete protonation states. *Proteins*, 79:3420–3436, 2011.
- [51] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314:141–151, 1999.
- [52] Errol G. Lewars. *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. Springer, New York, 2011.
- [53] 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.
- [54] Johannes Kästner. Umbrella sampling. *WIREs Comput. Mol. Sci.*, 1:932–942, 2011.
- [55] 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.
- [56] Peng Tao, Joseph D. Larkin, and Bernard R. Brooks. *Reaction Path Optimization and Sampling Methods and Their Applications for Rare Events*, chapter 14, pages 27–66. In Tech, 2012.
- [57] Groenhof G. *Introduction to QM/MM simulations*, volume 924 of *Methods Mol. Biol.*, chapter 3, pages 43–66. Humana Press, 2013.
- [58] Fernanda Duarte, Beat A. Amrein, David Blaha-Nelson, and Shina C.L. Kamerlin. Recent advances in QM/MM free energy calculations using reference potentials. *BBA - General Subjects*, 1850:954–965, 2015.
- [59] Hans Matrin Senn and Walter Thiel. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Edit.*, 48:1198–1229, 2009.

- [60] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, 103: 227–249, 1976.
- [61] A. Warshel and M. Karplus. Calculation of Ground and Excited State Potential Surfaces of Conjugated Molecules.’ I. Formulation and Parametrization. *J. Am. Chem. Soc.*, 94:5612–5625, 1972.
- [62] Michael J. S. Dewar, Eve Zoebisch, 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.
- [63] 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.
- [64] James J. P. Stewart. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.*, 10:209–220, 1989.
- [65] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th. Frauenheim, S. Suhai, and G. Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58(11):7260–7268, 1998.
- [66] J. J. P. Stewart. Optimization of parameters for semiempirical methods II. applications. *J. Comput. Chem.*, 10:221–264, 1989.
- [67] Marcus Elstner. The SCC-DFTB method and its application to biological systems. *Theor. Chem. Acc.*, 116(1-3):316–325, 2006.
- [68] Jiyoung Kang and Masaru Tateno. *Recent Applications of Hybrid Ab Initio Quantum Mechanics-Molecular Mechanics Simulations to Biological Macromolecules*, chapter 14, pages 359–384. In Tech, 2012.
- [69] Yaoquan Tu and Aatto Laaksonen. *Implementing Quantum Mechanics into Molecular Mechanics-Combined QM/MM Modeling Methods*, volume 59, pages 1–15. Elsevier Inc., 2010.
- [70] Dirk Bakowies and Walter Thiel. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J. Phys. Chem.*, 100:10580–10594, 1996.

- [71] 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.
- [72] U. C. Singh and P. A. Kollman. A combined *ab initio* quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: applications to the $\text{CH}_3\text{Cl}+\text{Cl}^-$ exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.*, 7:718–730, 1986.
- [73] Thom Vreven, Benedetta Mennucci, Clarissa O. da Silva, Keiji Morokuma, and Jacopo Tomasi. The ONIOM-PCM method: Combining the hybrid molecular orbital method and the polarizable continuum model for solvation. Application to the geometry and properties of a merocyanine in solution. *J. Chem. Phys.*, 115(1):62–72, 2001.
- [74] Mats Svensson, Stephane Humbel, Robert D. J. Froese, Toshiaki Matsubara, Stefan Sieber, and Keiji Morokuma. Oniom: A multilayered integrated mo + mm method for geometry optimizations and single point energy predictions. a test for diels-alder reactions and $\text{pt}(\text{p}(\text{t-bu})_3)_2 + \text{h}_2$ oxidative addition. *J. Phys. Chem.*, 100(50):19357–19363, 1996. URL <http://pubs.acs.org/doi/abs/10.1021/jp962071j>.
- [75] F. Maseras and K. Morokuma. IMOMM: A new integrated *ab initio* + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.*, 16:1170, 1995.
- [76] Walter Thiel and Alexander A. Voityuk. Extension of the MNDO formalism to d Orbitals: Integral approximations and preliminary numerical results. *Theor. Chim. Acta*, 81:391–404, 1992.
- [77] Walter Thiel and Alexander A. Voityuk. Extension of MNDO to d Orbitals: Parameters and Results for the Second-Row Elements and for the Zinc Group. *J. Phys. Chem.*, 100:616–626, 1996.
- [78] Xabier Lopez and Darrin M. York. Parameterization of semiempirical methods to treat nucleophilic attacks to biological phosphates: AM1/d parameters for phosphorus. *Theor. Chem. Acc.*, 109:149–159, 2003.
- [79] Michael J.S. Dewar and Walter Thiel. Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. *J. Am. Chem. Soc.*, 99(15):4899–4907, 1977.

