# COMPUTATIONAL STUDIES OF THE STRUCTURE, DYNAMICS, AND CATALYSIS IN THE HEPATITIS DELTA VIRUS RIBOZYME

BY

ABIR GANGULY

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Sharon Hammes-Schiffer, Chair
Professor Martin Gruebele
Professor Emad Tajkhorshid
Professor Susan Martinis

# Abstract

Ribozymes represent a unique class of RNA that is capable of catalyzing biochemical reactions. Since their discovery three decades ago, ribozymes have been shown to be involved in numerous crucial biological processes, such as gene regulation, translation, and RNA splicing. Similar to protein enzymes, these catalytic RNAs fold into well-defined tertiary structures with organized active sites to carry out catalysis. While their side chains do not possess the variety in functional groups accessible to the side chains of their protein counterparts, their efficiency as biological catalysts is often comparable to protein enzymes, making their catalytic strategies highly interesting. The study of ribozymes is also motivated by their possible significance in evolutionary biology.

The hepatitis delta virus (HDV) ribozyme is a small nucleolytic RNA that performs a phosphodiester self-cleavage reaction generating products with a 2', 3'-cyclic phosphate and a 5' -hydroxyl termini. This ribozyme was initially discovered in the RNA genome of the human pathogen HDV, where it played a crucial role in the viral life cycle, cleaving site-specifically the multimeric copies of the RNA genome into monomeric pieces. More recently, HDV-like ribozymes have been shown to be widespread across all kingdoms of life. Since their discovery, the HDV ribozymes have been intensely studied from structural as well as mechanistic perspectives, and much is known about their structure and catalytic strategies. The HDV ribozyme has a compact double-pseudoknot structure and uses a combination of metal ion and nucleobase catalysis to effect its self-cleavage reaction. An active site cytosine C75 is thought to act as a general acid in the catalytic reaction by donating a proton to the 5'-hydroxyl of the leaving group, and an active site $Mg^{2+}$ ion has been purported to play the role of a Lewis acid and activate the O2' nucleophile. The exact role of this putative catalytic ion is still uncertain. A

crystal structure also revealed a rare reverse G•U wobble close to the active site interacting with the putative catalytic ion. This base pair has been hypothesized to play an important role in positioning the metal ion for catalysis.

In this dissertation, the HDV ribozyme was studied using a variety of computational approaches. Classical molecular dynamics (MD) simulations and non-linear Poisson-Boltzmann (NLPB) calculations were utilized to study the metal binding characteristics of the reverse G•U wobble close to the active site of the ribozyme. These studies revealed that the reverse wobble creates a highly negative pocket that allows it to interact with metal ions and helps to shift the $pK_a$ of the nucleobase C75, thereby facilitating its protonation. MD simulations were also used to investigate the impact of C75 protonation and $Mg^{2+}$ ion interaction at the reverse G•U wobble on the structure as well as the motions of the HDV ribozyme. The protonated state of C75 was found to be essential for keeping the active site organized for catalysis. A localized, 'chelated' metal ion interaction was observed at the reverse G•U wobble, in contrast to a 'diffused' metal ion interaction observed at a standard G•U wobble also located close to the active site. The effects of mutation of the reverse G•U wobble, as well as the standard G•U wobble, to a Watson-Crick GC base pair were also studied. The overall tertiary structure and thermal motions of the ribozyme were not found to be significantly affected by C75 protonation, mutation of the reverse and standard wobbles, or the metal ion interaction at the two wobbles, suggesting that small local motions at the active site, rather than large-scale global motions, dominate the ribozyme reaction pathway.

Quantum mechanical/molecular mechanical (QM/MM) calculations were used to study the HDV ribozyme self-cleavage reaction and elucidate the role of the catalytic metal ion. The calculations suggested a concerted mechanism of the catalytic reaction in the presence of a divalent ion at the active site but a sequential mechanism in the presence of a monovalent ion at the same position. The divalent ion at the active site was found to lower the p$K_a$ of the nucleobase C75, making its proton donation more facile and thereby favoring the concerted mechanism. QM/MM calculations were also used to study the effects of phosphorothioate substitutions of the non-bridging oxygens at the scissile phosphate, commonly known as the 'thio effects', in the HDV ribozyme. In the case of the $R_P$ sulfur substrate, the calculations revealed a reactant state with pronounced active site distortion and an unfavorable reaction pathway with a high energetic barrier. In contrast, the reactant state of the $S_P$ sulfur substrate showed minimal distortion at the active site compared to the oxo-substrate. The results from these studies are consistent with several biochemical experimental studies.

The mechanism of the HDV ribozyme was further investigated using QM/MM free energy simulations to include conformational sampling and entropic effects. Umbrella sampling simulations were combined with a finite temperature string method to generate the multidimensional free energy surface underlying the self-cleavage reaction. The results of these simulations were qualitatively consistent with the previous QM/MM calculations, indicating a concerted mechanism in the presence of a $Mg^{2+}$ ion at the catalytic site and a sequential mechanism in the presence of a $Na^+$ ion. However, several new mechanistic insights were provided by the QM/MM free energy simulations, including the observation of proton transfer from the exocyclic amine of protonated C75 to the nonbridging oxygen of the scissile phosphate to stabilize the phosphorane intermediate in the sequential mechanism. The free energy barrier along the concerted pathway in the presence of the catalytic $Mg^{2+}$ ion was consistent with the intrinsic reaction rate of the HDV ribozyme cleavage reaction measured experimentally. The differences in the reaction pathways of the cleavage reaction in the presence of the $Mg^{2+}$ ion and the $Na^+$ ion illustrated several key roles of the catalytic metal ion in the HDV ribozyme catalysis, including activation of the O2’ nucleophile, acidification of the general acid C75, and stabilization of the non-bridging oxygen of the scissile phosphate.

# ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Professor Sharon Hammes-Schiffer for her constant support and guidance during the course of my graduate school. The scientific training that I received while working in her group, the exciting projects that I was fortunate enough to work on, and the freedom that I was given to pursue my research are few of the many reasons for which I am thankful to her. She, therefore, is credited with much of the success of this thesis.

The second most influential person during my graduate research has been Dr. Philip C. Bevilacqua, Professor of Chemistry at the Pennsylvania State University. He has been an inspiring teacher, a fantastic experimental collaborator, and borderline co-advisor. Many, if not all, of my projects have benefitted from his keen experimental perspectives, and I cannot stress enough on his importance in the fruition of this dissertation. I would like to thank him deeply for the invaluable suggestions, supervision, and encouragement that he provided throughout my graduate studies.

I would like to thank all the graduate students and postdoctoral fellows with whom I had the good fortune to work with in the Sharon Hammes-Schiffer research group. Be it scientific inputs as colleagues or just great company as friends, their contribution to this thesis have been significant. I would specially like to thank my colleagues Dr. Narayanan Veeraraghavan and Dr. Philip Hanoian for their help and support at the beginning of my graduate school, and my batch mates Dr. Andrew Sirjoosingh and Dr. Brian Solis, if for nothing less, then for keeping me honest all this time.

I am also deeply indebted to Dr. Alexander Soudackov for the innumerable hours that he spent on me over various discussions during the last five years. He has been an indispensable resource of scientific and technical knowledge in the Hammes-Schiffer research group, and without the acknowledgment of his contribution to this thesis, this section will be incomplete. I would like to thank him sincerely for all the help and support that he provided during various stages of my graduate study.

My gratitude also goes to all the students in the Bevilacqua research group with whom I had the pleasure to collaborate. In particular, I would like to thank Pallavi Thaplyal who has been a dear friend and a fantastic colleague.

Lastly, I would like to thank my parents and my entire family for their support and affection. I have been extremely fortunate to have parents who stood by me during the highs and the lows, at every step of my life. Without them it would have been impossible to come this far.

*In my darkest night,*

*when the moon was covered*

*and I roamed through wreckage,*

*a nimbus-clouded voice*

*directed me:*

*"Live in the layers,*

*not on the litter."*

*Though I lack the art*

*to decipher it,*

*no doubt the next chapter*

*in my book of transformations*

*is already written.*

*I am not done with my changes.*

*-- "The Layers" by Stanley Kunitz*

To my parents, family and friends.

Caffeine. Scotch.

# TABLE OF CONTENTS

# CHAPTER 1: INTRODUCTION

The capability of RNA to perform catalysis was first demonstrated by Cech and coworkers in 1982,[1] when they reported the self-splicing nature of a pre-ribosomal RNA intron in absence of enzymes. The discovery of RNA enzymes, or *Ribozymes*, established RNA as a unique class of biological molecule capable of storing information as well as performing catalytic function, and hinted at the possible significance of RNA in evolutionary biology.[2] The finding also provided a new perspective on biological catalysis, a field that solely belonged to protein enzymes. Three decades later, while ribozymes remain a hot topic of research, a lot has been learned about their functions and properties. Several other ribozymes have been identified performing diverse chemistry and playing key roles in important biological processes.[3] Much like enzymes, certain ribozymes such as RNase P have been shown to catalyze reactions with multiple turnovers, without undergoing self-alteration.[4] Examples of ribozymes acting on exogenous substrates (trans-cleaving) have also been documented.[5,6] Therapeutic approaches, utilizing the highly sequence-specific cleavage reactions of the ribozymes, have been devised to regulate the gene expression of specific m-RNAs.[7,8]

Unlike the protein enzymes, the ribozymes do not have a large variety of functional groups at their disposal. The diversity in the RNA side chain is limited to the four nucleobases: guanosine, cytosine, adenosine, and uracil. The $pK_a$ values of these nucleobases in solution are far from neutrality – $pK_a$ of N3 of cytosine, and N1 of adenosine ~ 4; $pK_a$ of N3 of uracil, and N1 of guanosine ~ 9; $pK_a$ of O2′ of ribose ~ 12.5 – which suggests that they would be reluctant to participate in chemistry near physiological conditions.[9] Furthermore, the highly charged nature of the RNA and its backbone flexibility should, in principle, prevent the ribozyme from achieving a precise active site organization, as exhibited typically by the protein enzymes, to facilitate catalysis.[10] Yet, the ribozymes have been shown to be very efficient catalysts. The rate enhancements in certain ribozyme-catalyzed reactions relative to the corresponding reactions in solution are comparable to rate enhancements observed in protein enzyme-catalyzed reactions.[11-13] The ability of ribozymes to perform efficient catalysis with such limited resources generates keen interest in their mechanistic approaches.

As a common strategy, the ribozymes are known to use specific nucleobases advantageously in their catalytic reaction. The highly anisotropic environment within the ribozyme can shift the $pK_a$ of these nucleobases towards neutrality, often allowing them to take part in catalysis in form of a general base or an acid.[14-16] The hydrogen bonding capability of the nucleobases also play a critical role, and can assist catalysis in several ways including binding of the substrate, stabilizing the active site, and stabilizing the transition state of the reaction.[17-20] Another crucial aspect of ribozyme catalysis is the role of metal ions, which can be both structural as well as functional. The folding of the highly negative structure of the ribozyme into an active conformation requires metal ions for stabilization. These metal ions can also take part in catalysis, either in form of a Lewis acid or by providing negative charge stabilization or by shifting the $pK_a$ of nearby nucleobases.[21,22] Metal-coordinated water molecules have also been suggested to take part in general acid-base catalysis.[23]

The majority of the naturally occurring ribozymes carry out trans-esterification reactions in which a nucleophile attacks a phosphorus center in an $S_N2$ fashion and results in the cleavage of a phosphorus-oxygen bond.[24] Notable exceptions are the ribozymes RNase P and the ribosome that carry out a phosphate hydrolysis reaction and a peptidyl transfer reaction, respectively. [25,26] Depending on their size and function, they can be broadly classified into two classes: large ribozymes that function to unite flanking sequences, and small ribozymes, also referred to as nucleolytic ribozymes that perform site-specific self-cleavage and ligation reactions.[24] The large ribozymes require binding of substrates at their active site and uses an external nucleophile to cleave RNA substrates generating termini with 5′-phosphate and 3′-hydroxyl groups. They have been shown to employ metal ion catalysis in their chemical reactions. Divalent ions, in particular, have been shown to be critical for the functioning of these ribozymes. Examples of this class of ribozymes include the group I and II self-splicing introns. The nucleolytic ribozymes, on the other hand, does not rely on an external nucleophile. The vicinal 2′-hydroxyl attacks the phosphorus center producing a 2′, 3′-cyclic phosphate and 5′-hydroxyl termini. In contrast to the metal ion assisted catalysis in the large ribozymes, the catalysis in nucleolytic ribozymes have been shown to be mostly independent of metal ions. [27]

There are five known members in the nucleolytic ribozyme family: the hairpin, hammerhead, GlmS, VS (Varkud satellite), and the hepatitis delta virus (HDV) ribozyme.[24] The key strategy of catalysis in the majority of these small ribozymes has been shown to be nucleobase centered. For example, in the hairpin ribozyme, a guanine (G8) and an adenine (A38) have been identified as important for the functioning of the ribozyme.[17,18] While the exact roles of these bases are still inconclusive, the crystal structures reveal that the G8 and the A38 are appropriately positioned to act as a general base and acid respectively.[19] Similarly, a guanine (G40) is purported to play the role of the general base in the GlmS ribozyme,[28] while two guanines (G8 and G12) have been shown to be critical for catalysis in the hammerhead ribozyme.[29,30] The HDV ribozyme is particularly interesting because it utilizes a combination of metal ion and nucleobase catalysis to effect its self-cleavage reaction.[23,31]

The hepatitis delta virus ribozyme was first discovered in the hepatitis delta virus (HDV), a satellite virus of the hepatitis B virus (HBV); the presence of HDV has been shown to worsen the symptoms of hepatitis B.[32,33] HDV has a circular RNA genome that replicates via a rolling-circle mechanism and encodes the hepatitis delta antigen protein. The HDV ribozyme occurs as a ~85 nt sequence in the RNA genome and plays a critical role in the viral replication process. During the rolling-circle mechanism, linear multimeric antigenomic RNAs are synthesized that are site-specifically severed into monomers by the antigenomic HDV ribozyme. These linear antigenomic RNAs can undergo ligation and subsequently act as templates for the synthesis of multimeric genomic RNAs that are then cleaved into monomers by the genomic HDV ribozyme. The genomic and antigenomic forms of the HDV ribozyme are closely related. HDV-like ribozymes have now been shown to exist across all kingdoms of life, including the human genome, thus generating considerable interest in the understanding of the functioning of these molecules.

In the last two decades, a wealth of information has been gathered on the HDV ribozymes from biochemical as well as X-ray crystallographic studies. From the structural perspective, historically there have been three major crystals structures of the HDV ribozymes reported till date. In 1998, Doudna and coworkers reported the first structure of the cleaved (product) form of the ribozyme.[34] The structure revealed a double-pseudoknotted geometry of the ribozyme, with five pairing regions, P1-P4 and P1.1. The regions P1, P4, P1.1 form one coaxial stack, and P2, P3 form another. The cleavage site, flanked by a uracil (U-1) and a guanosine (G1), occurs between the regions P1 and P1.1. The structure also revealed an active site cytosine (commonly referred to as C75) forming a hydrogen bond with the 5′-oxygen of the leaving group (G1), suggesting a protonated state of C75 pre-cleavage and a general acid role of the nucleobase in the catalytic reaction. This observation was consistent with previous studies that showed the $pK_a$ of C75 to be shifted > 2 units towards neutrality. There was no evidence, however, of metal ions at the cleavage site. This was surprising since biochemical studies have shown that divalent ions are critical for the functioning of the ribozyme. Next,[35] a set of precleaved structures of the HDV ribozyme was published, in which the cleavage reaction was thwarted either by mutating the nucleobase C75 to a uracil, or by removing $Mg^{2+}$ ions from the solution. The C75U structure was consistent with the existing product structure, except that the U75 was positioned to act as a general base in the reaction, and a $Mg^{2+}$ ion was located at the general acid position. Further studies, however, have suggested that this structure could be result of a misfolding, and the C75U mutation can have serious repercussions for the active site geometry. More recently, a third crystal structure of the HDV ribozyme was solved without any C75 mutation or omission of $Mg^{2+}$ ions.[31] The cleavage reaction was inhibited by replacing the sugar of U-1 by a deoxyribose. The structure was solved at a low pH, where C75 should remain protonated and thus, catalytically active. Unlike the C75U structure, this structure was much more similar to the cleaved structure, with the C75 positioned to act as a general acid. Interestingly, a $Mg^{2+}$ ion was resolved very close to the active site, interacting with the 2′-hydroxyl of U-1. This positioning of the $Mg^{2+}$ ion was favorable for the ion to act as a Lewis acid, stabilizing the O2′ nucleophile.

Another key revelation in this recent pre-cleaved structure was the evidence of a reverse wobble close to the active site, involving the residues G25 and U20, in addition to the already known occurrence of a standard wobble between the residues G1 and U37. Wobbles are non-Watson-Crick base pairs consisting of two hydrogen boonds, and are known to interact with metal ions through their major and minor grooves. In a standard wobble, the participating bases are both in *anti* conformation that allow the pair to interact with metal ions through their major groove. In a reverse wobbles, which is a much rarer motif, one of the involved bases are in *syn* conformation, that enables the base pair to interact with metal ions through their minor groove.

The G25•U20 reverse wobble is conserved across all the HDV ribozymes,[36] and its mutation to a A•U base pair results in a ~1000-fold reduction in cleavage rate.[37] It interacts with the catalytic metal ion through its hydration shell, and possibly plays a role is positioning the metal ion for catalysis. In contrast, the conservation percentage of the standard G•U wobble in the HDV ribozymes is ~80%, and its mutation to a G•C pair has shown no significant effects on the rate of catalysis.[38]

In recent years, owing largely to the advancement of computers, computational methodologies have been used extensively to study biological systems, validate experiments, and make new predictions. A wide variety of computational tools have been developed to address the broad range of questions that are associated with a complex biomolecule or a biological process. These methods range from accurate quantum mechanical calculations to atomistic classical simulations to mixed quantum-classical approaches to coarse-graining techniques, and are intended to explore the different aspects of biological systems. While each of these methods has their own limitation, the large repertoire of available methods allows the choice of an appropriate tool for a specific question.

The intent of this dissertation is to explore the different aspects of the HDV ribozyme molecule using a variety of computational tools. The questions addressed in this thesis can be loosely classified into two categories that are interrelated at some level. In the first category, the structural aspects of the ribozyme are explored. Some of the issues addressed in this category include the identification of rigid and flexible regions in the ribozyme, importance of hydrogen-bonding networks, role of electrostatics, function of specific motifs in the ribozyme, and interaction of the ribozyme with metal ions. In the second category, the catalysis in the HDV ribozyme is investigated in detail. The mechanistic approaches of the catalytic reaction are explored, and the roles of the nucleobase C75 and the catalytic $Mg^{2+}$ ion in the reaction are highlighted.

Classical molecular dynamics (MD) simulations have been employed extensively to study the structural dynamics and functionally relevant conformational changes in biomolecules. In the traditional MD simulations, the system is modeled at an atomic level, and is evolved in time by integrating the equations of motion numerically. The interactions between the atoms are described by molecular mechanical (MM) potentials, which are atomistic, pairwise additive interaction potentials derived typically from experimental data and high-level quantum mechanical calculations (force field). In spite of obvious limitations such as approximate force fields, finite simulation time-scales and limited sampling, MD with MM force fields has been shown to be capable of providing meaningful and predictive information on the dynamics of biomoecules in the vicinity of a crystallographically defined energy minimum. Careful analysis of MD simulations can provide insight into the structural roles of specific residues, predict favorable spatial conformations, identify rigid and flexible regions within the biomolecule, elucidate the role of electrostatics and help refine unresolved regions of x-ray crystal structures.

In the past, several MD studies have been performed on the HDV ribozyme, based on the product (cleaved) and C75U mutant structures. Simulations based on the product structure have been used to study the long-distance communication between the active site of the ribozyme and a distal C41 base triple,[39] while simulations starting with the C75U structure have been used extensively to characterize the structure and dynamics of the precleaved form of the HDV ribozyme. Initial simulations based on the C75U mutant structure implicated C75 as a general base in the mechanism.[40] This result, while consistent with the C75U crystal structure, did not agree with the product structure as well as several other biochemical studies. However, more recent longer simulations on the C75U mutant structure have shown that the C75 moves from its initial position and reorients itself to a new position from where it can act as a general acid.[41]

In this thesis, the various studies of the HDV ribozyme using MD simulations are described in chapters 2, 3, and 4. In chapter 2, the metal binding characteristics of the reverse G•U wobble was studied in the product and precleaved forms.[42] MD simulations in conjunction with non-linear Poisson-Boltzmann (NLPB) calculations were used to understand the interaction of this rare base pair with metal ions. These studies revealed a negative pocket close to the site of the reverse wobble, that possibly serves to bind metal ions and help shift the $pK_a$ of nucleobase C75.

In chapter 3, MD simulations were performed to analyze metal ion interaction with the reverse and standard G•U wobbles and to investigate the importance of the C75 protonated state in the structure and motions of the ribozyme.[43] Two different kinds of metal ion interaction was observed: 'chelated' at the reverse G•U wobble, and 'diffused' at the standard G•U wobble, that appeared to play a role in catalysis and stability, respectively. In chapter 4, a double mutant of the HDV ribozyme, in which the reverse G•U wobble was mutated to an isosteric GC base pair, was studied to elucidate the role of the putative catalytic metal ion in the cleavage reaction.[44] The metal binding pocket of the GC base pair is expected to be less negative than the reverse G•U base pair, resulting in the ejection of the $Mg^{2+}$ ion at the active site and causing a subsequent decrease in the reaction rate. The MD simulations along with biochemical experiments supported a model of the cleavage reaction in which the Mg2+ ion bound to the reverse G•U wobble is catalytic and functions either as a Lewis acid, a Brønsted base, or both to activate the O2′ nucleophile.

While classical MD simulations provide an excellent approach to study the equilibrium properties of biomolecules, description of chemical reactions, however, is beyond the capability of classical MD with MM potentials. In order to explore the mechanism of a reaction, accurate description of the electrons of the relevant atoms and their redistribution during the reaction is required. Traditionally, quantum mechanical (QM) methods have been used to describe reactions. Since the quantum mechanical treatment of the entire ribozyme is computationally too expensive, one approach has been to study a reduced model of the active site, containing only dozens of atoms treated quantum mechanically, with constraints applied on selected peripheral atoms. While this is a reasonable approach in many systems, the exclusion of long-distance interactions, which can be particularly important in the case of highly charged systems such as the ribozymes, can lead to unphysical results.

The introduction of the hybrid quantum mechanical/molecular mechanical (QM/MM) approaches brought a paradigm shift in the study of biochemical reactions.[45] The idea of partitioning the system into QM and MM regions is based on the assumption that most of the electron redistribution during a reaction is limited to a small number of atoms, while the rest of the atoms in the system may be described adequately by classical means. The methodology provides the valuable combination of the accuracy of the QM calculations with the efficiency of the force field based classical methods. Within the QM/MM framework, the potential energy surface underlying the biochemical reaction can be characterized by identifying various minima, transition states, and minimum energy pathways of the reaction using traditional techniques.

There have been a handful of studies on the HDV ribozyme using QM calculations in the past. Most of these calculations have been based on small model systems. In one study, density functional theory (DFT) was used to perform geometry optimizations on two cutoff models of the ribozyme active site, immersed in a dielectric continuum aqueous solvent.[46] The model with C75 acting as a general acid revealed a phosphorane intermediate and was found to be more favorable than the model with C75 acting as a general base based on reaction energetics and estimated concentrations. In another similar DFT study of the ribozyme active site based models,[47] the results were consistent with the general acid role of C75, but showed no stable intermediate. In a third study,[48] QM/MM calculations were performed on the solvated HDV ribozyme, and the results favored the C75 general base mechanism. However, this study was based on the C75U mutant structure, which has been shown to be not reflective of the catalytic conformation of the ribozyme by previous studies.[49]

Chapter 5 of this dissertation discusses a QM/MM study of the HDV ribozyme self-cleavage reaction and the impact of the catalytic metal ion in the mechanism.[50] The QM/MM calculations, starting from the recent precleaved crystal structure, revealed a concerted mechanism in presence of a divalent ion at the catalytic site, but a sequential mechanism with a phosphorane intermediate in presence of a monovalent ion at that site. Electrostatic potential calculations suggested that the divalent ion at the catalytic site lowers the pKa of C75, making its proton donation more facile, thereby resulting in the concerted mechanism.

In chapter 6, QM/MM calculations and biochemical experiments were combined to study phosphorothioate substitutions at the scissile phosphate of the HDV ribozyme.[51] Site-specific thio-substitution is a common experimental technique used in ribozymes to probe interactions at specific sites. The change in reaction rate due to the sulfur substitution, called the ‘thio effect’, can provide key mechanistic insights. The QM/MM calculations indicated that a sulfur substitution of the *pro-*$R_P$ oxygen introduced active site distortions and weakened the interaction of that site with the catalytic $Mg^{2+}$ ion, in accordance with the experimentally measured thio-effect.

While these types of QM/MM calculations can be used to obtain valuable qualitative mechanistic insights, they have few obvious limitations. The potential energy surfaces of biological systems are known to be complex, containing numerous related minima and transition states. Due to the lack of conformational sampling from these calculations, minima and transition states that are identified are often biased towards the starting configurations. The energy barriers of the minimum energy pathways are also not quantitatively meaningful because of the exclusion of entropic effects. The logical approach to avoid these limitations, then, would be to include conformational sampling in these QM/MM methods and characterize the free energy surface governing the biochemical reaction. The free energy surface of the reaction dictates the thermodynamically favorable pathways of the reaction, and the free energy barriers along these pathways can be qualitatively compared to the experimentally measured reaction rate constants.

One possibility of exploring free energy surfaces underlying chemical reactions is to perform free energy MD simulations within the QM/MM framework. However, QM/MM MD simulations can be computationally demanding. Additionally, most biochemical reactions involve a large number of reaction coordinates whose identities are not straightforward beforehand. As a result, mapping out entire multi-dimensional free energy surfaces underlying biochemical reactions using QM/MM free energy simulations is mostly not a feasible option. In the recent past, several approaches have been developed that get around these limitations by focusing only on the regions of conformational space that are relevant to a given biochemical process. Many of these approaches are based on various implementations of the string method.

The string method was initially formulated to study rare events.[52] In this method, the reaction pathway is described by a string (a curve) passing through the multidimensional space of a set of chosen coarse coordinates that are involved in the reaction. The string is then evolved in the lower-dimensional space of the coarse coordinates either on the potential energy surface (the zero temperature string method)[52] or on the free energy surface (the finite temperature string method)[53] using various techniques, to generate the minimum energy path (MEP) or the minimum free energy path (MFEP) for the reaction, respectively.

One particular implementation of the string method involves combining the finite temperature string approach with umbrella sampling simulations.[54] In this implementation, the string is evolved based on umbrella sampling simulations. A key feature of this approach is that the process of identification of the MFEP of a reaction is accompanied by the characterization of its multi-dimensional free energy surface. Thus, the method not only identifies favorable pathways of the reaction, but also the free energy barriers along these pathways. The number of coarse coordinates considered in the simulations does not add significantly to its computational expense making the method suitable for studying complex biochemical reactions. Also, the inclusion of a set of possible coarse coordinates obviates the requirement of the *a priori* knowledge of the coordinates important for the reaction.

Chapter 7 describes the characterization of the multi-dimensional free energy surface underlying the HDV ribozyme self-cleavage reaction using a combination of umbrella sampling simulations and a finite temperature string method.[55] The results from these simulations indicated a concerted reaction pathway with the $Mg^{2+}$ ion at the active site. The free energy barrier of ~13 kcal/mol along this pathway was consistent with free energy barriers extrapolated from experiments. In contrast, the reaction pathway in presence of a $Na^+$ ion at the active site was found to be sequential passing through a phosphorane intermediate, with free energy barriers of ~2-4 kcal/mol for both steps. Analysis of the different mechanisms of the cleavage reaction with the divalent and monovalent ions at the active site lead to the identification of three distinct roles of the catalytic metal ion in the cleavage reaction: assisting in the activation of the O2′ nucleophile, acidifying the nucleobase C75, and stabilizing the non-bridging oxygen of the scissile phosphate.

## References

(1) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. *Cell* **1982**, *31*, 147-57.

(2) Cech, T. R. Self-Splicing RNA: Implications for Evolution. In *International Review of Cytology*; G.H. Bourne, J. F. D., Jeon, K. W., Eds.; Academic Press, 1985; Vol. Volume 93; pp 3-22.

(3) *Ribozymes and RNA Catalysis*; Royal Society Chemistry, 2008.

(4) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. *Cell* **1983**, *35*, 849-57.

(5) Zaug, A. J.; Cech, T. R. *Science* **1986**, *231*, 470-75.

(6) Uhlenbeck, O. C. *Nature* **1987**, *328*, 596-600.

(7) Weng, D.; Masci, P.; Radka, S.; Jackson, T. E.; Weiss, P.; Ganapathi, R.; Elson, P.; Capra, W.; Parker, V.; Lockridge, J.; Cowens, J. W.; Usman, N.; Borden, E. *Molecular cancer therapeutics* **2005**, *4*, 948-55.

(8) Li, M.-J.; Kim, J.; Li, S.; Zaia, J.; Yee, J.-K.; Anderson, J.; Akkina, R.; Rossi, J. *Molecular therapy* **2005**, *12*, 900-9.

(9) Bevilacqua, P. C.; Brown, T. S.; Nakano, S.; Yajima, R. *Biopolymers* **2004**, *73*, 90-109.

(10) Narlikar, G. J.; Herschlag, D. *Annu. Rev. Chem.* **1997**, *66*, 19-59.

(11) Herschlag, D.; Cech, T. R. *Biochemistry* **1990**, *29*, 10159-71.

(12) Hertel, K. J.; Herschlag, D.; Uhlenbeck, O. C. *Biochemistry* **1994**, *33*, 3374-85.

(13) Radzicka, A.; Wolfenden, R. *Science* **1995**, *267*, 90-93.

(14) Bevilacqua, P. C.; Yajima, R. *Curr. Opin. Chem. Biol.* **2006**, *10*, 455-64.

(15) Cochrane, J. C.; Strobel, S. A. *Acc. Chem. Res.* **2008**, *41*, 1027-35.

(16) Lilley, David M. J. *Biochem. Soc. T* **2011**, *39*, 641-46.

(17) Kuzmin, Y. I.; Da Costa, C. P.; Fedor, M. J. *J. Mol. Biol.* **2004**, *340*, 233-51.

(18) Kuzmin, Y. I.; Da Costa, C. P.; Cottrell, J. W.; Fedor, M. J. *J. Mol. Biol.* **2005**, *349*, 989-1010.

(19) Salter, J.; Krucinska, J.; Alam, S.; Grum-Tokars, V.; Wedekind, J. E. *Biochemistry* **2005**, *45*, 686-700.

(20) Bevilacqua, P.; Yajima, R. *Curr. Opin. Chem. Biol.* **2006**, *10*, 455-64.

(21) Fedor, M. J. *Curr. Opin. Struc. Biol.* **2002**, *12*, 289-95.

(22) Woodson, S. A. *Curr. Opin. Chem. Biol.* **2005**, *9*, 104-09.

(23) Nakano, S.; Chadalavada, D. M.; Bevilacqua, P. C. *Science* **2000**, *287*, 1493-97.

(24) Lilley, David M. J.; Eckstein, F. *The Royal Society of Chemistry, Cambridge* **2007**.

(25) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 920-30.

(26) Evans, D.; Marquez, S. M.; Pace, N. R. *Trends in Biochemical Sciences* **2006**, *31*, 333-41.

(27) Lilley, D. M. J. *Trends in Biochemical Sciences* **2003**, *28*, 495-501.

(28) Klein, D. J.; Been, M. D.; Ferre-D'Amare, A. R. *J. Am. Chem. Soc.* **2007**, *129*, 14858-+.

(29) Uebayasi, M.; Uchimaru, T.; Sawata, S.; Shimayama, T.; Taira, K.; Koguma, T. *J. Org. Chem.* **1994**, *59*, 7414-20.

(30) Lee, T.-S.; Silva-Lopez, C.; Martick, M.; Scott, W. G.; York, D. M. *J. Chem. Theory. Comput.* **2007**, *3*, 325-27.

(31) Chen, J.-H.; Yajima, R.; Chadalavada, D. M.; Chase, E.; Bevilacqua, P. C.; Golden, B. L. *Biochemistry* **2010**, *49*, 6508-18.

(32) Lai, M. M. C. *Annu. Rev. Chem.* **1995**, *64*, 259-86.

(33) Been, M. D. HDV Ribozymes. In *Curr. Top. Microbiol. Immunol.*; Casey, J. L., Ed.; Springer Berlin Heidelberg, 2006; Vol. 307; pp 47-65.

(34) Ferre-D'Amare, A. R.; Doudna, J. A. *J. Mol. Biol.* **2000**, *295*, 541-56.

(35) Ke, A.; Zhou, K.; Ding, F.; Cate, J. H. D.; Doudna, J. A. *Nature* **2004**, *429*, 201-05.

(36) Webb, C.-H. T.; Riccitelli, N. J.; Ruminski, D. J.; Lupták, A. *Science* **2009**, *326*, 953-53.

(37) Tanner, N. K.; Schaff, S.; Thill, G.; Petit-Koskas, E.; Crain-Denoyelle, A.-M.; Westhof, E. *Current Biology* **1994**, *4*, 488-98.

(38) Chadalavada, D. M.; Cerrone-Szakal, A. L.; Bevilacqua, P. C. *RNA* **2007**, *13*, 2189-201.

(39) Veeraraghavan, N.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Mol. Biol.* **2010**, *402*, 278-91.

(40) Krasovska, M. V.; Sefcikova, J.; Spackov·, N. a.; Sponer, J.; Walter, N. G. *J. Mol. Biol.* **2005**, *351*, 731-48.

(41) Lee, T.-S.; Giambasu, G. M.; Harris, M. E.; York, D. M. *J. Phys. Chem. Lett.* **2011**, *2*, 2538-43.

(42) Veeraraghavan, N.; Ganguly, A.; Chen, J.-H.; Bevilacqua, P. C.; Hammes-Schiffer, S.; Golden, B. L. *Biochemistry* **2011**, *50*, 2672-82.

(43) Veeraraghavan, N.; Ganguly, A.; Golden, B. L.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Phys. Chem. B* **2011**, *115*, 8346-57.

(44) Chen, J.; Ganguly, A.; Miswan, Z.; Hammes-Schiffer, S.; Bevilacqua, P. C.; Golden, B. L. *Biochemistry* **2013**, *52*, 557-67.

(45) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227-49.

(46) Wei, K.; Liu, L.; Cheng, Y.-H.; Fu, Y.; Guo, Q.-X. *J. Phys. Chem. B* **2007**, *111*, 1514-16.

(47) Liu, H.; Robinet, J. J.; Ananvoranich, S.; Gauld, J. W. *J. Phys. Chem. B* **2007**, *111*, 439-45.

(48) Banas, P.; Rulisek, L.; Hanosova, V.; Svozil, D.; Walter, N. G.; Sponer, J.; Otyepka, M. *J. Phys. Chem. B* **2008**, *112*, 11177-87.

(49) Nakano, S.; Bevilacqua, P. C. *Biochemistry* **2007**, *46*, 3001-12.

(50) Ganguly, A.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Phys. Chem. Lett.* **2011**, *2*, 2906-11.

(51) Thaplyal, P.; Ganguly, A.; Golden, B. L.; Hammes-Schiffer, S.; Bevilacqua, P. C. *Biochemistry* **2013**, *52*, 6499-514.

(52) E, W.; Ren, W.; Vanden-Eijnden, E. *Phys. Rev. B* **2002**, *66*, 052301.

(53) Ren, W. Q.; Vanden-Eijnden, E. *J. Phys. Chem. B* **2005**, *109*, 6688-93.

(54) Rosta, E.; Nowotny, M.; Yang, W.; Hummer, G. *J. Am. Chem. Soc.* **2011**, *133*, 8934-41.

(55) Ganguly, A.; Thaplyal, P.; Rosta, E.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Am. Chem. Soc.* **2014**, *136*, 1483-96.

# CHAPTER 2: METAL BINDING MOTIF IN THE ACTIVE SITE OF THE HDV RIBOZYME BINDS DIVALENT AND MONOVALENT IONS[1]

## Abstract

The hepatitis delta virus (HDV) ribozyme uses both metal ion and nucleobase catalysis in its cleavage mechanism. A reverse G•U wobble was observed in a recent crystal structure of the precleaved state. This unusual base pair positions a $Mg^{2+}$ ion to participate in catalysis. Herein, we used molecular dynamics (MD) and X-ray crystallography to characterize the conformation and metal binding characteristics of this base pair in product and precleaved forms. Beginning with a crystal structure of the product form, we observed formation of the reverse G•U wobble during MD trajectories. We also demonstrated that this base pair is compatible with the diffraction data for the product-bound state. During MD trajectories of the product form, $Na^+$ ions interacted with the reverse G•U wobble in the RNA active site, and a $Mg^{2+}$ ion, introduced in certain trajectories, remained bound at this site. Beginning with a crystal structure of the precleaved form, the reverse G•U wobble with bound $Mg^{2+}$ remained intact during MD simulations. When we removed $Mg^{2+}$ from the starting precleaved structure, $Na^+$ ions interacted with the reverse G•U wobble. In support of the computational results, we observed competition between $Na^+$ and $Mg^{2+}$ in the precleaved ribozyme crystallographically. Non-linear Poisson-Boltzmann calculations revealed a negatively charged patch near the reverse G•U wobble. This anionic pocket likely serves to bind metal ions and to help shift the p$K_a$ of the catalytic nucleobase, C75. Thus, the reverse G•U wobble motif serves to organize two catalytic elements, a metal ion and catalytic nucleobase, within the active site of the HDV ribozyme.

---

[1] This chapter was published in its entirety in the journal *Biochemistry* in 2011. The complete reference to the journal article is as follows: Veeraraghavan, N.; Ganguly, A.; Chen, J-.H.; Bevilacqua, P. C.;

RNA is involved in many aspects of biology, where it serves both informational and functional roles (*1-3*). Indeed, RNA can act as a riboswitch, binding small molecules and regulating gene expression (*4, 5*), and as an enzyme, cleaving phosphodiester bonds during catalysis and driving peptide bond formation on the ribosome (*6*). These functions require the RNA to attain a precise three-dimensional structure (*7-12*) and utilize key catalytic strategies, including general acid-base and metal ion catalysis (*13-16*).

A challenge in studying RNA is determining functionally relevant structures at high resolution (*17-22*). Over the course of the last 15 years, X-ray crystallography has revolutionized our understanding of RNA structure and function Crystal structures, however, provide only a snapshot of a molecule, usually trapped in a catalytically incompetent state. Furthermore, crystals of RNAs often diffract X-rays to only moderate resolution (2.8 Å or worse), and key catalytic regions within these structures can be disordered (*17-22*). As a result, fitting structural models to electron density data can be ambiguous, which can lead to uncertainties in the RNA structure. Even high-resolution crystal structures often have local regions of disorder, and highly ordered regions can change in conformation during the course of a reaction. Theoretical approaches such as molecular dynamics (MD)[1] (*24-28*) have the potential to both decrease ambiguities in the available RNA crystal structures and provide insight into motions inherent to RNA molecules.

The structure and function of small ribozymes are of growing interest (*6, 29*). The hepatitis delta virus (HDV) ribozyme occurs as two closely related ~85 nt double-pseudoknotted genomic and antigenomic versions (Figure 2.1A) that function to linearize the RNA concatemers that form during replication of the genome (*30-32*). A closely related, highly reactive version of this

---

Hammes-Schiffer, S.; Golden, B. L., *Biochemistry*, **2011**, *50*, 2762-2682. N.V. performed the computational analysis, A. G. performed the simulations, and J. C. H. performed the experiments.

ribozyme also occurs in the human genome in an intron of the *CPEB3* gene (*33, 34*). Moreover, HDV and HDV-like ribozymes are widespread, occurring in plants, fish, and insects (*35*), making their chemical mechanisms of heightened interest.

The HDV ribozyme self-cleaves using a combination of metal ion and general acid-base catalysis, in which a cytosine nucleobase, C75, acts as a general acid, and a divalent metal ion acts as a Lewis acid (*36-42*) (Figure 2.1B). The $pK_a$ of C75 is shifted toward neutrality in the precleaved but not the product state (*36, 37, 43-45*), and C75 appears to donate a proton in the cleavage reaction (*39, 41, 46-48*). The role of the divalent metal ion in the reaction is less clear; however, a $Mg^{2+}$ ion in position to interact directly with the 2'-hydroxyl nucleophile is clearly resolved in the crystal structure of the inhibited precleaved HDV ribozyme (PDB ID 3NKB) (*42*). This metal ion is in position to interact with the pro-$R_P$ oxygen of the scissile phosphate, the pro-$S_P$ oxygen of U23 and, through its hydration shell, the Hoogsteen face of G25. This $Mg^{2+}$ ion appears to interact with the 2'O of U-1, serving as a Lewis acid. However, the binding of $Mg^{2+}$ ions within the HDV ribozyme active site is not highly specific. A wide range of divalent ions, including all alkaline earth and even certain transition metals, will react with similar or slightly greater activity (*49, 50*). In addition, under certain reaction conditions, the ribozyme will react in the absence of divalent ions through a channel in which monovalent ions promote the reaction (*38, 51*).

In the crystal structure of the inhibited precleaved HDV ribozyme, the nucleotide G25 is in the *syn* conformation, forming a rare reverse wobble base pair with U20 (*42*). Divalent metal ions are often observed to interact with canonical G•U wobble base pairs through their hydration shells (*52-54*). In contrast to Watson-Crick base pairs, canonical G•U wobble pairs provide a concentration of negative dipoles in the major groove that attracts cations. When a reverse G•U wobble is formed with a *syn* G base, a negatively charged surface is also formed, but it is found on the more accessible minor groove face of the helix. Thus, reverse wobbles represent a strategy to create minor-groove metal binding motifs, which could be used to facilitate tertiary contacts or the binding of catalytic metal ions.

In this study, we combine crystallographic experiments and all-atom MD calculations to characterize the HDV ribozyme's G25•U20 reverse wobble. We observe that this base pair is stable in both precleaved and product forms of the HDV ribozyme. In contrast, the inactive C75U variant of the ribozyme does not support formation of a reverse G25•U20 wobble pair. Our investigations indicate that the reverse wobble base pair contributes to a negatively charged pocket capable of binding either $Mg^{2+}$ or $Na^{+}$ ions in both the precleaved and product states of the ribozyme. Moreover, this negatively charged pocket is likely to contribute to the shifted $pK_a$ of the catalytic nucleobase, C75 (*43, 44*). Identification of this unusual reverse G•U wobble in the active sites of both the precleaved and product structures has key mechanistic implications for both metal ion- and nucleobase-mediated catalytic strategies.

## MATERIALS AND METHODS

*Molecular Dynamics Simulations*. We computed MD trajectories starting with product and precleaved crystal structures, using reactant and product states derived from the genomic HDV ribozyme and consistent with the proposed mechanism (Figure 2.1B). The product starting structure was obtained from PDB ID 1CX0 (*55, 56*), with P4 truncated by removing residues 48-69 according to standard numbering (*32*) (B148-B157 in PDB ID 1CX0). C75 was left unprotonated, as $pK_a$ measurements on the product form suggest this is the predominant state under biologically relevant conditions (*43*). The precleaved starting structure was derived from

PDB ID 3NKB, with the upstream nucleotide and scissile phosphate built as described previously (*42*); deoxynucleotides at positions 1 and 2 were converted to ribonucleotides by addition of 2'-hydroxyls with ideal bond lengths and bond angles. For this structure, C75 was protonated at N3, as suggested by $pK_a$ measurements on the precleaved form (*44*). The resulting product and precleaved models contained 62 and 73 nucleotides, respectively. Hydrogen atoms were added using Accelrys Discover Studio Visualizer 2.0.

All HDV ribozyme models were solvated with rigid TIP3P waters (*57*) in a periodically replicated orthorhombic box. $Mg^{2+}$ ions resolved in the crystal structure were included, although divalent metal ions near the U1A binding domain in the product RNA were excluded; in total, 10 and 11 $Mg^{2+}$ ions were included for the product and precleaved states, respectively. Structures were neutralized with $Na^+$ ions, and physiological monovalent ionic strength was added to the solvent via ~0.15 M NaCl. The forcefield does not distinguish between $Na^+$ and $K^+$ with quantitative accuracy, as discussed previously (*58*); thus, the simulations with $Na^+$ are viewed as qualitatively representative of either monovalent cation.

Calculations were performed with the Desmond MD program (*59, 60*) using the AMBER99 forcefield (*61, 62*), as in recent calculations from our labs (*63*). Partial charges for protonated cytosine were calculated with the RESP method (*64, 65*) using the RED-II program (*66*) as described previously (*63*). Updated charges for protonated cytosine are provided in Supporting Information. Long-range electrostatic interactions were calculated using the Smooth Particle Mesh Ewald method (*67*) with a cut-off of 12 Å, and SHAKE (*68*) constraints were applied to bonds involving hydrogen. Following a comprehensive simulated annealing equilibration procedure described in Supporting Information, we collected at least 25 ns of data at 298 K in the canonical ensemble (*i.e.*, constant NVT) for each system. A Nosé-Hoover thermostat (*69, 70*) was used to maintain temperature and pressure, and the time step was 1 fs for all MD trajectories.

*Non-linear Poisson-Boltzmann Analysis*. Electrostatic potential calculations were carried out using numerical solutions to the nonlinear Poisson-Boltzmann (NLPB) equation, as described previously (*38, 71-73*). The calculations were performed with the Adaptive Poisson Boltzmann Solver (APBS) (*74*). Structural coordinates for the precleaved form were obtained from the starting structure for MD, as described above. Structural coordinates for the product form were obtained after extensive equilibration and 20 ns of MD, as described above, in order to allow the reverse GU to form. Hydrogen atoms were added as described for MD. Bound metal ions and water molecules were omitted from NLPB calculations, as per standard protocols (*38, 71-73*). In addition, C75 was not protonated at N3 in either precleaved or product forms, although C41 was protonated using the updated partial charges provided in the Supporting Information. We chose not to include the catalytic $Mg^{2+}$ ion and not to protonate C75 to enable the assessment of negative potentials that attract these cationic species; we chose to protonate C41 to allow it to maintain its structural triple (*45, 63, 75*). Because C41 is ~15 Å from the active site, its electrostatic influence on the protonation of C75 is expected to be small.

For these calculations, the ribozyme was placed in a medium with a dielectric constant of 2 within the solvent accessible surface-enclosed volume, which was obtained using a probe radius of 1.4 Å. External solvent was treated as a continuum with a dielectric constant of 80, containing a 1:1 electrolyte. A 2.0 Å ion exclusion radius was added to the surface of the RNA to approximate a hydrated sodium ion. A salt concentration of 0.15 M was used in these calculations to mimic physiological conditions (*76, 77*). Atomic radii and partial charges were defined using the Amber99 parameter set, except for C41$^+$, which was defined as described in the Supporting Information. The calculations were performed with a 97 x 97 x 65 cubic lattice for the product state and a 129 x 65 x 97 cubic lattice for the precleaved state. The electrostatic potentials were calculated using a sequential focusing procedure (*78*). Initial potentials were approximated analytically at lattice points on the boundary of the grid using the Debye-Hückel equation (*79*), and solutions were obtained using the sequential focusing method. Three-dimensional structures and electrostatic potentials were rendered using PyMOL (*80*).

*Crystallographic Refinement of the HDV Ribozyme Postcleavage*. Coordinates and data for the product form of the HDV ribozyme were obtained from the protein data bank (PDB ID 1CX0) (*55, 56*). The structure of the ribozyme was adjusted by slightly rotating U20 to allow formation of a reverse G•U wobble pair between G25 and U20. These coordinates were subjected to two rounds of positional and B-factor refinement in Phenix (*81*) to generate a model with reasonable statistics. The original test set was retained in all calculations.

*Crystallographic Analysis of the HDV Ribozyme Precleavage*. RNA was synthesized and crystallized as described (*42*). Prior to data collection, crystals were transferred in a single step to a solution containing 50% 2-methyl-2,4-pentane-diol, 50 mM $MgCl_2$, 2 mM spermine, and 50 mM sodium acetate (pH 5.0) for 2–3 h.

Data were collected at GM/CA-CAT of the Advanced Photon Source, beamline 23-ID-D; processed using SCALEPACK2000 (*82*); and indexed in space group $C222_1$. An $F_o(K^+)$-$F_o(Na^+)$ map was calculated using CNS (*83, 84*). The structure factors $F_o(K^+)$ were obtained from a crystal soaked in potassium-containing cryostabilizing buffer described previously (*42*), while the structure factors $F_o(Na^+)$ were obtained from a crystal soaked in sodium-containing cryostabilizing buffer described above (data collection statistics are given in Table A.3). Phases were back-calculated from the coordinates of the HDV ribozyme precleavage (PDB ID 3NKB) after removal of the active site $Mg^{2+}$ and solvent molecules.

## RESULTS

*Overview*. We recently solved the structure of the active form of the HDV ribozyme in the presence of C75 and $Mg^{2+}$ at 1.9 Å resolution and pH 5.0 (*42*). This molecule was trapped pre-cleavage by substituting the 2'-OH of U–1 with a 2'-H (PDB ID: 3NKB), thereby removing the nucleophile. In this pre-cleavage structure, a reverse wobble between G25 and U20 was observed, with hydrogen bonding distances of 2.9–3.0 Å and angles of 172–176° (Figure 2.2A). This non-Watson-Crick base pair is of interest because it occurs rarely in other RNA structures (J.E. Sokoloski, S.A. Godfrey, and P. C. Bevilacqua, in prep), the nucleotides involved are conserved in all known HDV and HDV-like ribozymes (*35*), and it is located in the active site where it makes key contributions to binding a catalytic metal ion. Moreover, site-directed mutagenesis of these nucleotides has been shown to severely compromise catalytic activity (*85*).

We first examined other existing HDV ribozyme crystal structures to determine whether formation of the G25•U20 reverse wobble was a conserved feature. Structural biology of the HDV ribozyme is extensive and also includes crystal structures of the product (post-cleavage) form (PDB ID: 1CX0) (*55, 56*) and a structure of the catalytically inactive C75U mutant bound to an all RNA substrate (PDB ID: 1SJ3) (*86, 87*). The G25•U20 reverse wobble is not present in either of these crystal structures. In the product structure (1CX0), G25 is *syn*, but the relative angle between U20 and G25 results in long distances between hydrogen bond donors and acceptors: (U20(O4)-G25(N1) is 5.2 Å and U20(N3)-G25(O6) is 4.2 Å (Figure 2.2B, magenta structure) (*56*)). In the C75U mutant structure (1SJ3) G25 is *anti*, precluding reverse wobble pair formation (*86*). As the previous crystal structures do not provide evidence for a G25•U20 reverse wobble, we sought herein to explore the stability and metal binding properties of this key structural feature.

*MD Analysis of the Reverse G•U Wobble in Product and Precleaved Forms.* Previous crystallographic and computational studies on the HDV ribozyme suggested that conformational switching accompanies catalysis (*86, 88, 89*). These studies, however, drew heavily on the structure of the catalytically inactive C75U mutant ribozyme, which has a reorientation of residue 75. To ascertain whether the reverse G25•U20 wobble changes conformation during catalysis, we computed MD trajectories starting with product and precleaved crystal structures, both solved with the wild-type, C75, nucleobase. Given that C75 appears to act as the general acid in cleavage (*37, 39, 41*), we used protonated C75 for the reactant, precleaved state, and deprotonated C75 for the product, cleaved state (Figure 2.1B). Moreover, these protonation states represent the predominant C75 species under biologically relevant buffer conditions on the basis of $pK_a$ measurements (*43, 44*), and are consistent with microscopic reversibility.

As shown in Figure 2.3B, when beginning with the product form of the ribozyme, during equilibration and throughout the 25 ns trajectory, U20 and G25 form a stable two-hydrogen bond reverse wobble that undergoes only occasional fluctuation of one or the other of the hydrogen bonds. The hydrogen bonding distances are typically between ~2.8 and ~3.1 Å, and the thermally averaged angles U20(O4)-G25(H1)-G25(N1) and U20(N3)-U20(H3)-G25(O6) are ~160 ± 10°. Notably, these distances and angles are very similar to those observed in the precleaved structure (*42*). Formation of this reverse wobble was also observed in an independent MD trajectory, with occasional fluctuation of one of the two hydrogen bonds (Figure A.2A), and was observed in our previous simulations of the product form (*63*). Note that this product crystal structure includes U27, which was omitted in the recent precleaved structure because it is dispensable and U27D variants are fast-folding (*90*). To ascertain the impact of this nucleotide, we propagated an MD trajectory for which U27 was removed from the product crystal structure. We observed formation of the reverse wobble in this case as well (Figure A.2C).

We also computed similar trajectories for the C75U precleaved structure. The trajectories on this mutant ribozyme, however, did not result in such hydrogen bonding interactions between G25 and U20. Instead, over the course of several independent 4.5 ns trajectories, only a one-hydrogen bond interaction involving the N3 of U20 and the Hoogsteen face (either O6 or N7 of G25) formed (data not shown). Sponer, Walter, and colleagues previously conducted MD simulations using the AMBER 99 forcefield on product and C75U precleaved crystal structures and obtained similar results: a reverse G•U wobble formed in the product structure but not in the C75U mutant structure (*88*). Thus, these data suggest that the reverse G•U wobble is compatible with the wild-type base at position 75 but not a U.

To further the analysis, we propagated MD trajectories based on the new precleaved structure, which contains C75 and $Mg^{2+}$, with the U–1 nucleotide and scissile phosphate as previously described (*42*). As mentioned above, MD trajectories were propagated with protonated C75, using the updated atomic charges provided in Supporting Information; in this state, the ribozyme is poised for general acid catalysis. As shown in Figure 2.3C, the trajectory in the precleaved form indicates that the two hydrogen bonds of the reverse G•U wobble are exceptionally stable, with hydrogen bonding distances between 2.8 and 3.0 Å and hydrogen bonding angles of 159–163°, similar to those in the starting structure. Moreover, this stability of the reverse G•U wobble in the precleaved form was reproduced in an independent trajectory (Figure A.2B). Overall, the MD data suggest that the wild-type ribozyme does not change conformation at the G25•U20 reverse wobble pair during catalysis and that it is exceptionally stable in the precleaved state.

*Crystallographic Analysis of the Reverse G•U Wobble in Product and Precleaved Forms.* Lack of hydrogen bonding between U20 and G25 reported in the crystal structure of the product ribozyme could be the result of absence of the scissile phosphate, an artifact of crystal contacts, or the conformation of these nucleotides may be inadequately restrained by the diffraction data. We therefore examined the crystal structure of the product state, adjusted the positions of G25 and U20 to form the reverse wobble pair, and refined this model using the deposited structure factors (see Materials and Methods). A model with 3.0-3.4 Å hydrogen bonds between the appropriate non-hydrogen atoms in the reverse G•U wobble was attained (Figure 2.2B), with similar values for $R_{free}$ and $R_{work}$ (Table A.6). The composite simulated annealing omit map supports the modified conformation of U20 (Figure 2.2B). We conclude, therefore, that the original X-ray diffraction data on the product structure are compatible with formation of the reverse G•U wobble, even if they were insufficient to readily define it. Significantly, these results suggest that restraints derived from MD studies could enhance accuracy of crystal structures in regions not adequately restrained by high-quality electron density.

*MD Analysis of Metal Ion Binding in Product and Precleaved Forms.* In the crystal structure of the precleaved HDV ribozyme, G25 is a second-shell ligand, interacting with the catalytic $Mg^{2+}$

ion through solvent molecules. This ion is absent in the crystal structure of the product HDV ribozyme. We therefore examined MD trajectories of precleaved and product forms to assess whether metal cations can interact with the active site.

For the product structure, $Na^+$ ions bound near the reverse wobble during equilibration in two independent MD trajectories. A $Na^+$ ion interacted stably within ~5 Å from the O6 and N7 of G25 and the O2 of U20 (Figure A.3; Table A.5) for ~96% of the time during these trajectories (Table 2.1). These distances are consistent with a water-mediated interaction between the reverse wobble and the $Na^+$ ion (*91*). Additionally, sometimes two $Na^+$ ions bound to this site (~7% and ~40% of the time for the first and second independent trajectories, respectively). Average distances of the $Na^+$ ion to the reverse G•U wobble showed a relatively broad distribution, with RMSD values of 1.1-2.7 Å (Figure A.3A; Table A.5).

$Mg^{2+}$ ions are much less likely to bind to this site in these simulations because of the relatively low concentration of $Mg^{2+}$ ions during MD. Moreover, the $Mg^{2+}$ ions that were included in the above simulations are expected to exhibit low mobility because they were bound to the ribozyme in the product crystal structure. To probe whether this site is capable of stably binding $Mg^{2+}$, we therefore propagated two independent trajectories in which we removed two $Na^+$ ions from the bulk solvent and inserted a $Mg^{2+}$ ion into the active site so that it was bound to the reverse G•U wobble. We found that the $Mg^{2+}$ remained bound 86% and 100% of the time, with $Na^+$ binding 12% and 0% of the time, for the first and second independent trajectories, respectively (Table 2.1). Average distances of the $Mg^{2+}$ ion to the reverse G•U wobble were again ~5 Å but displayed a narrower distribution, with RMSD values of only 0.3-1.2 Å (Figure A.3B; Table A.5). Together, these two sets of results suggest that either $Na^+$ or $Mg^{2+}$ could potentially bind to the reverse G•U wobble in the product state. Due to limitations of the MD simulations, however, we are unable to predict the probabilities or strength of such binding.

When trajectories were propagated with the precleaved structure containing the observed active site $Mg^{2+}$ ion at the reverse wobble (*42*), the $Mg^{2+}$ ion remained bound 100% of the time for both MD trajectories (Table 2.1), and its distances of ~4.3 Å were relatively static, with RMSD values of 0.2 Å (Figure 2.4A; Table A.5). For comparison, MD trajectories were computed in which this $Mg^{2+}$ ion was removed from the active site and two additional $Na^+$ ions were placed in the bulk. We observed that $Na^+$ ions bound to the reverse wobble during the majority of these trajectories. Specifically, at least one $Na^+$ ion interacted stably with the reverse wobble for ~95% of the time (Table 2.1). The distribution of distances between the $Na^+$ ion and the reverse G•U wobble was broader than that observed for the $Mg^{2+}$ ion, with average distances ranging from ~3-7 Å and RMSD values of 0.4-1.5 Å (Figure 2.4B; Table A.5).

In summary, both monovalent and divalent ions bound to both product and precleaved forms of the HDV ribozyme in the MD simulations. There appears to be a preference for binding of divalent ions to the precleaved state because residency was 100% and there was essentially no movement of the $Mg^{2+}$ ion as compared to other metal ion/ribozyme state combinations. This observation is qualitatively consistent with experimental data indicating stronger binding of $Mg^{2+}$ than $Na^+$ (*49, 50*). However, limited sampling in the MD trajectories prevents a quantitative analysis of the relative probabilities and binding strengths.

*Crystallographic Analysis of Metal Ion binding in the Precleaved Form*. The MD trajectories described in the preceding section exhibited binding of either $Na^+$ or $Mg^{2+}$ in the HDV ribozyme active site. We wanted to assess the extent to which these calculations are supported by in vitro experimental data. Prior results from our lab indicated that increasing the concentration of $Na^+$ above 200 mM inhibits self-cleavage in the presence of 1 mM $Mg^{2+}$, consistent with competition between these ions (*50*). Moreover, Hill analysis of these data, which were conducted at high ionic strength of 0.3–2 M to separate out folding effects, supported displacement of the $Mg^{2+}$ ion by 1.7 ± 0.2 $Na^+$ ions. Moreover, monovalent ions are known to support cleavage by an alternative reaction mechanism (*50, 51*), making it possible that $Na^+$ ions observed herein participate in catalysis by this alternative reaction channel.

We therefore examined whether competition between the $Mg^{2+}$ and $Na^+$ ions could be observed crystallographically. Two crystals of the HDV ribozyme that varied only in the counterion used in the cryostabilization buffer were compared. An $F_o(K^+)$-$F_o(Na^+)$ difference Fourier map should contain positive peaks representing features present in $K^+$, but absent or smaller in the presence of $Na^+$. Likewise features present in $Na^+$ buffers but absent in $K^+$ will be revealed by negative peaks in the map (Figure 2.5). We observed that, in the presence of $Na^+$, the active site $Mg^{2+}$ has reduced occupancy and a $Na^+$ binding site lies nearby. Simultaneous binding of $Mg^{2+}$ and $Na^+$ to this site is precluded sterically as well as electrostatically. The positive peak in the $F_o(K^+)$-$F_o(Na^+)$ difference Fourier map (Figure 2.5) cannot be due to binding of $K^+$ at this position as the metal-ligand distances (2.1-2.4 Å) are smaller than those expected for a $K^+$ ion (~2.7 Å). In addition, the octahedral coordination geometry is consistent with this ion being a $Mg^{2+}$. The larger ionic radius of $K^+$ presumably prevents binding of this ion in the HDV ribozyme active site. Indeed, displacement of $Mg^{2+}$ by $K^+$ in the active site of group I introns induces significant rearrangements to accommodate the larger cation (*8, 92*). Thus, these experimental data support competition of $Na^+$ and $Mg^{2+}$ for binding to the HDV ribozyme site, consistent with MD and biochemical observations. No additional $Na^+$ binding sites were observed in the $F_o(K^+)$-$F_o(Na^+)$ difference Fourier map.

*Nonlinear Poisson-Boltzmann Calculations*. In an effort to discern a physical basis for cation binding to the reverse G•U wobble, we performed NLPB electrostatic calculations on product and precleaved states (Figures 2.6, A.5). These calculations provide qualitative insight by identifying negatively charged regions of the ribozyme that can serve as possible $Na^+$ and $Mg^{2+}$ binding sites. We first performed NLPB calculations on the product state with deprotonated C75 at various time steps along the 25 ns trajectory, during which the reverse G•U wobble is present and a $Na^+$ ion binds (see above). The results for the snapshot at 20 ns are shown in Figure A.5A. A negatively charged binding pocket was observed near the reverse G•U wobble, which persisted throughout the simulation.

Next, we carried out NLPB calculations on the precleaved state in which C75 was left deprotonated and $Mg^{2+}$ ions were removed (see Materials and Methods). The charged pocket observed here was significantly more negative (<–100 kT/e in precleaved versus ~ –40 kT/e in product) (Figure 2.6). The extremely negative potential appearing in the precleaved active site is due in part to positioning of a constellation of phosphates, including the scissile phosphate, in this region (Figure A.5B). The negatively charged character of this site probably accounts for the high occupancy of $Mg^{2+}$ in the active site of the precleaved RNA, both in the crystals and in the MD, and may assist in positioning and protonating C75 for catalysis (Figure 2.6B).

## DISCUSSION

The MD and experimental results suggest that a reverse G•U wobble base pair, in conjunction with a constellation of surrounding phosphates, provides a cation binding site within a structured RNA. The reverse G•U wobble motif is unusual in that, outside of UNCG hairpin loops, it has been observed in only one other large RNA (23S rRNA), based on a systematic study of ribozymes in the PDB (J.E. Sokoloski, S.A. Godfrey, and P. C. Bevilacqua, in prep). Furthermore, this G and U are universally conserved in HDV and HDV-like ribozymes (*35*). We observed herein that the reverse G•U wobble in the HDV ribozyme active site is stably formed in both functionally relevant states of the HDV ribozyme (precleaved and product), but is disrupted in the inactive C75U mutant. This conclusion is derived from analyses of crystal structures and MD simulations. Proper formation of this base pair is critical in that it contributes to both the three dimensional structure of the active site (*42*) and a highly negatively charged patch within the HDV ribozyme active site. The resulting cation binding site appears to recruit a catalytic $Mg^{2+}$ ion that participates directly in the cleavage reaction. Moreover, the negatively charged character of this site facilitates binding of monovalent ions and protonation of C75 (*43, 44*) and may aid in the accurate positioning of this catalytic nucleobase.

*A Motif that Contributes to a Negatively Charged Patch in the Minor Groove*. Large ribozymes, including group I introns, group II introns and RNase P, have active site pockets created by the junctions of multiple helices and 'single-stranded' RNAs (*8, 10, 92-94*). These complex active sites bring together clusters of 2-3 negatively charged phosphate groups from the ribozyme core that serve to bind and orient catalytic metal ions. On the other hand, small ribozymes, such as the hammerhead, hairpin, and HDV ribozymes, are only ~70 nts in length. Their active sites are usually formed by the juxtaposition of two base-paired helices and typically lack the phosphate clusters observed in large ribozymes. It was therefore anticipated that these ribozymes would function by $Mg^{2+}$-free mechanisms, and indeed all can react in the absence of $Mg^{2+}$ (*37, 95*).

The crystal structures of the HDV ribozyme provided the first glimpses of a metal ion within a small ribozyme active site positioned to interact with the scissile phosphate (*42*). Consistent with the lack of phosphate clusters in small ribozyme active sites, only a single phosphate from the HDV ribozyme core, that of U23, interacts with the catalytic metal ion. This metal binding site is buttressed by a reverse G•U wobble, a motif compatible with the duplex RNA structures often found in small ribozyme active sites. The negative potential of the metal binding pocket is due, in part, to the O6 and N7 of G25 and the O2 of U20 within the minor groove of the RNA helix. This observation is qualitatively consistent with previous observations that canonical G•U wobble pairs bind $Mg^{2+}$ ions in the major groove of RNA helices (*52, 53, 96-98*). A Watson-Crick A-U base pair cannot substitute for the G25-U20 base pair (*85*); however, additional work will be required to determine if the G25A mutation disrupts metal binding, the three-dimensional organization of the active site, folding, or a combination of these factors.

In addition to the G25-U20 reverse wobble, the pro-$S_P$ oxygen of U23 and, in the precleavage state, the pro-$R_P$ oxygen of the scissile phosphate cluster to create a negatively charged region within the HDV ribozyme active site. The negative potential in this pocket reaches a value of <–100 kT/e (Figure 2.6), which contributes to metal binding (*71, 72*). The value of this potential is similar to that for the metal ion binding core of P4-P6 (*71*), a ~160 nt independent folding domain of the *Tetrahymena thermophila* ribozyme (*99*). Thus, the HDV ribozyme, although having only approximately half the number of nucleotides, can assume a similar negatively charged binding motif in its core. Given that the P4-P6 motif functions in metal binding (*52*), it is reasonable that a similarly negatively charged motif in the HDV ribozyme could also function in metal binding.

Despite these similarities, there are also clear differences in catalytic metal ion requirements for the large ribozymes and HDV ribozymes. The former can use $Mg^{2+}$ and sometimes $Mn^{2+}$ for catalysis, while other ions such as $Ca^{2+}$ are inhibitory (*100*). In contrast, the HDV ribozymes operate proficiently with a broad range of divalent and monovalent ions, including all alkaline earth and alkali metals, and several transition metals (*49-51*). This loosening of specificity may be the result of fewer inner sphere contacts between the ribozyme and the catalytic metal, as well as the flexibility afforded by the water-mediated interactions between the catalytic metal and G25.

*Implications for HDV Ribozyme Catalysis***.** Crystallographic and MD studies support the presence of a cation binding site near the G25•U20 reverse wobble that is capable of binding either $Mg^{2+}$ or $Na^{+}$. This anionic pocket likely binds other divalent and monovalent ions that support catalysis. Based on these studies and others, we propose that this site binds a catalytic metal ion capable of participating in catalysis, most likely through a Lewis acid mechanism (Figure 2.1B) (*42*). These data are consistent with observed inhibition of the $Mg^{2+}$-dependent ribozyme reaction by sodium or cobalt hexammine ions (*37, 50, 101*). $Na^{+}$ ions presumably are capable of displacing the catalytic $Mg^{2+}$ ion, leading to a reduction in ribozyme activity as the ribozyme shifts to a $Mg^{2+}$-independent reaction channel. Cobalt hexammine is capable of displacing $Mg^{2+}$ from the HDV ribozyme (*86*); however, its ligands are non-labile, and it cannot serve as a Lewis acid. This ion therefore inhibits the reaction by competing with the catalytic $Mg^{2+}$.

In the presence of $Na^+$, the ribozyme is still capable of $Mg^{2+}$-free catalysis (*37, 38, 75*). The overall rate of reaction is ~3,000-fold faster in saturating $Mg^{2+}$ than 1 M NaCl, with ~125-fold due to $Mg^{2+}$ facilitating folding and ~25-fold due to $Mg^{2+}$ facilitating chemistry. Thus, high concentrations of $Na^+$ are reasonably effective in facilitating catalysis, a characteristic also observed in the hammerhead ribozyme (*102, 103*), although the mechanistic involvement of $Na^+$ ions is unclear in that system. Several lines of evidence suggest that the HDV ribozyme reaction catalyzed by monovalent ions proceeds through an alternative but related mechanism, described previously as a multichannel mechanism (*50*). First, the pH-rate profile for the reaction is inverted under $Mg^{2+}$-free conditions (*37, 41, 45*). Second, proton inventory experiments, which monitor the number of proton transfers in the rate limiting step, are 2 in low or no $Mg^{2+}$, but approach 1 in high $Mg^{2+}$ concentration (*47*); moreover, similar results—proton inventories of 1 in 10 mM $Mg^{2+}$—were reported for the antigenomic ribozyme by the Been laboratory (*46*). These data are consistent with a model in which $Mg^{2+}$ binding to the deprotonated U–1 2'-hydroxyl stabilizes deprotonation of the nucleophile, leaving only a single proton transfer, from $C75H^+$ to the O5' of G1, in the rate limiting step. When $Na^+$ displaces $Mg^{2+}$, the active site cation is bound in a similar location, but it cannot play the role of a Lewis acid. Indeed, solvent isotope effects under these conditions support deprotonation of the 2'OH of U-1 by a hydroxide ion from solution (*41*). $Na^+$ ions in the HDV ribozyme active site have the potential to contribute to the reaction by other catalytic means, such as stabilizing appropriate active site and substrate geometries and contributing to favorable electrostatics.

*Contribution of the Motif to Organization of the Active Site.* The negatively charged pocket near the G25•U20 reverse wobble not only binds $Mg^{2+}$ and $Na^+$ ions, but also likely contributes to the catalytic properties of C75. The $pK_a$ of C75 is shifted toward 7 in the precleaved state of the ribozyme; however, the $pK_a$ is similar to that of isolated cytosine in the product form of the ribozyme (*43-45*). Given that electrostatics help drive protonation of the nucleobases in general (*104*), the positioning of the phosphates, including the scissile phosphate in the precleaved state, likely helps drive the $pK_a$ of C75 toward neutrality (Figure 6B, A.5B). Moreover, protonation of C75 is coupled anticooperatively with binding of $Mg^{2+}$ ions (*37, 44, 45*). The negative potential of this pocket appears extreme enough to engage both a divalent metal ion and a positively charged base, helping to explain these previous experimental results.

Lastly, this motif may help position C75 properly for catalysis. According to our calculations of partial charges for protonated cytosine, the H4s on $C75^+$ carry a total net charge of +0.93, while H3 has a charge of just +0.30 (Table A.2). These partial charges are consistent with crystal structures of the HDV ribozyme (*42, 55*), where the H4s of $C75^+$, rather than H3, orient toward this patch. This positioning is critical because it leaves the H3 positioned for proton transfer to the O5' of G1. Previous MD simulations based on an alternative starting structure, in which the C75U crystal structure was mutated back to $C75^+$, and using different partial charges for $C75^+$ resulted in the H3 of C75 being recruited to the negatively charged phosphate of G1 rather than its O5' (*88*).

*Conclusions*. In this Chapter, we presented evidence that a reverse G•U wobble interaction capable of binding $Na^+$ and $Mg^{2+}$ ions is present in both precleaved and product states of the HDV ribozyme. This evidence was obtained from analyses of crystallographic data and MD simulations. Moreover, electrostatic potential calculations indicated that formation of the reverse G•U wobble contributes a highly negatively charged patch within the HDV ribozyme active site, thereby facilitating the binding of metal ions and the positioning of the catalytic protonated C75. The existence of the metal-binding reverse G•U wobble in both the precleaved and product states supports the mechanistic hypothesis that the catalytic metal is retained throughout the reaction. Thus, appropriate combinations of experimental data, structural models, and MD simulations have the potential to uncover catalytically relevant interactions from RNA crystal structures.

**Acknowledgements** We thank Alexander Soudackov for helpful discussions.

## Supporting Information

Procedures for MD, protonated cytosine partial charges, additional MD trajectories, metal ion movement during MD, heavy-atom RMSD plots, and additional NLPB results are provided in Appendix A. Also provided are tables of crystallographic data collection statistics, crystallographic $Na^+$ coordinates for the pre-cleaved ribozyme, and average distances between the reverse G•U wobble and metal ions from MD.

## References

(1) Copley, S. D., Smith, E., and Morowitz, H. J. *Bioorg. Chem.* **2007**, *35*, 430-443.
(2) Cech, T. R. *Cell* **2009**, *136*, 599-602.
(3) Turk, R. M., Chumachenko, N. V., and Yarus, M. *Proc. Natl. Acad. Sci. U S A* **2010**, *107*, 4585-4589.
(4) Henkin, T. M. *Genes Dev.* **2008**, *22*, 3383-3390.
(5) Roth, A., and Breaker, R. R. *Annu. Rev. Biochem.* **2009**, *78*, 305-334.
(6) Fedor, M. J. *Annu. Rev. Biophys.* **2009**, *38*, 271-299.
(7) Holbrook, S. R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 302-308.
(8) Golden, B. L., Kim, H., and Chase, E. *Nat. Struct. Mol. Biol.* **2005**, *12*, 82-89.
(9) Martick, M., and Scott, W. G. *Cell* **2006**, *126*, 309-320.
(10) Toor, N., Keating, K. S., Taylor, S. D., and Pyle, A. M. *Science* **2008**, *320*, 77-82.
(11) Montange, R. K., and Batey, R. T. *Annu. Rev. Biophys.* **2008**, *37*, 117-133.
(12) Jinek, M., and Doudna, J. A. *Nature* **2009**, *457*, 405-412.
(13) Narlikar, G. J., and Herschlag, D. *Annu. Rev. Biochem.* **1997**, *66*, 19-59.
(14) DeRose, V. J. *Curr. Opin. Struct. Biol.* **2003**, *13*, 317-324.
(15) Sigel, R. K., and Pyle, A. M. *Chem. Rev.* **2007**, *107*, 97-113.
(16) Bevilacqua, P. C. (Lilley, D. M., and Eckstein, F., Eds.) **2008**, pp 11-36, Royal Society of Chemistry, Cambridge.
(17) Cate, J. H., and Doudna, J. A. *Methods Enzymol.* **2000**, *317*, 169-180.
(18) Ferre-D'Amare, A. R., and Doudna, J. A. *Curr. Protoc. Nucleic Acid Chem.* **2001**, *Chapter 7*, Unit 7 6.
(19) Egli, M. *Curr. Opin. Chem. Biol.* **2004**, *8*, 580-591.
(20) Golden, B. L. *Methods Mol. Biol.* **2007**, *363*, 239-257.
(21) Yajima, R., Proctor, D. J., Kierzek, R., Kierzek, E., and Bevilacqua, P. C. *Chem. Biol.* **2007**, *14*, 23-30.
(22) Spitale, R. C., and Wedekind, J. E. *Methods* **2009**, *49*, 87-100.
(23) Nilsen, T. W. *Mol. Cell* **2007**, *28*, 715-720.
(24) Vaiana, A. C., Westhof, E., and Auffinger, P. *Biochimie* **2006**, *88*, 1061-1073.
(25) Kormos, B. L., Baranger, A. M., and Beveridge, D. L. *J. Am. Chem. Soc.* **2006**, *128*, 8992-8993.
(26) McDowell, S. E., Spackova, N., Sponer, J., and Walter, N. G. *Biopolymers* **2007**, *85*, 169-184.
(27) Lee, T. S., Lopez, C. S., Giambasu, G. M., Martick, M., Scott, W. G., and York, D. M. *J. Am. Chem. Soc.* **2008**, *130*, 3053-3064.
(28) Lee, T. S., Giambasu, G. M., Sosa, C. P., Martick, M., Scott, W. G., and York, D. M. *J. Mol. Biol.* **2009**, *388*, 195-206.
(29) Ferre-D'Amare, A. R., and Scott, W. G. *Cold Spring Harb. Perspect. Biol.* **2010**
(30) Lai, M. M. *Annu. Rev. Biochem.* **1995**, *64*, 259-286.
(31) Taylor, J. M. (Handa, H., and Yamaguchi, Y., Eds.) **2006**, pp 20-37, Landes Bioscience, Georgetown, TX.
(32) Been, M. D. (2006) HDV ribozymes. *Curr. Top. Microbiol. Immunol.* **2006**, *307*, 47-65.
(33) Salehi-Ashtiani, K., Luptak, A., Litovchick, A., and Szostak, J. W. *Science* **2006**, *313*, 1788-1792.

(34) Chadalavada, D. M., Gratton, E. A., and Bevilacqua, P. C. *Biochemistry* **2010**, *49*, 5321-5330.

(35) Webb, C.-H. T., Riccitelli, N. J., Ruminski, D. J., and Lupták, A. *Science* **2009**, *326*, 953.

(36) Perrotta, A. T., Shih, I., and Been, M. D. *Science* **1999**, *286*, 123-126.

(37) Nakano, S., Chadalavada, D. M., and Bevilacqua, P. C. *Science* **2000**, *287*, 1493-1497.

(38) Nakano, S., Proctor, D. J., and Bevilacqua, P. C. *Biochemistry* **2001**, *40*, 12022-12038.

(39) Das, S. R., and Piccirilli, J. A. *Nat. Chem. Biol.* **2005**, *1*, 45-52.

(40) Koo, S., Novak, T., and Piccirilli, J. A. (Lilley, D. M., and Eckstein, F., Eds.) **2008**, pp 92-122, RSC Publishing, Cambridge, UK.

(41) Cerrone-Szakal, A. L., Siegfried, N. A., and Bevilacqua, P. C. *J. Am. Chem. Soc.* **2008**, *130*, 14504-14520.

(42) Chen, J. H., Yajima, R., Chadalavada, D. M., Chase, E., Bevilacqua, P. C., and Golden, B. L. *Biochemistry* **2010**, *49*, 6508-6518.

(43) Luptak, A., Ferre-D'Amare, A. R., Zhou, K., Zilm, K. W., and Doudna, J. A. *J. Am. Chem. Soc.* **2001**, *123*, 8447-8452.

(44) Gong, B., Chen, J. H., Chase, E., Chadalavada, D. M., Yajima, R., Golden, B. L., Bevilacqua, P. C., and Carey, P. R. *J. Am. Chem. Soc.* **2007**, *129*, 13335-13342.

(45) Nakano, S., and Bevilacqua, P. C. *Biochemistry* **2007**, *46*, 3001-3012.

(46) Shih, I. H., and Been, M. D. *Proc. Natl. Acad. Sci. U S A* **2001**, *98*, 1489-1494.

(47) Nakano, S., and Bevilacqua, P. C. *J. Am. Chem. Soc.* **2001**, *123*, 11333-11334.

(48) Bevilacqua, P. C. *Biochemistry* **2003**, *42*, 2259-2265.

(49) Suh, Y. A., Kumar, P. K., Taira, K., and Nishikawa, S. *Nucleic Acids Res.* **1993**, *21*, 3277-3280.

(50) Nakano, S., Cerrone, A. L., and Bevilacqua, P. C. *Biochemistry* **2003**, *42*, 2982-2994.

(51) Perrotta, A. T., and Been, M. D. *Biochemistry* **2006**, *45*, 11357-11365.

(52) Cate, J. H., and Doudna, J. A. *Structure* **1996**, *4*, 1221-1229.

(53) Kieft, J. S., and Tinoco, I., Jr. *Structure* **1997**, *5*, 713-721.

(54) Keel, A. Y., Rambo, R. P., Batey, R. T., and Kieft, J. S. *Structure* **2007**, *15*, 761-772.

(55) Ferre-D'Amare, A. R., Zhou, K., and Doudna, J. A. *Nature* **1998**, *395*, 567-574.

(56) Ferre-D'Amare, A. R., and Doudna, J. A. *J. Mol. Biol.* **2000**, *295*, 541-556.

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

(58) Krasovska, M. V., Sefcikova, J., Reblova, K., Schneider, B., Walter, N. G., and Sponer, J. *Biophys. J.* **2006**, *91*, 626-638.

(59) Bowers, K. J., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossvary, I. K., Moraes, M. A., Sacerdoti, F. D., Salmon, J. K., Shan, Y., and Shaw, D. E. *in Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)* **2006**.

(60) D. E. Shaw Research, N. Y. Desmond Molecular Dynamics System, **2008**.

(61) Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. J., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(62) Wang, J. M., Cieplak, P., and Kollman, P. A. *J. Comp. Chem.* **2000**, *21*, 1049-1074.

(63) Veeraraghavan, N., Bevilacqua, P. C., and Hammes-Schiffer, S. *J. Mol. Biol.* **2010**, *402*, 278-291.

(64) Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.

(65) Cieplak, P., Cornell, W. D., Bayly, C., and Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357-1377.

(66) Pigache, A., Cieplak, P., and Dupradeau, F.-Y. in 227th ACS National Meeting, Anaheim, CA **2004**.

(67) Darden, T., York, D., and Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

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

(69) Nosé, S. *Mol. Phys.* **1984**, *52*, 255-268.

(70) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695-1697.

(71) Chin, K., Sharp, K. A., Honig, B., and Pyle, A. M. *Nat. Struct. Biol.* **1999**, *6*, 1055-1061.

(72) Misra, V. K., and Draper, D. E. *J. Mol. Biol.* **2000**, *299*, 813-825.

(73) Maderia, M., Hunsicker, L. M., and DeRose, V. J. *Biochemistry* **2000**, *39*, 12113-12120.

(74) Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. *Proc. Natl. Acad. Sci. U S A* **2001**, *98*, 10037-10041.

(75) Wadkins, T. S., Shih, I., Perrotta, A. T., and Been, M. D. *J. Mol. Biol.* **2001**, *305*, 1045-1055.

(76) Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. in Molecular Biology of the Cell pp 508, Garland Publishing, Inc., New York **1994**.

(77) Feig, A. L., and Uhlenbeck, O. The RNA World, 2nd Ed. (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) **1999**, pp 287-319, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

(78) Gilson, M. K., Sharp, K. A., and Honig, B. *J. Comput. Chem.* **1987**, *9*, 327-335.

(79) Gilson, M. K., and Honig, B. *Proteins* **1988**, *4*, 7-18.

(80) DeLano, W. L. DeLano Scientific, San Carlos, CA, USA **2002**.

(81) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213-221.

(82) Otwinowski, Z., and Minor, W. (Carter, C. W. J., and Sweet, R. M., Eds.) **1997**, pp 307-326, Academic Press, New York.

(83) Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. **1998**, *Acta Crystallogr. D Biol. Crystallogr.* *54*, 905-921.

(84) Brunger, A. T. *Nat Protoc* **2007**, *2*, 2728-2733.

(85) Tanner, N. K., Schaff, S., Thill, G., Petit-Koskas, E., Crain-Denoyelle, A. M., and Westhof, E. *Curr. Biol.* **1994**, *4*, 488-498.

(86) Ke, A., Zhou, K., Ding, F., Cate, J. H., and Doudna, J. A. *Nature* **2004**, *429*, 201-205.

(87) Ke, A., Ding, F., Batchelor, J. D., and Doudna, J. A. *Structure* **2007**, *15*, 281-287.

(88) Krasovska, M. V., Sefcikova, J., Spackova, N., Sponer, J., and Walter, N. G. *J. Mol. Biol.* **2005**, *351*, 731-748.

(89) Banas, P., Rulisek, L., Hanosova, V., Svozil, D., Walter, N. G., Sponer, J., and Otyepka, M. *J. Phys. Chem. B* **2008**, *112*, 11177-11187.

(90) Brown, T. S., Chadalavada, D. M., and Bevilacqua, P. C. *J. Mol. Biol.* **2004**, *341*, 695-712.

(91) Megyes, T., Bálint, S., Grósz, T., Randnai, T., and Bakó, I. *J. Chem. Phys.* **2008**, *128*, 044501.

(92) Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J., and Strobel, S. A. *Nature* **2004**, *430*, 45-50.

(93) Guo, F., Gooding, A. R., and Cech, T. R. *Mol. Cell* **2004**, *16*, 351-362.

(94) Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T., and Mondragon, A. *Nature* **2010**, *468*, 784-789.

(95) Murray, J. B., Seyhan, A. A., Walter, N. G., Burke, J. M., and Scott, W. G. *Chem. Biol.* **1998**, *5*, 587-595.

(96) Colmenarejo, G., and Tinoco, I., Jr. *J. Mol. Biol.* **1999**, *290*, 119-135.

(97) Montange, R. K., and Batey, R. T. *Nature* **2006**, *441*, 1172-1175.

(98) Stefan, L. R., Zhang, R., Levitan, A. G., Hendrix, D. K., Brenner, S. E., and Holbrook, S. R. *Nucleic Acids Res.* **2006**, *34*, D131-134.

(99) Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. *Science* **1996**, *273*, 1678-1685.

(100) Hougland, J. L., Piccirilli, J. A., Forconi, M., Lee, J., and Herschlag, D. 3rd Ed. (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) **2006,** pp 133-205, Cold Spring Harbor Press, Cold Spring Harbor, New York.

(101) Gong, B., Chen, J. H., Bevilacqua, P. C., Golden, B. L., and Carey, P. R. *Biochemistry* **2009**, *48*, 11961-11970.

(102) Curtis, E. A., and Bartel, D. P. *RNA* **2001**, *7*, 546-552.

(103) O'Rear, J. L., Wang, S., Feig, A. L., Beigelman, L., Uhlenbeck, O. C., and Herschlag, D. *RNA* **2001**, *7*, 537-545.

(104) Tang, C. L., Alexov, E., Pyle, A. M., and Honig, B. *J. Mol. Biol.* **2007**, *366*, 1475-1496.

## FIGURES AND TABLES

***Table 2.1.*** Metal ion residency at the reverse G•U wobble site for product and precleaved forms.

| Structure | Trajectory | $Mg^{2+}$ present initially | $Mg^{2+}$ residency (%) | $Na^{+}$ residency (%) |
|---|---|---|---|---|
| Product (C75$^{o}$) | | | | |
| | 1 | No | 0 | 94 |
| | 2 | No | 0 | 99 |
| | 1 | Yes | 86 | 12 |
| | 2 | Yes | 100 | 0 |
| Precleaved (C75$^{+}$) | | | | |
| | 1 | Yes | 100 | 0 |
| | 2 | Yes | 100 | 0 |
| | 1 | No | 0 | 93 |
| | 2 | No | 0 | 99 |

Percentage residency of metal ions at the reverse G•U wobble site for product C75$^{o}$ and precleaved C75$^{+}$ states from MD simulations is presented. Metal ion residency is defined as the percentage of time during the trajectory in which a $Mg^{2+}$ or $Na^{+}$ ion, as appropriate, is located within 5 Å from at least one of the following three atoms: G25(N7), G25(O6), or U20(O2). This distance is consistent with observed second shell ligands (*91*). Results are obtained from 25 ns trajectories, with data points taken every 5 ps.

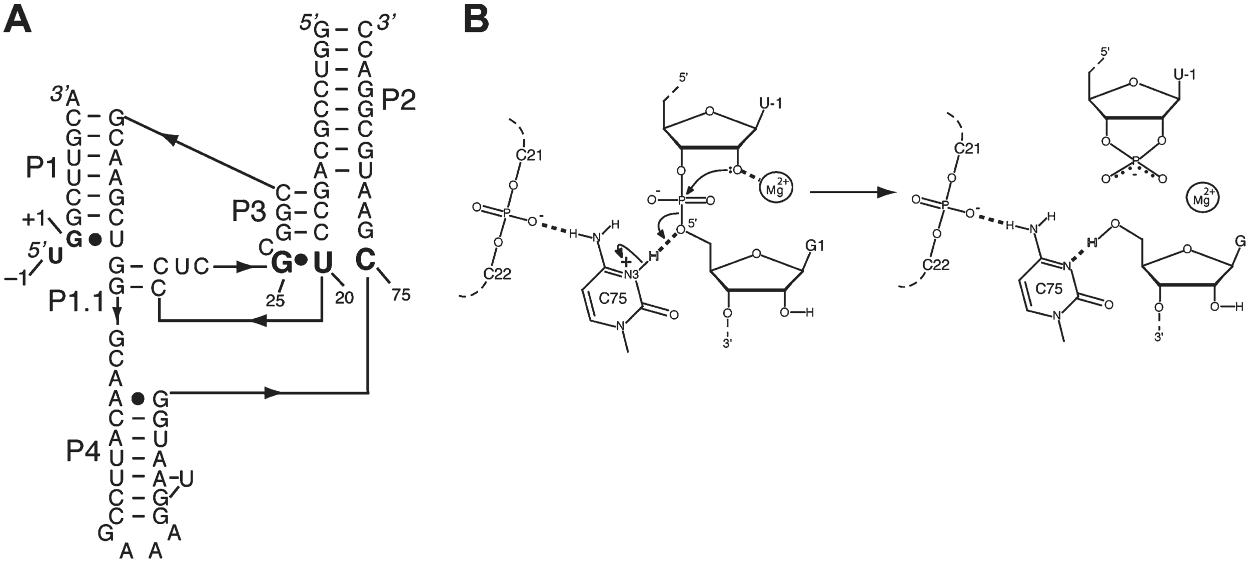

**Figure 2.1.** Structure and proposed mechanism of the HDV ribozyme. (A) Secondary structure of the precleaved HDV ribozyme used in a recent crystallography study and herein for MD (PDB ID 3NKB). Numbering is based on the genomic HDV ribozyme (*42*). This is a two-piece, fast-folding U27D variant (*90*), in which P4 has been truncated and modified to facilitate crystallography (*42*). The reverse G•U wobble and catalytic C75 residue are in large bold font. The cleavage site is between U–1 and G1, and the five pairing regions are noted as P1–P4 and P1.1. The product ribozyme sequence used for MD is similar to the one shown, but it has a further truncated P4, a joining region between P1 and P2, a single additional nucleotide insertion, U27, and it lacks the –1 nucleotide (*63*). (B) Proposed mechanism of HDV ribozyme self-cleavage in which $Mg^{2+}$ serves as a Lewis acid and C75 as a general acid (*37, 39, 42*). The C75 protonation states depicted in the precleaved and product states are used in the MD simulations.