- [80] Michael J.S. Dewar and Walter Thiel. Ground States of Molecules. 39. MNDO Results for Molecules Containing Hydrogen, Carbon, Nitrogen, and Oxygen. *J. Am. Chem. Soc.*, 99(15): 4907–4917, 1977.
- [81] Kwangho Nam, Jiali Gao, and Darrin M. York. Quantum mechanical/molecular mechanical simulation study of the mechanism of hairpin ribozyme catalysis. *J. Am. Chem. Soc.*, 130(14): 4680–4691, 2008.
- [82] Henry Eyring. The Activated Complex in Chemical Reactions. *J. Chem. Phys.*, 3:107–115, 1935.
- [83] M. G. Evans and M. Polanyi. Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans. Faraday. Soc.*, 31:875–894, 1935.
- [84] David Chandler. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J. Chem. Phys.*, 68:2959–2970, 1978.
- [85] Yang Yang and Qiang Cui. Does water relay play an important role in phosphoryl transfer reactions? Insights from theoretical study of a model reaction in water and *tert*-butanol. *J. Phys. Chem.*, 113:4930–4939, 2009.
- [86] Jonathan K. Lassila, Jesse G. Zalatan, and Daniel Herschlag. Biological phosphoryl-transfer reactions: Understanding mechanism and catalysis. *Annu. Rev. Biochem.*, 80:669–702, 2011.
- [87] Elizabeth A. Doherty and Jennifer A. Doudna. Ribozyme Structures and Mechanisms. *Annu. Rev. Biophys. Biomol. Struct.*, 30:457–475, 2001.
- [88] J. J. Beintema. Introduction: the ribonuclease A superfamily. *Cell.Mol.Life.Sci.*, 54:763–765, 1998.
- [89] Kimberly D. Dyer and Helene F. Rosenberg. The RNase a superfamily: Generation of diversity and innate host defense. *Molecular Diversity*, 10:585–597, 2006.
- [90] H F Rosenberg. RNase A ribonucleases and host defense: an evolving story. *J.Leukoc. Biol.*, 83:1079–1087, 2008.
- [91] Claudi M. Cuchillo, M. Victòria Nogués, and Ronald T. Raines. Bovine pancreatic ribonuclease: Fifty years of the first enzymatic reaction mechanism. *Biochemistry*, 50:7835–7841, 2011.

- [92] D.G Herries, A. P Mathias, and B. R Rabin. The active site and mechanism of action of bovine pancreatic ribonuclease. 3. The pH-dependence of the kinetic parameters for the hydrolysis of cytidine 2',3'-phosphate. *Biochem.J.*, 85:127–134, 1962.
- [93] D Findlay, A. P Mathias, and B. R Rabin. The active site and mechanism of action of bovine pancreatic ribonuclease. 4. The activity in inert organic solvents and alcohols. *Biochem. J.*, 85:134–138, 1962.
- [94] D. Findlay, A. P Mathias, and B.R. Rabin. The active site and mechanism of action of bovine pancreatic ribonuclease. 5. The charge types at the active centre. *Biochem. J.*, 85:139–144, 1962.
- [95] C. A. Ross, A. P. Mathias, and B. R. Rabin. The active site and mechanism of action of bovine pancreatic ribonuclease. 6. Kinetic and spectrophotometric investigation of the interaction of the enzyme with inhibitors and p-nitrophenyl acetate. *Biochem. J.*, 85:145–151, 1962.
- [96] D. Findlay, D. G. Herries, A. P. Mathias, B. R. Rabin, and C. A. Ross. The active site and mechanism of action of bovine pancreatic ribonuclease. 7. The catalytic mechanism. *Biochem. J.*, 85:152–153, 1962.
- [97] Herbert Witzel. *Progress in Nucleic Acid Research*, volume 2, chapter The Function of the Pyrimidine Base in the Ribonuclease Reaction, pages 221–258. Academic Press Inc, 1963.
- [98] Jui H. Wang. Facilitated proton transfer in enzyme catalysis. It may have a crucial role in determining the efficiency and specificity of enzymes. *Science*, 161:328–333, 1968.
- [99] Eric Anslyn and Ronald Breslow. On the mechanism of catalysis by ribonuclease: cleavage and isomerization of the dinucleotide UpU catalyzed by imidazole buffers. *J. Am. Chem. Soc.*, 111:4473–4482, 1989.
- [100] Brian D. Wladkowski, Morris Krauss, and Walter J. Stevens. Ribonuclease A catalyzed transphosphorylation: An *ab Initio* theoretical study. *J. Phys. Chem.*, 99:6273–6276, 1995.
- [101] Ernesto J. del Rosario and Gordon G. Hammes. Kinetic and equilibrium studies of the ribonuclease-catalyzed hydrolysis of uridine 2',3'-cyclic phosphate. *Biochemistry*, 8:1884–1889, 1969.
- [102] John L. Markley. . *Acc. Chem. Res.*, 8:70–80, 1975.