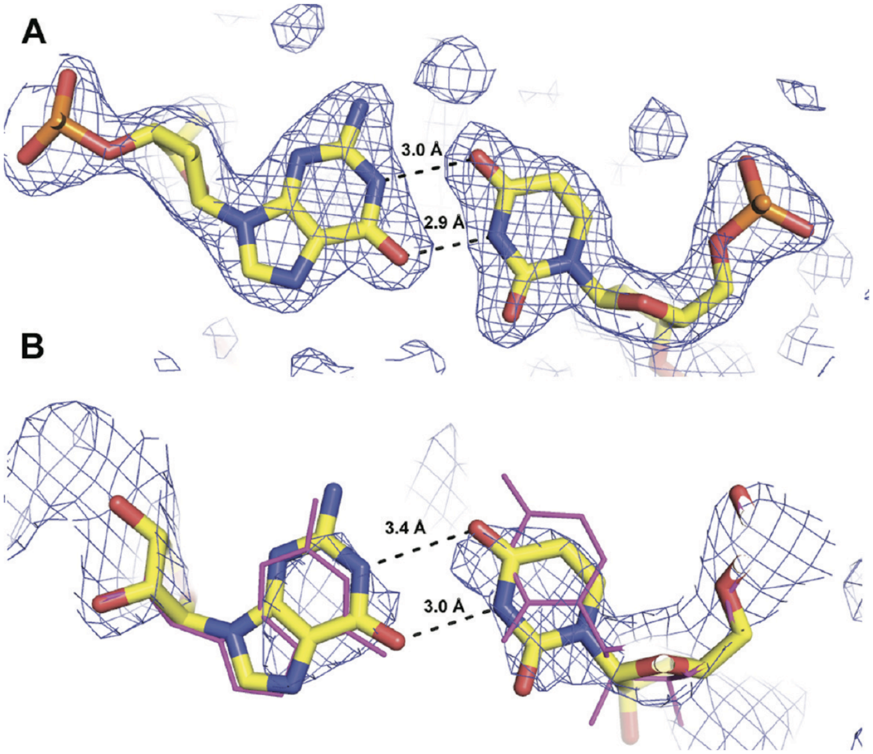

**Figure 2.2.** Crystallographic data from the HDV ribozyme product are consistent with formation of a G25•U20 reverse wobble. (A) The G25•U20 reverse wobble from the 1.9 Å crystal structure of the precleaved HDV ribozyme (PDB ID 3NKB) (*42*). The 2F$_o$-F$_c$ electron density map is contoured at 1 s and drawn within 3 Å of atoms shown. (B) The coordinates of the HDV ribozyme product (PDB ID 1CX0) (magenta lines) do not indicate a G25•U20 reverse wobble pair. This structure was adjusted by a slight rotation of U20 to form a G•U reverse wobble, and the structure was refined as described to generate the model shown. A 2.3 Å composite simulated annealing omit map contoured at 1 s within 5 Å of the illustrated atoms shows that the crystallographic data are compatible with the G•U reverse wobble geometry.

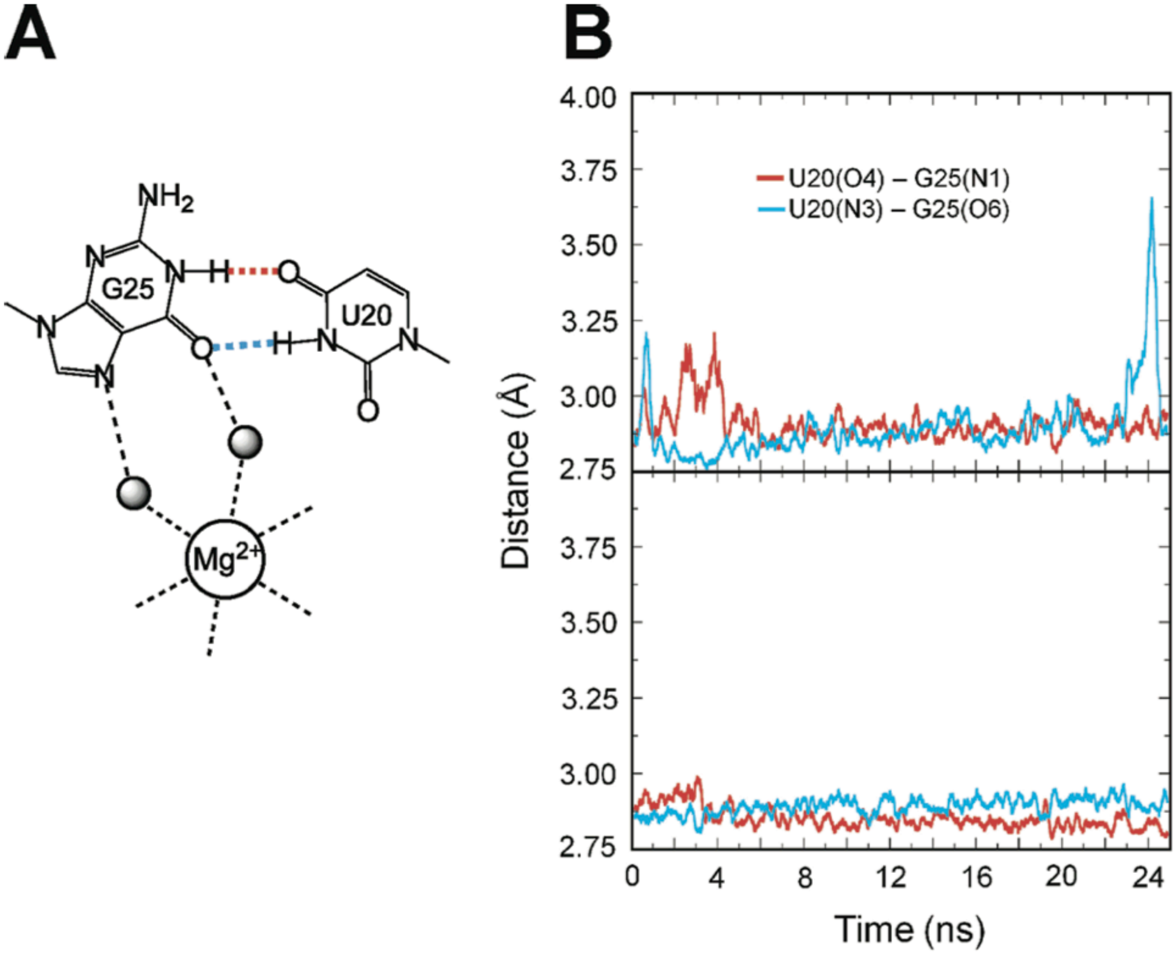

**Figure 2.3.** Molecular dynamics simulations of reverse G•U wobble in the product and precleaved forms of HDV ribozyme. (A) Schematic depiction of reverse G•U wobble observed in crystal structure of the precleaved form (*42*). Color-coding of the hydrogen bonds is maintained in panels B and C. Hydrogen atoms at hydrogen-bonding interface are not shown. (B) Hydrogen bonding distances in MD trajectory of product C75° state. At neutral pH, C75 is deprotonated in the product state (*43*). (C) Hydrogen bonding distances in MD trajectory of precleaved $C75^+$ state. In the precleaved state, C75 has a p$K_a$ near neutrality and is largely protonated (*44*). See Supporting Information for additional independent MD trajectories.

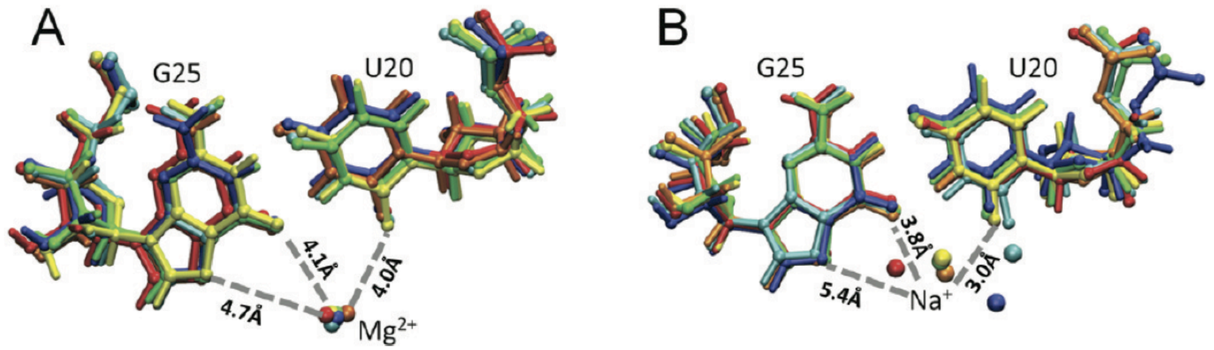

**Figure 2.4.** Snapshots of movement of metal ion during MD trajectories of the precleaved C75$^+$ state of the HDV ribozyme. (A) Coordination of Mg$^{2+}$ for a trajectory with Mg$^{2+}$ bound to this site throughout the trajectory. (B) Coordination of Na$^+$ for a trajectory with the bound Mg$^{2+}$ ion removed from the active site and replaced by two Na$^+$ ions in the bulk prior to equilibration. The colors show snapshots at various time steps: blue – 0 ns, cyan – 5 ns, green – 10 ns, yellow – 15 ns, orange – 20 ns, red – 25 ns. These results were obtained for a single independent trajectory of each type. (For part B, the 0 ns snapshot is actually a 1.5 ns snapshot to satisfy the criteria for bound Na$^+$.) The snapshots at these time steps were aligned to minimize the RMSD for the heavy atoms of residues G25 and U20. Average distances between key atoms of the reverse G•U wobble and these metal ions are indicated by dashed lines. These distances, as well as their standard deviations, are provided as the first trajectory in Table A.5. See Figure A.3 for the analogous figure for the product C75$^o$ state.

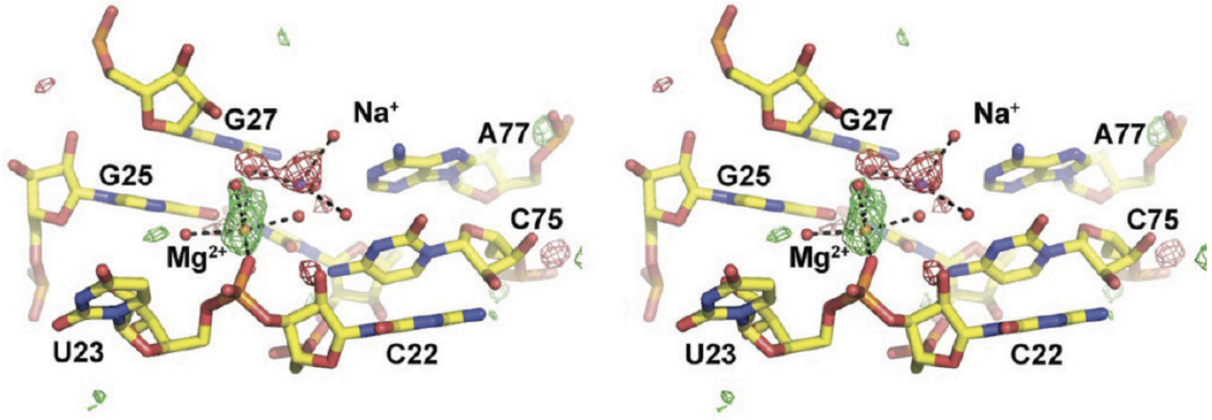

**Figure 2.5.** Stereoview of the F$_o$(K$^+$)-F$_o$(Na$^+$) difference Fourier map of the precleaved HDV ribozyme. Data to 1.9 Å resolution were used to calculate the map. The map is contoured at 3 s (green) and at –3 s (red) within 5 Å of every atom shown. Positive peaks represent atoms that are

present at higher occupancy in the absence of $Na^+$ ion, while negative peaks represent atoms that are present at higher occupancy in the presence of $Na^+$. $Mg^{2+}$ and $Na^+$ ions, colored orange and purple, respectively, are observed at competing sites within the HDV ribozyme active site.

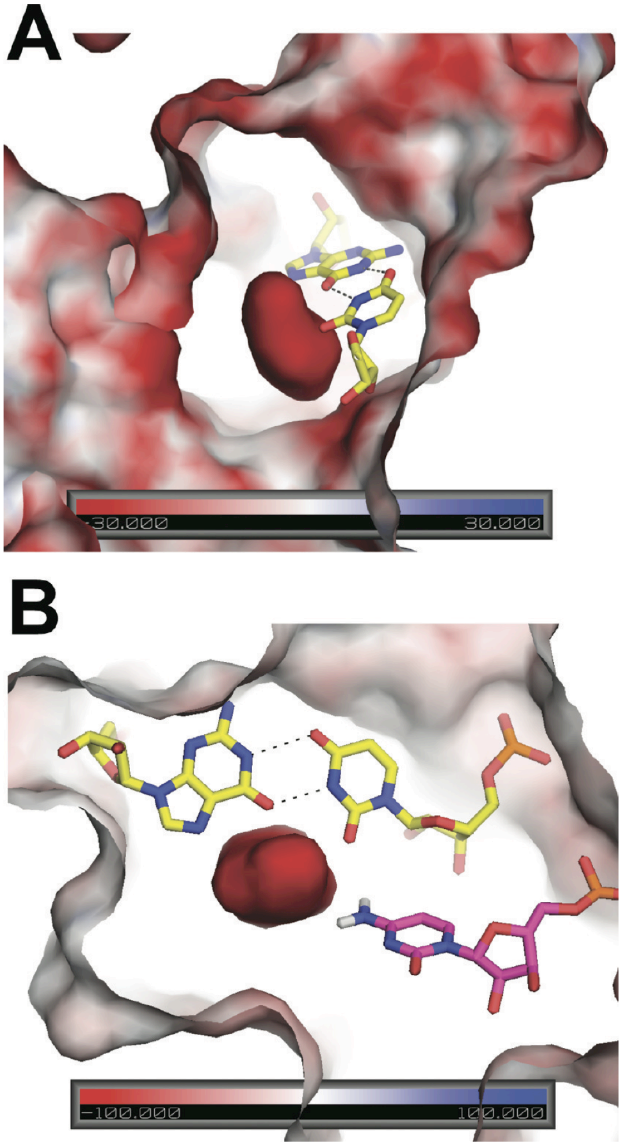

**Figure 2.6.** Surface electrostatic potential of precleaved HDV ribozyme. The potential is colored according to the scales provided in the base of each panel. Views shown here are near the reverse G•U wobble, which is depicted in yellow sticks with hydrogen bonding in black. (A) Orientation showing large negative potential in the center, which reaches values greater in magnitude than –100 kT/e. Analogous figure for the product state of the HDV ribozyme

provided in Figure A.5A. (B) Orientation showing interaction of C75 and reverse G•U wobble with negative potential in the precleaved state. C75 is in magenta with N4 hydrogens shown explicitly. Note that the C75 is oriented toward the highly negatively charged pocket generated near the reverse G•U wobble. The amine of C75 is a second shell ligand to the catalytic $Mg^{2+}$ (not shown here) and is hydrogen bonded to the scissile phosphate. View is rotated counterclockwise by ~90° from panel (A) and viewed from the top. Details of phosphates contributing to the potential of the precleaved state are provided in Figure A.5B.

# CHAPTER 3: MECHANISTIC STRATEGIES IN THE HDV RIBOZYME: CHELATED AND DIFFUSE METAL ION INTERACTIONS AND ACTIVE SITE PROTONATION[1]

## Abstract

The crystal structure of the precleaved form of the hepatitis delta virus (HDV) ribozyme reveals two G•U wobbles near the active site: a rare reverse G•U wobble involving a *syn* G base, and a standard G•U wobble at the cleavage site. The catalytic mechanism for this ribozyme has been proposed to involve a $Mg^{2+}$ ion bound to the reverse G•U wobble, as well as a protonated C75 base. We carried out molecular dynamics simulations to analyze metal ion interaction with the reverse and standard G•U wobbles and to investigate the impact of C75 protonation on the structure and motions of the ribozyme. We identified two types of $Mg^{2+}$ ions associated with the ribozyme, chelated and diffuse, at the reverse and standard G•U wobbles, respectively, which appear to contribute to catalysis and stability, respectively. These two metal ion sites exhibit relatively independent behavior. Protonation of C75 was observed to locally organize the active site in a manner that facilitates the catalytic mechanism, in which $C75^+$ acts as a general acid and $Mg^{2+}$ as a Lewis acid. The simulations also indicated that the overall structure and thermal motions of the ribozyme are not significantly influenced by the catalytic $Mg^{2+}$ interaction or C75 protonation. This analysis suggests that the reaction pathway of the ribozyme is dominated by small local motions at the active site rather than large-scale global conformational changes. These results are consistent with a wealth of experimental data.

---

[1] This chapter was published in its entirety in the Journal of Physical Chemistry B in 2011. The complete reference to the journal article is as follows: Veeraraghavan, N.; Ganguly, A.; Golden, B. L.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Phys. Chem. B* **2011**, *115*, 8346-8357. N. V. performed the analysis and A. G. performed the simulations.

# I. INTRODUCTION

RNA enzymes, or ribozymes, can use both metal ions and nucleobases in their catalytic mechanisms. Large ribozymes tend to rely on the positioning of divalent ions to stabilize their transition states,[1,2] while smaller ribozymes primarily use nucleobases to stabilize their transition states, often via proton transfer.[3-5] The genomic hepatitis delta virus (HDV) ribozyme, depicted in Figure 3.1A, is unusual in that it uses both metal ion and nucleobase catalysis in its reaction mechanism.[6-8] This ribozyme is comprised of five pairing regions, P1-P4 and P1.1, and the active site is found in the cleft formed between the P1/P1.1/P4 and P2/P3 coaxial stacks. Recent studies reveal that the HDV ribozyme occurs in all three kingdoms of life and is found in the human genome.[9,10] Study of the HDV ribozyme is thus of special interest both for how it integrates two diverse catalytic mechanisms, as well as for its importance to fundamental life processes.

We have described the mechanism of the ribozyme as multichannel, in which three parallel channels carry out the same chemistry of phosphodiester bond cleavage using the 2'-hydroxyl of U-1 as a nucleophile.[6,11] All three channels appear to use C75 as the general acid to protonate the 5'-bridging oxygen of G1 but differ in their use of $Mg^{2+}$ ions.[12-15] In channel 1, $Mg^{2+}$ is not present, and high concentrations of $Na^+$ promote the reaction; in channel 2, $Mg^{2+}$ ions serve solely structural roles, accounting for a rate acceleration of ~125-fold; and in channel 3, $Mg^{2+}$ serves a catalytic role, accounting for an additional ~25-fold in rate acceleration, or an overall contribution of $Mg^{2+}$ ions to reactivity of ~3,000 fold.[6] Figure 3.1B provides a schematic depiction of the mechanism in channel 3.[8,16] The $pK_a$ of C75 is shifted toward neutrality in all three reaction channels, enhancing its ability to serve as a general acid, although this $pK_a$ shift is opposed somewhat by $Mg^{2+}$ addition.[13,15,17,18]

The structural biology of the HDV ribozyme includes three major structures. The first was a structure of the cleaved (product) form of the ribozyme.[12,19] This structure revealed that C75 accepts a hydrogen bond from the leaving group 5'OH of G1, which suggested a general acid role for C75H$^+$ in the cleavage reaction according to microscopic reversibility. The second was a set of structures of the precleaved form, in which self-cleavage was inhibited by a C75U mutation or by omission of Mg$^{2+}$ ions.[20,21] The C75U precleaved structure was similar in overall topology to the product structure, but it showed U75 in position to serve as a general base for cleavage, with Mg$^{2+}$ in the general acid position. Recent evidence suggests, however, that this particular ribozyme construct is misfolded owing to the amine-to-keto change in C to U and the inability of U to be cationic.[8,14] Recently we solved a third structure of the HDV ribozyme, which was also of the precleaved state, but in the presence of Mg$^{2+}$ ions and with C75. Moreover, the structure was at a low pH where C75 should remain in the protonated catalytically active state. In these crystals, the reaction was inhibited by changing the 2'-hydroxyl nucleophile of U-1 to a hydrogen.[8] The overall topology of the C75 precleaved structure is also similar to the product. However, unlike the C75U mutant structure, the active site of this precleaved structure was much more similar to that of the product structure, with C75 in position to serve as the general acid. Moreover, the C75 precleaved structure also revealed a Mg$^{2+}$ ion positioned to serve as a Lewis acid by stabilizing a deprotonated 2'-hydroxyl of U-1. In general, the C75 precleaved structure is more consistent with a large literature of experimental biochemical studies than is the C75U structure.[8,14]

The HDV ribozyme active site contains both standard and reverse G•U wobbles. A standard G•U wobble (Figure 3.2, G1•U37) has two hydrogen bonds and is largely compatible with the A-form helix,[22] especially for terminal wobbles that induce less deformation of the helix.[23] Much less common is a reverse G•U wobble (Figure 3.2, G25•U20), also with two-hydrogen bonds, in which the G is in the rare *syn* conformation wherein the base resides over the ribose sugar.[24] Notably, both of these wobbles expose functional groups that possess negative dipoles and are capable of interacting with metal ions in the major and minor grooves for standard and reverse wobbles, respectively. In the HDV ribozyme, the G and U at positions 25 and 20 are absolutely conserved,[10] and they interact with the putative catalytic $Mg^{2+}$ ion through its hydration shell.[8] The standard wobble between G1 and U37 is at the base of P1, where it is present in the majority (~80%) of HDV ribozymes,[25] although it is sometimes replaced with Watson-Crick GC and AU base pairs in nature[9,10,25] with little effect on reaction kinetics or metal dependence.[26] Solution kinetic and Raman spectroscopic experiments suggest a structural $Mg^{2+}$ ion may interact with the G1•U37 wobble, although this ions was not unambiguously identified in the recent crystal structure.[8] In addition, a G•U to AU change at the reverse wobble leads to more than a 1000-fold reduction in cleavage rate,[27] whereas the same change, as well as a G•U to GC change, at the standard wobble has no effect.[26]

We have recently combined experiments and molecular dynamics (MD) calculations on both product and precleaved HDV ribozyme structures. Studies on the product form of the ribozyme revealed long-distance communication between a structural base triple and the active site.[28] Subsequent studies on the C75 precleaved form of the ribozyme revealed that the rare reverse G•U wobble helps form an anionic motif in the active site that is capable of interacting with both divalent and monovalent ions.[16] The electrostatic potential of this site is exceptionally negative, which accommodates the experimental observations that this site can both interact strongly with divalent ions and drive protonation of C75 with a shifted $pK_a$, despite repulsive electrostatic interaction between the two cations. Moreover, this reverse G•U wobble was also observed in MD simulations of the product form of the ribozyme and was found to be consistent with the crystallographic data for the product form.[16] These studies on both the C75 precleaved and product forms of the HDV ribozyme further indicated that the reverse G•U wobble is maintained throughout the catalytic reaction.

In the present study, we investigate the impact of C75 protonation and $Mg^{2+}$ interaction at both of the G•U wobbles on the structure as well as the motions of the HDV ribozyme. We computationally alter protonation of C75 and presence of the catalytic metal ion at the reverse G•U wobble, as well as investigate the effects of mutation of the standard G•U wobble to a Watson-Crick GC base pair. Our simulations indicate that protonation of C75 provides local organization of the active site and that the metal ion interactions are qualitatively different at the reverse and standard G•U wobbles. In addition, we find that these two metal ion sites exhibit relatively independent behavior. Finally, our calculations indicate that the overall structure and thermal motions of the ribozyme are not significantly affected by C75 protonation or metal ion interaction.

## II. METHODS

We computed MD trajectories starting with the coordinates from the recently solved C75 precleaved crystal structure PDB ID 3NKB.[8] The deoxynucleotides at positions 1 and 2 were converted to ribonucleotides by addition of 2'-hydroxyls with ideal bond lengths and bond angles. The upstream nucleotide and scissile phosphate were built as described previously,[8,16] leading to a total of 73 nucleotides in the system. Hydrogen atoms were added using Accelrys Discover Studio Visualizer 2.0. C75 was protonated at N3 for the trajectories denoted '$C75^{+}$' and deprotonated for the trajectories denoted '$C75^{o}$'. Residue C41 was protonated in all trajectories to allow the native base triple to form.[19,28] The partial charges used for protonated cytosine are provided in Ref.[16]. The ribozyme was solvated with rigid TIP3P waters[29] in a periodically replicated orthorhombic box. The 10 $Mg^{2+}$ ions resolved in the crystal structure were included. The system was neutralized with $Na^+$ ions, and physiological monovalent ionic strength was added to the solvent to give a physiological ionic strength of ~0.15 M NaCl.

The MD calculations were performed with the Desmond MD program[30,31] using the AMBER99 forcefield.[32,33] Long-range electrostatic interactions were calculated using the Smooth Particle Mesh Ewald method[34] with a cut-off of 12 Å, and SHAKE constraints[35] were applied to bonds involving hydrogen. Following the comprehensive simulated annealing equilibration procedure described in Supporting Information, we collected 25 ns of data at 298 K in the canonical ensemble (*i.e.*, at constant NVT) for each trajectory. A Nosé-Hoover thermostat[36,37] was used to maintain constant temperature, and the time step was 1 fs for all MD trajectories. For each state of the system studied, we propagated at least two independent trajectories and, in some cases, additional independent trajectories.

Electrostatic potential calculations were carried out using numerical solutions to the nonlinear Poisson-Boltzmann (NLPB) equation using the Adaptive Poisson Boltzmann Solver (APBS), as described previously.[16,38,39] Structural coordinates were obtained from the starting structure used for MD. Metal ions and water molecules were omitted from the NLPB calculations, as is customary. Both C75 and C41 were protonated in NLPB calculations. The atomic radii and partial charges were defined using the Amber99 parameter set except for $C41^+$ and $C75^+$, which were defined as described previously.[28] Three-dimensional structures were rendered using VMD,[40] and electrostatic potentials were rendered using PyMOL.[41]

We also performed additional analyses of the MD data. The root mean squared deviations (RMSDs) and radial distribution functions (RDFs) were calculated using VMD,[40] and the root mean square fluctuations (RMSFs) of the residues, including only the heavy atoms, were calculated using GROMACS.[42] To examine the statistical correlation between the thermal fluctuations of pairs of atoms, we calculated the cross-correlation matrices with elements

$$S_{jk} = \frac{C_{jk}}{\sqrt{C_{jj}C_{kk}}} . \tag{3.1}$$

The covariance matrix has elements defined as

$$C_{jk} = \left\langle \left(\mathbf{r}_j - \langle \mathbf{r}_j \rangle\right) \bullet \left(\mathbf{r}_k - \langle \mathbf{r}_k \rangle\right) \right\rangle , \tag{3.2}$$

where the brackets denote an ensemble average over configurations. The cross-correlation matrix describes pairwise atomic motions that are either in phase (i.e., moving in the same direction) or out of phase (i.e., moving in the opposite direction).

To analyze the charge density near the G•U wobbles, we generated charge isodensity plots using VMD.[40] For this purpose, a cube with sides of length 12 Å was defined centered on the G25(O6) atom for the reverse G•U wobble and on the G1(O6) atom for the standard G•U wobble. The volume of this cube was divided into a grid of cubes with sides of length 1 Å. The total charge per 1 Å cube was determined by summing the charges of $Mg^{2+}$, $Na^{+}$, and $Cl^{-}$ in each 1 Å cube. The charge isodensity was plotted by filling the 1 Å grid cubes representing the greatest 30% and 80% of the charge density in the entire 12 Å cube with dark and light colors, respectively.

## III. RESULTS

The goal of our MD simulations is to determine the impact of active site nucleobase protonation and $Mg^{2+}$ ion interactions on the structure and motions of the HDV ribozyme, as well as to gain insight into the catalytic mechanism. Specifically, we examine the impact of C75 protonation because C75 is thought to serve as the general acid in the catalytic mechanism.[6,13,14] We also study the influence of the crystallographically resolved $Mg^{2+}$ ion at the reverse G25•U20 wobble because this $Mg^{2+}$ ion, denoted '$Mg_{rev}$' in this Chapter, is also thought to be directly involved in the catalytic mechanism, acting as a Lewis acid.[8] To perform this analysis, we computed three types of trajectories. In the trajectories denoted '$C75^+/Mg_{rev}$', C75 is protonated, and the $Mg_{rev}$ is included in the starting structure. In the trajectories denoted '$C75^+/noMg_{rev}$', C75 is protonated, but the $Mg_{rev}$ is removed from the starting structure prior to equilibration; in this case, two additional $Na^+$ ions are added to the bulk solvent to maintain overall charge neutrality of the system. In the trajectories denoted '$C75^o/Mg_{rev}$', C75 is deprotonated, one $Na^+$ ion is added to the bulk solvent, and the $Mg_{rev}$ is included in the starting structure. We analyze the active site hydrogen-bonding structure in each of these trajectories. In addition, we analyze the interaction of ions with both the reverse G25•U20 wobble and the nearby standard G1•U37 wobble by generating radial distribution functions and charge isodensity plots, as well as electrostatic potentials via NLPB calculations. Finally, we analyze the atomic fluctuations and correlated motions of the ribozyme by calculating the RMSFs and cross-correlation matrices.

### A. Active site hydrogen bonding interactions

Three key hydrogen-bonding interactions are observed for C75 in the HDV ribozyme in the precleaved crystal structure: C75(N3) to G1(O5'), C75(N4) to G1(O2P) and C75(N4) to

C22(O2P) (Figure 3.2).[8] The time evolution of these distances is shown in Figures 3.3A, B, and C for trajectories of type C75$^+$/Mg$_{rev}$, C75$^+$/noMg$_{rev}$, and C75$^o$/Mg$_{rev}$, respectively. A second set of independent trajectories for each type is provided in Supporting Information (Figure B.1A, B, and C). As illustrated in Figure 3.3A, all three of the hydrogen bonds involving C75 are maintained throughout the first C75$^+$/Mg$_{rev}$ trajectory, and similar behavior was observed for the second independent trajectory (Figure B.1A). These trajectories represent the catalytically competent ribozyme because the C75 is protonated and the crystallographically resolved catalytic Mg$^{2+}$ is bound.

For the first C75$^+$/noMg$_{rev}$ trajectory, the two hydrogen bonds between C75 and G1 are weakened for the first 10 ns but are stably formed for the remainder of the trajectory (red and green lines in Figure 3.3B). In the second independent trajectory (Figure B.1B), these hydrogen bonds are predominantly formed for the first 5 ns but are weakened somewhat for the remainder of the trajectory, although the distances between the heavy atoms still remain less than 4 Å. The fluctuations in hydrogen bonding observed in the early portions of the trajectories may indicate that the systems are not fully equilibrated until ~5 ns. The hydrogen bond between C75(N4)-C22(O2P), which is the one not involving G1 (Figure 3.2, blue line), is the one hydrogen bond that is maintained throughout both trajectories. This analysis suggests that Mg$_{rev}$ may help stabilize the G1 phosphate group to position C75$^+$ for catalysis. This conclusion is also consistent with the inner-sphere contact between the pro-$R_P$ oxygen of G1 and Mg$_{rev}$ identified in the crystal structure of the C75 precleaved ribozyme.[8] Nevertheless, even in the absence of Mg$_{rev}$, the hydrogen-bonding interactions involving C75 are maintained, although slightly weakened in some cases.

The C75$^{o}$/Mg$_{rev}$ trajectory exhibits substantial movement of C75 away from G1 (Figure 3.3C). When C75 is deprotonated, there can be no hydrogen bonding between C75(N3) and G1(O5') as both atoms are hydrogen bond acceptors. In addition, the hydrogen bond between C75(N4) and G1(O2P) is completely absent (red and green lines in Figure 3.3C), and the hydrogen bond between C75(N4) and C22(O2P) is lost after ~6 ns (blue line in Figure 3.3C) in the first trajectory. In the second independent trajectory (Figure B.1C), the hydrogen bonds between C75 and G1 are again completely absent, although the single hydrogen bond between C75(N4) and C22(O2P) is maintained throughout the trajectory. Thus, removal of the proton on N3 of C75 eliminates important hydrogen-bonding and electrostatic interactions, leading to local structural rearrangements that diminish the stability of the other two hydrogen bonds. Overlay of thermally averaged structures from the C75$^{+}$/Mg$_{rev}$ and C75$^{o}$/Mg$_{rev}$ trajectories (Figure B.2) reveals that protonated C75 remains in a reactive alignment, whereas deprotonated C75 does not. When C75 is deprotonated, the interaction between the nucleophilic O2' of U-1 and the catalytic Mg$^{2+}$ and is also lost (Figure B.3. The average distance between the catalytic Mg$^{2+}$ and the nucleophilic O2' of U-1 (red hydrogen bond in Figure 3.1B) is ~2 Å when C75 is protonated, or but ~5.5 Å when C75 is deprotonated (Figure B.3, red line). For both protonation states, however, the distance between the catalytic Mg$^{2+}$ and G1(O2P) (green hydrogen bond in Figure 3.1B) remains ~2 Å throughout the trajectory (Figure B.2, green line). This analysis suggests that protonation of C75 plays an important role in aligning the active site for catalysis.

Hydrogen-bonding interactions in the reverse G•U wobble were also monitored. The time evolution of these two distances is depicted in Figures 3.3D, E, and F for the three types of trajectories. A second set of independent trajectories for each type is provided in Supporting Information (Figure B.1D, E, and F). In the two independent C75$^{+}$/Mg$_{rev}$ trajectories, these hydrogen bonds are stably maintained at ~2.8 Å (Figures 3.3D and B.1D). The $Mg_{rev}^{2+}$ ion maintains its crystallographically modeled position throughout both trajectories, interacting with U-1(O2'), G1(O2P) and U23(O1P) directly, as well as with the reverse G•U wobble via water molecules (Figure 3.3, S3). The reverse G•U wobble hydrogen bonds are also maintained in the two independent $C75^+$/$noMg_{rev}$ trajectories (Figures 3.3E and B.1E), with occasional excursions in the initial part of the first trajectory. In this case, the $Mg_{rev}$ eliminated in the starting structure is replaced by one or more $Na^+$ ions during equilibration and throughout the trajectories, as reported previously.[16] This analysis indicates that the absence of $Mg_{rev}$ does not significantly impact the hydrogen-bonding interactions within the reverse G•U wobble.

The two $C75^o$/$Mg_{rev}$ trajectories exhibited qualitatively different behavior in this region. In the first $C75^o$/$Mg_{rev}$ trajectory, the hydrogen bond between G25(O6) and U20(N3) is maintained throughout the trajectory, but the hydrogen bond between G25(N1) and U20(O4) is lost after ~6 ns, although it increases to only ~3.7 Å (magenta line in Figure 3.3F). Detailed investigation reveals that deprotonation of C75 leads to slight rearrangement of the backbone near residues 21-24 and hence slight movement of U20. In the second independent trajectory, both reverse G•U wobble hydrogen bonds are maintained throughout the trajectory, but the $Mg_{rev}$ moves away and is replaced by a $Na^+$ ion (not shown). Thus, the hydrogen bonds within the reverse G•U wobble are maintained when C75 is deprotonated, although the hydrogen-bonding interactions may be slightly weakened.

### B. Metal ion interactions with the reverse and standard GU wobbles

In this subsection, we analyze and compare the interaction of metal ions with the reverse G25•U20 wobble and with the major groove near the standard G1•U37 wobble. These two metal ion interaction sites are separated by a distance of ~15Å in the C75 precleaved structure. We showed previously that the crystallographically resolved $Mg^{2+}$ interacting with the reverse G25•U20 wobble remains bound and is relatively stationary during MD trajectories.[16] In the present work, we observe that $Mg^{2+}$ interactions with the standard G1•U37 wobble are qualitatively different. In some trajectories, a $Mg^{2+}$ ion near the Hoogsteen face of G35, ~7 Å from the major groove of the standard G•U wobble in the C75 precleaved crystal structure, moves closer to the standard G•U wobble, where it is relatively mobile, while in other instances $Na^+$ ions occupy this region. Thus, while $Mg_{rev}$ may be viewed as "chelated" at the reverse G•U wobble, $Mg^{2+}$-binding in the region near the standard G•U wobble may be viewed as "diffuse", as defined by Draper.[43]

Radial distribution functions between metal ions and the reverse G•U wobble atoms G25(N7), G25(O6), and U20(O2) for the three types of trajectories help define these interactions and are depicted in Figure 3.4. These radial distribution functions, as well as the corresponding running coordination numbers, were generated for both $Mg^{2+}$ and $Na^+$ ions. Charge isodensity plots representing the spatial distribution of the total amount of positive charge for the three types of trajectories are provided in Figure 3.5. The combination of these two types of analyses provides insight into the metal ion interactions.

For the $C75^+/Mg_{rev}$ trajectory, a single $Mg^{2+}$ was persistently located within ~4.5 Å of all three specified atoms of the reverse G•U wobble, as indicated by the radial distribution plot and the running coordination number with sharp step-like behavior to unity at this distance (Figures 3.4A and B, red lines). Since $Mg^{2+}$ remained bound for the entire trajectory, no $Na^+$ was observed in this region (Figures 3.4C and D, red lines). The charge isodensity analysis provides further confirmation of this behavior. For the $C75^+/Mg_{rev}$ trajectory (Figure 3.5A), the $Mg^{2+}$ was highly localized near the O6 of G25 and the O2 of U20. The second independent C75$^+$/Mg$_{rev}$ trajectory exhibited qualitatively similar behavior (Figures B.4, B.5A, and B.5D for stereoview).

For the C75$^+$/noMg$_{rev}$ trajectory, the Mg$^{2+}$ was removed from this region prior to equilibration, and no other Mg$^{2+}$ was observed to move into this region. Instead, a Na$^+$ ion moved into this vicinity and was located within ~2.5 Å of all three specified atoms of the reverse G•U wobble for a significant portion of the trajectory (Figures 3.4C and D, blue lines). In this case, the running coordination number (Figure 3.4D) does not reach unity until larger distances and does not exhibit sharp step-like behavior as in Figure 3.4B. This behavior indicates that Na$^+$ did not remain in this region throughout the entire trajectory. Thus, the Na$^+$ ion was relatively mobile compared to the Mg$^{2+}$ in the C75$^+$/Mg$_{rev}$ trajectory. This behavior was confirmed by analysis of the charge isodensity plot. For the first C75$^+$/noMg$_{rev}$ trajectory (Figure 3.5B), the positive charge density was relatively delocalized, corresponding to the more mobile Na$^+$ that replaced the removed Mg$_{rev}$. In the second independent C75$^+$/noMg$_{rev}$ trajectory (Figure B.5B), the Na$^+$ ion was more localized near the O2 of U20 than in the first trajectory. This observation is consistent with the radial distribution functions for this second trajectory (Figures B.4C and D, blue lines), which exhibit a peak at ~2.5 Å for U20(O2) (solid blue line), with a corresponding running coordination number of nearly unity at this distance, but no peaks for G25(N7) (dashed blue line) or G25(O6) (dotted blue line). Thus, the Na$^+$ ion is localized near O2 of U20 for the second trajectory but is delocalized over a much larger region for the first trajectory.

The two C75$^o$/Mg$_{rev}$ trajectories exhibited different behavior. In particular, the Mg$^{2+}$ remained bound to the reverse G•U wobble throughout the first trajectory (Figures 3.4A and B, green lines), but it moved away and was replaced by a Na$^+$ ion in the second trajectory (Figure B.4, green lines). This behavior is also illustrated by the charge isodensity plots. For the first

C75°/Mg$_{rev}$ trajectory (Figure 3.5C), the Mg$^{2+}$ was highly localized, with a distribution similar to that for the C75$^{+}$/Mg$_{rev}$ trajectory. In the second independent C75°/Mg$_{rev}$ trajectory (Figure B.5C), the Mg$^{2+}$ was replaced by the more mobile Na$^{+}$, leading to a more delocalized positive charge distribution (compare Figures B.5C and 5C). As discussed above, deprotonation of C75 leads to structural rearrangements of the active site, including the loss of hydrogen bonding at C75 to the leaving group, displacement of the O2' of U-1 from Mg$_{rev}$, and the slight weakening of hydrogen-bonding interactions within the reverse G•U wobble (Figures 3.3, B.1). The replacement of the Mg$_{rev}$$^{2+}$ by Na$^{+}$ in the second trajectory may be linked to these structural rearrangements at the active site. Nevertheless, the reverse G•U wobble still appears to effectively interact with either a divalent or monovalent metal ion, even when C75 is deprotonated.

We also performed the same analyses for the standard G1•U37 wobble (radial distribution functions in Figures 3.6 and B.6, and charge isodensity plots in Figure 3.7 and S7). In this case, both hydrogen bonds within the standard wobble were maintained for all trajectories, which is consistent with our experiments that supported importance of a stable base pair at this position.[26] The region containing the standard G•U wobble and P1.1 (G38, G39) just below provides a favorable negative pocket for cationic metal ions (see below). The radial distribution functions and running coordination numbers for the C75$^{+}$/Mg$_{rev}$ and C75$^{+}$/noMg$_{rev}$ trajectories (Figure 3.6 and S6, red and blue lines, respectively) indicate intermittent presence of Mg$^{2+}$ and Na$^{+}$ in this region. When present (Figures 3.6A-B), the Mg$^{2+}$ moved into the standard G•U wobble region from its crystallographic position near the G35 Hoogsteen face, which is further up P1. Movement of this ion is shown with an arrow in Figure 3.7C, where ending points of the arrow also happen to correspond to crystallographic Mg$^{2+}$ ions, albeit from two different structures (3NKB and 2OJ3).[8,21] When the $Mg^{2+}$ ion was absent, $Na^{+}$ monovalent ions were in this region instead, which is characteristic of diffuse sites.[43] For the $C75^{o}/Mg_{rev}$ trajectories (Figure 3.6, S6A-D, green lines), the region near the standard G•U wobble was occupied mainly by $Na^{+}$ ions, again characteristic of diffuse sites.

The charge isodensity plots for the standard G1•U37 wobble trajectories (Figure 3.7A) illustrate that the positive charge density was delocalized over the region near the standard G•U wobble for all trajectories (Figure B.7D for stereoview), indicating diffuse metal ions. Qualitatively similar behavior was exhibited for the second set of trajectories (Figure B.7). In addition, the charge isodensity plot for $Mg^{2+}$ overlaid well with positioning of $Mg^{2+}$ from crystallography, and a similar plot for $Na^{+}$ overlaid well with position of $Tl^{+}$, a heavy monovalent ion, from crystallography (Figure 3.7C); this observation suggests that charge isodensity calculations are reflective of experimentally determined favorable ion interactions. In summary, behavior near the reverse and standard G•U wobbles contrasts strongly (compare $C75^{+}/Mg_{rev}$ trajectories in Figures 3.5A and 3.7A), suggestive of chelation in the former and a diffuse ion atmosphere in the latter.

Non-linear Poisson-Boltzmann (NLPB) calculations of these two G•U wobbles provide a plausible explanation for these two qualitatively different types of interactions. As shown in Figure 3.8, the potential near the standard G•U wobble is ~ –20 kT/e, which is similar to previously reported values for tandem G•U wobbles and to that of the phosphate backbone,[44] and this negatively charged region extends up P1 into the vicinity of the G2 and down P1.1 into the vicinity of G38 and G39. This potential is very different from that at the reverse G•U wobble, where the potential was much more negative (~ –150 kT/e) and more localized.[16] These qualitatively different potentials are consistent with a more mobile and diffuse ion atmosphere near the standard G•U wobble and a less mobile and chelating ion near the reverse G•U wobble.

As mentioned above, the standard G•U wobble is sometimes replaced by a standard Watson-Crick GC or AU base pair in nature. We therefore analyzed the effect of a U37C mutation (=GC base pair) on hydrogen-bonding interactions at the active site and on the metal interaction with both the reverse and standard G•U wobbles. For this purpose, we propagated trajectories of the U37C mutant of the type $C75^{+}/Mg_{rev}$, where the crystallographic catalytic $Mg^{2+}$ was included and the C75 was protonated. As shown in Figure B.7, all three of the hydrogen bonds involving C75 remained stable in the mutant; in addition, the standard hydrogen bonding in the GC base pair was retained (Figure 3.7B). We also found that the $Mg_{rev}$ remained bound, as in the analogous wild-type trajectories. As shown in Figure 3.7B, the positive charge density near the G1-C37 region was qualitatively similar to that observed in the wild-type trajectories (Figure 3.7A), except for a shift in the positive charge density away from the N4 atom of residue 37, as expected for the keto to amine change associated with the mutation of U to C. This shift in the positive charge density was also observed for a second independent trajectory of the U37C mutant (Figure B.8), where a lack of charge density is found near the 4 position of C37.

## C. Atomic fluctuations and correlated motions

We tested whether changes in C75 protonation and $Mg^{2+}$ interactions at the reverse G•U wobble affect the global structure or motion of the ribozyme. The root mean square deviation (RMSD) between the thermally averaged structures for the $C75^{+}/Mg_{rev}$ and $C75^{+}/noMg_{rev}$ trajectories and for the $C75^{+}/Mg_{rev}$ and $C75^{o}/Mg_{rev}$ trajectories is relatively small, as shown in Tables S1 and S2. These RMSDs thus indicate that the global structure is not significantly affected by C75 protonation or $Mg_{rev}$. Note that deprotonation of C75 has a slightly larger effect on the RMSD (Table S1) of the active site than does removal of $Mg_{rev}$ (Table S2), consistent with the above analysis.

The root mean square fluctuations (RMSFs) of the residues for the $C75^{+}/Mg_{rev}$, $C75^{+}/noMg_{rev}$ and $C75^{o}/Mg_{rev}$ trajectories are depicted in Figure 3.9. The RMSFs are qualitatively similar for all three trajectories, as well as for the second set of independent trajectories (Figure B.10). In all cases, the terminal residues at the top of P1 (magenta region), P2 (cyan region), and P4 (orange region) exhibit the greatest fluctuations. The solvent exposed and bulged residues (violet arrows identifying C27, A42, and G76) also exhibited significant fluctuations. The other regions did not fluctuate significantly during these trajectories, other than typical thermal motions of 1-2 Å.

The cross-correlation matrix for the heavy atoms in the $C75^{+}/Mg_{rev}$ trajectory is plotted in Figure 3.10. Cross-correlation matrices for the $C75^{+}/noMg_{rev}$ and $C75^{o}/Mg_{rev}$ trajectories are qualitatively similar and are provided in Figure B.11. Red regions indicate pairs of atoms that are moving in the same direction (*i.e.* correlated), and blue regions indicate pairs of atoms that are moving in the opposite direction (*i.e.* anticorrelated). The red diagonal regions perpendicular to the main diagonal represent strands of helical regions that are moving together. Helical regions corresponding to P1, P2, P3 and P4 are evident in Figure 3.10 (white ellipses). In addition, correlated motions extend along the two aforementioned coaxial helical stacks: P2 and P3, which form one coaxial stack, are correlated, as are P1 and P1.1, and P1.1 and the top of P4, which together form an extended coaxial stack. In terms of anticorrelated motions, the P1 helix moves in the opposite direction from the P2 helix (Figure 3.10, yellow circles). Anticorrelated motions are also found near L4 (the loop of P4); however, this region is not part of the native structure.[8]

Focusing on the reverse and standard G•U wobbles, we observed strongly anticorrelated motions between G25 and U37 for all three types of trajectories (Figures 3.11 and B.12). This observation suggests that the reverse and standard G•U wobbles are moving in opposite directions, despite the distance of 23 Å between residues G25 and U37. This behavior is also observed in the absence of $Mg_{rev}^{2+}$ and C75 protonation (Figure B.12), suggesting that it is not due solely to the positive charge on the metal or C75. The possible connection of these anticorrelated motions to catalysis is a direction for future research.

## IV. DISCUSSION

In the Discussion, we compare the simulations and calculations to experimental data. Overall, we find broad consistency between theory and experiment. The Discussion is organized into the same three subsections as the Results.

### A. Active site hydrogen bonding interactions

The hydrogen-bonding interactions at $C75^+$ and the reverse G•U wobble presented in Figures 3.3 and B.1 suggest that the catalytic $Mg^{2+}$ ion and C75 protonation provide structural stability to the active site. For the $C75^+/Mg_{rev}$ trajectories, observation that the hydrogen bonds involving $C75^+$ and within the reverse G•U base pair are well-formed throughout the trajectories suggests that this state of the ribozyme is particularly stable. This observation is consistent with catalytic competence of this state, in which $C75^+$ acts as the general acid to protonate the 5'-bridging oxygen of G1 and the $Mg^{2+}$ ion at the reverse wobble acts as the Lewis acid (Figure

3.1B).[8] The initial structure of the HDV ribozyme bound to its substrate was generated by superposing the cleavage site dinucleotide from the hammerhead ribozyme onto the HDV ribozyme active site. In the resulting model it was not adjusted further and, therefore, the C75(N3)-G1(O5') bond was long at 3.6 Å. During the MD trajectories, C75(N3) and G1(O5') moved to within hydrogen bonding distance of each other. Thus, the MD supports and refines the model generated from the crystal structure.

For the $C75^+/noMg_{rev}$ trajectories, the behavior of $C75^+$ and the reverse G•U base pair in the absence of $Mg_{rev}$ has implications for the mechanism of channel 1 of the HDV ribozyme, wherein reaction is promoted in the presence of $Na^+$ ions without divalent ions. Observation that C75 is positioned in a catalytically competent fashion at least part of the time without $Mg_{rev}$ is consistent with experiments that support C75 acting as the general acid under these conditions with a reduced rate.[15] Additionally, these observations are largely consistent with product and precleaved crystal structures being very similar in overall geometries despite absence of the catalytic $Mg^{2+}$ ion in the crystal structure of the HDV ribozyme product.[8,12] For the $C75^o/Mg_{rev}$ trajectories, observation that C75 does not remain in the crystallographically-observed hydrogen-bonding patterns and that the O2' of U-1 is no longer engaged with the catalytic $Mg^{2+}$ ion suggests that protonation of C75 in the precleaved state assists in organizing the active site. This point is also illustrated by an overlay of the thermally averaged structures from the $C75^+/Mg_{rev}$ and $C75^o/Mg_{rev}$ trajectories, which revealed loss of catalytic positioning upon deprotonation (Figure B.2). Walter and co-workers also provided evidence that C75 positioning changes in going from precleaved to product.[45]

These changes in conformation afforded by C75 protonation are likely local accommodations rather than global, as supported by three observations. First, comparison of the thermally averaged structures in the C75$^+$ and C75$^o$ states reveals only small RMSDs in atom positions, especially in the active site (Table S1). Second, the overall geometry of the product and precleaved crystal structures, which are likely deprotonated and protonated respectively,[18,46] are similar.[8,12] Third, there is a lack of large changes in Raman signal upon protonation of C75 in precleaved HDV ribozyme crystals.[18]

The recent crystal structure of the HDV ribozyme reveals a $Mg^{2+}$ ion in proximity to the cleavage site. Its position relative to C75 and the G1(O5') led us to propose that this ion participated in catalysis by acting as a Lewis acid, facilitating the deprotonation of the U-1(O2') nucleophile.[8] In this model of the HDV ribozyme in the precleaved state, $Mg_{rev}$ interacts with both U-1(O2') and G1(O2P). The interaction between $Mg_{rev}$ and the scissile phosphate is supported by the deleterious effect of sulfur substitution G1(O2P).[47] Proton inventory experiments in the presence and absence of $Mg^{2+}$ are consistent with $Mg^{2+}$-facilitated deprotonation of the U-1(O2'),[15,48] although direct interaction between U-1(O2') and the active site $Mg^{2+}$ ion has never been directly probed. The trajectories described here explore the feasibility of this model by examining the stability of the interactions between $Mg_{rev}$ and both U-1(O2') and G1(O2P). The stable interaction maintained throughout both C75$^+$/$Mg_{rev}$ trajectories (Figure B.3) is consistent with the crystallographic model corresponding to a reactive conformation.

### B. Metal ion interactions with the reverse and standard G•U wobbles

Analysis of metal ion interaction at the reverse and standard G•U wobbles is consistent with a chelated ion at the reverse G•U wobble and a diffuse monovalent or divalent ion interaction at the standard G•U wobble. Presence of both monovalent and divalent ions in the vicinity of the standard G•U wobble is diagnostic of diffuse sites.[43] For the reverse G•U wobble, the electrostatic potential was exceptionally negative at –150 kT/e[16] as expected for chelated sites,[44] the radial distribution functions exhibited sharp peaks with running coordination numbers of unity at binding distances (Figure 3.4), and the charge isodensity plots indicated highly localized interaction (Figure 3.5). In contrast, for the standard G•U wobble, the electrostatic potential was only ~ –20 kT/e (Figure 3.8) as expected for this type of motif,[44] the radial distribution functions were less sharp with fractional running coordination numbers at binding distances (Figure 3.6), and the charge isodensity plots indicated diffuse ion interaction (Figure 3.7). These differences in metal ion interactions at the two types of G•U wobbles arise from the greater magnitude and higher localization of the negative charge in the pocket at the reverse G•U wobble, as compared to the weaker and more delocalized negative charge near the standard G•U wobble, which extends up P1 and down into P1.1 (Figure 3.8).

A diffuse ion atmosphere near the standard G•U wobble is also supported by prior experiments from our labs that used 7-deazaguanosine (*i.e.,* N7 changed to a C-H) single and double substitutions at G1 and G2 in the HDV ribozyme substrate.[49] These variant RNAs, which have lessened negative potential in the major groove, reacted more than 10 times slower than wild-type in the double mutant. Raman spectroscopic characterization of a ribozyme crystal with a single 7-deazaguanosine substitution at G1 revealed a loss of interaction between N7 and a $Mg^{2+}$ ion; however, the $Mg^{2+}$ dependence of the ribozyme reaction was only subtly perturbed. That the 7-deazaguanosine-modified RNAs still bound $Mg^{2+}$ avidly in spite of the loss of interaction with G1(N7) is again consistent with a diffuse site. Loss of a factor of 10 in rate with the double deazaguanosine mutant suggests that the metal ion at the standard G•U position could contribute to rate enhancement, possibly by affecting folding (see below).

Our MD simulations indicated qualitatively similar behavior for the U37C mutant and the wildtype HDV ribozyme in terms of the active site hydrogen-bonding interactions, as well as the metal ion interactions with the reverse and standard G•U wobbles. This observation is consistent with the known occurrence of GC and AU Watson-Crick base pairs in HDV and HDV-like ribozymes at this position.[9,10,25] It is also consistent with the experimental determination that AU or GC at the cleavage site gives indistinguishable reaction kinetics and $Mg^{2+}$-dependence.[26]

Draper has argued that diffuse metal ions are major factors in stabilizing RNA tertiary structures.[43] Thus, one possibility is that the ion at the junction of P1 and P1.1 plays a role in stabilizing the tertiary fold of the RNA, where it imparts stability from entropy gain in delocalizing along the RNA.[43,50] In contrast, the ion at the reverse G•U wobble plays a role in catalyzing the reaction as a Lewis acid, where fixing its position is likely critical for chemistry to occur. The two-base pair P1.1 helix is one of the smallest helices observed in any functional RNA, and it is often just one base pair in the HDV-like ribozymes;[10] thus, it seems that this region could benefit from added tertiary structure stabilization. The HDV ribozyme is known to be exceptionally stable, with a melting temperature for tertiary structure of ~75 °C in just 0.1 mM $Mg^{2+}$.[11] Such exceptional stability is consistent with the relative independence of C75 protonation and metal ion interaction to the reverse and standard G•U wobbles.

### C. Atomic fluctuations and correlated motions

Essentially no differences in the RMSF and cross-correlation plots were found among the various C75 protonation and $Mg^{2+}$-bound states of the ribozyme. The RMSF analysis indicates that the larger thermal motions are limited to the periphery of the ribozyme, while the core of the ribozyme exhibits only relatively small thermal motions. Correlated motions were found within each helix, as well as between the coaxial stacks of P2/P3 and P1/P1.1/P4. These observations suggest that the ribozyme exists as at least two somewhat stiff rods. Interestingly, portions of these two stacks appear to be anticorrelated with each other. In particular, P1 and P2 are strongly anticorrelated. Moreover, the motions of G25 and U37 also appear to be anticorrelated. Whether these motions contribute to catalysis, perhaps by aligning reactive functionalities, is unclear at present.

## V. CONCLUSIONS

We carried out MD simulations to analyze metal ion interaction with a standard and a reverse G•U wobble and to investigate the impact of C75 protonation on the structure and motions of the HDV ribozyme. Importantly, the calculations and simulations herein agree with a wealth of experimental data. We identified two classes of $Mg^{2+}$ ions associated with the ribozyme, chelated and diffuse, which appear to contribute to catalysis and stability, respectively. We observed that protonation of C75 locally organizes the active site in a manner that facilitates the catalytic mechanism in which $C75^{+}$ acts as a general acid and $Mg^{2+}$ as a Lewis acid. On the other hand, protonation of C75 does not alter the global structure and equilibrium thermal motions of the ribozyme. We also found that the catalytic $Mg^{2+}$ ion does not significantly impact C75 hydrogen-bonding interactions or the global structure and thermal motions of the ribozyme. Furthermore, our simulations indicate that C75 protonation and the chelated interaction of $Mg^{2+}$ to the reverse G•U wobble have virtually no impact on the diffuse metal ion interactions with the standard G•U wobble. In a previous study,[8] we determined that the active site structure, including the reverse G•U wobble with the bound catalytic $Mg^{2+}$, is structurally similar in the precleaved and post-cleaved forms, and that protonation has a minimal effect on Raman spectra.[18] These findings, along with the present analysis, suggests that the chemical reaction pathway of the ribozyme is dominated by small local motions at the active site, rather than large-scale global conformational changes.

## ACKNOWLEDGMENTS

We thank Alexander Soudackov and Phil Hanoian for helpful discussions. This project was supported by NIH grant R01GM095923 (B.L.G. & P.C.B), NIH grant GM56207 (S.H.S.), and instrumentation funded by the NSF through grant OCI-0821527.

## Supporting Information

Details of equilibration procedure, tables of RMSDs, and additional figures depicting data from MD trajectories are provided in Appendix B.

## References

(1) DeRose, V. J. *Curr. Opin. Struct. Biol.* **2003**, *13*, 317-324.

(2) Hougland, J. L.; Kravchuk, A. V.; Herschlag, D.; Piccirilli, J. A. *PLoS Biol.* **2005**, *3*, 1536-1548.

(3) Ferre-D'Amare, A. R.; Scott, W. G. *Cold Spring Harb. Perspect. Biol.* **2010**.

(4) Bevilacqua, P. C.; Yajima, R. *Curr. Opin. Chem. Biol.* **2006**, *10*, 455-464.

(5) Bevilacqua, P. C. Proton Transfer in Ribozyme Catalysis. In *Ribozymes and RNA Catalysis*; Lilley, D. M., Eckstein, F., Eds.; Royal Society of Chemistry: Cambridge, 2008; pp 11-36.

(6) Nakano, S.; Proctor, D. J.; Bevilacqua, P. C. *Biochemistry* **2001**, *40*, 12022-12038.

(7) Perrotta, A. T.; Been, M. D. *Biochemistry* **2006**, *45*, 11357-11365.

(8) Chen, J. H.; Yajima, R.; Chadalavada, D. M.; Chase, E.; Bevilacqua, P. C.; Golden, B. L. *Biochemistry* **2010**, *49*, 6508-6518.

(9) Salehi-Ashtiani, K.; Luptak, A.; Litovchick, A.; Szostak, J. W. *Science* **2006**, *313*, 1788-1792.

(10) Webb, C.-H. T.; Riccitelli, N. J.; Ruminski, D. J.; Lupták, A. *Science* **2009**, *326*, 953.

(11) Nakano, S.; Cerrone, A. L.; Bevilacqua, P. C. *Biochemistry* **2003**, *42*, 2982-2994.

(12) Ferre-D'Amare, A. R.; Zhou, K.; Doudna, J. A. *Nature* **1998**, *395*, 567-574.

(13) Nakano, S.; Chadalavada, D. M.; Bevilacqua, P. C. *Science* **2000**, *287*, 1493-1497.

(14) Das, S. R.; Piccirilli, J. A. *Nat. Chem. Biol.* **2005**, *1*, 45-52.

(15) Cerrone-Szakal, A. L.; Siegfried, N. A.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2008**, *130*, 14504-14520.

(16) Veeraraghavan, N.; Ganguly, A.; Chen, J. H.; Bevilacqua, P. C.; Hammes-Schiffer, S.; Golden, B. L. *Biochemistry* **2011**, in press.

(17) Nakano, S.; Bevilacqua, P. C. *Biochemistry* **2007**, *46*, 3001-3012.

(18) Gong, B.; Chen, J. H.; Chase, E.; Chadalavada, D. M.; Yajima, R.; Golden, B. L.; Bevilacqua, P. C.; Carey, P. R. *J. Am. Chem. Soc.* **2007**, *129*, 13335-13342.

(19) Ferre-D'Amare, A. R.; Doudna, J. A. *J. Mol. Biol.* **2000**, *295*, 541-556.

(20) Ke, A.; Zhou, K.; Ding, F.; Cate, J. H.; Doudna, J. A. *Nature* **2004**, *429*, 201-205.

(21) Ke, A.; Ding, F.; Batchelor, J. D.; Doudna, J. A. *Structure* **2007**, *15*, 281-287.

(22) Saenger, W. Principles of Nucleic Acid Structure; Springer-Verlag: New York, 1984.

(23) Siegfried, N. A.; Metzger, S. L.; Bevilacqua, P. C. *Biochemistry* **2007**, *46*, 172-181.

(24) Sokoloski, J. E.; Godfrey, S. A.; Dombrowski, S. E.; Bevilacqua, P. C. **submitted**.

(25) Chadalavada, D. M.; Cerrone-Szakal, A. L.; Bevilacqua, P. C. *RNA* **2007**, *13*, 2189-2201.

(26) Cerrone-Szakal, A. L.; Chadalavada, D. M.; Golden, B. L.; Bevilacqua, P. C. *RNA* **2008**, *14*, 1746-1760.

(27) Tanner, N. K.; Schaff, S.; Thill, G.; Petit-Koskas, E.; Crain-Denoyelle, A. M.; Westhof, E. *Curr. Biol.* **1994**, *4*, 488-498.

(28) Veeraraghavan, N.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Mol. Biol.* **2010**, *402*, 278-291.

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

(30) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I. K.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. *in Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)* **2006**.

(31) D. E. Shaw Research, N. Y. Desmond Molecular Dynamics System, 2008.

(32) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(33) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comp. Chem.* **2000**, *21*, 1049-1074.

(34) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

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

(36) Nosé, S. *Mol. Phys.* **1984**, *52*, 255-268.

(37) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695-1697.

(38) Gilson, M. K.; Sharp, K. A.; Honig, B. *J. Comput. Chem.* **1987**, *9*, 327-335.

(39) Gilson, M. K.; Honig, B. *Proteins* **1988**, *4*, 7-18.

(40) Humphrey, W.; Dalke, A.; Schulten, K. *J. Molec. Graphics* **1996**, *14*, 33-38.

(41) DeLano, W. L. The PyMOL molecular graphics system; DeLano Scientific, San Carlos, CA, USA, 2002.

(42) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306-317.

(43) Draper, D. E. *RNA* **2004**, *10*, 335-343.

(44) Chin, K.; Sharp, K. A.; Honig, B.; Pyle, A. M. *Nat. Struct. Biol.* **1999**, *6*, 1055-1061.

(45) Harris, D. A.; Rueda, D.; Walter, N. G. *Biochemistry* **2002**, *41*, 12051-12061.

(46) Luptak, A.; Ferre-D'Amare, A. R.; Zhou, K.; Zilm, K. W.; Doudna, J. A. *J. Am. Chem. Soc.* **2001**, *123*, 8447-8452.

(47) Fauzi, H.; Kawakami, J.; Nishikawa, F.; Nishikawa, S. *Nucleic Acids Res.* **1997**, *25*, 3124-3130.

(48) Nakano, S.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2001**, *123*, 11333-11334.

(49) Chen, J. H.; Gong, B.; Bevilacqua, P. C.; Carey, P. R.; Golden, B. L. *Biochemistry* **2009**, *48*, 1498-1507.

(50) Amzel, L. M. *Proteins* **1997**, *28*, 144-149.

## FIGURES

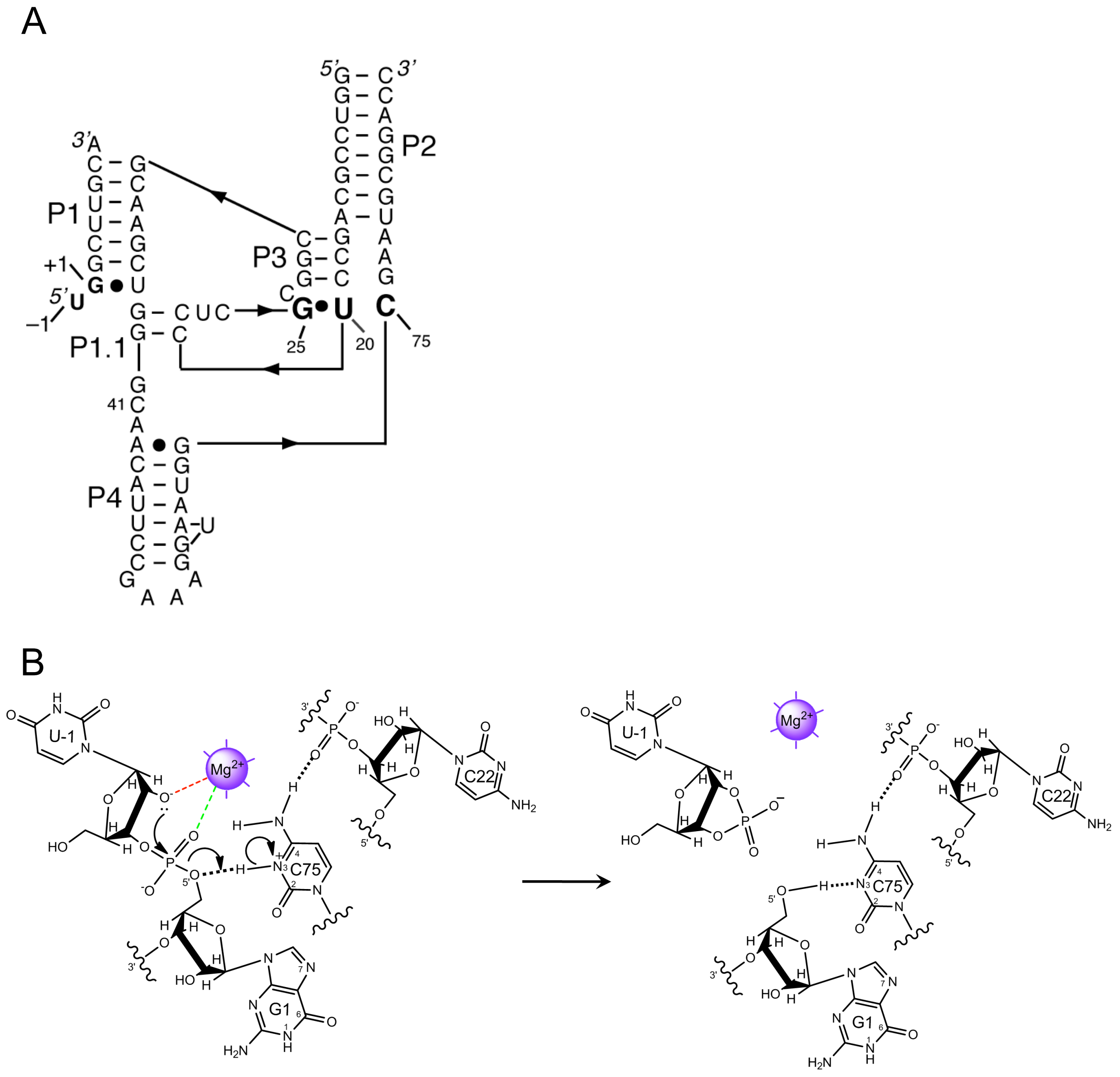

**Figure 3.1.** Secondary structure of the precleaved genomic HDV ribozyme and proposed mechanism. (A) Secondary structure of the precleaved HDV ribozyme (PDB ID 3NKB)[8] used for the MD studies. There are five pairing regions, P1 to P4, and P1.1. Cleavage site is between U–1 and G1. Numbering is based on the genomic HDV ribozyme. The reverse G•U wobble and

catalytic C75 residue are in large bold font. (B) Proposed mechanism of HDV ribozyme self-cleavage in which $Mg^{2+}$ serves as a Lewis acid and C75 as a general acid.[8]

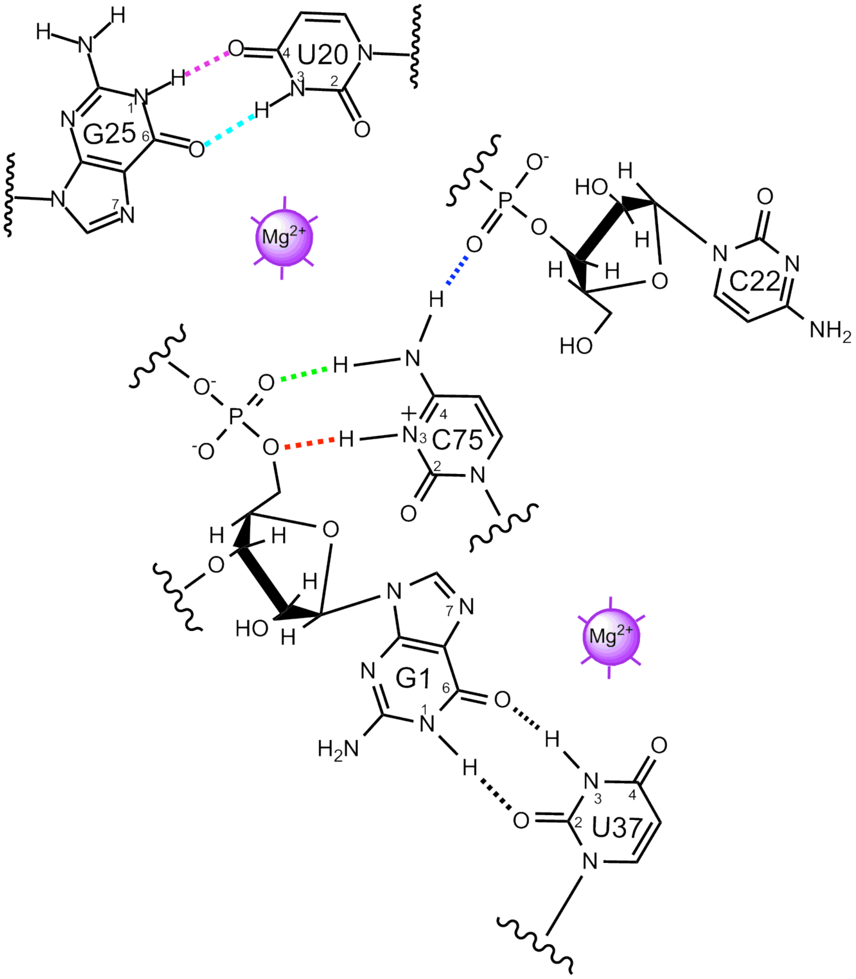

**Figure 3.2.** Key active site interactions in the genomic HDV ribozyme. The wobble between G25 and U20 is a reverse G•U wobble, and the wobble between G1 and U37 is a standard G•U wobble. The precleaved crystal structure shows the presence of the $Mg^{2+}$ ion at the reverse wobble, and the $Mg^{2+}$ ion at the standard wobble is included for illustrative purposes.

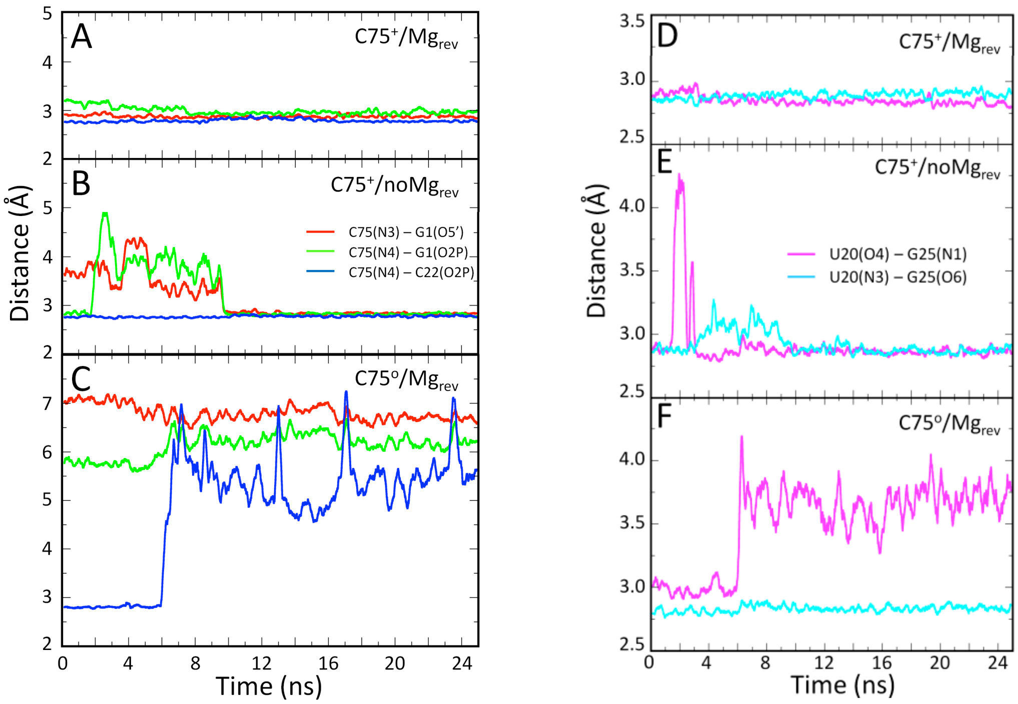

**Figure 3.3.** Hydrogen bonding interactions in the active site of the HDV ribozyme. The distances shown are between heavy atoms of the hydrogen-bonding interactions involving C75 (panels A, B, and C) and within the reverse G•U wobble (panels D, E, and F). Results are provided for the three different types of trajectories: C75$^+$/Mg$_{rev}$ (panels A and D), C75$^+$/noMg$_{rev}$ (panels B and E), and C75$^o$/Mg$_{rev}$ (panels C and F). The color scheme is the same as in Figure 3.2.

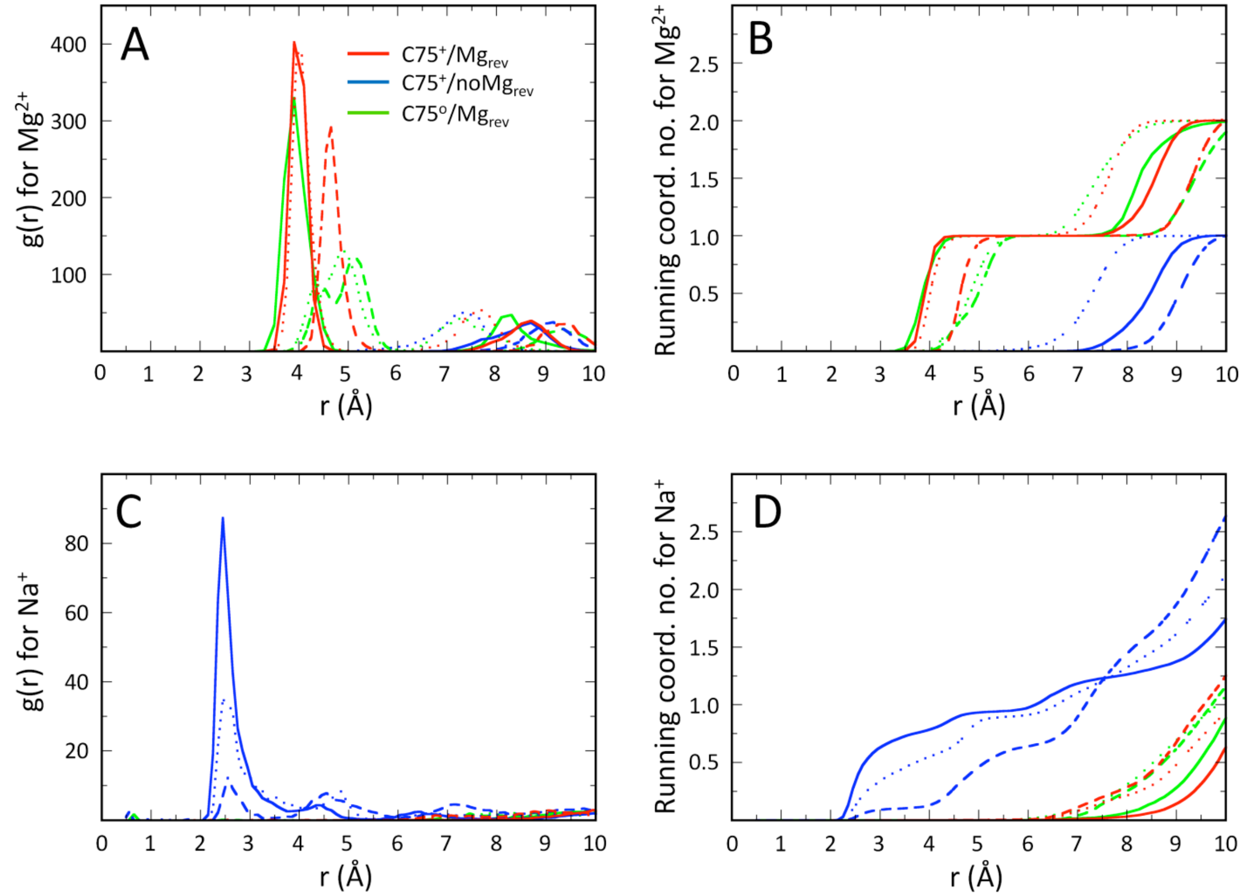

**Figure 3.4.** Analysis of metal ion interaction at the reverse G25•U20 wobble. Radial distribution functions (panels A and C) and running coordination numbers (panels B and D) are plotted for $Mg^{2+}$ ions (panels A and B) and $Na^{+}$ ions (panels C and D) with respect to G25(N7) (dashed line), G25(O6) (dotted line) and U20(O2) (solid line) atoms of the reverse wobble. Results are provided for $C75^{+}/Mg_{rev}$ (red), $C75^{+}/noMg_{rev}$ (blue) and $C75^{o}/Mg_{rev}$ (green) trajectories.

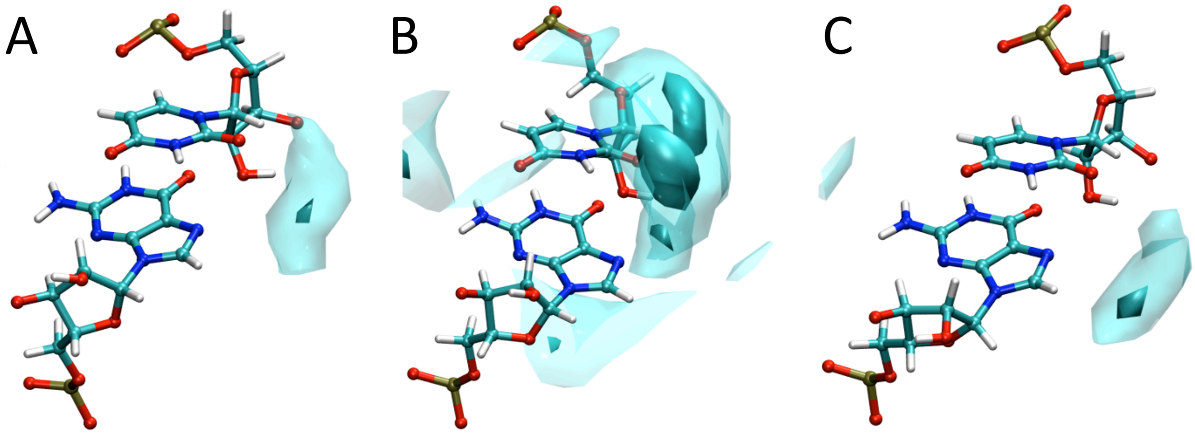

**Figure 3.5.** Charge isodensity plots in the region of the reverse G25•U20 wobble for (A) C75$^+$/Mg$_{rev}$ , (B) C75$^+$/noMg$_{rev}$, and (C) C75°/Mg$_{rev}$ trajectories. The darker cyan represents the greatest 30% of the positive charge density, and the lighter cyan represents the greatest 80% of the positive charge density. The positive charge density is due predominantly to Mg$^{2+}$ in panels A and C, and to Na$^+$ in panel B. Stereoview of panel A is provided as Supporting Figure B.5D.

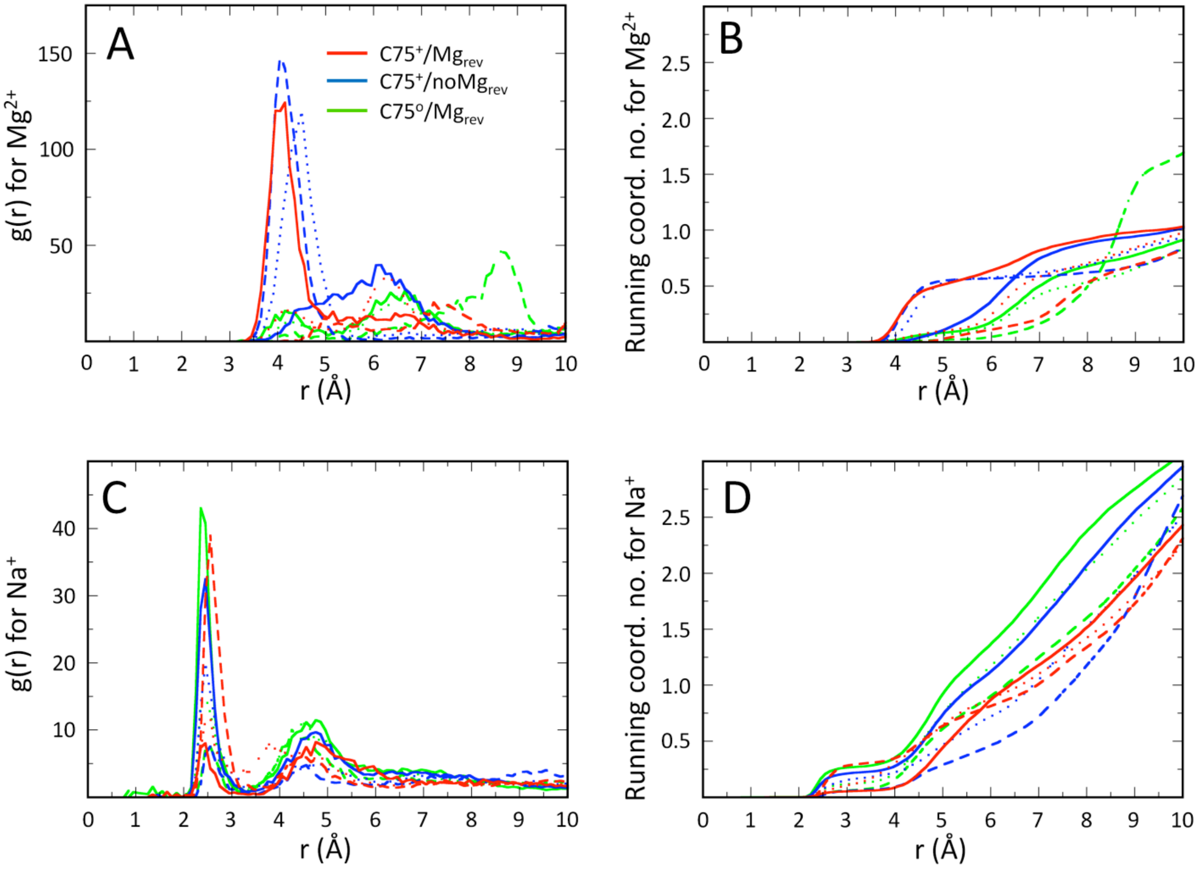

**Figure 3.6.** Analysis of metal ion interaction at the standard G1•U37 wobble. Radial distribution functions (panels A and C) and running coordination numbers (panels B and D) are plotted for Mg$^{2+}$ ions (panels A and B) and Na$^+$ ions (panels C and D) with respect to G1(N7) (dashed line), G1(O6) (dotted line) and U37(O2) (solid line) atoms of the standard wobble. Results are provided for C75$^+$/Mg$_{rev}$ (red), C75$^+$/noMg$_{rev}$ (blue) and C75$^o$/Mg$_{rev}$ (green) trajectories.

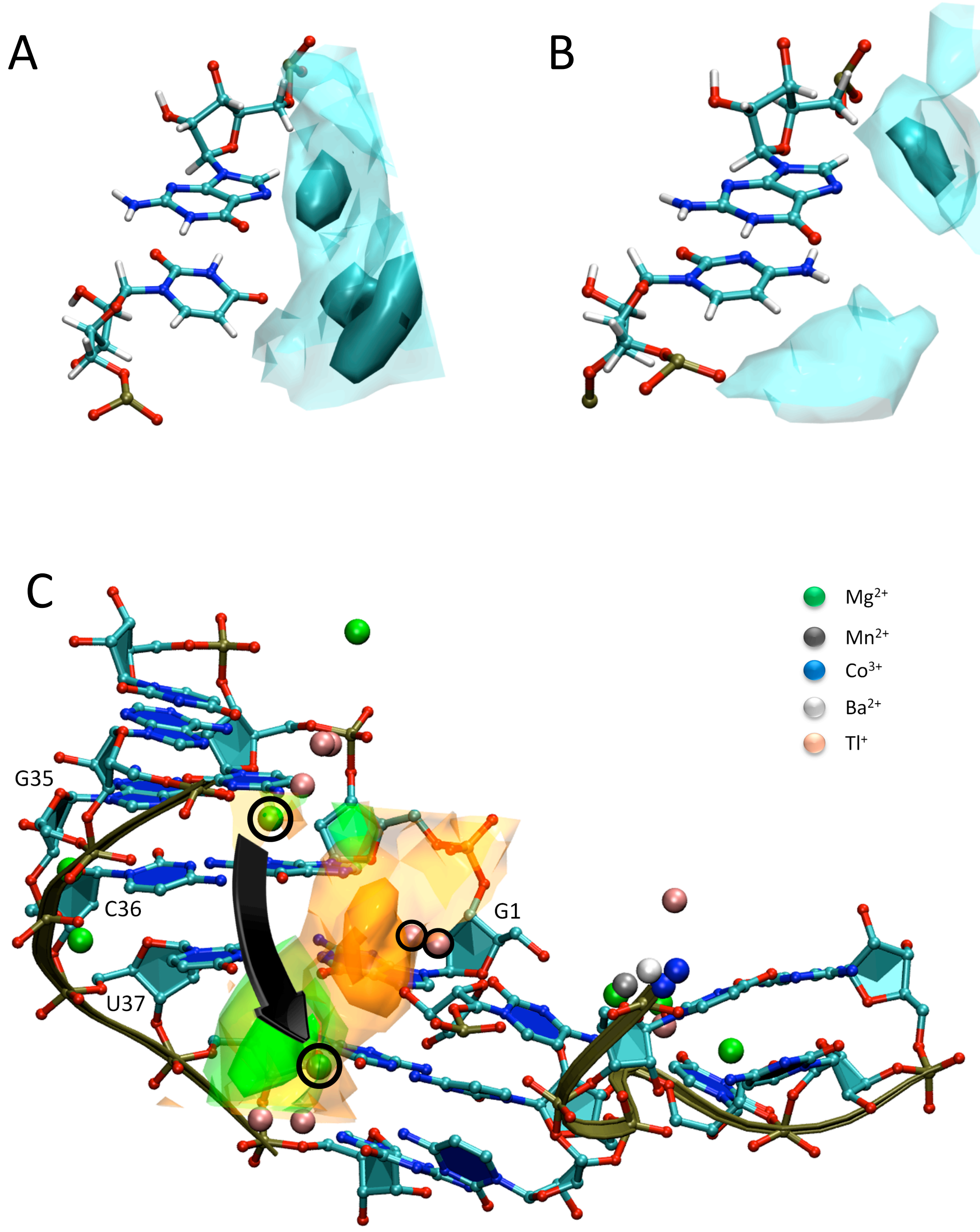

**Figure 3.7.** Metal ion interactions near the standard G•U wobble. Charge isodensity plots in the region of the (A) G1•U37 standard wobble in the wild-type and (B) G1C37 Watson-Crick base pair in the U37C mutant for the C75$^{+}$/Mg$_{\text{rev}}$ trajectories. The darker cyan represents the greatest

30% of the positive charge density, and the lighter cyan represents the greatest 80% of the positive charge density. The positive charge density is due to both $Mg^{2+}$ and $Na^{+}$. Stereoview of panel A is provided as Supporting Figure B.6D. (C) Comparison of divalent and monovalent ions from theoretical and experimental methods overlap. Charge isodensity plots for $Mg^{2+}$ (green) and $Na^{+}$ (amber) ions in the region of the G1•U37 normal wobble are shown in the center. This portion of the figure is the same as panel A except that $Mg^{2+}$ and $Na^{+}$ are depicted separately in green and amber, respectively, and the view is rotated ~90° about the vertical axis to better view the major groove of P1/P1.1. These plots are superimposed with crystallographically resolved metal ions from multiple HDV ribozyme structures, showing location of divalent and monovalent ions, which are colored similarly (see key for all ion identities). The arrow depicts motion of a $Mg^{2+}$ ion observed during the early stages of the $C75^{+}/Mg_{rev}$ trajectory. The $Mg^{2+}$ ion at the start of the arrow is present at G35 in crystal structure 3NKB, and the $Mg^{2+}$ ion at the end of the arrow is present at P1.1 in crystal structure 1CX0. The monovalent ions are $Tl^{+}$ ions present in crystal structure 2OJ3.