- [103] D H Meadows, O Jardetzky, R M Epan, H H Ruterjans, and H A Scheraga. Assignment of the histidine peaks in the nuclear magnetic resonance spectrum of ribonuclease. *Proc. Natl. Acad. Sci. USA*, 60:766–772, 1968.
- [104] J. H. Bradbury and J. S. Teh. Reassignment of the Histidine ^1H Nuclear Magnetic Resonances of Ribonuclease-A. *Chem. Commun.*, 00:936–937, 1975.
- [105] Dorothy Findly, D.G. Herries, B.R. Rabin A.P. Mathias, and C.A. Ross. The Active Site and Mechanism of Action of Bovine Pancreatic Ribonuclease. *Nature*, 190:781–784, 1961.
- [106] Chiwook Park, L. Wayne Schultz, and Ronald T. Raines. Contribution of the active site histidine residues of ribonuclease A to nucleic acid binding. *Biochemistry*, 40:4949–4956, 2001.
- [107] S. Capasso, A. Di Donato, L. Esposito, F. Sica, G. Sorrentino, L. Vitagliano, A. Zagari, and L. Mazzarella. Deamidation in Proteins: The Crystal Structure of Bovine Pancreatic Ribonuclease with an Isoaspartyl Residue at Position 67. *J. Mol. Biol.*, 257:492–496, 1996.
- [108] A Di Donato, M A Ciardiello, M de Nigris, R Piccoli, L Mazzarella, and G D’Alessio. Selective deamidation of ribonuclease A. Isolation and characterization of the resulting isoaspartyl and aspartyl derivatives. *J. Biol. Chem.*, 268:4745–4751, 1993.
- [109] June M. Messmore, Dana N. Fuchs, and Ronald T. Raines. Ribonuclease A: Revealing Structure-Function Relationships with Semisynthesis. *J. Am. Chem. Soc.*, 117(31):8057–8060, 1995.
- [110] James E. Thompson and Ronald T. Raines. Value of general acid-base catalysis to ribonuclease a. *J. Am. Chem. Soc.*, 116:5467–5468, 1994.
- [111] Konstantin I. Panov, Andrei L. Okorokov Elena Yu. Kolbanovskaya, Tatiana B. Panova, Anke C. Terwisscha van Scheltinga, Marat Ya. Karpeisky, and Jaap J. Beintema. Ribonuclease A mutant His¹¹⁹ Asn: the role of histidine in catalysis. *FEBS Lett.*, 398:57–60, 1996.
- [112] Eri Chatani, Rikimaru Hayashi, Hideaki Moriyama, and Tatzuo Ueki. Conformational strictness required for maximum activity and stability of bovine pancreatic ribonuclease A as revealed by crystallographic study of three Phe120 mutants at 1.4 Å resolution. *Protein Sci.*, 11:72–81, 2002.
- [113] Stephen B. delCardayre and Ronald T. Raines. A Residue to Residue Hydrogen Bond Mediates the Nucleotide Specificity of Ribonuclease A. *J. Mol. Biol.*, 252:328–336, 1995.