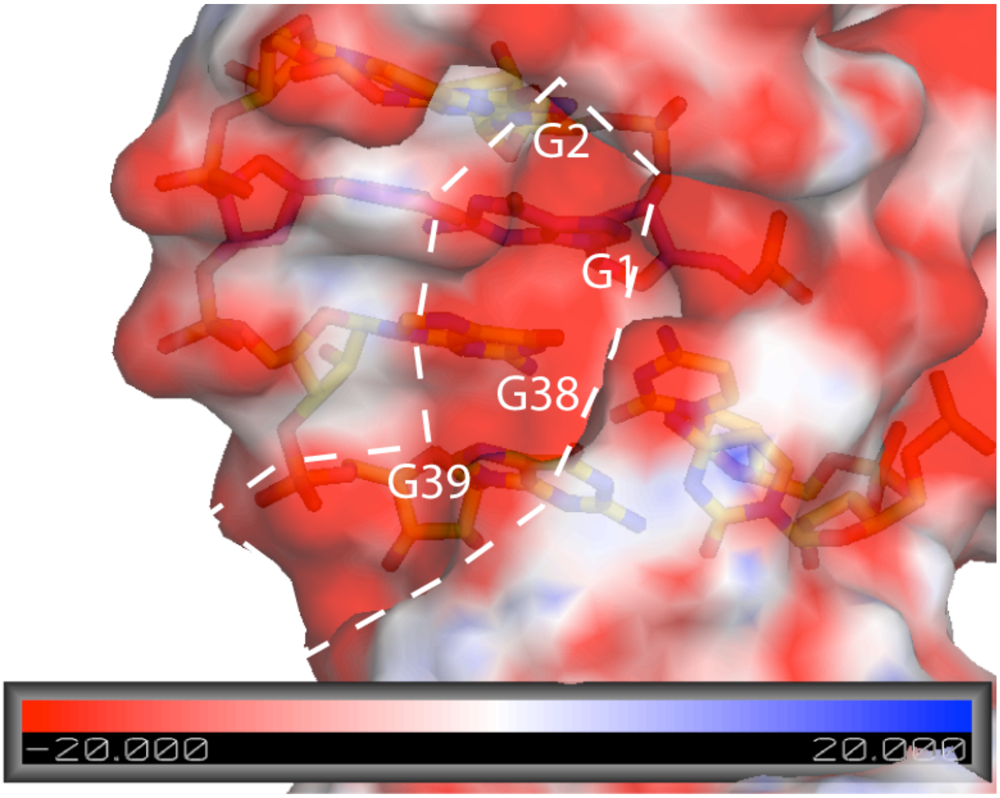

**Figure 3.8.** Nonlinear Poisson Boltzmann (NLPB) electrostatic potential for a snapshot obtained after 25 ns of simulation for a C75$^{+}$/Mg$_{rev}$ trajectory. Potential is colored according to the gradient scale at bottom of the panel (–20 kT/e to +20 kT/e). Surface is shown with 20% transparency. View is of the major groove of the G1•U37 wobble, the G2C36 base pair above, and the two P1.1 GC base pairs below involving G38 and G39. A diffuse negative patch involving these four base pairs is observed, outlined with a dashed line.

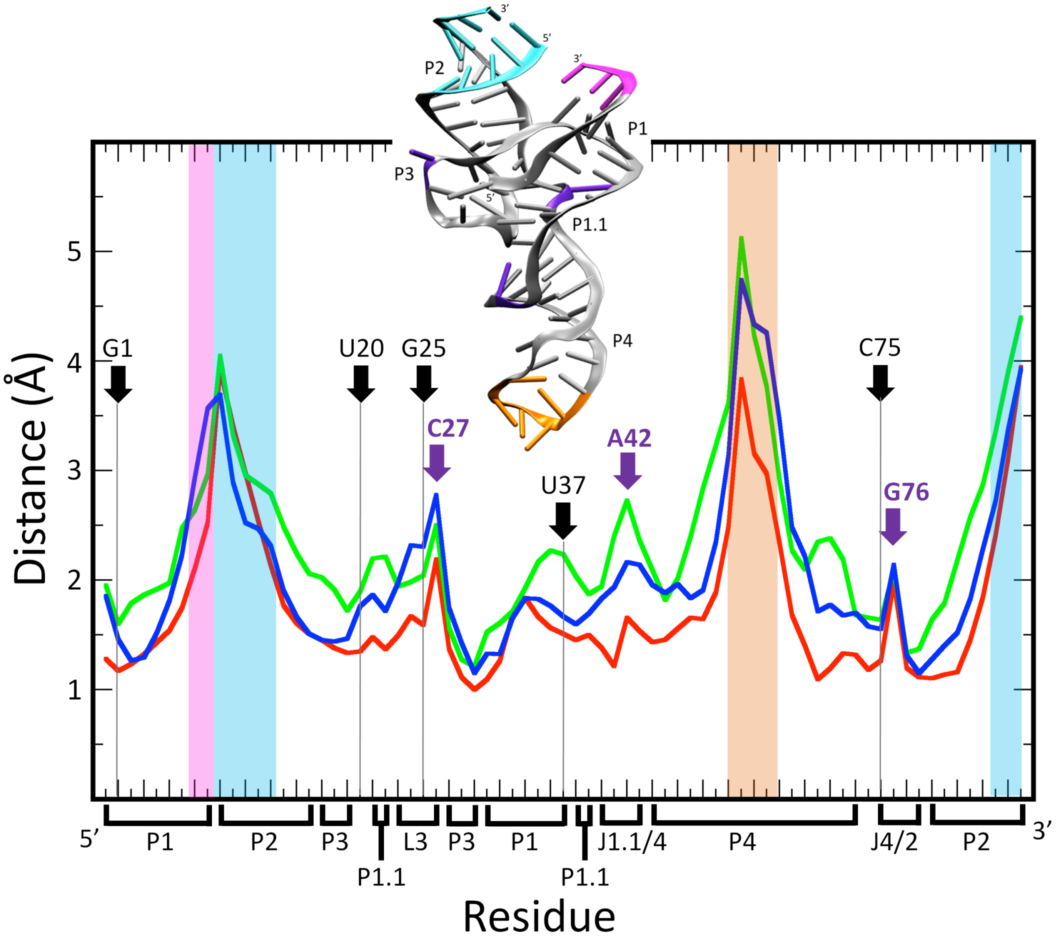

**Figure 3.9.** Root Mean Square Fluctuation (RMSF) of heavy atoms for each residue of the HDV ribozyme. Results are provided for C75$^{+}$/Mg$_{\text{rev}}$ (red solid line), C75$^{+}$/noMg$_{\text{rev}}$ (blue solid line), and C75°/Mg$_{\text{rev}}$ (green solid line) trajectories. Regions of the ribozyme, as shown in Figure 3.1, are identified on the x-axis. Residues exposed to the solvent are indicated by violet arrows. Key residues of the active site are indicated by black arrows. Regions of the ribozyme displaying a particularly large RMSF are shown by pink, cyan and orange regions. Corresponding regions are also marked in the three-dimensional cartoon representation.

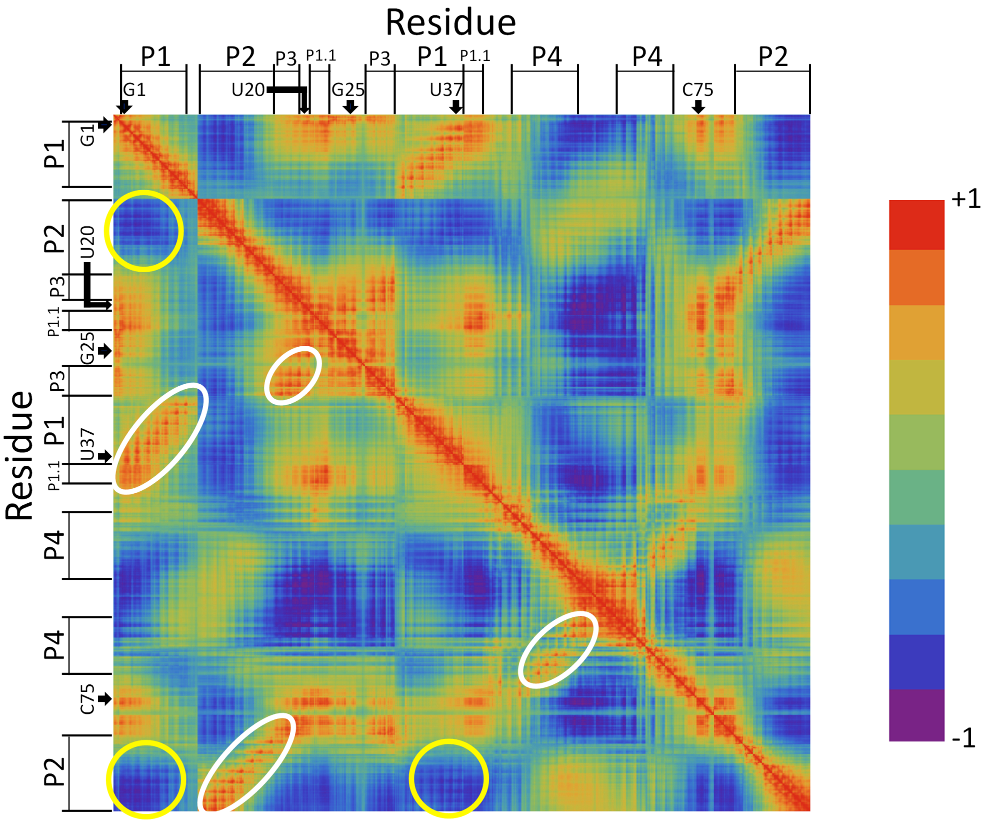

**Figure 3.10.** Cross-correlation matrix of the heavy atom motions for the C75$^+$/Mg$_{rev}$ trajectory. Regions of the ribozyme, as shown in Figure 3.1, are identified on both axes. Certain key residues are highlighted on the axes with black arrows. Red regions indicate correlated motions, and blue regions indicate anticorrelated regions. Helical regions, whose 5' and 3' strands exhibit correlated motions, are identified by white ovals. Certain anticorrelated regions are highlighted by yellow circles.

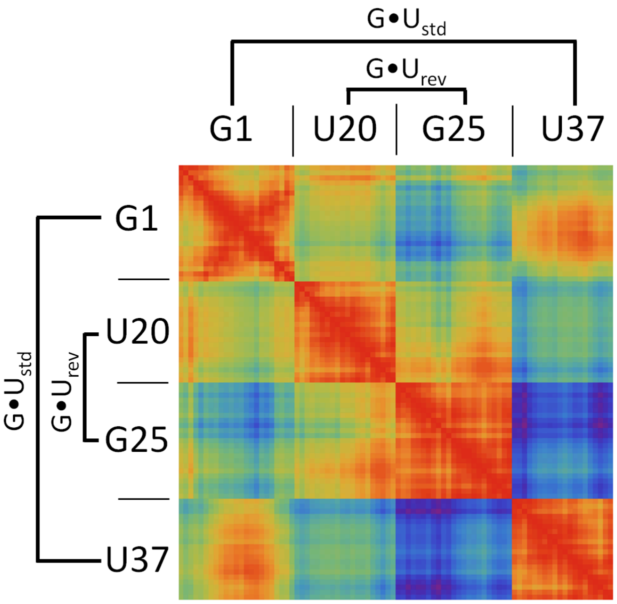

**Figure 3.11.** Cross-correlation matrix of the heavy atom motions in the reverse (G25•U20) and standard (G1•U37) wobbles for the C75$^+$/Mg$_{rev}$ trajectory. As in Figure 3.10, red regions indicate correlated motions, and blue regions indicate anticorrelated regions.

# CHAPTER 4: IDENTIFICATION OF THE CATALYTIC $Mg^{2+}$ ION IN THE HEPATITIS DELTA VIRUS RIBOZYME[1]

## Abstract

The hepatitis delta virus ribozyme catalyzes an RNA cleavage reaction using a catalytic nucleobase and a divalent metal ion. The catalytic base, C75, serves as a general acid and has a p$K_a$ shifted towards neutrality. Less is known about the role of metal ions in the mechanism. A recent crystal structure of the pre-cleavage ribozyme identified a $Mg^{2+}$ ion that interacts through its partial hydration sphere with the G25•U20 reverse wobble. In addition, this $Mg^{2+}$ ion is in position to directly coordinate the nucleophile, the 2'-hydroxyl of U(-1), suggesting it can serve as a Lewis acid to facilitate deprotonation of the 2'-hydroxyl. To test the role of the active site $Mg^{2+}$ ion, we replaced the G25•U20 reverse wobble with an isosteric A25•C20 reverse wobble. This change was found to significantly reduce the negative potential at the active site, as supported by electrostatics calculations, suggesting that active site $Mg^{2+}$ binding could be adversely affected by the mutation. Kinetic analysis and molecular dynamics of the A25•C20 double mutant suggest that this variant stably folds into an active structure. However, pH-rate profiles of the double mutant are inverted relative to the profiles for wild-type ribozyme, suggesting that the A25•C20 double mutant has lost the active site metal ion. Overall, these studies support a model wherein the partially hydrated $Mg^{2+}$ positioned at the G25•U20 reverse wobble is catalytic and could serve as a Lewis acid, a Brønsted base, or both to facilitate deprotonation of the nucleophile.

---

[1] This chapter was published in its entirety in the journal Biochemistry in 2013. The complete reference to the journal article is as follows: Chen, J.; Ganguly, A.; Miswan, Z.; Hammes-Schiffer, S.; Bevilacqua, P. C.; Golden, B. L. *Biochemistry*, **2013**, *52*, 557-567. C. J. and M. Z. performed the experimental studies and A. G. performed the computational studies.

## INTRODUCTION

The hepatitis delta virus (HDV)[1] ribozymes were originally identified in the genomic and antigenomic RNAs produced by the human hepatitis delta virus and are integral to the viral life cycle. These ribozymes have a conserved secondary structure comprised of five short basepaired double helices arranged in a double-pseudoknoted topology (Figure 4.1).[1,2] Key nucleotides in the active site are conserved (Figure 4.1). For many years, these ribozymes were believed to be rare in nature and found only within HDV RNAs. Recently, however, HDV-like ribozymes have been discovered in a wide variety of organisms, where they may play essential roles in regulation of gene expression.[3-7]

Similar to other small nucleolytic ribozymes, the HDV ribozyme activates the 2'-hydroxyl of the nucleotide upstream of the scissile phosphate for nucleophilic attack at the scissile phosphate. This reaction results in the formation of products that contain 2', 3'-cyclic phosphate and 5'-hydroxyl termini. Many of the small nucleolytic ribozymes use nucleobases in their catalytic mechanisms.[8] These nucleobases often appear to serve as general acids, donating a proton to the 5'-hydroxyl leaving group. Both the HDV ribozyme and the hairpin ribozyme shift the $pK_a$ of the general acid (C75 for the HDV ribozyme and A38 for the hairpin ribozyme) toward neutrality, thereby increasing reactivity. [9,10] In addition, hammerhead, hairpin and VS ribozymes have guanosine bases in good position to serve as a general base,[11-13] although other roles such as positioning and electrostatic stabilization remain possible and may be mutually compatible with general base catalysis.[14,15]

Metal ion catalysis is common in large ribozymes, including group I and group II introns, RNase P and, likely, the spliceosome.[16-20] The smaller, nucleolytic ribozymes can function in the absence of divalent cations, a characteristic that initially suggested metal ions may not play catalytic roles in their mechanism.[21] However, there is significant evidence that a $Mg^{2+}$ ion participates in the cleavage reaction of the HDV ribozyme in biologically relevant buffer conditions. For example, the rate-pH profile is inverted when $Mg^{2+}$ is removed from the reaction, suggesting a change in reaction mechanism in the absence of metal ions.[22-24] The catalytic metal ion has been estimated to contribute at least 25-fold to the rate of the HDV ribozyme cleavage step,[23] and similar values are estimated for the hammerhead ribozyme reaction.[25,26]

Although the HDV ribozyme can function in the absence of divalent metal ions, under physiological buffer conditions catalysis depends on the presence of divalent metal cations such as $Mg^{2+}$.[27] These divalent cations both facilitate folding of the RNA and appear to participate in catalysis.[23,28] A recent crystal structure resolved a $Mg^{2+}$ ion within the active site of the HDV ribozyme.[29,30] When the cleavage site dinucleotide was modeled into this crystal structure, the $Mg^{2+}$ ion was observed to directly bind to the 2'-O of U(-1), which serves as the nucleophile, and to the *pro*-$R_P$ oxygen of the scissile phosphate (Figure 4.1B). This suggested that the active site $Mg^{2+}$ ion could act as a Lewis acid to activate the 2'-OH nucleophile and stabilize negative charge on the non-bridging oxygen of the scissile phosphate. Testing the functional relevance of this $Mg^{2+}$ ion is especially important because the conformation of the HDV ribozyme cleavage site was not unambiguously determined by the electron density.[29]

Although the solution biochemical experiments have suggested that a $Mg^{2+}$ ion may play a direct role in the HDV ribozyme cleavage reaction, no data functionally linking the active site $Mg^{2+}$ ion observed in the crystal structure to a role in the cleavage reaction are yet available. A phosphorothioate substitution at the *pro*-$R_P$ oxygen of G1, a ligand to the active site $Mg^{2+}$, is disruptive to catalysis.[31] However, this modification was not rescued by a thiophilic metal ion, and thus a functional interaction between the scissile phosphate and the active site $Mg^{2+}$ ion has not yet been demonstrated.

The active site $Mg^{2+}$ ion has additional interactions with the ribozyme. It forms a single inner-sphere ligand to the ribozyme core, the *pro*-$S_P$ oxygen of U23. In addition, a rare but conserved reverse wobble pair, G25•U20, helps position this $Mg^{2+}$ ion: atoms from this base pair serve as second shell ligands that hydrogen bond to the hydration shell of the $Mg^{2+}$ ion (Figure 4.1B). To address the question of whether the active site $Mg^{2+}$ ion plays a catalytic role in HDV ribozyme, we have characterized the effects of perturbing these outer sphere ligands. The G25•U20 reverse wobble base pair was mutated to an isosteric A25•C20 base pair (Figure 4.1C). We find that this mutation is sufficient to disrupt binding of the active site $Mg^{2+}$ ion and that the cleavage reaction of the A25•C20 double mutant uses a $Mg^{2+}$-free mechanism. These findings suggest that the active site $Mg^{2+}$ ion observed in the crystal structure plays a direct role in the cleavage reaction of the WT HDV ribozyme.

## MATERIALS AND METHODS

*RNA Preparation*. The 74 nt *trans*-acting HDV ribozyme was designed based on a fast-folding version whose crystal structure has been determined trapped in the pre-cleavage state.[29] A 63 nt RNA (termed the 'ribozyme strand') was made by large-scale *in vitro* transcription and purified by urea denaturing gel electrophoresis as previously described.[29] A 5'-Dy547-labeled 11 nucleotide fluorescent RNA (termed the 'substrate strand') was prepared by chemical synthesis. The sequence of this strand is 5'-(DY547)-UAU*GGCUUGCA, where '*' indicates the cleavage site. The substrate strand was purchased from Thermo Scientific, deprotected, and desalted by following the manufacturer's protocols.

*Ribozyme Kinetics and Data Fitting*. Rate constants for ribozyme reactions were obtained using single-turnover cleavage assays. In these, the ribozyme and DY547-tagged substrate RNAs were mixed to achieve final concentrations of 5 μM ribozyme and 50 nM substrate in 50 mM of the indicated buffer in a total volume of 100 μL. Buffers used were potassium acetate (pH 5.0 and 5.5), potassium MES (pH 6.0 and 6.5), potassium MOPS (pH 7.0), or Tris-HCl (pH7.5). The reaction mixture was heated at 90°C for 2 min, cooled to room temperature for 10 min, and equilibrated at 37°C for 2 min before removing a 5 μL aliquot to serve as the zero time point. The ribozyme reaction was initiated by adding sufficient concentrated $MgCl_2$ solutions, buffered with 50 mM of one of the buffers described above, to bring the reaction to the desired final $MgCl_2$ concentration. At appropriate time points, a 5 μL sample was removed and quenched with a 5 μL~25 μL volume of quenching buffer containing 5-200 mM EDTA and 57%-90% formamide. Volumes were chosen such that final concentration of EDTA was 2 to 10 times in excess over the free $Mg^{2+}$ concentration, ensuring that the reaction was terminated by the addition of the quenching buffer. Samples were immediately placed on ice and stored at -20ºC before they were analyzed on a 10% acrylamide, 7 M urea denaturing gel. Identical results were obtained when samples were placed on dry ice and stored at -80ºC. The fluorescent signal from the DY547 fluorophore was detected by Typhoon 8600 variable mode imager (Molecular Dynamics) and quantified by ImageQuant 5.1 (Molecular Dynamics). Plots of fraction cleaved as a function of time were analyzed using KaleidaGraph 4.1 (Synergy Software).

The DY547 fluorophore is bulky and carries a positive charge. As a result, the labeled 3 nucleotide product migrates more slowly in the polyacrylamide gel than the full length substrate. To verify that the product was the expected length and contained a 2',3'-cyclic phosphate, the product was treated with T4 polynucleotide kinase and calf intestinal phosphatase and found to co-migrate an authentic DY547-UAU RNA standard (Figure C.1).

Reactions were fit either to a single (equation 4.1) or double (equation 4.2) exponential reaction as appropriate from the observed data. When reactions were fit to a double exponential equation, the rate constant of the predominant, fast-reacting phase was used to calculate kinetic and thermodynamic parameters described below.

$$F_t = F_0 + (F_\infty - F_0)\left(1 - e^{-k_{obs}t}\right) \qquad (4.1)$$

$$F_t = F_0 + (F_\infty - F_0)\left[1 - f_1 e^{(-k_{obs,1}t)} - \left(1 - f_1\right) e^{(-k_{obs,2}t)}\right] \qquad (4.2)$$

where $F_t$ is the fraction of ribozyme cleaved at time $t$, $F_0$ is the fraction cleaved at time zero, $F_\infty$ is the fraction cleaved at infinite time, $k_{\text{obs}}$ is the observed first-order rate constant; for a double exponential process (eq 2), $f_1(F_\infty - F_0)$ is the fraction cleaved in an initial fast phase with a rate constant $k_{obs,1}$, and $\left(1 - f_1\right)(F_\infty - F_0)$ is the fraction cleaved in a second slow phase with a rate constant $k_{obs,2}$ . For $Mg^{2+}$-titration studies, values as a function of $Mg^{2+}$ concentrations were fit to:

$$k_{obs} = k_{\max} \frac{\left(\left[Mg^{2+}\right] / K_{D,Mg^{2+}}\right)^{n_{Hill}}}{1+\left(\left[Mg^{2+}\right] / K_{D,Mg^{2+}}\right)^{n_{Hill}}} \tag{4.3}$$

where $K_{D,Mg^{2+}}$ is the apparent dissociation constant, $k_{\max}$ is the maximal observed rate constant, and $n_{Hill}$ is the Hill coefficient for $Mg^{2+}$ binding. For the rate-pH profile of WT HDV ribozyme, the observed $pK_a$ value was given by

$$k_{obs} = \frac{k_{\max}}{1+10^{pK_a-pH}} \tag{4.4}$$

For the AC variant, the observed $pK_{a1}$ and $pK_{a2}$ values were obtained by

$$k_{obs} = \frac{k_{\max}}{1+10^{pK_{a_1}-pH}+10^{pH-pK_{a_2}}+10^{pK_{a_1}-pK_{a_2}}} \tag{4.5}$$

*Molecular Dynamics*. We computed Molecular Dynamics (MD) trajectories starting with coordinates based on the structure of the HDV ribozyme trapped prior to cleavage (PDB ID 3NKB), as previously described.[29] The 2'-H of the deoxynucleotides at positions 1 and 2 were replaced by 2'-OH to convert to the corresponding ribonucleotides. The G25A and U20C mutations were performed *in silico* using Accelrys Discover Studio Visualizer 2.0, followed by optimization of only the A25•C20 reverse wobble base pair, with the remainder of the ribozyme fixed. For MD studies, C75 was protonated at the N3 position to represent the active state for general acid catalysis, and C41 was protonated at N3 to maintain the structural integrity of a C-protonated base triple.[32-35] The partial charges used for the protonated cytosines were derived using the RESP method,[36] as described previously.[37] The ribozyme was solvated in an orthorhombic box of rigid TIP3P waters[38] with periodic boundary conditions. The system was neutralized with $Na^+$ ions, and 0.15 M NaCl was added to the solvent to give a physiologically meaningful ionic strength. The solvation procedure was carried out using the Maestro program (Schrödinger, New York, NY). Details regarding the equilibration and simulation protocol used for the system have been described elsewhere.[37,39] Molecular dynamics simulations were performed with the Desmond MD program (D.E. Shaw Research, New York, NY) using the AMBER99 force field.[40] At least two independent 25 ns trajectories were propagated at 298 K in the canonical ensemble (i.e., at constant NVT) for each case studied.

### *Electrostatics calculations*

Electrostatic potential calculations were performed using the Adaptive Poisson-Boltzmann Solver (APBS),[41] which uses the Finite Element ToolKit[41-43] to solve the non-linear Poisson-Boltzmann (NLPB) equations numerically. Structural coordinates were obtained from the starting structure of the ribozyme used for the MD simulations. All water molecules and crystallographic metal ions, except the metal ion at the catalytic site when specified, were omitted from the NLPB calculations, and C75 was left unprotonated; these settings allow the extent of the negative potential to be sensed and are customary for such calculations.[23,44-46] In addition, C41, which is distal from the active site, was protonated as in the MD studies in order to maintain the aforementioned base triple. [35] Charges on C75 and C41 are identical to those used in our earlier electrostatics calculations.[37] The atomic radii and partial charges were defined using the AMBER99 parameter set, except for protonated C41, for which the partial charges were derived using RESP calculations as described previously.[35]

The following parameters of the electrostatics calculations were chosen to be consistent with similar previous calculations.[45,47,48] The interior dielectric constant of the ribozyme was set to 2, a value previously shown to be physically meaningful,[49] and the solvent dielectric constant was set to 80. All calculations were performed on a grid centered on the crystal coordinates of the catalytic metal ion with a grid spacing of 0.35 Å. The dimensions of the ribozyme, centered on the position of the catalytic metal ion, were calculated in the X, Y, and Z directions, and the final grid dimensions were chosen after leaving 20 Å distance between the ribozyme boundary and the grid boundary. A 2.0 Å exclusion radius was added to the ribozyme surface to incorporate hydrated sodium ions. The calculations were performed in the absence of any salt in the bulk solvent, as well as in the presence of different concentrations of mixed 1:1 and 2:1 salts, *i.e.* mixture of NaCl and $MgCl_2$, as described elsewhere.[45] The radius of the catalytic $Mg^{2+}$ or $Na^+$ ion was set to 1.45 Å or 1.68 Å, respectively, values that were shown previously to reproduce their corresponding experimental hydration free energies.[50,51] The trends in the relative electrostatic binding free energies were maintained over a range of physically reasonable radii.

To estimate the binding energy of the metal ion at the reverse wobble, we adopted the methodology of Misra and Draper.[51] We considered the case of the WT and the G25A•U20C double mutant structures, with a single metal ion, either $Mg^{2+}$ or $Na^+$, located at the crystallographic position of the $Mg^{2+}$ at the reverse wobble. The metal ion binding free energy is calculated from the difference between the total electrostatic free energy of the metal ion-bound ribozyme in solution and the sum of the total electrostatic free energies of the isolated ribozyme and the isolated metal ion in solution.

$$\Delta G^{\text{el}}_{\text{bind}} = \Delta G^{\text{el}}\left(\text{HDV+ion}\right) - \left[\Delta G^{\text{el}}\left(\text{HDV}\right) + \Delta G^{\text{el}}\left(\text{ion}\right)\right] \qquad (4.6)$$

The difference between $\Delta G_{\text{bind}}^{\text{el}}$ of the WT and G25A•U20C double mutant structures, as indicated in equation 4.7, provides an indication of the relative favorability of metal ion binding at the reverse G•U wobble in comparison to the reverse A•C wobble:

$$\Delta\Delta G_{\text{bind}}^{\text{el}} = \Delta G_{\text{bind}}^{\text{el}}\left(\text{WT}\right) - \Delta G_{\text{bind}}^{\text{el}}\left(\text{DM}\right) \tag{4.7}$$

Note that the site-specific binding free energy of a metal ion binding to a particular construct includes additional contributions to those given in equation 4.7, as discussed by Draper, but these contributions, as well as $\Delta G^{el}\left(ion\right)$, will cancel for the calculation of $\Delta\Delta G_{\text{bind}}^{\text{el}}$ in equation 4.7.

We emphasize that these electrostatics calculations provide only qualitative information and are based on several approximations. For example, no conformational sampling of the ribozyme is included, no explicit water molecules are included (i.e., the ligation of water molecules to the $Mg^{2+}$ is neglected), and for calculations involving $Na^+$, the $Na^+$ ion in the WT and G25A•U20C double mutant is assumed to be located at the same position as the $Mg^{2+}$. Moreover, we found that the quantitative results depend on the dielectric constant and the ionic radii, although the trends in $\Delta\Delta G_{\text{bind}}^{\text{el}}$ between WT and the G25A•U20C double mutant are reproducible within the physically reasonable regimes for these parameters. The electrostatic potentials were rendered using PYMOL. [52] The dxmath tool within APBS was used to obtain the difference between two electrostatic potential maps.

## RESULTS

Here, we examine the impact of substituting the G25•U20 reverse wobble by mutating G25 to an adenine and U20 to a cytosine. This double mutant has the potential to allow formation of a A25•C20 reverse wobble base pair that is isosteric with a G•U reverse wobble (Figure 4.1B,C). Previous studies have tested the impact of single mutations of this G25□U20 base pair on ribozyme activity. At pH 8.0 and saturating Mg$^{2+}$ concentration, ~3000- and ~1600-fold reaction rate reductions were measured for the G25A single mutation of a genomic HDV ribozyme and the U23C single mutation of an antigenomic version (equivalent of U20 in the genomic ribozyme), respectively.[53,54]. These data suggest that the G25•U20 pair is essential for the ribozyme activity and that a reverse wobble conformation, rather than a Watson-Crick base pair, is required at this position. A double mutation would allow an isosteric reverse wobble to be adopted by an A25•C20 base pair.[55] While this base pair is similar in shape to a G•U reverse wobble pair, the presence of the N6 amino of adenosine in the minor groove is expected to alter the electrostatic properties of the binding site for the active site Mg$^{2+}$. The G25A•U20C double mutant of the HDV ribozyme thus has the potential to specifically disrupt binding of the Mg$^{2+}$ ion without perturbing the overall architecture of the ribozyme. We therefore characterized the G25A•U20C double mutant using solution biochemical and computational methods, in an effort to assess the role of the 25•20 base pair in binding the active site Mg$^{2+}$ ion and to characterize the role of this ion in the chemical reaction.

### *The G25A•U20C Double Mutant of the HDV Ribozyme Retains an Active Structure*

While an A•C base pair in the reverse wobble geometry is isosteric with a G•U reverse wobble base pair, the possibility exists that in the context of the HDV ribozyme an A25•C20 reverse wobble may not form stably. This would lead to structural changes at the active site. We therefore computed MD trajectories of the G25A•U20C double mutant. In these trajectories, $Mg^{2+}$ ions present in the 3NKB crystal structure coordinates, including the ion at the reverse wobble, were retained. In our simulations, we found that the A25•C20 reverse wobble is stable over two independent 25 ns trajectories, with average A25(N1)-C20(N4) and A25(N6)-C20(N3) distances of 3.05 ± 0.19 Å and 2.90 ± 0.09 Å, respectively, consistent with hydrogen bonds (Table C.1). There were no noticeable conformational changes observed in the active site, including the general acid, C75, and the cleavage site dinucleotides G1 and U-1 (data not shown).

In both of these trajectories, the active site $Mg^{2+}$ ion remained bound to the A25•C20 reverse wobble, and the interaction of the metal ion with the A•C motif was found to be similar to its interaction with the reverse G•U wobble. As these simulations were much shorter than the timescale of the reaction, they may not address whether $Mg^{2+}$ can stably bind to the G25A•U20C double mutant variant ribozyme. Limitations in both the molecular mechanical force field and the conformational sampling may prevent observation of any diffusion of the $Mg^{2+}$ from this site. As the G25A•U20C double mutant has the potential to disrupt binding of the active site $Mg^{2+}$, we chose to next explore the stability of the A25•C20 reverse wobble in the *absence* of this $Mg^{2+}$ ion.

We ran three independent trajectories of the G25A•U20C double mutant ribozyme in the absence of $Mg^{2+}$ at the active site (Figure 4.2). In two out of the three trajectories, we found that a $Na^+$ ion moves from bulk solution to the active site and that the A25•C20 reverse wobble as well as the active site are stable (Table C.1). In the third trajectory, no $Na^+$ ion moved into the region to bind to the reverse wobble, and one hydrogen bond in the A25•C20 reverse wobble was lost at ~2.5 ns. This observation probably resulted from starting the trajectory without any ion in this region, leading to breaking of the hydrogen bond before the

$Na^+$ was able to move into the region to stabilize the reverse wobble. This situation is not likely relevant to the ribozyme in solution. Thus, we did not further analyze this trajectory. As shown in Table C.1, for the first two trajectories, the two hydrogen bonds in the A25•C20 reverse wobble have similar distances as they did in the presence of bound $Mg^{2+}$.

Overall, these results suggest that the structure of the G25A•U20C double mutant of the HDV ribozyme is stable in the presence or absence of a $Mg^{2+}$ ion in the active site. However, monovalent ions may be required to stabilize the geometry of the A25•C20 wobble in the absence of $Mg^{2+}$ ions.

### *The A25•C20 Variant of the HDV Ribozyme is Catalytically Active.*

Our computational studies suggested that the G25A•U20C double mutant of the HDV ribozyme would fold correctly. Therefore, we introduced a G25A•U20C double mutation into the HDV ribozyme. The sequence of the WT ribozyme and the G25A•U20C double mutant were derived from the ribozyme whose structure we had recently solved.[29] This is a two-piece RNA containing a large RNA strand that spans most of the active site and is formally a ribozyme. A small substrate RNA was annealed to this ribozyme to reconstitute the three-dimensional structure prior to initiating the reaction with $Mg^{2+}$. For this study, a fluorophore, DY547, was linked to the 5’-end of the substrate. Upon incubation with ribozyme, this substrate is cleaved to generate the expected products (Figure C.1). We would predict that this modification would not affect catalysis because the crystal structure of this RNA reveals a sharp turn at the scissile phosphate that places the nucleotides upstream of U(-1) out of the active site and into bulk solution.[29] In support of this, we find that the single-turnover kinetic parameters $k_{\max}$, $K_{D,Mg^{2+}}$ and $n_{Hill}$ obtained in this study are similar to those obtained with a radiolabelled substrate lacking the fluorophore (Table 4.1).[56]

Next, we conducted single-turnover studies on the AC variant. These studies demonstrated that, at pH 7 and lower, the observed rate constant $k_{obs}$ of the AC variant is at most ~100-fold lower than WT (Figures 4.3 and 4.4). Thus, the G25A•U20C double mutation partially rescued the activity. In these reactions, the G25A•U20C variant cleaved its substrate monophasically to at least 80% of completion (Figure 4.3A, triangles). This extent of reaction is similar to what is observed with the WT ribozyme (Figure 4.3A, squares) and suggests that the majority of the AC variant is in an active conformation. Recently, Perrault and co-workers demonstrated a similar finding on the antigenomic ribozyme. [57]

## *The G25A•U20C Mutation does not Affect the Apparent $Mg^{2+}$ Ion-Binding Properties*

To further examine the folding properties of the G25A•U20C double mutant, $Mg^{2+}$ titrations were conducted on both the WT and G25A•U20C double mutant ribozyme reactions. At every $Mg^{2+}$ concentration tested, the reaction rate of the G25A•U20C double mutant decreased by 80-120 fold relative to WT (Figure 4.3). As a result, the apparent $K_{D,Mg^{2+}}$ value is the same, within error, for both the G25A•U20C double mutant and the WT ribozymes. We observed apparent $K_{D,Mg^{2+}}$ values of 1.4 ± 0.1 mM and 1.3 ± 0.1 mM for the G25A•U20C double mutant and the WT ribozymes, respectively (Table 4.1). In addition, the Hill coefficient, $n_{Hill}$, does not change as a result of the mutation (1.8 ± 0.2 for the WT versus 2.0 ± 0.2 for the mutant) (Table 4.1). This similarity suggests that the same concentration of $Mg^{2+}$ ion is needed to attain a folded conformation and that the decreased activity of the G25A•U20C double mutant is most likely not caused by misfolding.

### *The G25A•U20C Double Mutant in the Presence of $Mg^{2+}$ has the Same Rate-pH Profile as the WT Ribozyme in the Absence of $Mg^{2+}$*

To assess the impact of the G25A•U20C double mutant on active site folding and catalysis, we characterized the rate-pH profile of this variant. We first determined the rate-pH profile for this WT ribozyme in 5 mM $Mg^{2+}$, and we obtained the expected profile with a log-linear increase in rate between pH 5.0 and 6.0 and a $pK_a$ of 6.1 ± 0.1 (Figure 4.4, squares). This is in good agreement with the $pK_a$ previously determined using a radiolabelled oligonucleotide of 6.4 under similar conditions.[56]

The rate-pH profile of the G25A•U20C double mutant was studied in 50 mM $Mg^{2+}$, well above the apparent $K_{D,Mg^{2+}}$ for this ribozyme (Figure 4.3B). We found that the reaction is nearly independent of pH in the range of pH 5.5-6.0 (Figure 4.4, triangles), likely because the general acid, C75, remains largely protonated in this pH range.

If the active site were misfolded and a structural rearrangement were the rate-limiting step for the G25A•U20C double mutant, then the change in $k_{obs}$ as a function of pH might be minimal. However, above pH 6.0, the reaction rate decreases log-linearly, suggesting that chemistry, not folding, is rate-limiting and that the cleavage reaction rate is dominated by deprotonation of C75 (Figure 4.4). The apparent $pK_a$ value is 6.2 ± 0.2 under these conditions, and as discussed below, likely represents the $pK_a$ of C75. We observed that the cleavage rate drops by ~2-fold when the pH is decreased below 5.0. Loss of activity at low pH may be due to acid denaturation and might suggest that the G25A•U20C double mutant is less stable than the WT ribozyme as both A and C start to protonate in this pH range. [58] Such a protonation at A25, might allow the formation of a positively-charged standard A•C wobble between A25 and C20. This conformation is not isosteric with a reverse wobble and would be predicted to be disruptive to the ribozyme's active site structure. Another possibility is a slight contribution from a general base with a p$K_a$ near 5.4 or one that is lost near this pH (see Discussion).

The rate-pH profile of the G25A•U20C double mutant is inverted from that of the WT ribozyme in $Mg^{2+}$-containing buffers (Figure 4.4, compare triangles and squares, respectively). Qualitatively, it resembles the rate-pH profile of the WT ribozyme reacting in the absence of $Mg^{2+}$.[22,23] The data on the WT ribozyme were interpreted to suggest that a $Mg^{2+}$-dependent proton transfer is involved in the cleavage reaction under biologically-relevant, $Mg^{2+}$-containing conditions. In an effort to restore a WT-like pH rate profile, we repeated this analysis in 500 mM $Mg^{2+}$. However, these data are not significantly different than the data obtained in 50 mM $Mg^{2+}$(not shown), suggesting that the active site $Mg^{2+}$ binds especially weakly or in an alternate binding pocket within the G25A•U20C double mutant.

### *The G25A•U20C Double Mutant in the Absence of $Mg^{2+}$ has the Same Rate-pH Profile as the WT Ribozyme in the Absence of $Mg^{2+}$*

As the G25A•U20C double mutant rate-pH profile appeared to resemble the rate-pH profile of the WT HDV ribozyme in the absence of $Mg^{2+}$, we wanted to compare the reaction of both ribozymes in the absence of $Mg^{2+}$. Unfortunately, the two-piece HDV ribozyme used here does not react in the absence of $Mg^{2+}$, possibly because substrate does not stably anneal to the ribozyme in the high concentration of NaCl used in these reactions. We therefore introduced the G25A•U20C double mutant into a *cis*-acting, self-cleaving version of the HDV ribozyme characterized previously. This allowed a side-by-side comparison of the WT ribozyme and G25A•U20C double mutant in the absence of $Mg^{2+}$ ion (supplementary material).

We first studied the pH dependence of the G25A•U20C double mutant reaction in the absence of divalent metal ions (See Supplementary Materials and Methods). These reactions were performed at 1M NaCl and 100 mM EDTA (final concentration) was included in the reaction buffer to chelate any trace amount of contaminating polyvalent metal ions, especially at low pH (see Materials and Methods). We observed the G25A•U20C double mutant was catalytically active under all the pH conditions tested. At least 80% of completion was achieved in all but the reaction at pH8.0 (Figure C.2). The reaction progress can be fit to equation (4.1). This suggests that the $Mg^{2+}$ ion is nonessential to the G25A•U20C double mutant catalysis when high concentration of monovalent salt is present. The reaction is pH-independent between pH 6.0 and pH 6.5 and an inhibitory effect is seen at pH below 6.0 (Figure C.3). The rate-pH curve yields an apparent $pK_a$ of 7.3 ± 0.1, which is assigned to the general acid C75.

We next compared the rate-pH profile of the WT HDV ribozyme in the absence of divalent metal ions to that obtained for the G25A•U20C double mutant. The rate-pH profile is bell-shaped, which suggests the involvement of two ionizable groups (Figure C.3). Two apparent $pK_a$ values were obtained from the rate-pH profile. The log-linear decrease of reaction rate between pH 7.5 and pH 8.5 are attributed to the deprotonation of general acid C75 as discussed elsewhere. Thus, the high $pK_a$ value of 7.6 ± 0.1 likely represents the $pK_a$ of C75. Between pH 6.0 and pH 6.5, the reaction rate is largely independent of pH, which is consistent with that a constant amount of functional form of C75 is present over this pH range. An apparent $pK_a$ value of 5.3 ± 0.2 was also obtained, similar to what is observed with the WT self-cleavage reaction (Figure C.3) and the two piece G25A•U20C double mutant HDV ribozyme (Figure 4.4). In general the rates and shape of the curve agree well with those previously reported. {Nakano, 2001 #218}

Comparing the two rate-pH profiles (Figure C.3), we observed that the G25A•U20C double mutant reaction shows a similar pH-dependence as the WT ribozyme. A log-linear decrease at high pH range is almost parallel to that observed in the WT ribozyme. The reaction rate of G25A•U20C double mutant is only ~5-fold slower than the WT under the same reaction condition, which corresponds to a $\Delta G^\circ_{37}$ of just 1 kcal/mol. This indicates that the G25A•U20C double mutant might is only slightly less active than the WT. The ~5-fold rate difference observed here, in $Mg^{2+}$-containing buffers, is also observed when $Mg^{2+}$-mediated base catalysis is minimal at pH 5.0 suggesting that folding of the G25A•U20C double mutant ribozyme is largely intact both in the presence and absence of $Mg^{2+}$ (Figure 4.4).