- [114] L. Wayne Schultz, David J. Quirk, and Ronald T. Raines. His Asp Catalytic Dyad of Ribonuclease A: Structure and Function of the Wild-Type, D121N, and D121A Enzymes. *Biochemistry*, 37:8886–8898, 1998.
- [115] Claudi M. Cuchillo, Xavier Parés, Alícia Guasch, Tom Barmanc, Franck Traverse, and M. Victòria Nogués. The role of 2', 3'-cyclic phosphodiester in the bovine pancreatic ribonuclease A catalysed cleavage of RNA: intermediates or products? *FEBS Lett.*, 333:207–210, 1993.
- [116] M. Kunitz. Isolation from beef pancreas of a crystalline protein possessing ribonuclease activity. *Science*, 90:112–113, 1939.
- [117] M. Kunitz. Crystalline ribonuclease. *J. Gen. Physiol.*, 24:15–32, 1940.
- [118] Witzel H. The function of the pyrimidine base in the ribonuclease reaction. *Prog. Nucleic Acid Res.*, 2:221–258, 1963.
- [119] D. A. Usher. On the mechanism of ribonuclease action. *Proc. Natl. Acad. Sci. USA*, 62:661–667, 1969.
- [120] Eric Anslyn and Ronald Breslow. On the Mechanism of Catalysis by Ribonuclease: Cleavage and Isomerization of the Dinucleotide UpU Catalyzed by Imidazole Buffers. *J. Am. Chem. Soc.*, 111:4473–4482, 1989.
- [121] Brian D. Wladkowski, L. Anders Svensson, Lennart Sjölin, Jane E. Ladner, and Gary L. Gilliland. Structure (1.3Å) and Charge States of a Ribonuclease A-Uridine Vanadate Complex: Implications for the Phosphate Ester Hydrolysis Mechanism. *J. Am. Chem. Soc.*, 120:5488–5498, 1998.
- [122] 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.
- [123] Carol A. Deakyne and Leland C. Allen. Role of active-site residues in the catalytic mechanism of ribonuclease A. *J. Am. Chem. Soc.*, 101:3951–3959, 1979.
- [124] Ronald Breslow and Ruo Xu. Recognition and catalysis in nucleic acid chemistry. *Proc. Natl. Acad. Sci. USA*, 90:1201–1207, 1993.

- [125] Brigitta Elsässer, Gregor Fels, and John H. Weare. QM/MM Simulation (B3LYP) of the RNase A Cleavage-Transesterification Reaction Supports a Triester AN+DN Associative Mechanism with an O2' H Internal Proton Transfer. *J. Am. Chem. Soc.*, 136:927–936, 2014.
- [126] Bostjan Kobe and Johann Deisenhofer. Mechanism of ribonuclease inhibition by ribonuclease inhibitor protein based on the crystal structure of its complex with ribonuclease A. *J. Mol. Biol.*, 264:1028–1043, 1996.
- [127] Ingrid Zegers, Dominique Maes, Minh-Hoa Dao-Thi, Freddy Poortmans, Rex Palmer, and Lode Wyns. The structures of RNase A complexed with 3'-CMP and d(CpA): Active site conformation and conserved water molecules. *Protein Sci.*, 3:2322–2339, 1994.
- [128] Enguerran Vanquelef, Sabrina Simon, Gaelle Marquant, Elodie Garcia, Geoffroy Klimerek, Jean Charles Delepine, Piotr Cieplak, and François-Yves Dupradeau. R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.*, 39:W511–W517, 2011.
- [129] Hans W. Horn, William C. Swope, Jed W. Pitera, Jeffry 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.
- [130] 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.
- [131] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.*, 23:327–341, 1977.
- [132] 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.
- [133] 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.
- [134] Na Zhang, Shenglong Zhang, and Jack W. Szostak. Activated ribonucleotides undergo a sugar pucker switch upon binding to a single-stranded RNA template. *J. Am. Chem. Soc.*, 134(8): 3691–3694, 2012.

- [135] Jane E. Ladner, Brian D. Wlaskowski, L. Anders Svensson, Lennart Sjolin, and Gary L. Gilliland. X-ray Structure of a Ribonuclease A-Uridine Vanadate Complex at 1.3 Å Resolution. *Acta. Cryst.*, D53:290–301, 1997.
- [136] David S. Cerutti, Isolde Le Trong, Ronald E. Stenkamp, and Terry P. Lybrand. Simulations of a Protein Crystal: Explicit Treatment of Crystallization Conditions Links Theory and Experiment in the Streptavidin-Biotin Complex. *Biochemistry*, 47(46):12065–12077, 2008.
- [137] David S. Cerutti, Isolde Le Trong, Ronald E. Stenkamp, and Terry P. Lybrand. Dynamics of the Streptavidin-Biotin Complex in Solution and in Its Crystal Lattice: Distinct Behavior Revealed by Molecular Simulations. *J. Phys. Chem. B*, 113(19):6971–6985, 2009.
- [138] David S. Cerutti, Peter L. Freddolino, Robert E. Duke, Jr., and David A. Case. Simulations of a protein crystal with a high resolution X-ray structure: evaluation of force fields and water models. *J. Phys. Chem. B*, 114:12811–12824, 2010.
- [139] Pawel A. Janowski, David S. Cerutti, James Holton, and David A. Case. Peptide Crystal Simulations Reveal Hidden Dynamics. *J. Am. Chem. Soc.*, 135(21):7938–7948, 2013.
- [140] Anthony John Kirby. *Encyclopedia of Life Sciences (ELS)*, chapter Acid-Base Catalysis by Enzymes. John Wiley & Sons, Ltd, Chichester, 2010. doi: 10.1002/9780470015902.a0000602.pub2.
- [141] Jennifer A. Doudna and Thomas R. Cech. The chemical repertoire of natural ribozymes. *Nature*, 418:222–228, 2002.
- [142] 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.
- [143] Jason A. Wallace and Jana K. Shen. Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant pH. *J. Chem. Phys.*, 137:184105, 2012.
- [144] Garrett B. Goh, Jennifer L. Knight, and Charles L. Brooks. Constant pH molecular dynamics simulations of nucleic acids in explicit solvent. *J. Chem. Theory Comput.*, 8:36–46, 2012.
- [145] Garrett B. Goh, Jennifer L. Knight, and Charles L. Brooks III. pH-dependent dynamics of complex RNA macromolecules. *J. Chem. Theory Comput.*, 9:935–943, 2013.

- [146] Serena Donnini, Florian Tegeler, Gerrit Groenhof, and Helmut Grubmüller. Constant pH molecular dynamics in explicit solvent with λ -dynamics. *J. Chem. Theory Comput.*, 7:1962–1978, 2011.
- [147] Ulf Börjesson and Philippe H. Hünenberger. Explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines. *J. Chem. Phys.*, 114(22):9706–9719, 2001.
- [148] Maurice R. Eftink and Rodney L. Biltonen. Energetics of ribonuclease A catalysis. 1. pH, ionic strength, and solvent isotope dependence of the hydrolysis of cytidine cyclic 2',3'-phosphate. *Biochemistry*, 22:5123–5134, 1983.
- [149] Chiwook Park and Ronald T. Raines. Catalysis by ribonuclease A is limited by the rate of substrate association. *Biochemistry*, 42:3509–3518, 2003.
- [150] 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. doi: 10.1021/jp505777n.
- [151] R. Berisio, F. Sica, V. S. Lamzin, K. S. Wilson, A. Zagari, and L. Mazzarella. Atomic resolution structures of ribonuclease A at six pH values. *Acta Crystallogr. D.*, 58:441–450, 2002.
- [152] D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Salomon-Ferrer, R. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman. *AMBER 12*. University of California, San Francisco, San Francisco, CA, 2012.
- [153] Francisco Figueirido, Gabriela S. Del Buono, and Ronald M. Levy. On finite-size effects in computer simulations using the Ewald potential. *J. Chem. Phys.*, 103(14):6133–6142, 1995.
- [154] Stephen Bogusz, Thomas E. Cheatham III, and Bernard R. Brooks. Removal of pressure and free energy artifacts in charged periodic systems via net charge corrections to the Ewald potential. *J. Chem. Phys.*, 108(17):7070–7084, May 1998.
- [155] J. L. Markley. Correlation proton magnetic resonance studies at 250 MHz of bovine pancreatic ribonuclease. I. Reinvestigation of the histidine peak assignments. *Biochemistry*, 14(16):3546–3554, 1975.