### *Electrostatics of the Active Site are Affected by the G25A□U20C Double Mutant*

Because the solution biochemical studies strongly suggested that the G25A•U20C double mutant was not able to bind the active site $Mg^{2+}$, we sought to characterize the electrostatics of the active site. MD simulations can be a useful tool for understanding important motions in biological systems; however, there are known limitations associated with MD for investigating interactions between divalent metal ions and macromolecules. These limitations include inadequate description of divalent ions by molecular mechanical force fields and issues related to trapping in local minima due to insufficient sampling during simulations.[59,60] We therefore performed NLPB electrostatics calculations on the WT and G25A•U20C double mutant structures to obtain qualitative insight into the metal ion binding at the reverse G•U and A•C reverse wobbles, respectively. For both ribozymes, an intense negative surface is found in the vicinity of the reverse wobble (Figure 4.5), similar to that reported previously.[37] To directly compare the results from the WT and the G25A•U20C double mutant ribozymes, we calculated a surface electrostatic difference plot (G•U minus A•C electrostatic surface). This plot clearly reveals a negative patch in the vicinity of the reverse wobble (Figure 4.5C), indicating that the metal ion binding pocket near the reverse A•C wobble is less negatively charged and suggesting weaker binding of the metal ion at the reverse A•C wobble. These observations are consistent with the changes in the rate-pH profile we observed in the G25A•U20C double mutant ribozyme.

We also estimated the relative binding free energy for the metal ion to the catalytic sites of the WT and variant ribozymes using Eqs. (4.6) and (4.7). These calculations were performed for either a $Mg^{2+}$ or a $Na^{+}$ ion binding to the catalytic site. Negative values for $\Delta\Delta G_{bind}^{el}$ were found, indicating that metal ion binding at this site is more thermodynamically favorable for the WT than for the G25A•U20C double mutant ribozyme, as expected from the NLPB electrostatic surface potentials in Figure 4.5. In particular, for a $Mg^{2+}$ ion and a dielectric constant of 2, the binding is predicted to be ~12 kcal/mol more favored at the WT G•U reverse wobble than at the A•C reverse wobble of the double mutant. Likewise, binding $Na^{+}$ ion in the active site of the WT ribozyme is ~5 kcal/mol more favored than in the active site of the double mutant. It is important to note, however, that these values are only qualitative and are not quantitatively accurate (e.g., the magnitude of these differences decreases with the choice of a higher dielectric constant). The lower magnitude of $\Delta\Delta G_{bind}^{el}$ for $Na^{+}$ suggests that the metal binding pocket at the G•U reverse wobble is more selective toward $Mg^{2+}$ ions, consistent with the above experiments. The results were similar for calculations in the presence of varying concentrations of diffuse 1:1 and 2:1 salts (data not shown).

## DISCUSSION

The HDV ribozyme uses a rare reverse G•U wobble pair to help position an active site $Mg^{2+}$ ion. Hydrated $Mg^{2+}$ ions have long been known to interact with the negative dipoles of canonical G•U base pairs.[61,62] However, these interactions are typically in the *major* groove, which in the case of A-form RNA is deep and inaccessible; such $Mg^{2+}$ ions are thus unlikely to be catalytic. In contrast, the active site $Mg^{2+}$ ion of the HDV ribozyme is in the minor groove and is therefore in excellent position to participate in the cleavage reaction. This can occur because the G25•U20 reverse wobble has G25 in the *syn* geometry, allowing the N7 and O6 of the G to reside in the shallow and accessible minor groove. G25 is thus able to help position the active site $Mg^{2+}$ ion in such a way that it could participate in catalysis. In this study, we assessed the catalytic potential of this ion using functional assays.

### *The G25A•U20C Double Mutant does not Disrupt the Structure of the HDV Ribozyme*.

The G25•U20 reverse wobble pair of the HDV ribozyme contributes significantly to the stability of the active site $Mg^{2+}$ ion by providing outer sphere ligands and by enhancing the magnitude of the negatively charge in the metal ion binding pocket.[29,37] An A•C reverse wobble base pair is isosteric with the G•U reverse wobble, but substitutes a carbonyl with an amino group.[55] In this manner, this double mutation has the potential to disrupt active site metal ion binding without perturbing the tertiary structure of the RNA.

When introducing mutations into RNA molecules, it is possible to generate RNA molecules that are misfolded into non-catalytic conformations. There are several pieces of evidence, however, suggesting that the G25A•U20C double mutant is not misfolded. First, under single turnover conditions, the mutant ribozyme reacts with monophasic kinetics and to ~80% completion, similar to the WT ribozyme (Figure 4.3). Such behavior is expected of a single population of ribozymes reacting in a similar manner. Second, characterization of the rate-pH profile for the G25A•U20C double mutant suggests that chemistry, not folding, is rate limiting. If a conformational change were the rate-limiting step, we would expect to see a profile that was largely independent of pH. Instead, we observe a profile in which the rate constant decreases in a log-linear fashion above pH ~6.0 (Figure 4.4). Third, the $Mg^{2+}$-rate profile of the G25A•U20C double mutant is strikingly similar to that of the WT ribozyme, both with respect to the apparent $K_{D,Mg^{2+}}$ and the Hill coefficient, $n_{Hill}$ (Figure 4.3). This suggests that the tertiary structure forms similarly, that the metal ions contributing to the three-dimensional structure bind with similarly affinity, and that the thermodynamics of RNA folding are largely unaffected by this mutation.

The lack of structural disruption in the double mutant is also supported by the calculations. In molecular dynamics simulations of the G25A•U20C double mutant, the A25•C20 reverse wobble remains largely intact, even in the absence of $Mg^{2+}$. In addition, the key active site nucleotides, including the general acid C75 and the cleavage site G1 and U(-1), are largely unperturbed. This is consistent with the biochemical analyses, strongly suggesting that the structure of the HDV ribozyme is not significantly changed in the G25A•U20C double mutant.

Lastly we note that upon lowering the pH to 5, the difference in rate constants between the WT ribozyme and the G25A•U20C double mutant ribozyme is lessened. In fact, when the plateau region of the rate-pH profile of the AC variant is extrapolated to pH 4.5, where the $Mg^{2+}$-dependent deprotonation in the WT ribozyme hardly contributes, similar rate constants are obtained for the WT and G25A•U20C double mutant (Figure 4.4). This observation further supports native folding of the G25A•U20C double mutant.

### *The G25A•U20C Double Mutant Disrupts Binding of the Active Site $Mg^{2+}$ Ion.*

The rate-pH profile of the WT HDV ribozyme depends critically on the presence of $Mg^{2+}$ ions. In the presence of $Mg^{2+}$ ions, the rate constant increases log-linearly with pH until about pH 6 where it plateaus (Figure 4.4, squares). When the $Mg^{2+}$ ions are removed from the reaction, the pH rate profile is inverted (Figure 4.4, triangles):[22,24] it is flat until ~pH 6, where it decreases log-linearly with pH. This inversion is consistent with loss of a proton transfer with a p$K_a$ greater than 10 and supports a mechanism in which C75 serves as the general acid and the base either has a p$K_a$ of 5.4, or proton transfer from the 2'OH occurs in a non rate-determining step (Figures C.4 and C.5). (The general base with a p$K_a$ of 5.4 seems unlikely as this is a low p$K_a$ for a general base, and also because the low pH arm of the double mutant rate-pH profile could be due to acid denaturation or contribution of a catalytic metal ion only at low pH.) As a result, transfer of a proton from the N3 of C75 to the 5'-hydroxyl leaving group represents the only p$K_a$ observable near neutrality in this pH range in the absence of $Mg^{2+}$.

The rate-pH profile of the G25A•U20C double mutant in $Mg^{2+}$ ion looks very much like that for a ribozyme that lacks or mispositions the active site $Mg^{2+}$ ion.[22,24] These reactions were performed at a $Mg^{2+}$ ion concentration of 50 mM, well above the $Mg^{2+}$ ion concentration in which the G25A•U20C double mutant achieves maximal velocity. Similar to the WT ribozyme reaction in the absence of $Mg^{2+}$ ions, the reaction rate of the G25A•U20C double mutant decreases with pH, suggesting that the $Mg^{2+}$-dependent proton transfer observed in the WT ribozyme is missing (Figures C.4 and C.5). These data suggest that in the G25A•U20C double mutant, the interaction between the active site $Mg^{2+}$ ion and the cleavage site is lost. Addition of $Mg^{2+}$ at concentrations as high as 500 mM is not sufficient to restore the pH-rate profile to that seen in the WT ribozyme.

This model is supported by shifts in the $pK_a$ of C75 observed in this study. The $pK_a$ of C75 is dependent on the concentration of $Mg^{2+}$ ion in the reaction buffer. As the concentration of $Mg^{2+}$ is increased, the $pK_a$ of C75 drops.[9,22] This anticooperativity suggests that one or more $Mg^{2+}$ ions are bound in the active site of the WT HDV ribozyme close enough to C75 that the protonated base and the ion can interact electrostatically. This is consistent with the crystal structure of the HDV ribozyme where the N4 of C75 is within 3.5 Å of the hydration shell of the active site $Mg^{2+}$ ion and may therefore serve as an outer sphere ligand.[29] In the WT ribozyme, we observed that the $pK_a$ of C75 is 6.1 ± 0.1 in the presence of 5 mM $Mg^{2+}$ (Figure 4.4). In the G25A•U20C double mutant, however, this $pK_a$ is not shifted lower when the $Mg^{2+}$ concentration is raised to 50 mM. In our studies, $pK_a$ values of 6.2 ± 0.2 are obtained for C75 (Figure 4.4). In contrast, Nakano et al observed the $pK_a$ of C75 within the WT ribozyme shifted from 6.4 to 5.8 when the $Mg^{2+}$ ion concentration was raised from 5 mM to 50 mM. [22] This suggests that in the G25A•U20C double mutant the environment near C75 is less positively charged. Again, this is consistent with a model in which the active site $Mg^{2+}$ ion no longer occupies the binding site observed in the crystal structure.

Electrostatics calculations further suggest that the G25A•U20C double mutant specifically disrupts active site $Mg^{2+}$ ion binding. NLPB calculations reveal a significant loss of negative charge in the active site $Mg^{2+}$ ion binding pocket in the double mutant (Figure 4.5). Furthermore, the DDG between WT and the G25A•U20C double mutant for $Mg^{2+}$ binding is ~12 kcal/mol, suggesting that active site $Mg^{2+}$ binding is more thermodynamically favorable to the WT than to the G25A•U20C double mutant. Note that the quantitative value for DDG depends on certain choices of parameters, such as dielectric constant, but the sign of this quantity is reproducible.

These results all point to a scenario where the active site $Mg^{2+}$ ion in the G25A•U20C double mutant is displaced or altered such that it can no longer interact with the 2'-hydroxyl of U(-1) and C75 (Figures C.4 and C.5). Overall, these data indicate that the active site $Mg^{2+}$ ion observed in the crystal structure of the WT HDV ribozyme is indeed positioned at the reverse G•U wobble and appears to participate directly in catalysis.

### *$Mg^{2+}$ binding isotherm does not reflect binding of an active site $Mg^{2+}$ ion*

The dependence of attainment of maximal ribozyme activity on $Mg^{2+}$ ion concentration is the same in both the WT and G25A•U20C double mutant ribozymes. Although the G25A•U20C double mutant never matches the velocity of the WT ribozyme, the apparent $K_{D,Mg^{2+}+}$ and Hill coefficient are, within error, the same (Table 4.1). As the G25A•U20C double mutant appears to disrupt active site $Mg^{2+}$ binding as discussed above, these data suggest that structural $Mg^{2+}$ ion binding, not active site $Mg^{2+}$ binding, dominate the apparent $K_{D,Mg^{2+}}$ and observed Hill coefficients. The catalytic $Mg^{2+}$ ion must therefore have higher affinity than the structure-stabilizing $Mg^{2+}$ ions. Consistent with this interpretation, high affinity $Mg^{2+}$ ion binding sites in RNAs are predicted to have dissociation constants in the 1-25 mM range (depending on ionic strength).[28,49]

### *What is activating the nucleophile?*

If the G25A•U20C double mutant ribozyme lacks an active site $Mg^{2+}$ ion, are there mechanisms by which this ribozyme could activate the 2'-hydroxyl nucleophile? There are several possibilities. First, there is a $Na^+$ ion-binding pocket in the ribozyme active site that is observed when crystals of the HDV ribozyme are soaked in buffers containing $Na^+$ ion.[37] While this $Na^+$ binding site partially overlaps the active site $Mg^{2+}$ ion, its binding is distinct from that of the active site $Mg^{2+}$ ion. The $Na^+$ ion interacts through its water ligands with nucleotides G27 and A77 of the ribozyme. This ion could facilitate the cleavage reaction through electrostatic effects, stabilizing the developing negative charge on the nucleophile and the scissile phosphate, but it is not in position to interact directly with the 2'-hydroxyl nucleophile or the scissile phosphate.

In both the WT ribozyme and the G25A•U20C double mutant, the acceptor of the proton from the 2'-hydroxyl of U(-1) in the cleavage reaction, is unknown. There are several possibilities. The 2'-hydroxyl of U(-1) is hydrogen bonded to the 2'-hydroxyl of G27, and this hydrogen bond is likely to be retained in the G25A•U20C double mutant as the active site structure appears to remain intact. Alternatively, the proton acceptor could be a water or hydroxide from the surrounding solvent, which may not be detected in the double mutant if it occurs in a step that is not rate-determining.  In the case of the WT ribozyme, one of the water ligands from the catalytic $Mg^{2+}$ ion could serve this role. Thus, the active site $Mg^{2+}$ ion could play two roles in the WT, interacting with the nucleophile as a Lewis acid, coordinating a hydroxide molecule to serve as a Brønsted base, or both (Figure C.4).

This study represents the first set of data to functionally link the active site $Mg^{2+}$ ion observed in the HDV ribozyme active site to a catalytic role in the cleavage reaction. Key questions still remain, however. The identity of the high $pK_a$ that affects the rate-pH profile in the WT ribozyme, but not in the G25A•U20C double mutant is still unknown. It could arise from ionization of a water bound to the active site $Mg^{2+}$ that is deprotonated to serve as a general base (Figure C.5A), from hydroxide in solution, or from the $pK_a$ of the 2'-OH of U(-1) itself (Figure C.5B). In addition, while the present data associate the reverse G•U wobble with the catalytic metal ion, contributions of other ligands, such as the pro-$R_P$ oxygen of the scissile phosphate, to metal ion catalysis have not yet been evaluated.

## Acknowledgements

We would like to thank Pallavi Thaplyal and Aamir Mir for lively discussions and for helpful suggestions on this manuscript.

## Supporting Information

Identification of the HDV ribozyme cleavage product, self-cleavage reaction of the G25A•U20C double mutant, pH–rate profiles for the self-cleavage reactions of the G25A•U20C double mutant and WT in the absence of the $Mg^{2+}$ ion, summary of the population–pH profiles for generating observed rate–pH profiles, possible mechanisms for deprotonation of the 2′-hydroxyl, and statistics from MD simulation of the G25A•U20C double mutant are provided in Appendix D.

## References

(1) Perrotta, A. T., and Been, M. D. *Nature* **1991**, *350*, 434-436.

(2) Ferre-D'Amare, A. R., Zhou, K. H., and Doudna, J. A. *Nature* **1998**, *395*, 567-574.

(3) Salehi-Ashtiani, K., Luptak, A., Litovchick, A., and Szostak, J. W. *Science* **2006**, *313*, 1788-1792.

(4) Webb, C. H., Riccitelli, N. J., Ruminski, D. J., and Luptak, A. *Science* **2009**, *326*, 953.

(5) Eickbush, D. G., and Eickbush, T. H. *Mol. Cell. Biol.* **2010**, *30*, 3142-3150.

(6) Webb, C. H., and Luptak, A. (2011) HDV-like self-cleaving ribozymes. *RNA Biol.* **2011**, *8*,719-727.

(7) Ruminski, D. J., Webb, C. H., Riccitelli, N. J., and Luptak, A. *J. Biol. Chem.* **2011**, *286*, 41286-41295.

(8) Bevilacqua, P. C., and Yajima, R. *Curr. Opin. Chem. Biol.* **2006**, *10*, 455-464.

(9) Gong, B., Chen, J. H., Chase, E., Chadalavada, D. M., Yajima, R., Golden, B. L., Bevilacqua, P. C., and Carey, P. R. *J. Amer. Chem. Soc.* **2007**, *129*, 13335-13342.

(10) Guo, M., Spitale, R. C., Volpini, R., Krucinska, J., Cristalli, G., Carey, P. R., and Wedekind, J. E. *J. Am. Chem. Soc.* **2009**, *131*, 12908-12909.

(11) Rupert, P. B., and Ferre-D'Amare, A. R. *Nature* **2001**, *410*, 780-786.

(12) Martick, M., and Scott, W. G. *Cell* **2006**, *126*, 309-320.

(13) Wilson, T. J., and Lilley, D. M. *RNA* **2011**, *17*, 213-221.

(14) Liu, L., Cottrell, J. W., Scott, L. G., and Fedor, M. J. *Nature Chem. Biol.* **2009**, *5*, 351-357.

(15) Viladoms, J., Scott, L. G., and Fedor, M. J. *J. Am. Chem. Soc.* **2011**, *133*, 18388-18396.

(16) DeRose, V. J. *Curr. Opin. Struct. Biol.* **2003**, *13*, 317-324.

(17) Schnabl, J., and Sigel, R. K. *Curr. Opin. Chem. Biol.* **2010**, *14*, 269-275.

(18) Donghi, D., and Sigel, R. K. *Methods Mol. Biol.* **2012**, *848*, 253-273.

(19) Johnson-Buck, A. E., McDowell, S. E., and Walter, N. G. *Met Ions Life Sci* **2011**, *9*, 175-196.

(20) Wedekind, J. E. *Met Ions Life Sci* **2011**, *9*, 299-345.

(21) Murray, J. B., Seyhan, A. A., Walter, N. G., Burke, J. M., and Scott, W. G. *Chem. Biol.* **1998**, *5*, 587-595.

(22) Nakano, S., Chadalavada, D. M., and Bevilacqua, P. C. *Science* **2000**, *287*, 1493-1497.

(23) Nakano, S., Proctor, D. J., and Bevilacqua, P. C. *Biochemistry* **2001**, *40*, 12022-12038.

(24) Cerrone-Szakal, A. L., Siegfried, N. A., and Bevilacqua, P. C. *J. Am. Chem. Soc.* **2008**, *130*, 14504-14520.

(25) O'Rear, J. L., Wang, S., Feig, A. L., Beigelman, L., Uhlenbeck, O. C., and Herschlag, D. *RNA* **2001**, *7*, 537-545.

(26) Curtis, E. A., and Bartel, D. P. *RNA* **2001**, *7*, 546-552.

(27) Been, M. D. *Curr. Top. Microbiol. Immunol.* **2006**, *307*, 47-65.

(28) Nakano, S., Cerrone, A. L., and Bevilacqua, P. C. *Biochemistry* **2003**, *42*, 2982-2994.

(29) Chen, J. H., Yajima, R., Chadalavada, D. M., Chase, E., Bevilacqua, P. C., and Golden, B. L. *Biochemistry* **2010**, *49*, 6508-6518.

(30) Golden, B. L. *Biochemistry* **2011**, *50*, 9424-9433.

(31) Fauzi, H., Kawakami, J., Nishikawa, F., and Nishikawa, S. *Nucleic Acids Res.* **1997**, *25*, 3124-3130.

(32) Ferre-D'Amare, A. R., and Doudna, J. A. *J. Mol. Biol.* **2000**, *295*, 541-556.

(33) Wadkins, T. S., Shih, I., Perrotta, A. T., and Been, M. D. *J. Mol. Biol.* **2001**, *305*, 1045-1055.

(34) Nakano, S., and Bevilacqua, P. C. *Biochemistry* **2007**, *46*, 3001-3012.

(35) Veeraraghavan, N., Bevilacqua, P. C., and Hammes-Schiffer, S. *J. Mol. Biol.* **2010**, *402*, 278-291.

(36) Wang, J., Cieplak, P., and Kollman, P. *J. Comp. Chem* **2000**, *21*, 1049-1074.

(37) Veeraraghavan, N., Ganguly, A., Chen, J. H., Bevilacqua, P. C., Hammes-Schiffer, S., and Golden, B. L. *Biochemistry* **2011**, *50*, 2672-2682.

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

(39) Veeraraghavan, N., Ganguly, A., Golden, B. L., Bevilacqua, P. C., and Hammes-Schiffer, S. *J Phys Chem B.* **2011**, *115*, 8346-8347.

(40) Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

(41) Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037-10041.

(42) Holst, M. *Adv. Comp. Math.* **2001**, *15*, 139-191.

(43) Bank, R., and Holst, M. *SIAM Review* **2003**, *45*, 291-323.

(44) Chin, K., Sharp, K. A., Honig, B., and Pyle, A. M. *Nat. Struct. Biol.* **1999**, *6*, 1055-1061.

(45) Misra, V. K., and Draper, D. E. *J. Mol. Biol.* **2000**, *299*, 813-825.

(46) Maderia, M., Hunsicker, L. M., and DeRose, V. J. *Biochemistry* **2000**, *39*, 12113-12120.

(47) Garcia-Garcia, C., and Draper, D. E. *J. Mol. Biol.* **2003**, *331*, 75-88.

(48) Wang, T., Tomic, S., Gabdoulline, R. R., and Wade, R. C. *Biophys. J.* **2004**, *87*, 1618-1630.

(49) Misra, V. K., and Draper, D. E. *Biopolymers* **1998**, *48*, 113-135.

(50) Hummer, G., Pratt, L. R., and Garcia, A. E. *J. Chem. Phys.* **1997**, *107*, 9275-9277.

(51) Misra, V. K., and Draper, D. E. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 12456-12461.

(52) Qsite, v. 5.0; Schrödinger, LLC: New York **2008**.

(53) Tanner, N. K., Schaff, S., Thill, G., Petit-Koskas, E., Crain-Denoyelle, A. M., and Westhof, E. *Curr. Biol.* **1994**, *4*, 488-498.

(54) Perrotta, A. T., and Been, M. D. *Nucleic Acids Res.* **1996**, *24*, 1314-1321.

(55) Leontis, N. B., Stombaugh, J., and Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497-3531.

(56) Chen, J. H., Gong, B., Bevilacqua, P. C., Carey, P. R., and Golden, B. L. *Biochemistry* **2009**, *48*, 1498-1507.

(57) Levesque, D., Reymond, C., and Perreault, J. P. (2012) *PLoS One* **2012**, *7*, e40309.

(58) Izatt, R. M., Christensen, J. J., and Rytting, J. H. *Chem. Rev.* **2012**, *71*, 439-481.

(59) Gresh, N., Sponer, J. E., Spackova, N., Leszczynski, J., and Sponer, J. *J. Phys. Chem. B* **2003**, *107*, 8669-8681.

(60) Reblova, K., Spackova, N., Koca, J., Leontis, N. B., and Sponer, J. *Biophys J* **2004**, *87*, 3397-3412.

(61) Cate, J. H., and Doudna, J. A. *Structure* **1996**, *4*, 1221-1229.

(62) Kieft, J. S., and Tinoco, I., Jr. *Structure* **1997**, *5*, 713-721.

## FIGURES AND TABLES

**Table 4.1.** Kinetic parameters for cleavage of WT and double-mutant ribozymes measured in this study.

| | $k_{max}$ (min$^{-1}$) | $K_{D,Mg2+}$ (mM) | $n_{Hill}$ | p$K_a$ |
|---|---|---|---|---|
| WT (fluorophore) | 6.2 ± 0.2 [a] | 1.3 ± 0.1 [a] | 1.8 ± 0.2 [a] | 6.1 ± 0.1 [b] |
| WT (radiolabel) | 7.0 ± 0.2 [c] | 2.1 ± 0.1 [c] | 1.6 ± 0.1 [c] | 6.4 [d] |
| A25•C20 (fluorophore) | 0.078 ± 0.002 [a] | 1.4 ± 0.1 [a] | 2.0 ± 0.2 [a] | 6.2 ± 0.2 [b] |

[a] Reported here. Reactions were performed at 37°C in 50 mM potassium MOPS, pH 7.0.

[b] Reported here. Reactions of WT(fluorophore) and A25•C20 (fluorophore) were performed in 5 mM $Mg^{2+}$ and 50 mM $Mg^{2+}$, respectively at 37 °C. Buffers used were potassium acetate (pH 5.0 and 5.5), potassium MES (pH 6.0 and pH 6.5), potassium MOPS (pH7.0) or Tris-HCl (pH 7.5). All were 50 mM in concentration.

[c] Values from [56] in 25 mM Tris-HCl, pH 7.0 at 37°C.

[d] Values from [22] in 5 mM $Mg^{2+}$. Buffers were 25 mM MES (pH 4.5-6.5) or 25 mM Hepes (pH 6.75-9.0).

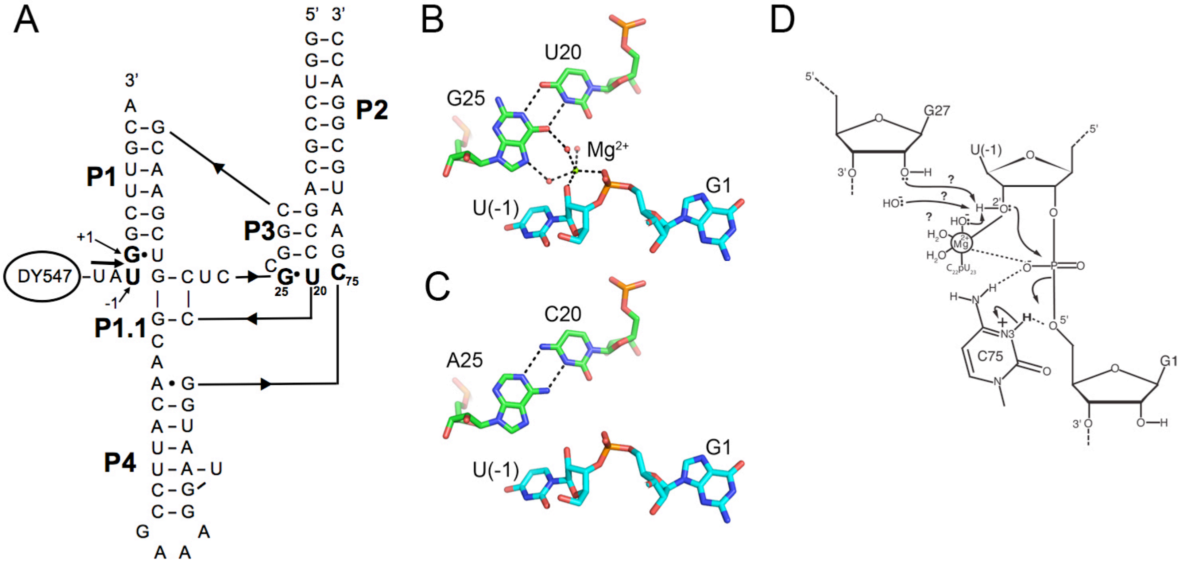

**Figure 4.1.** The structure of the HDV ribozyme**. A.** Secondary structure of the two-piece WT HDV ribozyme used in this study, showing the substrate strand with its attached fluorophore, DY547. Nucleotides U-1, G1, G25, U20 and C75 are in bold and numbered. This is the same RNA used in our pre-cleavage crystal structure.[29] Numbering restarts in the truncated P4 to be consistent with the WT ribozyme. **B.** Tertiary structural features of the WT ribozyme pdbid 3NKB[29] showing the G25•U20 reverse wobble, the cleavage site dinucleotide, and the active site $Mg^{2+}$ ion. **C.** An isosteric reverse wobble that could form in the A25•C20 double mutant. Note that the $Mg^{2+}$ ion is present in the G•U reverse wobble but appears to be absent in the G25A•U20C double mutant (see text). Panels B and C were generated with Pymol.[52] **D.** HDV ribozyme cleavage mechanism. C75 has a $pK_a$ that is shifted towards neutrality and is in good position to serve as a general acid. The active site $Mg^{2+}$ ion is positioned to activate the nucleophile. The identity of the functional group that accepts the proton from the 2'-OH of U(-1) is unknown, but candidates include hydroxide from the solvent, a $Mg^{2+}$-bound hydroxide, and the 2'-OH of G27. See Figure C.5 for further details.

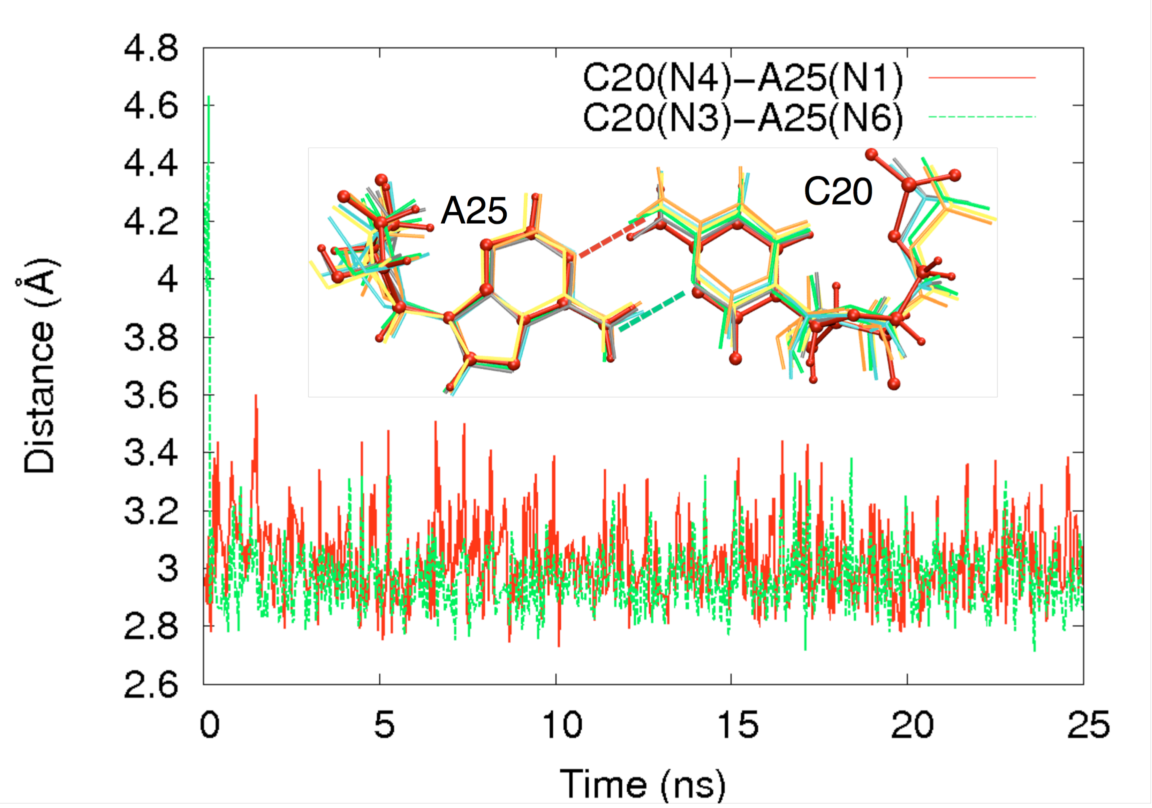

**Figure 4.2.** Stability of the A25•C20 reverse wobble along a MD trajectory. In this particular simulation, the active site $Mg^{2+}$ ion was removed, and two $Na^+$ ions were placed in the bulk solvent. The C20(N4)-A25(N1) and C20(N3)-A25(N6) distances are plotted along the trajectory in red and green respectively. The inset figure shows snapshots of the A•C reverse wobble at various time steps along the trajectory: red, 0ns (also ball and stick representation); cyan, 5ns; gray, 10ns; orange, 15ns; yellow, 20ns; green, 25ns. Very little motion of the A•C reverse wobble occurs.

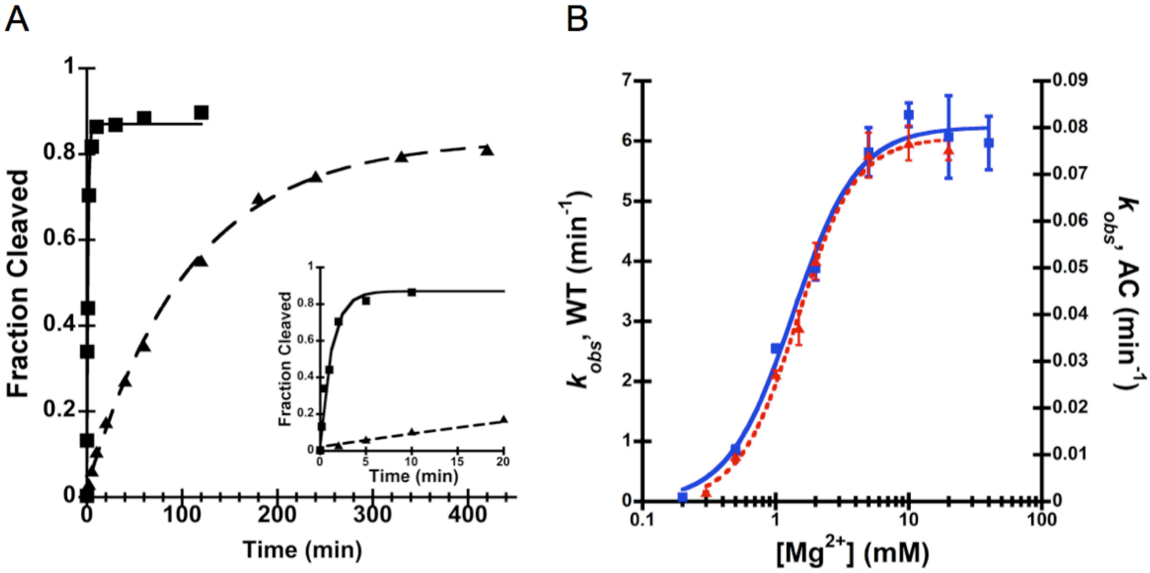

**Figure 4.3.** Magnesium dependence of the HDV ribozyme reaction. **A**. Comparison of the WT and A25•C20 double mutant ribozymes. Cleavage of a fluorescently labeled RNA substrate by WT (squares) and A25•C20 double mutant ribozymes (triangles). Cleavage reactions contain 0.5 mM Mg$^{2+}$, 50 mM MOPS (pH 7.0). The inset shows fraction cleaved at early time points. **B.** The cleavage rate constants as a function of Mg$^{2+}$ concentration. Rate constants for WT (squares) and A25•C20 double mutant ribozymes (triangles) are plotted as a function of Mg$^{2+}$ ion concentration. All reactions contain 50 mM MOPS (pH 7.0). Although the maximal velocity is higher for the WT ribozyme than the A25•C20 double mutant ribozyme, the apparent $K_{D,Mg^{2+}}$ and $n_{Hill}$ are the same for the two ribozymes within error (Table 4.1).

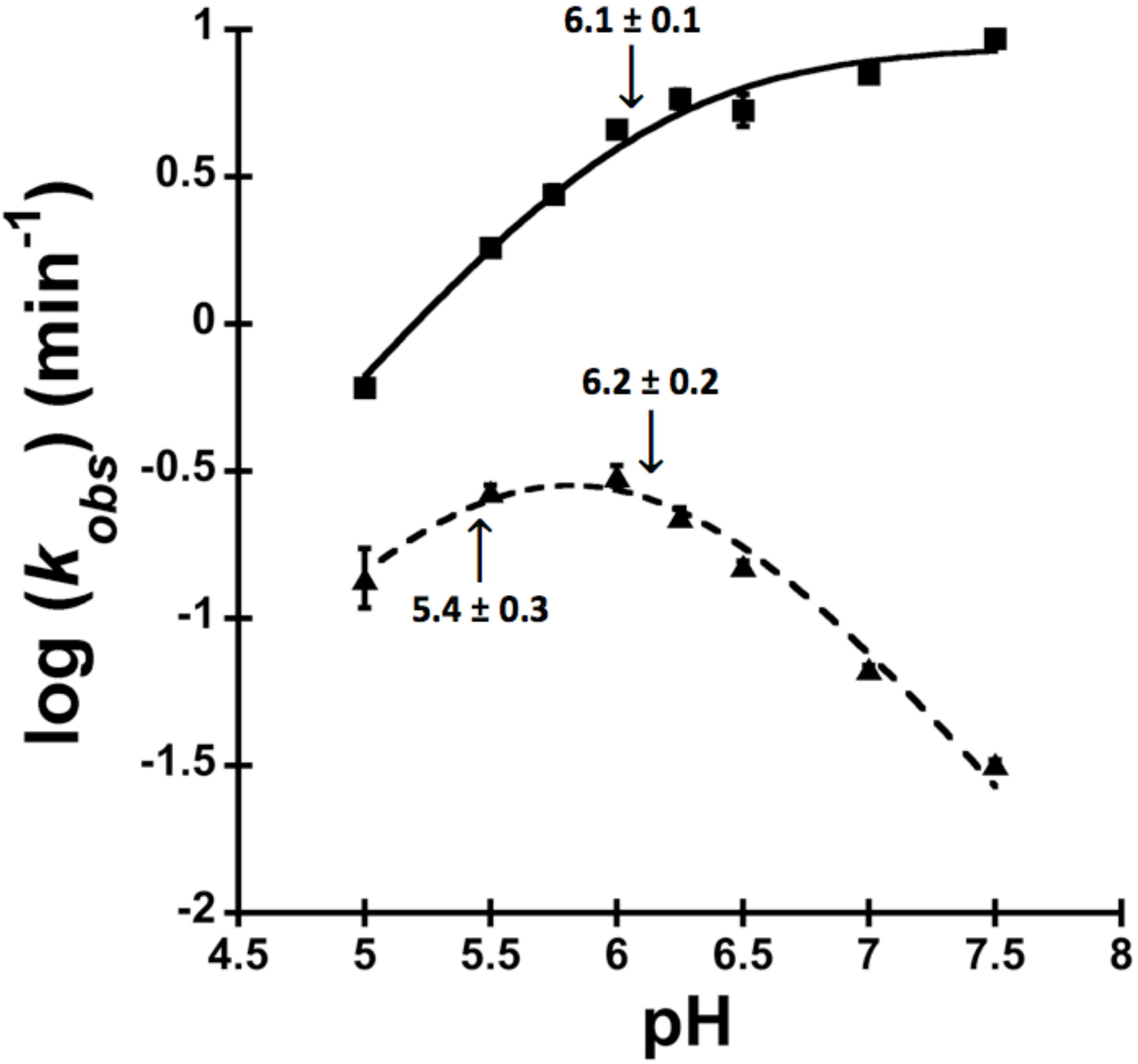

**Figure 4.4.** The pH-rate profiles of A25•C20 double mutant ribozyme as compared to that of the WT. The reactions were performed in 5 mM $Mg^{2+}$ for the WT (squares) and 50 mM $Mg^{2+}$ for the A25•C20 double mutant (triangles). The curves were fit by using equation (4.4) for the WT and equation (4.5) for the A25•C20 double mutant ribozyme. The apparent p$K_a$s are indicated with an arrow and summarized in Table 4.1. Note that the p$K_a$ of 5.4 ± 0.3 might be originated from acid denaturation or it could be due to contribution of a catalytic metal ion only at low pH (See Discussion).

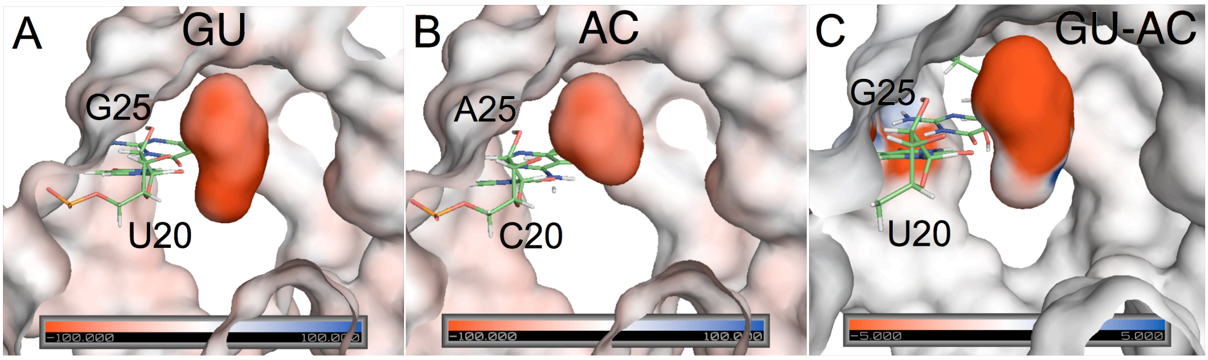

**Figure 4.5.** Surface electrostatic potential of the precleaved HDV ribozyme. **A.** The WT structure, and **B.** the G25A•U20C double mutant, labeled DM. **C.** A difference map showing changes in the electrostatic potential between the WT and G25A•U20C double mutant ribozymes mapped onto the WT structure. The scale for panels **A** and **B** is -100 to +100 kT/e, and for panel **C** is -5 to +5 kT/e. The pocket near the G•U reverse wobble in the WT structure is more negative than the pocket near the A•C reverse wobble in the G25A•U20C double mutant structure. Panels were rendered using Pymol. [52]

# CHAPTER 5: QUANTUM MECHANICAL/MOLECULAR MECHANICAL STUDY OF THE HDV RIBOZYME: IMPACT OF THE CATALYTIC METAL ION ON THE MECHANISM[1]

## Abstract

A recent crystal structure of the precleaved state of the HDV ribozyme along with biochemical data support a mechanism for phosphodiester bond self-cleavage in which C75 acts as a general acid and bound $Mg^{2+}$ ion acts as a Lewis acid. Herein this precleaved crystal structure is used as the basis for quantum mechanical/molecular mechanical calculations of the solvated ribozyme. These calculations indicate that the self-cleavage reaction is concerted with a phosphorane-like transition state when a divalent ion, $Mg^{2+}$ or $Ca^{2+}$, is bound at the catalytic site but is sequential with a phosphorane intermediate when a monovalent ion, such as $Na^{+}$, or no ion is at this site. Electrostatic potential calculations suggest that the divalent metal ion at the catalytic site lowers the $pK_a$ of C75, leading to the concerted mechanism in which the proton is already partially transferred to the leaving group in the phosphorane-like transition state. These observations are consistent with experimental data, including $pK_a$ measurements, reaction kinetics, and proton inventories with divalent and monovalent ions.

---

[1] This chapter was published in its entirety in the Journal of Physical Chemistry Letters in 2013. The complete reference to the journal article is as follows: Ganguly, A.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Phys. Chem. Lett.* **2011**, *2*, 2906-2911.

RNA enzymes, or ribozymes, participate in a number of biologically critical reactions, including tRNA maturation, splicing, and translation. The hepatitis delta virus (HDV) ribozyme is widespread in biology and functions in viral processing in human cells through its phosphodiester self-cleavage.[1] The HDV ribozyme undergoes a reaction in which a phosphodiester bond is formed between O2′(U-1) and P(G1) and broken between P(G1) and O5′(G1). A recent structure of the HDV ribozyme trapped in the precleaved state was solved in the presence of $Mg^{2+}$ with a C at position 75, and at low pH of 5.0 to maintain C75 in the protonated state.[2] In this structure, an active site $Mg^{2+}$ ion was clearly resolved. The structure, along with a variety of biochemical data, suggest a plausible mechanism for this reaction: C75 is positioned to act as a general acid and transfer a proton to the 5′-bridging oxygen leaving group of G1, and a catalytic $Mg^{2+}$ ion serves as a Lewis acid to stabilize the 2′-oxyanion of U-1 (Figure 5.1).[2,3] We recently conducted several molecular dynamics (MD) studies on this reaction,[4,5] leading to two main conclusions: (1) the catalytic $Mg^{2+}$ ion remains localized (chelated) near the reverse G25•U20 wobble, in contrast to a metal ion that is more diffuse near the standard G1•U37 wobble, and (2) protonation of C75 keeps the active site primed for the catalytic reaction through key hydrogen-bonding interactions and binding of $Mg^{2+}$. To our knowledge, this precleaved crystal structure has not yet been used to investigate the mechanism of self-cleavage with theoretical methods.