- [156] Joyce E. Jentoft, Thomas A. Gerken, Neil Jentoft, and Dorr G. Dearborn. [^{13}C]Methylated Ribonuclease A, ^{13}C NMR studies of the interaction of lysine 41 with active site ligands. *J. Biol. Chem.*, 256:231–236, 1981.
- [157] D. H. Meadows and O. Jardetzky. Nuclear magnetic resonance studies of the structure and binding sites of enzymes. IV. Cytidine 3'-monophosphate binding to ribonuclease. *Proc. Natl. Acad. Sci. USA*, 13:406–413, 1968.
- [158] David J. Quirk and Ronald T. Raines. His . . . Asp catalytic dyad of ribonuclease A: Histidine pKa values in the wild-type, D121N, and D121A enzymes. *Biophys. J.*, 76:1571–1579, 1999.
- [159] A. Neuberger and K. Brocklehurst, editors. *Hydrolytic Enzymes*. New Comprehensive Biochemistry. Elsevier, Amsterdam, The Netherlands, 1987.
- [160] 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.
- [161] Junmei Wang, Wei Wang, Peter A. Kollman, and David A. Case. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.*, 25:247–260, 2006.
- [162] Pierre Turq, Frédéric Lantelme, and Harold L. Friedman. Brownian dynamics: Its application to ionic solutions. *J. Chem. Phys.*, 66:3039–3044, 1977.
- [163] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.
- [164] Alexey Onufriev, Donald Bashford, and David A. Case. Exploring protein native states and large-scale conformational changes with a modified Generalized Born model. *Proteins*, 55: 383–394, 2004.
- [165] 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.
- [166] Cuchillo C. M., Vilanova M., and Nogués M. V. *Ribonucleases: structures and functions*. Academic Press, New York, 1997.
- [167] R.T. Raines. Active site of ribonuclease a. *Nucl. Ac. Mol. Bio.*, 13:19–32, 2004.

- [168] Alvan C. Hengge, Aleksandra E. Tobin, and Cleland W. W. Studies of transition-state structures in phosphoryl transfer reactions of phosphodiester of *p*-nitrophenol. *J. Am. Chem. Soc.*, 117:5919–5926, 1995.
- [169] Gail Mitchell Emilsson, Shingo Nakamura, Adam Roth, and Ronald R. Breaker. Ribozyme speed limits. *RNA*, 9:907–918, 2003.
- [170] Brigitta Elsässer, Marat Valiev, and John H. Weare. A Dianionic Phosphorane Intermediate and Transition States in an Associative $A_N + D_N$ Mechanism for the RibonucleaseA Hydrolysis Reaction. *J. Am. Chem. Soc.*, 131:3869–3871, 2009.
- [171] Axel Brünger, Charles L. Brooks III, and Martin Karplus. Stochastic boundary conditions in molecular dynamics simulations of ST2 water. *Chem. Phys. Lett.*, 105:495–500, 1984.
- [172] 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.
- [173] Jacob Bigeleisen and Max Wolfsberg. Theoretical and experimental Aspects of Isotope Effects in Chemical Kinetics. *Adv. Chem. Phys.*, 1:15–76, 1958.
- [174] 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.
- [175] Zuzana Vokáčová, Miloš Buděšínský, Ivan Rosenberg, Bohdan Schneider, Jiří Šponer, and Vladimír Sychrovský. Structure and Dynamics of the ApA, ApC, CpA, and CpC RNA Dinucleoside Monophosphates Resolved with NMR Scalar Spin-Spin Couplings. *J. Phys. Chem. B*, 113:1182–1191, 2009.
- [176] Carmay Lim and Philip Tole. Endocyclic and exocyclic cleavage of phosphorane monoanion: a detailed mechanism of the RNase A transphosphorylation step. *J. Am. Chem. Soc.*, 114:7245–7252, 1992.
- [177] Ronald Breslow and William H. Chapman, Jr. On the mechanism of action of ribonuclease A: Relevance of enzymatic studies with a *p*-nitrophenylphosphate ester and a thiophosphate ester. *Proc. Natl. Acad. Sci. USA*, 93:10018–10021, 1996.
- [178] Timothy M. Glennon and Arieh Warshel. Energetics of the Catalytic Reaction of Ribonuclease A: A Computational Study of Alternative Mechanisms. *J. Am. Chem. Soc.*, 120:10234–10247, 1998.

- [179] Brigitta Elsässer, Marat Valiev, and John H. Weare. A Dianionic Phosphorane Intermediate and Transition States in an Associative $A_N + D_N$ Mechanism for the RibonucleaseA Hydrolysis Reaction. *J. Am. Chem. Soc.*, 131:3869–3871, 2009.
- [180] Reed M. Izatt, James J. Christensen, and J. Howard Rytting. Sites and thermodynamic quantities associated with proton and metal ion interaction with ribonucleic acid, deoxyribonucleic acid, and their constituent bases, nucleosides, and and nucleotides. *Chem. Rev.*, 71:439–481, 1971.
- [181] David Metzler. *Biochemistry: The Chemical Reactions Of Living Cells*. Academic Press, Inc, 1977.
- [182] Wolfram Saenger. *Principles of Nucleic Acid Structure*. Springer New York, 1984.
- [183] Philip C. Bevilacqua, Trevor S. Brown, Shu-ichi Nakano, and Rieko Yajima. Catalytic roles for proton transfer and protonation in ribozymes. *Biopolymers*, 73:90–109, 2004.
- [184] Martha J. Fedor. Comparative enzymology and structural biology of RNA self-cleavage. *Annu. Rev. Biophys.*, 38:271–299, 2009.
- [185] 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.
- [186] Shu-ichi Nakano, Durga M. Chadalavada, and Philip C. Bevilacqua. General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science*, 287:1493–1497, 2000.
- [187] Adegboyega K. Oyelere, Julia R. Kardon, and Scott A. Strobel. pK_a perturbation in genomic hepatitis delta virus ribozyme catalysis evidenced by nucleotide analogue interference mapping. *Biochemistry*, 41:3667–3675, 2002.
- [188] Sapna Ravindranathan, Samuel E. Butcher, and Juli Feigon. Adenine Protonation in Domain B of the Hairpin Ribozyme. *Biochemistry*, 39:16026–16032, 2000.
- [189] Neocles B. Leontis, Jesse Stombaugh, and Eric Westhof. The non-watson-crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, 30:3497–3531, 2002.
- [190] Jung C. Lee and Robin R. Gutell. Diversity of Base-pair Conformations and their Occurrence in rRNA Structure and RNA Structural Motifs. *J. Mol. Biol.*, 344:1225–1249, 2004.

- [191] Nathan A. Siegfried, Bernie OHare, and Philip C. Belivacqua. Driving Forces for Nucleic Acid pKa Shifting in an A⁺C Wobble: Effects of Helix Position, Temperature, and Ionic Strength. *Biochemistry*, 49:3225–3236, 2010.
- [192] Se Bok Jang, Li-Wei Hung, Young-In Chi, Elizabeth L. Holbrook, Richard J. Carter, and Stephen R. Holbrook. Structure of an RNA Internal Loop Consisting of Tandem C.A+ Base Pairs. *Biochemistry*, 37:11726–11731, 1998.
- [193] Jeffrey D. Levensgood, Carrie Rollins, Clay H.J. Mishler, Charles A. Johnson, Grace Miner, Prashant Rajan, Brent M. Znosko, and Blanton S. Tolbert. Aromatic rings act as hydrogen bond acceptors. *J. Mol. Biol.*, 415:680–698, 2012.
- [194] R. Parthasarathy, M. Malik, and S. M. Fridey. X-ray structure of a dinucleoside monophosphate A2'p5'C that contains a 2'-5' link found in (2'-5')oligo(A)s induced by interferons: single-stranded helical conformation of 2'-5'-linked oligonucleotides. *Proc. Natl. Acad. Sci. USA*, 79:7292–7296, 1982.
- [195] William J. Cook Howard Einspahr and Charles E. Bugg. Conformational flexibility in single-stranded oligonucleotides: crystal structure of a hydrated calcium salt of adenylyl-(3'-5')-adenosine. *Biochemistry*, 20:5788–5794, 1981.
- [196] Christopher L. Tang, Emil Alexov, Anna Marie Pyle, and Barry Honig. Calculation of pK_as in RNA: On the Structural Origins and Functional Roles of Protonated Nucleotides. *J. Mol. Biol.*, 366:1475–1496, 2007.