In this chapter, we perform quantum mechanical/molecular mechanical (QM/MM) calculations based on this precleaved crystal structure to examine the self-cleavage mechanism of the HDV ribozyme. Our objective is to determine if this reaction is sequential, potentially involving a phosphorane intermediate, or concerted and to examine the role of the catalytic $Mg^{2+}$ in the mechanism. Our calculations indicate that the self-cleavage reaction in the HDV ribozyme is concerted with a phosphorane-like transition state and that the catalytic $Mg^{2+}$ strongly influences the mechanism. We find that the mechanism becomes sequential when the catalytic $Mg^{2+}$ is replaced by a monovalent ion but remains concerted when the catalytic $Mg^{2+}$ is replaced by another divalent ion, $Ca^{2+}$. Our analysis suggests that the divalent metal ion lowers the $pK_a$ of C75, leading to a concerted mechanism in which the proton is already partially transferred to the leaving group in the phosphorane-like transition state. These results are consistent with previous crystallographic and biochemical studies on the HDV ribozyme.[6-9]

Three prior quantum mechanical calculations differ fundamentally from this study in that they used a starting model based on the crystal structure of the HDV ribozyme with a C75U mutation,[10] albeit mutated back to C *in silico*. This C75U mutant does not appear to be reflective of the catalytic conformation of the ribozyme for multiple reasons involving mis-positioning of functionalities and metal ions, as discussed elsewhere.[11] Moreover, these studies have provided conflicting results regarding a stable phosphorane intermediate. In one study, density functional theory (DFT) was used to perform geometry optimizations on two cutoff models, each containing 42 non-hydrogen atoms, with the environment modeled as a dielectric continuum aqueous solvent.[12] A stable phosphorane intermediate was found in a model with C75 acting as the general acid but not in a model with C75 acting as the general base, and analysis based on the energy barriers and estimated concentrations favored the C75 general acid mechanism. In a second study, DFT geometry optimizations of similar, relatively small model systems favored the C75 general acid mechanism but with no stable intermediate.[13] In a third study, QM/MM calculations of the solvated ribozyme favored the C75 general base mechanism with no stable intermediate.[14] Thus, there is no consensus among the three theoretical studies to date on the formation of stable intermediates or whether C75 serves as the general acid or base in the mechanism when the structure derived from the C75U mutant is probed.

In contrast to these previous studies, the QM/MM calculations in the present study are based on the precleaved crystal structure (PDB ID 3NKB) solved with wild-type sequences (i.e., with the C75 present and protonated) and with $Mg^{2+}$ bound.[2] The upstream nucleotide and scissile phosphate were built as described previously.[2] The resulting model of the precleaved ribozyme contains 73 nucleotides. Hydrogen atoms were added to the entire structure using Accelrys Discover Studio Visualizer 2.0. C41 was protonated at N3 to maintain its hydrogen-bonded base triple,[15] and C75 was protonated at N3, as suggested by $pK_a$ measurements on the precleaved form.[7] The 11 $Mg^{2+}$ ions resolved in the crystal structure were included. The ribozyme was solvated with rigid TIP3P waters[16] in a periodically replicated orthorhombic box. The system was neutralized with $Na^+$ ions, and physiological monovalent ionic strength was added to the solvent to give ~0.15 M NaCl.

The starting configurations for the QM/MM optimizations were obtained from molecular dynamics (MD) simulations of the solvent and $Na^+$ and $Cl^-$ ions according to the equilibration procedure described in the Supporting Information. This MD equilibration was performed with the Desmond[17,18] program using the AMBER99[19,20] force field. During equilibration, the ribozyme and the $Mg^{2+}$ ions were fixed at the coordinates from the crystal structure. Following the MD equilibration, the solvent-equilibrated system was truncated by deleting all solvent molecules outside a cutoff distance of 10 Å from the ribozyme. In addition, the O2′(U-1) was deprotonated to drive its nucleophilic attack on the scissile phosphate in the subsequent QM/MM calculations. This study does not focus on the mechanism of deprotonation of this oxygen or the identity of the general base.

The QM/MM calculations were performed using the Qsite program.[21] The QM region was treated with density functional theory (DFT) at the B3LYP/6-31G** level of theory, including the LACVP* pseudopotential for the metal ions $K^+$, $Cs^+$ and $Ca^{2+}$, with an ultrafine grid. The hydrogen-capping method was used for the residues at the boundary of the QM and MM regions.[21] The MM region was described by the OPLS2005 force field, and a 100 Å residue-based nonbonded cutoff was used. Because we are interested mainly in the changes that occur in the active site, all atoms outside a 20 Å radius sphere centered at the scissile phosphate were held fixed throughout the QM/MM calculations, although their energetic influence on all atoms (QM and MM) was maintained throughout the study. Prior to the full QM/MM geometry optimization, an initial MM geometry optimization of the atoms within this 20 Å radius sphere was performed with the QM region fixed.

Four different QM regions were considered for this system. Each of these regions contains the following key residues: C75 nucleobase, U-1 sugar and phosphate, U23 phosphate, and the active site $Mg^{2+}$ ion with its three crystallographic waters. These residues, as well as the additional residues included in each region, are depicted in Figure D.1. QM/MM geometry optimizations were performed with each of these QM regions to enable comparison of key distances and angles in the active site. Table D.1 provides a comparison of the important active site distances of the ribozyme after QM/MM geometry optimization using the different QM regions. Based on this comparison, the QM region chosen for the subsequent calculations contained 87 atoms and consisted of the U-1 nucleotide, G1 sugar and scissile phosphate, protonated C75 nucleobase, U23 phosphate, C22 sugar, and catalytic $Mg^{2+}$ ion with its three water ligands (Figure 5.1).

While these types of QM/MM calculations provide useful qualitative mechanistic information, they are also associated with known limitations. In complex systems, a large number of related minima and transition states (i.e., first-order saddle points) are present on the potential energy surface. The specific reactant, transition state, and product identified with QM/MM calculations depends on the starting configuration, which corresponds to the crystal structure coordinates of the precleaved ribozyme immersed in equilibrated solvent in the present work. The resulting reaction pathway is therefore not unique. Nonetheless, the underlying assumption of these calculations is that the qualitative mechanism is representative of a set of such reaction pathways. The energetics along the reaction pathway are not quantitatively meaningful because entropic contributions are neglected, and conformational sampling is required to average over the entire set of relevant reaction pathways. Future work will utilize *ab initio* molecular dynamics methods to generate free energy profiles along the dominant reaction coordinates[22] or an approach such as transition path sampling to obtain additional information about the mechanisms and the relative rate constants.

In the present study, QM/MM calculations, in conjunction with experimental data from the literature, provide insight into the qualitative mechanism of the self-cleavage reaction of the HDV ribozyme. The specific objectives of this study were to determine if this self-cleavage mechanism is sequential, potentially with a phosphorane intermediate, or concerted and to understand the impact of the catalytic $Mg^{2+}$ ion on the mechanism. The reactant (precleaved) and product (cleaved) structures were determined by placing the transferring hydrogen from the general acid C75 near its donor or acceptor, respectively, prior to the QM/MM geometry optimization using the methods described above. The reactant and product structures, which are minima on the QM/MM potential energy surface, are shown in Figures 5.2a and 5.2c, respectively. In the reactant structure, the proton (H3) is bonded to N3, and the P-O5′ bond is formed, while in the product structure, the proton is bonded to O5′, the P-O2′ bond is formed, and the P-O5′ bond is broken.

Starting from the QM/MM-optimized reactant and product structures, we identified a transition state (TS) using the quasi-synchronous transit (QST)[23] method. The transition state obtained has a phosphorane-like structure, with the scissile phosphate penta-coordinated, the P-O2′ bond partially formed, and the P-O5′ bond partially broken (Figure 5.2b). To confirm that this TS corresponds to the reactant and product of interest, we calculated the mass-weighted minimum energy path (MEP) from the TS back to the reactant and product structures along the intrinsic reaction coordinate (IRC). The energy along this MEP is depicted in Figure 5.3a. The calculated energy barriers along the MEP are 25 kcal/mol and 15 kcal/mol in the forward and reverse directions, respectively. We emphasize that these energies cannot be used to estimate rate constants because they do not include entropic effects and only represent a single reaction pathway.

We also investigated the degree of conformational changes along the reaction pathway. For this purpose, we calculated the root-mean-squared deviation (RMSD) between the active site atoms, defined as the QM region atoms, for specified reactant, TS, or product structures. The RMSD between the starting structure and the optimized reactant structure was 0.41 Å. These structural differences arise predominantly from a decrease in the N3(C75)-O5′(G1) distance (i.e., the proton donor-acceptor distance) from 3.6 Å to 3.01 Å. The RMSDs between the optimized TS and reactant and between the optimized product and reactant are 0.15 Å and 0.19 Å, respectively. These values indicate that no significant conformational changes occur along this portion of the reaction pathway for self-cleavage in the HDV ribozyme.

According to the MEP shown in Figure 5.3, the mechanism of self-cleavage in the HDV ribozyme is a concerted reaction with a phosphorane-like TS (i.e. there is no intermediate along this reaction pathway). The relevant reaction coordinates in this process are the P(G1)-O5′(G1) and P(G1)-O2′(U-1) distances, as well as the difference between the N3(C75)-H3(C75) and O5′(G1)-H3(C75) distances. Figure 5.3b depicts the values of these distances along the IRC. Although the overall self-cleavage reaction is concerted, with a single TS connecting the reactant and product, it can be visualized in three stages along the IRC. Initially the O2′ attacks the scissile phosphate to bring the local geometry into a favorable conformation for the subsequent proton transfer and the bond formation/cleavage processes. This initial attack is observed in the P-O2′ distance decreasing without much change in P-O5′ or proton transfer distances (Figure 5.3b). Then the proton (H3) is transferred from N3(C75) to O5′(G1), which is observed in the (N3-H3)-(O5’-H3) distance changing dramatically and the P(G1)-O2′(U-1) and P(G1)-O5′(G1) distances remaining virtually unchanged. In the last stage, the P(G1)-O2′(U-1) and P(G1)-O5′(G1) bonds are completely formed and broken, respectively.

Although no stable intermediate was observed along the reaction pathway depicted in Figure 5.3, suggesting a concerted mechanism, we cannot rule out the possibility that a sequential pathway also exists for this system. Formation of intermediate penta-coordinate phosphorane species has been proposed in some other ribozyme catalyzed mechanisms.[12] We searched for these intermediate minima by performing a series of constrained optimizations along the P-O5′ and P-O2′ coordinates. These constrained optimizations were performed for three different configurations at the proton transfer interface corresponding to the reactant state, transition state, and product state. As shown in Figure D.2, the three resulting constrained potential energy surfaces do not show evidence of additional minima corresponding to intermediate structures. Thus, although an exhaustive search of the many-dimensional potential energy surface is not possible, the evidence provided by our calculations is consistent with a concerted mechanism in the presence of a bound $Mg^{2+}$ ion.

As mentioned above, the structure of the precleaved HDV ribozyme contains a catalytic $Mg^{2+}$ ion interacting with one of the non-bridging (*pro-*$R_P$) oxygens of the scissile phosphate. This $Mg^{2+}$ ion is depicted as a purple sphere in Figure 5.2. We hypothesized that the absence of the phosphorane intermediate may be related to the presence of the $Mg^{2+}$ ion at that position. Biochemical experiments indicate that the HDV ribozyme is active with a wide range of divalent ions, especially $Ca^{2+}$,[24,25] as well as monovalent ions at high concentrations.[9] To explore our hypothesis, we replaced the catalytic $Mg^{2+}$ ion with monovalent ions or another divalent ion ($Li^+$, $Na^+$, $K^+$, $Cs^+$ and $Ca^{2+}$) and investigated the possibility of a phosphorane intermediate for each metal ion. We also explored the case with no ion at the catalytic position.

For each of these modified systems, the starting point for the QM/MM optimization was the TS obtained for the original system, converted to the modified system by replacing the catalytic $Mg^{2+}$ with another ion or by removing it. In each case, we searched for both a transition state and a minimum from this starting structure using a variety of constrained optimization procedures. The results are summarized in Table 5.1, with active site structural details in Table D.2. For all cases with a monovalent ion or no ion at the catalytic position, a phosphorane intermediate was obtained as a minimum on the potential energy surface. In contrast, for both cases with a divalent ion at the catalytic position, no such intermediate was obtained (i.e., no minima other than the reactant and product states were found), but a phosphorane-like transition state was identified. Although we found a phosphorane intermediate for all cases with a monovalent ion at the catalytic position, we were unable to find a phosphorane-like transition state for these cases. In other words, for each case studied, we found either a phosphorane-like transition state or a phosphorane intermediate, but not both.

A second active site $Mg^{2+}$ ion has been observed to interact diffusely with the standard G1•U37 wobble. In the calculations described above, only $Na^{+}$ ions were interacting with this standard G•U wobble. The results of calculations performed with a $Mg^{2+}$ ion at the G1•U37 wobble are qualitatively the same as those obtained in its absence. Thus, this diffusely interacting $Mg^{2+}$ ion does not appear to influence the qualitative mechanism of self-cleavage.

The structures of the phosphorane-like TS with a divalent ion at the catalytic position and the phosphorane intermediate with a monovalent ion at the catalytic position are very similar. The most prominent difference between these two structures is the position of H3(C75) (Figure D.3, Table D.2). In the phosphorane-like TS, this hydrogen is midway between N3(C75) and O5′(G1), and in the phosphorane intermediate, this hydrogen is still bonded to N3(C75). Quantitatively, the N3(C75)-H3(C75) and O5′(G1)-H3(C75) distances are 1.32 Å and 1.31 Å, respectively, for the phosphorane-like TS obtained with the catalytic $Mg^{2+}$ and 1.06 Å and 1.86 Å, respectively, for the phosphorane intermediate with $Na^{+}$ (Table D.2). Note that the N3(C75)-O5′(G1) distance is ~0.3 Å smaller in the phosphorane-like TS than in the phosphorane intermediate. One plausible explanation for these structural differences is that the higher charge of the $Mg^{2+}$ ion allows the divalent metal ion to withdraw more electron density from C75, thereby increasing the acidity of the proton H3(C75). Indeed, solution and crystallographic experiments reveal lowering of $pK_a$ values of C75 from 7.25 to 5.9 upon going from unbound to bound $Mg^{2+}$, respectively, but maintaining of a high $pK_a$ value, near 7.5, in the presence of up to

1 M $Na^+$.[6,7,26] (Note that all of these p$K_a$ values are perturbed toward neutrality from the unshifted value of 4.2 for free cytosine, and hence favor general acid catalysis.)

The lower p$K_a$ value of C75 with bound divalent metal ions as compared to monovalent ions is supported by comparison of the electrostatic potentials generated for the reactant state with a $Mg^{2+}$ ion or a $Na^+$ ion at the catalytic position (Figure 5.4). This figure suggests that the electrostatic potential at C75 is more positive with the catalytic $Mg^{2+}$ in the active site, resulting in a lower p$K_a$ value of C75. To determine whether this phenomenon was due to differences between the divalent and monovalent metal ions or between the optimized reactant structures, we also calculated the electrostatic potentials for each optimized structure with the catalytic $Mg^{2+}$ replaced by $Na^+$, or the bound $Na^+$ replaced by $Mg^{2+}$. These results, which are provided in Figure D.4, indicate that the differences between the electrostatic potentials shown in Figure 5.4 are due to the presence of the divalent versus monovalent ion, rather than to structural differences.

In conclusion, our QM/MM calculations indicate that the self-cleavage reaction of the HDV ribozyme is concerted with a phosphorane-like transition state in the presence of $Mg^{2+}$. When the catalytic $Mg^{2+}$ ion is replaced by a monovalent ion, the mechanism becomes sequential with a phosphorane intermediate. On the other hand, when this catalytic $Mg^{2+}$ ion is replaced by another divalent ion, the mechanism remains concerted. These results are consistent with the experimental observations that the HDV ribozyme functions with similarly fast reaction kinetics with $Mg^{2+}$ and $Ca^{2+}$,[24,25] suggesting that these reactions proceed by the same mechanism, but functions about 20-fold slower with $Na^+$, suggesting that this reaction proceeds by an altered mechanism.[26] In addition, proton inventory experiments support a stepwise reaction of the HDV ribozyme in the presence of $Na^+$ but a concerted reaction in physiological concentrations of $Mg^{2+}$.[6,8] According to our calculations, the main structural difference between the phosphorane-like TS and the phosphorane intermediate is that the proton is partially transferred to the leaving group in the former but not in the latter. Our analysis suggests that the catalytic $Mg^{2+}$ ion withdraws more electron density from C75 than does a monovalent metal ion bound at this site. As a result, the p$K_a$ value of C75 is lower in the presence of the divalent metal ion, leading to a concerted mechanism with a phosphorane-like transition state in which the C75 proton is partially transferred. This result is consistent with the experimental observation that the p$K_a$ value of C75 is lowered by more than 1 unit by bound $Mg^{2+}$ but is not affected by high concentrations of $Na^+$.[6] Furthermore, our calculations indicate that no significant conformational changes occur along the reaction pathway. Our studies herein suggest unique ways in which the HDV ribozyme uses communication between a metal ion and a nucleobase to help drive catalysis.

## Acknowledgments

We gratefully acknowledge assistance from Prof. Barbara Golden, Phil Hanoian, Dr. Alexander Soudackov, Chaehyuk Ko, and Dr. Victor Guallar. This project was supported by NIH grant GM56207 (S.H.-S.) and NIH grant R01GM095923 (P.C.B).

## Supporting Information

Description of MD equilibration procedure, analysis of QM regions considered for QM/MM calculations, potential energy surfaces obtained from constrained optimizations, comparison of structures and energetics of phosphorane-like transition states and phosphorane intermediates with divalent and monovalent ions, respectively, at catalytic site, analogous figure to Figure 5.3 with $Ca^{2+}$ instead of $Mg^{2+}$ at the catalytic site, and analogous figure to Figure 5.4 with $Mg^{2+}$ and $Na^+$ ions exchanged are provided in Appendix C.

**References**

1. Webb, C.-H. T.; Luptak, A. *RNA Biol.* **2011,** *8,* 719-727.
2. Chen, J.-H.; Yajima, R.; Chadalavada, D. M.; Chase, E.; Bevilacqua, P. C.; Golden, B. L. *Biochemistry* **2010,** *49*, 6508-6518.
3. Das, S.; Piccirilli, J. *Nat. Chem. Biol.* **2005,** *1*, 45-52.
4. Veeraraghavan, N.; Ganguly, A.; Chen, J.; Bevilacqua, P. C.; Hammes Schiffer, S.; Golden, B. L. *Biochemistry* **2011,** *50*, 2672-2682.
5. Veeraraghavan, N.; Ganguly, A.; Golden, B. L.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Phys. Chem. B* **2011,** *115*, 8346-8357.
6. Cerrone-Szakal, A. L.; Siegfried, N. A.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2008,** *130*, 14504-14520.
7. Gong, B.; Chen, J.-H.; Chase, E.; Chadalavada, D. M.; Yajima, R.; Golden, B. L.; Bevilacqua, P. C.; Carey, P. R. *J. Am. Chem. Soc.* **2007,** *129*, 13335-13342.
8. Nakano, S.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2001,** *123*, 11333-11334.
9. Nakano, S.; Chadalavada, D. M.; Bevilacqua, P. C. *Science* **2000,** *287*, 1493-1497.
10. Ke, A.; Zhou, K.; Ding, F.; Cate, J. H. D.; Doudna, J. A. *Nature* **2004,** *429*, 201-205.
11. Nakano, S.; Bevilacqua, P. C. *Biochemistry* **2007,** *46*, 3001-3012.
12. Wei, K.; Liu, L.; Cheng, Y.-H.; Fu, Y.; Guo, Q.-X. *J. Phys. Chem. B* **2007,** *111*, 1514-1516.
13. Liu, H.; Robinet, J. J.; Ananvoranich, S.; Gauld, J. W. *J. Phys. Chem. B* **2007,** *111*, 439-445.
14. Banas, P.; Rulisek, L.; Hanosova, V.; Svozil, D.; Walter, N. G.; Sponer, J.; Otyepka, M. *J. Phys. Chem. B* **2008,** *112*, 11177-11187.
15. Veeraraghavan, N.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Mol. Biol.* **2010***, 402*, 278-291.
16. Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983,** *79*, 926-935.
17. Desmond Molecular Dynamics System, v. 2.2. *D. E. Shaw Research, New York, NY* **2009**. Maestro-Desmond Interoperability Tools, v. 2.2. *Schrodinger, New York, NY, 2009.*
18. Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I. K.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, November 11-17* **2006**.
19. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995,** *117*, 5179-5197.
20. Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comp. Chem.* **2000,** *21*, 1049-1074.
21. Qsite v. 5.0. *Schrodinger, LLC, New York, NY, 2008.*
22. Rosta, E.; Nowotny, M.; Yang, W.; Hummer, G. *J. Am. Chem. Soc.* **2011,** *133*, 8934-8941.
23. Schlegel, H. B. *Isr. J. Chem.* **1993,** *33*, 449-454.
24. Suh, Y. A.; Kumar, P. K. R.; Taira, K.; Nishikawa, S. *Nucleic Acids Res* **1993,** *21*, 3277-3280.
25. Nakano, S.; Cerrone, A.; Bevilacqua, P. C. *Biochemistry* **2003,** *42*, 2982-2994.
26. Nakano, S.; Proctor, D. J.; Bevilacqua, P. C. *Biochemistry* **2001,** *40*, 12022-12038.

# FIGURES AND TABLES

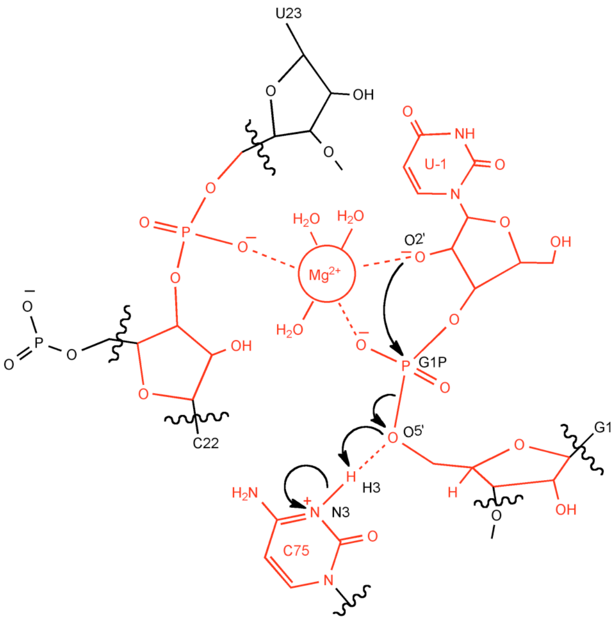

**Figure 5.1.** Schematic illustration of the QM region chosen for the QM/MM calculations used to generate the results in this study. The atoms in red are included in the QM region, and wavy lines indicate the QM/MM boundary. The region contains 87 QM atoms. The arrows describe the self-cleavage reaction.

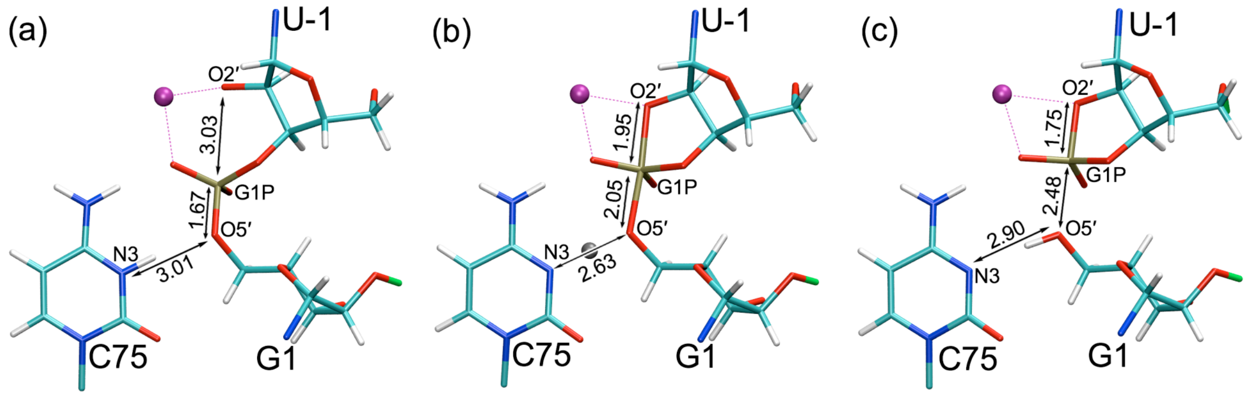

**Figure 5.2.** Cleavage site for the (a) reactant, (b) transition state, and (c) product states obtained from the QM/MM optimizations. The active site $Mg^{2+}$ (purple) is coordinated to the *pro*-$R_P$ oxygen of the scissile phosphate and the nucleophilic O2′ of U-1. In the transition state, the transferring hydrogen is depicted as a gray ball. Catalytically relevant bond distances are given.

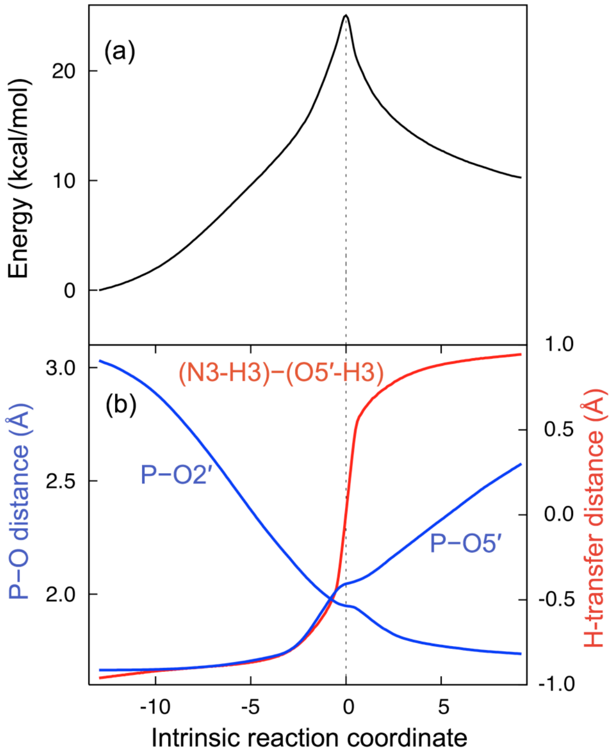

**Figure 5.3.** Minimum energy path (MEP) obtained for the self-cleavage reaction using QM/MM methods. The energy (a) and relevant distances (b) are shown along the intrinsic reaction coordinate.

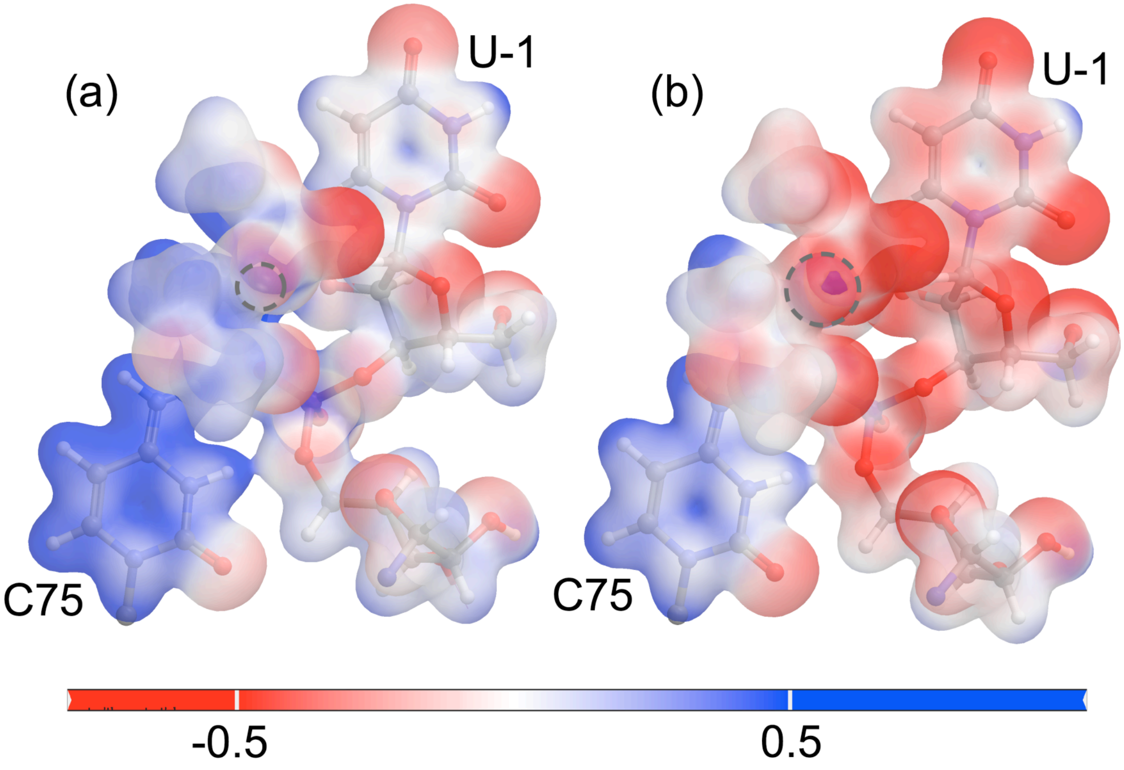

**Figure 5.4.** Molecular electrostatic potential mapped onto the electron density isosurface. The QM region is shown here for the reactant state with (a) $Mg^{2+}$ ion and (b) $Na^{+}$ ion at the catalytic position. In both panels, the scale is from -0.5 to 0.5 kT/e at 298 K. The dotted gray circle indicates the position of the metal ion in each case. These figures suggest that the $Mg^{2+}$ ion at the catalytic site withdraws greater electron density from C75, making the residue more electron deficient and the transferring proton more acidic as compared to the case of a $Na^{+}$ ion at the catalytic site.

| Metal ion at catalytic position | Phosphorane intermediate obtained | Phosphorane-like TS obtained |
|---|---|---|
| $Mg^{2+}$ | No | Yes |
| $Ca^{2+}$ | No | Yes |
| $Li^{+}$ | Yes | No |
| $Na^{+}$ | Yes | No |
| $K^{+}$ | Yes | No |
| $Cs^{+}$ | Yes | No |
| no metal ion | Yes | No |

**Table 5.1.** Results of exploring different cases of active site metal ions. Phosphorane intermediates were obtained in cases with monovalent ions at the catalytic position but not in cases with divalent ions at the same site.

# CHAPTER 6: THIO EFFECTS AND AN UNCONVENTIONAL METAL ION RESCUE IN THE GENOMIC HEPATITIS DELTA VIRUS RIBOZYME[1]

## Abstract

Metal ion and nucleobase catalysis are important for ribozyme mechanism, but the extent to which they cooperate is unclear. A crystal structure of the hepatitis delta virus (HDV) ribozyme suggested that the *pro*-$R_P$ oxygen at the scissile phosphate directly coordinates a catalytic $Mg^{2+}$ ion and is within hydrogen bonding distance of the amine of the general acid C75. Prior studies on the genomic HDV ribozyme, however, showed neither a thio effect nor metal ion rescue using $Mn^{2+}$. Herein, we combine experiment and theory to explore phosphorothioate substitutions at the scissile phosphate. We report significant thio effects at the scissile phosphate and metal ion rescue with $Cd^{2+}$. Reaction profiles with an $S_P$-phosphorothioate substitution are indistinguishable from those of the unmodified substrate in the presence of $Mg^{2+}$ or $Cd^{2+}$, supporting that the *pro*-$S_P$ oxygen does not coordinate metal ions. The $R_P$-phosphorothioate substitution, however, exhibits biphasic kinetics, with the fast- and slow-reacting phases displaying thio effects of 5- and 1,000-fold, respectively. Moreover, the fast- and slow-reacting phases give metal ion rescues in $Cd^{2+}$ of up to 10- and 330-fold, respectively. This suggests a direct interaction of the catalytic metal ion with the *pro*-$R_P$ oxygen. Experiments without divalent ions, with mutants that interfere with $Mg^{2+}$ binding, or with C75 deleted suggest that the

---

[1] This chapter was published in its entirety in the Journal of Biochemistry in 2013. The complete reference to the journal article is as follows: Thaplyal, P.; Ganguly, A.; Golden, B. L.; Hammes-Schiffer, S.; Bevilacqua, P. C. *Biochemistry* **2013**, 52, 6499-6514. P.T. performed the experiments and A.G. performed the computational studies.

*pro*-$R_P$ oxygen plays at most a redundant role in positioning C75. Quantum mechanical/molecular mechanical (QM/MM) studies indicate that the metal ion contributes to catalysis by interacting with both the *pro*-$R_P$ oxygen and the nucleophilic 2'- hydroxyl, supporting the experimental findings.

## Introduction

The hepatitis delta virus (HDV) ribozyme is a small ~85 nt RNA that occurs in closely related genomic and antigenomic forms and is important in the life cycle of the virus.[1,2] It self-cleaves at the phosphodiester bond between U-1 and G1 to generate products with 2',3'-cyclic phosphate and 5'- hydroxyl termini. The secondary structure has five pairings, P1-P4 and P1.1, which comprise two pseudoknots that provide a compact fold (Figure 6.1A).[3,4] The ribozyme has several conserved residues including the catalytic C75, which participates in proton transfer, and one standard and one reverse G•U wobble (Figure 6.1).[5,6] The HDV ribozyme is unique amongst the well-characterized ribozymes in its ability to employ both nucleobase and metal ion catalysis in its reaction mechanism.[7]

Several studies support C75 as the general acid in the reaction. Early experiments indicated that removal of this C can be rescued by the addition of imidazole[8] indicating a role in proton transfer, which was supported by Bronsted studies.[9,10] Solution kinetic studies revealed a shift of the p$K_a$ for C75 by 2-3 units toward neutrality and an inverted rate-pH profile in the absence of $Mg^{2+}$.[11–13] The conclusion of a p$K_a$ shifted toward neutrality was strengthened by direct measurements of the p$K_a$ using Raman studies on crystals of a pre-cleaved genomic ribozyme.[14] Importantly, a hyperactive 5'-phosphorothiolate mutant rescued the deleterious effect of the C75U mutation, which strongly implicated C75 as the general acid.[15] Structural studies on these crystals revealed that C75 is positioned to protonate the leaving group, 5'O of G1.[3,5]

Furthermore, both pH-rate profiles and pH-dependent Raman crystallography reveal anticooperative interaction between protonated C75 and $Mg^{2+}$, suggesting an important relationship between nucleobase and metal ion catalysis in the reaction mechanism.[11,14]

Metal ions can contribute to RNA catalysis through specific and non-specific modes of action.[16,17] With the HDV ribozyme in particular, a wide range of divalent ions can contribute to catalysis, including most alkaline earth and certain transition metals.[18,19] In addition, the genomic HDV ribozyme is partially active in the absence of divalent metal ions. According to a multichannel framework for the genomic ribozyme, divalent metal ions contribute 3,000-fold to the reaction: ~125-fold to folding and ~25-fold to catalysis.[12]

A recently solved crystal structure of the HDV ribozyme revealed that the catalytic metal ion is positioned, in part, by water-mediated interactions with a rare, but absolutely conserved reverse G•U wobble (Figure 6.1B).[5] However, the cleavage site was partially disordered, which necessitated modeling in this portion of the structure.  It is therefore critically important to test the metal ion interactions with mechanistic biochemical studies. Recent characterization of a mutant that replaces the reverse G•U wobble with a reverse A•C wobble functionally links the metal ion observed in the crystal structure with the catalytic metal ion. This metal ion is positioned with the help of the *pro*-$R_P$ oxygen of the scissile phosphate and the phosphate group of U23. These interactions allow the metal to coordinate the nucleophile in the reaction, the 2′O of U-1. Thus, the metal ion could potentially play a dual role in catalysis: a Lewis acid to facilitate deprotonation of the nucleophile and a Bronsted base to accept the proton from nucleophile.

Here, we characterize the network of interactions involving the *pro*-$R_P$ oxygen of the scissile phosphate. A previous analysis of sulfur substitutions at the scissile phosphate of the genomic HDV ribozyme suggested potential functional importance of the *pro*-$R_P$ oxygen, as cleavage reactions in the presence of an $R_P$ sulfur substitution did not proceed to completion.[20] However, those studies did not reveal a thio effect nor a metal ion rescue by $Mn^{2+}$, the only thiophilic metal ion used to attempt rescue.[20,21] Those studies may have been hampered by the strong propensity of the genomic HDV ribozyme to misfold.[22] In contrast to these results, the recent crystal structure of the HDV ribozyme did implicate the *pro*-$R_P$ oxygen in a direct interaction with the putative catalytic metal ion (Figure 6.1B).[5] The ribozyme construct used to solve this recent structure is a variant of the G11C/U27D ribozyme developed in our labs, which is especially fast reacting and folds homogeneously into its native structre.[23] Extensive characterization of this variant demonstrates that it has a rate-limiting step of bond cleavage, as supported by solvent isotope effects, proton inventories, theoretical studies, and a p$K_a$ shifted towards neutrality that is in agreement with Raman crystallographic studies.[11,24–26] We therefore used this variant to examine the thio effect at the scissile phosphate and to fully characterize the metal ion rescue with a wider array of thiophilic metal ions.

## Materials and Methods

*RNA Oligonucleotides and Constructs.*

The RNA was prepared in two pieces. The first strand, formally the enzyme, spans 62 nucleotides (Figure 6.1A) and was prepared by T7 transcription under standard conditions as previously described.[5] The sequence was cloned into pUC19 and confirmed using dideoxy sequencing. The same procedure was adopted for the transcription of the G25A•U20C double mutant ribozyme, where the reverse G25•U20 wobble was mutated to a reverse A•C wobble.[6] For the C75D ribozyme, the enzyme strand was transcribed using T7 RNA polymerase from a hemi-duplex consisting of the top strand with the sequence 5' TAA TAC GAC TCA CTA TAG, and the bottom strand containing the sequence 5'GGT CCC ATT CCC ATT ACC TTT CGG AAT GTT GCC CAG CTT GCG CCG CGA GGA GGC TGG GAC CTA TAG TGA GTC GTA TTA.

Following transcription, the enzyme pieces were purified on an 8% polyacrylamide/ 7 M urea gel. Bands were visualized by UV shadowing and excised from the gel. RNA was eluted overnight at 4 °C into 10 mM Tris-HCl, 0.1 mM EDTA, and 250 mM NaCl, precipitated with 3 volumes of ethanol, and stored at -20 ºC.

The second strand, formally the substrate, spans the 5'-portion of P1 and the upstream region and contains the cleavage site. This strand was prepared by solid-phase synthesis (Dharmacon Inc.). The sequence of the substrate is 5'UAU*GGCUUGCA, where '*' indicates the site of cleavage and phosphorothioate substitution. A synthetic oligonucleotide containing the single phosphorothioate substitution was separated into the $S_P$ and $R_P$ diastereomers by C18 reverse-phase HPLC (solvent A, 0.1 M $NH_4OAc$; solvent B, 50% A and 50% $CH_3CN$) starting with solvent A and employing a linear gradient (0 to 15%) of solvent B. The $R_P$ and $S_P$ isomers were well-resolved, as confirmed by re-injecting the purified material into the HPLC (Figure E.1). The $R_P$ isomer elutes first from the column, as described elsewhere.[20] In addition, the identities of the $R_P$ and $S_P$ isomers in this study were confirmed with snake venom phosphodiesterase (data not shown). Purified fractions were dried, redissolved in sterile water, and stored at -20 °C. The separated diastereomers were reanalyzed by HPLC regularly (on an approximately bimonthly basis), which confirmed no contamination, sulfur wash, or degradation. Purified oligonucleotides were treated with T4 polynucleotide kinase (New England Biolabs) and g-$^{32}$P ATP to add a radiolabeled phosphate to their 5′-ends prior to kinetic analysis.

*Kinetics Assays and Data Analysis.*

All reactions were single-turnover and performed at 37 ºC. To avoid precipitation of $Cd^{2+}$ as $Cd(OH)_2$, a 100 mM $CdCl_2$ stock solution was prepared in NaHEPES (12.5 mM, pH 5.6), and subsequently diluted. The ribozyme (3 mM) was annealed to trace amounts of end-labeled substrate (<2 nM) at 90 ºC for 2 min and then cooled at room temperature for 10 min followed by the addition of either NaMES (25 mM, pH 5.6) or NaHEPES (25 mM, pH 8.0). To confirm that all of the substrate is bound, we performed the reactions with an increasing enzyme concentration of up to 16 $\mu$M under the above conditions. Increasing the enzyme concentration above 3 mM did not affect the observed reaction rate, confirming that 3 mM of ribozyme was sufficient to bind all the substrate.

The reaction was initiated by adding a pre-equilibrated metal ion mix that led to a final concentration of 10 or 50 mM $Mg^{2+}$ with 0 to 10 mM $Cd^{2+}$. At regular intervals, 3 µL aliquots were removed and quenched by adding an equal volume of 96% formamide/20 mM EDTA and placing immediately on powdered dry ice. The efficacy of the quench was confirmed by adding the quench first and observing the absence of reaction. The quench efficacy was tested under the following reaction conditions: 10 mM $Mg^{2+}$ (pH 5.6 and 8.0), 10 mM $Mg^{2+}$ with 10 mM $Cd^{2+}$ (pH 5.6), and 10 mM $Mg^{2+}$ with 2 mM $Cd^{2+}$ (pH 8.0). The AC double mutant reactions were similarly performed with 5 mM final concentration of ribozyme to provide direct comparison with another study.[6]

For the C75D reaction, trace amounts of the labeled substrate (<2 nM) and 3 mM ribozyme were heated at 90 ºC for 2 min, allowed to cool to room temperature, and then incubated at 37 ºC for 10 min in 5 mM Tris-HCl (pH 7.5) and 0.1 M NaCl.[10] The renatured RNA was then diluted in a reaction buffer to a final concentration of 25 mM $MgCl_2$, 0.5 mM spermidine, and 0.4 M imidazole (pH 7.0), as per earlier studies on C75D.[10] Aliquots were removed at regular intervals and quenched with an equal volume of a high EDTA quench of 90% formamide/50 mM EDTA, and placed immediately on powdered dry ice. Control reactions on the C75D ribozyme carried out in the absence of imidazole showed no reaction. Reactions in the absence of divalent metal ions were performed as previously described.[25]

For all reactions, the 5'-end labeled cleavage product was separated on 16% acrylamide/ 7 M urea gels, and the fraction of the substrate cleaved versus time was quantitated using a Typhoon PhosphoroImager (Molecular Dynamics). Data from the oxo and $S_P$ substrate were fit to the single-exponential equation:

$$f_{\mathrm{cleaved}} = A + Be^{-k_{\mathrm{obs}}t} \tag{1}$$

where $f_{\mathrm{cleaved}}$ is the fraction of substrate cleaved, $k_{\mathrm{obs}}$ is the observed first-order rate constant, A is the fraction of the ribozyme cleaved at completion, –B is the amplitude of the observable phase, 1–A is the unreactive fraction, and A+B is the amplitude of the burst fraction, which was ~0 herein.

Data from the $R_P$ substrate were fitted to a double exponential equation,

$$f_{cleaved} = A + Be^{-k_1 t} + Ce^{-k_2 t} \tag{2}$$

where $k_1$ and $k_2$ are the observed first-order rate constants for the fast and slow phases, respectively, A is the fraction of the ribozyme cleaved at completion, –B and –C are the amplitudes of the observable phases, 1–A is the unreactive fraction, and A+B+C is the amplitude of the burst fraction which was ~0 herein. Error is reported as the standard deviation of two or more repeats.

For slower reactions, time courses were followed for up to several days; control experiments in which the enzyme was omitted showed that the cleaved product obtained was not from substrate degradation. In particular, at time points of 60 h in 1 M NaCl/100 mM EDTA and 20 h in C75D/25 mM $Mg^{2+}$, the oxo and $R_P$ substrates showed only ~5% and ~2% degradation, respectively. In contrast the actual reaction typically had ~15-20 % reaction at these time points.

*Calculations of Thio Effect and Metal Ion Rescue.*

Substitution of the non-bridging oxygen by sulfur alters the observed ribozyme kinetics by a factor noted as the 'thio effect', which is defined as follows:[27, 28]

$$\text{Thio effect} = \left(\frac{k_O}{k_S}\right)^{Mg^{2+}} \tag{3}$$

where '$k_O$' is $k_{obs}$ for the oxo-substrate and '$k_S$' is $k_{obs}$ for the phosphorothioate substrate. Because the thio effect is defined in the absence of thiophilic metal ion but presence of $Mg^{2+}$, we add a superscript $Mg^{2+}$ to the above equation. A normal thio effect is defined as a thio effect greater than unity, while an inverse thio effect is defined as one less than unity.

Rescue of a thio effect greater than unity by a thiophilic metal ion (denoted as '$M^{2+}$') is defined as 'metal ion rescue'. In general, addition of a thiophilic metal ion, typically in a background of $Mg^{2+}$ to aid native ribozyme folding, may affect the rate of the wild-type reaction too, for example by competing with $Mg^{2+}$ for an ion site or through an allosteric effect. To account for this possibility, metal ion rescue is generally defined as the ratio of $k_S$ in the presence of $M^{2+}$ to $k_S$ in the presence of $Mg^{2+}$-only, normalized to this ratio for the all-oxo substrate ($k_O$) (eq 4):[29]

$$\text{Metal Ion Rescue} = \left(\frac{k_S{}^{M^{2+}}}{k_S{}^{Mg^{2+}}}\right) \Big/ \left(\frac{k_O{}^{M^{2+}}}{k_O{}^{Mg^{2+}}}\right) = \left(\frac{k_O}{k_S}\right)^{Mg^{2+}} \Big/ \left(\frac{k_O}{k_S}\right)^{M^{2+}} \qquad (4)$$

A mathematically equivalent ratio of ratios is provided on the right-hand end of eq 4, which considers the thio effect, $\left(\frac{k_O}{k_S}\right)^{Mg^{2+}}$, divided by the ameliorated thio effect in the presence of $M^{2+}$, $\left(\frac{k_O}{k_S}\right)^{M^{2+}}$. Values greater than 5 for eq 4 are generally considered a metal ion rescue.[28] Error was propagated from relative error by standard means.[30] Because eq 4 involves four rate constants, error was larger, but still tolerable, for metal ion rescue.

*Calculation of $Cd^{2+}$ Inhibition Constant.*

The competition between $Cd^{2+}$ and $Mg^{2+}$ was evaluated by a Dixon plot[31] in which the reciprocal of $k_{obs}$ was plotted versus concentration of $Cd^{2+}$, parametric in $Mg^{2+}$ concentration. Intersection of these lines in the second quadrant is consistent with competitive interaction between the two ions, and the negative of the abscissa corresponding to their point of intersection provides a $K_i$ for $Cd^{2+}$ binding.

*Quantum Mechanical/ Molecular Mechanical (QM/MM) Calculations.*

The QM/MM calculations in this study are based on the pre-cleaved HDV ribozyme coordinates (PDB ID 3NKB).[5] Hydrogen atoms were added to the entire ribozyme using Accelrys Discover Studio Visualizer 2.0. Residues C41 and C75 were kept protonated at N3 during all calculations to provide structural stability and prepare C75 in its general acid functional form.[14,32] All 11 crystallographically resolved $Mg^{2+}$ ions were included, and the ribozyme was immersed in rigid TIP3P water in a periodically replicated orthorhombic box. Counterions were added to neutralize the system, and 0.15 M NaCl was added to replicate physiological ionic strength.

Starting configurations for the QM/MM calculations were obtained after molecular dynamics (MD) equilibration of solvent and ions, the protocol of which has been detailed elsewhere.[33] The MD simulations were performed with the Desmond program[34] using the AMBER99 force field.[35] The QM/MM calculations were performed using the Qsite program.[36] The QM region was treated with density functional theory at the B3LYP/6-31G**[37] level of theory using an ultrafine grid. The LACVP*[38,39] pseudopotential was used for the $Cd^{2+}$ metal ion, and the hydrogen-capping method was used for the residues at the QM region boundary to define the QM region. The MM region was described using the OPLS2005[40] force field, and the residue-based nonbonded cutoff was set to 100 Å. All atoms lying outside a radius of 20 Å from the scissile phosphate were held fixed throughout the QM/MM calculations. The QM region considered in this study is similar to that used in our previous study.[26] Each non-bridging oxygen of the scissile phosphate was replaced in-silico by sulfur to obtain the corresponding $R_P$ or $S_P$ phosphorothioate diastereomer.

## Results

*Phosphorothioate Substitution Reveals a Thio Effect for the $R_P$ Substrate.*

We began by examining the effects of sulfur substitution at each of the diastereotopic non-bridging oxygens of the scissile phosphate. These experiments were conducted with the fast-reacting G11C variant of the genomic HDV ribozyme, which has chemistry as the rate-limiting step.[13,19,20,23,25] Single-turnover kinetic profiles with the oxo substrate in the presence of 10 mM $Mg^{2+}$ were monophasic and went nearly to completion (Figure 6.2A). The rate-pH profile showed an increase in rate with pH followed by a plateau region (Figure 6.2B), similar to results previously described.[11] The $pK_a$ for the reaction with the oxo substrate was 5.5 ± 0.1, which is in reasonable agreement with the previously reported value of 6.1 ± 0.1[2].[11] The $S_P$ substrate reacted with a nearly identical monophasic kinetic profile, rate-pH profile, and $pK_a$ value (Figures 6.2A, B). This suggests that the $S_P$ oxygen does not contribute significantly to the reaction, consistent with the crystal structure.[20]

The $R_P$ substrate, in contrast, reacted with a biphasic kinetic profile, with each phase showing a thio effect (Figure 6.2A). The rate for the fast phase of the $R_P$ substrate increased with pH followed by a plateau, and gave a $pK_a$ of 5.1 ± 0.1 (Figure 6.2B). This value is somewhat smaller than that observed with oxo and $S_P$ substrates suggesting that the $R_P$ sulfur substitution results in a smaller $pK_a$ shift for C75. Nonetheless, similarity in the shapes and $pK_a$ values of these three rate-pH profiles suggests that chemistry remains rate limiting for these three substrates. The rate of the slow phase of the $R_P$ substrate, on the other hand, did not show a dependence on pH (Figure 6.2C). This observation suggests that the rate of chemistry has

---

[2] The slightly lower value may be due to experiments herein being on a two-piece construct and the referenced $pK_a$ being on a one-piece construct. Recent $pK_a$ values on the two-piece construct at 5 mM $Mg^{2+}$ were lower than the one-piece by 0.3 $pK_a$ units (Chen *et al.*, *Biochemistry*, **2013**, *52*, 557-567).

slowed to the point where another step in the mechanism has become rate-limiting or that metal no longer participates in the reaction as a general base (see Discussion).

Next we quantitated the thio effects of the $S_P$ and $R_P$ substrates. The thio effect is the ratio of $k_{obs}$ for the oxo and phosphorothioate substrate in the presence of $Mg^{2+}$ (eq 3). The cleavage rate of the $S_P$ substrate was the same as the oxo within error (Table 6.1), and so has no thio effect. The $R_P$ substrate, on the other hand, displayed significant thio effects. At pH 5.6, the fast and slow phases had thio effects of ~3- and 1000-fold, respectively, while at pH 8.0 they had thio effects of ~5- and 800-fold, respectively (Table 6.1, Figure 6.3A, B). (These two pH conditions were chosen in order to explore effects with and without protonation of C75.) In sum, sulfur substitution at the non-bridging oxygens of the scissile phosphate has a stereospecific effect, interfering with the reactivity of only the $R_P$ substrate.

### *$Cd^{2+}$ Rescues the Thio Effect of the $R_P$ Substrate.*

Loss of activity upon sulfur substitution at the *pro*-$R_P$ position could be due to disruption of an interaction between a metal ion and this ligand. To test this possibility, thiophilic metal ions were added to the reaction mixture and activity was re-measured, as previously described.[27,29,41–43] We studied effects of two thiophilic metal ions on the reaction kinetics, namely $Mn^{2+}$ and $Cd^{2+}$. Addition of $Mn^{2+}$ did not promote metal ion rescue. As shown in Table 6.1 and Figure E.2, addition of $Mn^{2+}$ to 2 mM, in a background of 10 mM $Mg^{2+}$ to promote native RNA folding, had no significant effect on $k_{obs}$ for the oxo and $S_P$ substrates, and gave modest *decreases* in the rate of the $R_P$ fast and slow phases, by 2- and 4-fold, respectively. This led to 1.7- and 3.3-fold inverse metal ion rescues for $R_P$ fast and slow phases, which provides no support for a direct metal-*pro*-$R_P$ oxygen interaction. Absence of rescue by $Mn^{2+}$ is consistent with the earlier studies of Nishikawa and co-workers[3] on the genomic HDV ribozyme.[20]

Absence of rescue by $Mn^{2+}$ does not, however, constitute absence of direct metal ion-nonbridging oxygen interaction. There have been several reports wherein $Mn^{2+}$ is unable to rescue loss in reactivity upon sulfur substitution. For example, in the *Tetrahymena* group I intron the double sulfur substitution of *pro*-$S_P$ oxygen and 3'O could not be rescued by $Mn^{2+}$, but $Cd^{2+}$ was able to rescue the activity.[41] We therefore measured $k_{obs}$ in the presence of $Cd^{2+}$, as the concentration of $Cd^{2+}$ was increased from 0.01 mM to 10 mM for pH 5.6, or from 0.01 mM to 0.5 mM for pH 8.0, all in the background of 10 mM $Mg^{2+}$ to promote native RNA folding (Figure 6.3).[4] Increasing the concentration of $Cd^{2+}$ *decreased* $k_{obs}$ for both the oxo and $S_P$ substrates, and to similar extents at both low and high pH (Figure 6.3A, B). This $Cd^{2+}$-induced inhibition of the reactivity of the oxo and $S_P$ substrates is further revealed in plots of $(k_O/k_S)^{Cd2+}$ versus $Cd^{2+}$ concentration for $S_P$ at low and high pH, which are essentially flat (Figure 6.3C, D). In contrast to the $S_P$ substrate, $Cd^{2+}$ had essentially no effect on $k_{obs}$ for the $R_P$ substrate fast or slow phases (Figure 6.3A, B). This suggests that $Cd^{2+}$ is enhancing the reactivity of the Rp substrate, thereby canceling the inhibitory effect of $Cd^{2+}$ that is observed with the oxo and Sp substrates. Thus, plots of $(k_O/k_S)^{Cd2+}$ versus $Cd^{2+}$ concentration for the $R_P$ fast and slow phases decrease with $Cd^{2+}$ concentration at low and high pH (Figure 6.3C, D). This trend indicates that the impediment to ribozyme catalysis caused by the $R_P$ sulfur substitution is assuaged to some extent by the thiophilic metal ion.

---

[3] These studies did not see the decrease in $R_P$ rate with $Mn^{2+}$, but they used more $Mn^{2+}$, without $Mg^{2+}$, and employed a slower reacting ribozyme. Nonetheless, there is overall agreement in the inability of $Mn^{2+}$ to rescue the reaction.

[4] $K_{Sp}(Cd(OH)_2) = 5.9 * 10^{-15}$ at 37°C; At pH 8.0; $[OH^-] \sim 2.5*10^{-6}$ M, using $Kw \sim 2.5*10^{-14}$ at 37°C. For $Cd(OH)_2$ to be soluble at pH 8.0, $Q = [Cd^{2+}][OH^-]^2 < 5.9*10^{-15}$. Thus $[Cd^{2+}]$ was kept less than 1 mM at pH 8.0

Next we consider the values for metal ion rescue upon addition of $Cd^{2+}$. Metal ion rescue is calculated as $(k_O/k_S)^{Mg2+}$ over $(k_O/k_S)^{Cd2+}$ to correct for background effects of inhibition on the oxo substrate (eq 4). No rescue is observed for the $S_P$ substrate with increasing concentrations of $Cd^{2+}$ at either pH 5.6 or 8.0; indeed, metal ion rescue values stay at or slightly below unity (Figure 6.4). Absence of rescue by $Cd^{2+}$ supports absence of interaction between a metal ion and the *pro*-$S_P$ oxygen. In contrast, metal ion rescue occurs for the $R_P$ substrate at both low and high pH and increases with $Cd^{2+}$ concentration (Figure 6.4). For the slow phase, maximal $Cd^{2+}$ rescues values of 330±100 (pH 5.6) and 31±9 (pH 8.0) are observed, while for the fast phase maximal $Cd^{2+}$ rescues of 7.5±4 (pH 5.6) and 10±3 (pH 8.0) are observed (Table 6.1). These values are significant when compared to typical literature values for rescue by $Cd^{2+}$ of 10-100-fold.[28] Overall, observation of significant metal ion rescue for the $R_P$ fast and slow phases provides support for a direct metal-*pro*-$R_P$ oxygen interaction as suggested by the recent crystal structure.

The above results suggest that $Cd^{2+}$ has at least two binding sites, one that is inhibitory, as revealed in decreasing $k_{obs}$ versus $Cd^{2+}$ concentration plots for the oxo and $S_P$ substrates (Figure 6.3A, B, blue and red bars), and one that is stimulatory, as revealed by lack of inhibition in the presence of the $R_P$ substrate (Figure 6.3A, B, dark and light green bars). (Note that if there were just a single site, then $k_{obs}$ for the $R_P$ substrate should increase with $Cd^{2+}$.) Presence of stimulatory and inhibitory sites for $Cd^{2+}$ in the genomic HDV ribozyme is consistent with reports that $Cd^{2+}$ can stimulate or inhibit reactivity depending on the HDV ribozyme construct.[18] The location of the inhibitory site is unclear. It could be any of the 11 $Mg^{2+}$-binding sites resolved in the crystal structure, or it could be a site unique to $Cd^{2+}$-binding. In contrast, the stimulatory

$Cd^{2+}$-binding site must be at the active site where the *pro*-$R_P$ sulfur is located as it is uniquely dependent on the sulfur substitution at this position.

We next investigated the nature of the loss of activity in the oxo and $S_P$ substrates upon $Cd^{2+}$ addition. Values of $k_{obs}$ were determined as a function of $Cd^{2+}$ concentration at various $Mg^{2+}$ concentrations. This was done at pH 5.6 because a wider range of $Cd^{2+}$ concentrations could be explored without precipitation of $Cd(OH)_2$, and because C75 is protonated in a significant fraction of molecules under these conditions, which leads the active site to be more stable and less dynamic.[44] Plotting these data in the form of a Dixon plot revealed lines that intersect in the second quadrant (Figure E.3). This graphical form is consistent with inhibition by $Cd^{2+}$ being competitive with respect to $Mg^{2+}$. Moreover, a value of 137±73 mM for the inhibition constant of $Cd^{2+}$, $K_{i,}$, is obtained from these plots, which is significantly tighter than the apparent $K_d$ of 16 mM observed for $Mg^{2+}$ at pH 5.5 in the wild-type ribozyme.[11] Observation that the $K_i$ is much tighter than the apparent $K_d$ is consistent with our recent study on the double mutant version of the ribozyme, wherein the catalytic metal ion bound tighter than the apparent $K_d$.[6] It thus appears that the inhibitory $Cd^{2+}$ competes with the $Mg^{2+}$ ions.[5]

Next, we investigated the $Cd^{2+}$-concentration-dependence of rescue of the $R_P$ substituted substrate. Interestingly, metal ion rescue by $Cd^{2+}$ at pH 5.6 begins to assume values significantly greater than unity when the concentration of $Cd^{2+}$ is between 1 and 2 mM (see rounding point Figure 6.4A). This value is smaller than the apparent $K_d$ of 16 mM for $Mg^{2+}$ binding to the oxo substrate at pH 5.5,[11] consistent this ion is binding at the active site as described in the preceding paragraph. We attempted to achieve saturation of the ribozyme with $Cd^{2+}$, however at $Cd^{2+}$

---

[5] Observation of inhibition led to the notion that we could simplify kinetics by interfering with binding of the inhibitory metal ion. Our previous studies suggested that this ion might be at the cleavage site G•U wobble. However, changing this wobble to an AU or GC Watson-Crick base pair did not eliminate the inhibitory effect of $Cd^{2+}$ (data not shown).

concentrations higher than 10 mM degradation of the RNA became problematic over the timescale of ribozyme cleavage (data not shown).

At pH 8.0 the metal ion rescue assumes values greater than unity when the concentration of $Cd^{2+}$ is between 0.2 and 0.6 mM (see rounding point in Figure 6.4B). We are unable to saturate the reaction with $Cd^{2+}$ ion at pH 8.0 owing to formation of cadmium hydroxide.**Error! Reference source not found.** Nonetheless, rescue at lower $Cd^{2+}$ concentrations in pH 8.0 experiments strongly suggests that the binding of $Cd^{2+}$ is tighter at higher pH; these trends parallel the observed pH dependence of $Mg^{2+}$ affinity for the unsubstituted ribozyme.[11] Observation that the $K_d$ for $Cd^{2+}$, like that of the active site $Mg^{2+}$,[11,14] decreases with increasing pH further supports that this $Cd^{2+}$ ion binding site is within the active site.

### *Role of the pro-$R_P$ Oxygen in Positioning C75.*

The non-bridging *pro-$R_P$* oxygen at the scissile phosphate is positioned to interact with both the catalytic metal ion and the exocyclic amine of C75. According to the crystal structure,[5] the metal ion is only 2.6 Å from the *pro-$R_P$* oxygen but 3.6 Å from the amine of C75, although MD studies in the presence of $Mg^{2+}$ and protonated C75 indicate that this distance decreases to ~3.0 Å on average.[5,44] The above phosphorothioate substitutions and subsequent metal ion rescues monitor interaction of the non-bridging *pro-$R_P$* oxygen with the catalytic metal ion but do not probe its interaction with C75. In an effort to evaluate interaction between the *pro-$R_P$* oxygen and C75, we measured thio effects in three variant backgrounds, summarized in Table 6.2: 1) In 1M NaCl/ 100 mM EDTA, which removes all divalent ions, 2) In a G25•U20 to A25•C20 double mutant that interferes with active site $Mg^{2+}$ binding, and 3) In a C75D mutant in which activity was rescued by imidazole. The latter two sets of experiments were conducted in the presence of $Mg^{2+}$ ion.

We first explored whether the thio effect observed with the $R_P$ substrate was still observable when the reaction was conducted in the absence of divalent metal ions. There are no divalent $Mg^{2+}$ ions in the active site of the ribozyme under these conditions, however the thio modification might still be predicted to interfere with the hydrogen bond between the exocyclic amine of C75 and the scissile phosphate. The ribozyme is known to react, albeit with a reduced rate, in the absence of divalent metal ions when high concentrations of NaCl are present.[45] We employed 1 M NaCl along with 100 mM EDTA to ensure no divalent metal ion contamination. (High concentrations of EDTA are especially important at low pH where EDTA binds contaminating trace divalent ions less tightly.) The $R_P$ substrate was found to react 2.5-fold more slowly than the oxo substrate, providing a small normal thio effect (Table 6.2). The $S_P$ substrate, on the other hand, reacted 3.4-fold *faster* than the oxo substrate, providing an inverse thio effect. The nature of these effects is possibly electrostatic in nature (see Discussion). Overall, these data provide no strong support for the *pro*-$R_P$ oxygen interacting directly with C75, or any other functionality.

Next, we investigated the effect of phosphorothioate substitution on reaction with the A25•C20 double mutant. The reverse G•U wobble found at the active site contributes to the electronegative pocket for the divalent metal ion.[5,6,33] Recently, we showed that mutating it to an isosteric A•C reverse wobble, which replaces the keto oxygen of guanosine with the amine of the adenosine, gives rise to a bell-shaped rate-pH profile with p$K_a$ values of 5.4±0.3 and 6.2±0.2, which suggests that any contribution of the catalytic metal ion to the reaction occurs at low pH only. When we analyzed the reaction of this double mutant with the thio substrates, no thio effect was observed for the $R_P$ substrate at neutral pH in 10 or 50 mM $Mg^{2+}$ (thio effects of 1.5±0.3 and 1.4±0.3, respectively, Table 6.2), consistent with absence of strong metal binding in the active site of the A25•C20 double mutant at biological pH as previously concluded.[6] Importantly, these data further provide no support for the *pro*-$R_P$ oxygen interacting directly with C75, as a thio effect should have been observed in the absence of metal ion if this were the case.

At the lower pH of 5.6 and 10 mM $Mg^{2+}$, no thio effect was observed for $R_P$ substrate in the A•C double mutant (Table 6.2), again providing no support for the *pro*-$R_P$ oxygen interacting with C75. However, at pH 5.6 and the elevated $Mg^{2+}$ concentration of 50 mM, a 4.2±0.8-fold thio effect was observed for the $R_P$ substrate (Table 6.2). This effect was stereospecific in that it was absent with the $S_P$ substrate, which gave just a 1.3±0.03-fold effect (Table 6.2). These data suggest that only at the lowest pH, where C75 is protonated and the active site is aligned[44] and at the highest $Mg^{2+}$ concentrations can the A•C double mutant support the direct metal ion interaction at the active site. Consistent with this notion, we observed a partial rescue of this thio effect by $Cd^{2+}$ (Table 6.2, footnote). These data support direct *pro*-$R_P$ oxygen-$Mg^{2+}$ interaction at the active site and are also consistent with the upward arm observed in the rate-pH profile for the A•C double mutant between pH 5 and 5.5 at 50 mM $Mg^{2+}$ in our recent study.[6]

Lastly, we measure the effect of phosphorothioate substitution in the C75D background. These experiments were done in the presence of 25 mM $Mg^{2+}$ and 0.4 M imidazole, which are optimal for folding and C75 rescue, respectively, under these conditions.[10] First, we note that all three substrates were cleaved in the presence of the C75D enzyme. A 3.2±0.08-fold thio effect was observed for the $R_P$ substrate that was absent with the $S_P$ substrate, which exhibited a 0.7±0.03-fold effect (Table 6.2). These data support the importance of interaction between the *pro*-$R_P$ oxygen and the catalytic metal ion, even in the absence of the amine of C75. This indicates that any interaction between the *pro-*$R_P$ oxygen and C75 is relatively unimportant for *pro-*$R_P$ oxygen function with metal ion. In sum, the thio effects observed in 1 M NaCl/100 mM EDTA, the A•C double mutant, and C75D support the direct *pro-*$R_P$ oxygen-$Mg^{2+}$ interaction but provide no support for direct interaction between the *pro-*$R_P$ oxygen and the N4 of C75 or any other functionality in the reaction.

*Insights from QM/MM Study on the Thio Effect at the Scissile Phosphate.*

In order to investigate the effect of phosphorothioate substitution at the scissile phosphate theoretically, we performed geometry optimizations on the phosphorothioate-substituted $R_P$ and $S_P$ substrates in the presence of the catalytic $Mg^{2+}$ ion. The optimized reactant structure of the $R_P$ substrate exhibited significant active site distortion (Figure 6.5A) compared to the optimized reactant structure of the oxo species. The most prominent effect was the increase in the C75(N3) and G1(O5') distance by ~0.7 Å. These two atoms serve as the donor and acceptor for the proton transfer reaction that is a key step in the self-cleavage reaction of the ribozyme.[26] Moreover, the distance between the catalytic ion and the *pro-*$R_P$ sulfur increased by ~0.5 Å compared to that distance in the oxo species. These two distortions are consistent with the significantly slower rates of the $R_P$ substrate observed experimentally in the presence of $Mg^{2+}$. In contrast, no such distortions were observed in the optimized reactant state of the $S_P$ substrate (Figure 6.5B).

For the $R_P$ substrate, we identified a transition state (TS) using the quasi-synchronous transit method and performed frequency calculations to confirm that the resulting stationary point has a single imaginary frequency (Figure 6.6). This transition state has a similar structure to that of the reactant structure, except for the position of the proton (H3) on N3(C75), which has partially transferred to O5'(G1). Starting from this TS structure, we calculated the mass-weighted minimum energy path (MEP) back to the reactant and product structures following the intrinsic reaction coordinate (IRC) (Figure 6.6A, B). When we followed this path in the product direction, we obtained a minimum corresponding to an intermediate in which the proton has fully transferred to G1(O5') (Figure 6.6C). The calculated energy barriers along this pathway are 37 and 5 kcal/mol in the forward and reverse directions, respectively. While these energies cannot be used to calculate reaction rate constants because they do not include entropic effects and only represent one possible reaction pathway, they can still be used to gain qualitative insight into the mechanism. This reaction path suggests that the reaction mechanism is sequential, with the proton transferring prior to phosphoros-oxygen bond formation and cleavage. In contrast, prior analogous QM/MM calculations on the oxo species suggested a concerted mechanism with a phosphorane-like transition state when a divalent ion is bound at the catalytic site.[26] Moreover, the calculated energy barrier from the reactant to the TS is 12 kcal/mol higher for the $R_P$ substrate than for the oxo substrate, although the energetics are not reliable due to incomplete sampling of conformational space.[26] We emphasize that the search for transition states was not fully comprehensive, and other transition states with different character may also exist.

Although this sequential reaction pathway with initial proton transfer does not seem probable based on the high energy barrier, we propose a possible explanation as to why it may be followed by the $R_P$ substrate. The *pro*-$R_P$ oxygen of the scissile phosphate interacts with both the catalytic metal ion and the exocyclic amine of C75. Replacing this oxygen by sulfur weakens both of these interactions. The weakening of the *pro*-$R_P$-N4(C75) interaction causes C75 to move away from G1, thus increasing the proton donor-acceptor distance from 3.01 Å in WT to 3.72 Å in the $R_P$ substrate for the optimized reactant state. Note that the proton donor-acceptor distance decreases to 2.75 Å at the transition state of the first step depicted in Figure 6.6 for the $R_P$ substrate to enable proton transfer. An explanation as to why the proton may transfer in the first step of a sequential mechanism for the $R_P$ substrate is that the weakening of the *pro*-$R_P$-metal ion interaction may strengthen the metal ion-O2'(U-1) interaction, thereby hindering the attack of the O2' nucleophile on the scissile phosphate.

### *Probing the Role of $Cd^{2+}$ in Metal Ion Rescue using QM/MM studies.*

To understand the role of $Cd^{2+}$ ion in rescuing the catalytic reaction, we performed QM/MM calculations after replacing the catalytic $Mg^{2+}$ ion by a $Cd^{2+}$ ion. For both the oxo and $S_P$ substrates in the presence of $Cd^{2+}$, we were unable to identify a transition state or intermediate associated with either a concerted or a sequential mechanism. Several constrained optimization techniques were utilized to search for both a phosphorane intermediate and an intermediate corresponding to initial proton transfer. Although these searches were not comprehensive, the calculations suggest that the oxo and $S_P$ substrate species behave similarly (i.e., with no stable intermediates) when $Cd^{2+}$ is bound at the active site.

For the $R_P$ substrate in the presence of $Cd^{2+}$, we tried to locate an intermediate similar to the one observed for the $R_P$ substrate with a $Mg^{2+}$ ion at the catalytic site (i.e., with the proton transferred to the O5'(G1)) using several constrained optimization techniques but were unsuccessful. However, we were able to locate a phosphorane intermediate as a minimum on the potential energy surface (Figure 6.7). Formation of a phosphorane intermediate species has been shown to be involved in similar ribozyme catalyzed reactions.[46] Moreover, in our previous study,[26] we reported a phosphorane intermediate for the oxo case with a monovalent ion ( $Li^+$, $Na^+$, $K^+$, $Cs^+$) at the catalytic site, in contrast to the concerted reaction pathway observed with a divalent ion at the catalytic site ($Mg^{2+}$, $Ca^{2+}$). The absence of an intermediate arising from initial proton transfer for the $R_P$ substrate in the presence of $Cd^{2+}$ can be attributed to the strengthening of the $Cd^{2+}$-*pro*-$R_P$ sulfur interaction and the weakening of the $Cd^{2+}$-O2' (U-1) interaction, thereby allowing the O2' nucleophile to attack the scissile phosphate to form the phosphorane intermediate structure.

For the $R_P$ substrate, the type of intermediate observed depends on the particular metal ion at the active site. In terms of energetics, the intermediate corresponding to proton transfer identified in the presence of $Mg^{2+}$ was 33 kcal/mol higher than the optimized reactant, whereas the phosphorane intermediate identified in the presence of $Cd^{2+}$ was only 10 kcal/mol higher than the optimized reactant. Although these values are not quantitatively meaningful, the large difference is consistent with the experimental observation that the overall reaction of the $R_P$ substrate is more facile in the presence of $Cd^{2+}$.

## DISCUSSION

The importance of nucleobase and metal ions in ribozyme mechanism has been established, but the extent of cooperation between them remains unclear. A recent crystal structure of the HDV ribozyme indicated interaction between the catalytic metal ion and the *pro*-$R_P$ oxygen of the scissile phosphate,[5] but previous functional studies did not provide evidence in support of this interaction.[20] Herein, we utilized experimental and theoretical methods to assess the functional importance of the *pro*-$R_P$ oxygen at the active site. The $R_P$ sulfur substitution at the scissile phosphate resulted in biphasic kinetics with a thio effect as large as ~1000 that was rescuable with the thiophilic metal ion $Cd^{2+}$. These data, in contrast to previously reported results, implicate the *pro*-$R_P$ oxygen of the scissile phosphate as a catalytically important metal ion ligand. Furthermore, QM/MM calculations support the importance of the *pro-R*$_P$ oxygen and help explain the experimentally observed thio effects for the *R*$_P$ substrate. Interaction between the *pro-R*$_P$ oxygen and C75 was also investigated by experiments conducted in the absence of bound divalent metal ion. These studies provided no normal thio effect and thus no evidence for catalytically important interaction between the exocyclic amine of C75 and the *pro-R*$_P$ oxygen, although the absence of thio effect cannot be equated to an absence of interaction.

*Comparison to Other Ribozyme Phosphorothioate Studies.*

The *pro-R*$_P$ oxygen on the scissile phosphate of the HDV ribozyme has received some study in both the genomic and antigenomic versions of the HDV ribozyme. At present, crystal structures are available only for the genomic ribozyme, in both cleaved and precleavage states.[3,5,47] The previous study of phosphorothioate modifications on the genomic version of the ribozyme was conducted in the presence of 10 mM $Mg^{2+}$ by Nishikawa and co-workers.[20] These authors reported nearly identical behavior for the *S*$_P$ and oxo substrates, similar to our findings (Table 6.1). They also reported that the *R*$_P$ substrate reacted to only ~10-20% completion and that this phase had no thio effect. This is similar, although not identical to our findings: we too found that ~30% *R*$_P$ substrate reacted in a fast phase, although this phase has a modest thio effect of 2.8±1.4 (see below). These differences can perhaps be attributed to our construct being rate-limited by chemistry and therefore faster reacting. We note that those authors did not include time points longer than 30 min; the slow phase in our reaction has $t_{1/2}$'s of ~175 and ~75 min at pH 5.6 and 8.0, respectively (Table 6.1)—long enough that the previous study may have missed this phase in their slower reacting constructs. Moreover, the previous experiments attempted rescue in 10 mM

$Mn^{2+}$—a mildly thiophilic metal ion[27]—and did not find any,[20] an observation similar to that herein in 10 mM $Mg^{2+}$/2 mM $Mn^{2+}$, although we see slight *inverse* metal ion rescues (Table 6.1).

In the antigenomic version of the ribozyme, phosphorothioate substitution of the *pro*-$R_P$ oxygen led to a significant 25-fold thio effect that was partially (4-fold) rescued by either 1 mM $Cd^{2+}$ or 1 mM $Mn^{2+}$, both in the background of 9 mM $Mg^{2+}$.[48] The authors thus concluded that the *pro*-$R_P$ oxygen participates in a metal ion interaction. These data support a direct interaction of the catalytic metal with the *pro*-$R_P$ oxygen in the antigenomic ribozyme, similar to that found herein for the genomic ribozyme. Although no crystal structure currently exists for the antigenomic ribozyme, this result is encouraging as the two ribozyme are closely related and their active site nucleotides are strongly conserved across evolution.[49,50] On the other hand, there are clearly some differences: for example, $Mn^{2+}$ rescues the thio effect in the antigenomic ribozyme but not in the genomic ribozyme.

In the study of phosphorothioate effects in the antigenomic HDV ribozyme, several mutations near the active site were also examined for thio effects and metal ion rescue. These led to reduced thio effects and small or inverse metal ion rescues. These mutants largely reacted slower than the wild-type with the natural oxo substrate. This illustrates that slowly-reacting ribozymes can lead to loss of thio effects, most likely when the rate limiting step is not chemistry. This may account for why Nishikawa and co-workers did not observe a thio effect on the $R_P$ substrate for the genomic ribozyme.[20,48]

Metal ion rescue studies have been employed extensively on the *Tetrahymena* group I ribozyme to investigate the nature and the extent of interactions between its catalytic metal ions and their potential ligands.[27–29,41–43,51] The *Tetrahymena* ribozyme is a metalloenzyme that uses two, or perhaps three, metal ions in its catalytic mechanism. Because it is a well-understood and absolutely divalent metal ion-dependent ribozyme, comparisons between it and the HDV ribozyme are of interest. Thio effects in the *Tetrahymena* ribozyme, which occur for both *pro-*$R_P$ and *pro-*$S_P$ oxygens, range from several-fold to 1,200-fold. These values are similar to the thio effects we observe for the *pro-*$R_P$ oxygen herein: up to 5-fold for the fast phase and 1,000-fold for the slow phase.

The identity of the metal used in metal ion rescue experiments is critical in the *Tetrahymena* ribozyme. Metal ion rescues for the *Tetrahymena* ribozyme were often specific to one particular thiophilic metal ion, such as $Mn^{2+}$, $Zn^{2+}$, or $Cd^{2+}$, and depended on the ion-ligand pair studied. This likely arises because metal-ligand interactions can be exquisitely sensitive to the ligand geometry. While one type of ion may be accommodated within a metal binding site and maintain contact with the ligand being probed, another may not. In the HDV ribozyme, we also observe that metal ion rescues are specific to particular thiophilic metal ions; for instance, $Cd^{2+}$ but not $Mn^{2+}$ rescues in the genomic ribozyme. Also, the magnitude of the rescues in the HDV ribozyme are similar to those in the group I ribozyme, with metal ion rescues for the $R_P$ substrate of 7.5- fold for the fast phase and 330-fold for the slow phase, similar to the 5- to 330-fold found for the group I ribozyme. Overall, observation that thio effects and metal ion rescues for the HDV ribozyme have qualitative and quantitative similarities to the *Tetrahymena* ribozyme supports the notion that the HDV ribozyme is a metalloenzyme.

### *Integration of Theory and Experiments Supports the Role of the Catalytic Metal.*

The main findings of the experiments are that the *pro-*$R_P$ oxygen is important for the reaction mechanism and directly coordinates the catalytic metal ion. This conclusion is supported by stereospecific thio effects and thiophilic metal ion rescue. These experimental results are further supported by various aspects of the calculations. The optimized structure in the presence of the catalytic $Mg^{2+}$ revealed significant active site distortion for the $R_P$ substrate but not the $S_P$ substrate (Figure 6.5). Distortions included increased distance between the general acid and the leaving group, and between the catalytic ion and the *pro*-$R_P$ sulfur. These distortions are consistent with the experimentally observed thio effect for the $R_P$ substrate. In addition, the calculations revealed different mechanisms for the oxo and $R_P$ substrates in the presence of $Mg^{2+}$ as well as the $R_P$ substrate in the presence of $Cd^{2+}$. The mechanism previously identified for the oxo species in the presence of $Mg^{2+}$ is concerted with a phosphorane-like transition state,[26] while the mechanism identified for the $R_P$ substrate in the presence of $Mg^{2+}$ has an intermediate arising from proton transfer with a very high energy barrier (Figure 6.6). The calculations suggested yet another mechanism for the $R_P$ substrate in the presence of $Cd^{2+}$, namely a sequential mechanism with a phosphorane intermediate. For the $R_P$ substrate, the energy of the optimized intermediate relative to the optimized reactant is ~23 kcal/mol lower in the presence of $Cd^{2+}$ than in the presence of $Mg^{2+}$. This observation is consistent with the experimentally observed thiophilic metal ion rescue.

### *The pro-$R_P$ Oxygen is Not Necessary for Positioning C75.*

Experiments conducted in the absence of divalent metal ion provided no support for direct interaction between the *pro*-$R_P$ oxygen at the active site and the N4 of C75 or any other functional group. In particular, experiments in 1 M NaCl/100 mM EDTA showed either a small normal thio effect or a small inverse thio effect. The sulfur in a non-bridging phosphorothioate bears most of the negative charge of the phosphate.[52] Thus one interpretation of these observations is that pointing negative charge away from the negative active site, which does not have a divalent ion to neutralize it, accelerates the rate of the reaction, while pointing it towards the active site decreases the rate. The oxo, with its intermediate charge distribution, gives an intermediate effect. These results support the importance of the $Mg^{2+}$ in balancing the negative charge of the active site. They do not, however, support the *pro-*$R_P$ oxygen interacting directly with C75.

It is interesting to note that an inverse thio effect was observed by Perrotta and Been in both the genomic and antigenomic ribozymes.[45] In the presence of NaCl, the genomic ribozyme gave rise to a 3.3-fold inverse thio effect for the $S_P$ substrate[6], identical to that seen herein (Table 6.2). Curiously, in the presence of $NH_4Cl$, large inverse thio effects of 18- and 34-fold were observed for the genomic and antigenomic ribozymes for the $S_P$ substrate. The origin of this effect remains unclear.

Experiments in the A•C double mutant showed lack of a thio effect in all but the low pH, high metal ion conditions. These observations strongly support the conclusion that the *pro-*$R_P$ oxygen is not necessary for positioning C75. Lastly, experiments with C75D indicated that any interaction between the *pro-*$R_P$ oxygen and C75 is not critical for *pro-*$R_P$-metal ion interaction.

These observations raise the question as to what interactions are important for positioning C75 for proton transfer to the leaving group. In the crystal structure, the *pro-*$R_P$ oxygen of C22 forms a 3.0 Å hydrogen bond with C75N4, and MD studies in the presence of $Mg^{2+}$ and protonated C75 indicate an average distance of ~2.8 Å.[5,44] This distance is somewhat shorter than the distance between C75N4 and the *pro-*$R_P$ oxygen at the scissile phosphate both in the crystal structure and MD, suggesting that the phosphate of C22 might be relatively more important in positioning of C75. In addition, C75 is positioned by a hydrogen bond between the

---

[6] The diastereomers were left as a mixture in this study, but the authors provide evidence that the more reactive portion comes from the Sp substrate.

O5' of G1 and C75N3 found both in the crystal structure and in MD (Figure 6.1B), and C75 forms a strong stacking interaction with A77. This multitude of interactions involving C75 may make any one contact nonessential to its positioning. Such principles have been applied to nonadditivity in enzymology.[53] Indeed, we have argued for low cooperativity of RNA folding in a tetraloop hairpin, where the multitude of stacking, hydrogen bonding, and hydration interactions provides an extensive and overdetermined network of interactions that make any single interaction nonessential.[54,55] Thus, the relative unimportance of the scissile phosphate *pro-$R_P$* oxygen in interacting with C75 does not mean that these two species do not interact.

*The $R_P$ Substrate has Biphasic Kinetics.*

The studies herein used a two-piece ribozyme construct, which facilitated introduction of the phosphorothioate at the cleavage site and separation of the diastereomers. We studied the oxo, $S_P$, and $R_P$ substrates with the same enzyme strand. The oxo and $S_P$ substrates gave simple, single exponential kinetic behavior, as reflected in the traces in Figure 6.2. The $R_P$ substrate, however, gave double exponential kinetics under most conditions[7] despite using the same enzyme strand as for the other substrates. The origin of the biphasic behavior is not entirely clear, although simple results with the oxo and $S_P$ substrates suggest that it is unlikely due to problems with heterogeneity in the ribozyme strand, which has beset other ribozymes.[56]

The fast and slow phases of the $R_P$ substrate have quite different behavior. The thio effects are only up to ~5-fold for the fast phase, but are up to ~1000-fold for the slow phase. Moreover, metal ion rescue is 7.5-fold and 330-fold for the fast and slow phases, respectively. Thus, in general larger thio effects and metal ion rescues are found for the slower phase. In

---

[7] The only exception was higher $Cd^{2+}$ conditions of 4 to 10 mM.

addition, the fast phase showed a typical rate-pH profile, albeit with a slightly lower $pK_a$ (Figure 6.2B), while the slow phase did not show a dependence on pH (Figure 6.2C). The latter observation is consistent with loss of participation of $Mg^{2+}$ in the reaction,[57] which would be consistent with larger thio effects and the potential for greater rescue by thiophilic metal ion, as observed.

## Supporting Information

Data on the HPLC separation of the thio substrates, iodine cleavage, ILP, MD simulations of the thio substrates with the enzyme strand, metal ion rescue with Mn2+, and inhibition of the oxo substrates by Cd2+ is provided in Appendix E.

## Acknowledgements

The authors would like to thank Squire Booker and Tyler Grove for use of their HPLC for the separation of the phosphorothioate diastereomers, Ji Chen for providing the clone for the A25•C20 double mutant ribozyme, and Andrew Sirjoosingh for help with gnuplot.

## References

1. Lai, M. M. C. *Annu. Rev. Biochem.* **1995**, *64*, 259-286.

2. Tseng, C.-H., and Lai, M. M. C. *Viruses 1*, **2009**, 818-831.

3. Ferré-D'Amaré, A. R., Zhou, K., and Doudna, J. A. *Nature* **1998**, *395*, 567-574.

4. Perrotta, A. T., and Been, M. D. *Nature* **1991**, *350*, 444-446.

5. Chen, J.-H., Yajima, R., Chadalavada, D. M., Chase, E., Bevilacqua, P. C., and Golden, B. L. *Biochemistry* **2010**, *49*, 6508-6518.

6. Chen, J., Ganguly, A., Miswan, Z., Bevilacqua, P. C., Hammes-Schiffer, S., and Golden, B. L. *Biochemistry* **2010***, 52,* 557-567.

7. Golden, B. L. *Biochemistry* **2011**, *50*, 9424-9443.

8. Perrotta, A. T., Shih, I., and Been, M. D. *Science* **1999**, *286*, 123-126.

9. Shih, I. H., and Been, M. D. *Proc. Natl. Acad. Sci.* **2001**, *98*, 1489-1494.

10. Perrotta, A. T., Wadkins, T. S., and Been, M. D. *RNA* **2006**, *12*, 1282-1291.

11. Nakano, S., Chadalavada, D. M., and Bevilacqua, P. C. *Science* **2000**, *287*, 1493-1497.

12. Nakano, S., Proctor, D. J., and Bevilacqua, P. C. *Biochemistry* **2001**, *40*, 12022-12038.

13. Nakano, S., and Bevilacqua, P. C. *Biochemistry* **2007**, *46*, 3001-3012.

14. Gong, B., Chen, J.-H., Chase, E., Chadalavada, D. M., Yajima, R., Golden, B. L., Bevilacqua, P. C., and Carey, P. R. *J. Am. Chem. Soc.* **2007**, *129*, 13335-13342.

15. Das, S. R., and Piccirilli, J. A. *Nat. Chem. Biol.* **2005**, *1*, 45-52.

16. Draper, D. E. *RNA* **2004**, *10*, 335-344.

17. Fedor, M. J., and Williamson, J. R. *Nat. Rev.: Mol. Cell Biol.* **2005**, *6*, 399-412.

18. Suh, Y. A., Kumar, P. K., Taira, K., and Nishikawa, S. *Nuc. Acids Res.* **1993**, *21*, 3277-3280.

19. Nakano, S., Cerrone, A. L., and Bevilacqua, P. C. *Biochemistry* **2003**, *42*, 2982-2994.

20. Fauzi, H., Kawakami, J., Nishikawa, F., and Nishikawa, S. *Nuc. Acids Res.* **1997**, *25*, 3124-3130.

21. Jeoung, Y. H., Kumar, P. K., Suh, Y. A., Taira, K., and Nishikawa, S. *Nuc. Acids Res.* **1994**, *22*, 3722-3727.

22. Chadalavada, D. M., Knudsen, S. M., Nakano, S., and Bevilacqua, P. C. *J. Mol. Biol.* **2000**, *301*, 349-367.

23. Gong, B., Chen, Y., Christian, E. L., Chen, J.-H., Chase, E., Chadalavada, D. M., Yajima, R., Golden, B. L., Bevilacqua, P. C., and Carey, P. R. *J. Amer. Chem. Soc.* **2008**, *130*, 9670-9672.

24. Nakano, S., and Bevilacqua, P. C. *J. Amer. Chem. Soc.* **2001**, *123*, 11333-11334.

25. Cerrone-Szakal, A. L., Siegfried, N. A., and Bevilacqua, P. C. *J. Am. Chem. Soc.* **2008**, *130*, 14504-14520

26. Ganguly, A., Bevilacqua, P. C., and Hammes-Schiffer, S. *J. Phys. Chem. Lett.* **2011**, *2*, 2906-2911.

27. Herschlag, D., Piccirilli, J. A., and Cech, T. R. *Biochemistry* **1991**, *30*, 4844-4854.

28. Frederiksen, J. K., and Piccirilli, J. A. *Methods* **2009**, *49*, 148-166.

29. Szewczak, A. A., Kosek, A. B., Piccirilli, J. A., and Strobel, S. A. *Biochemistry* **2002**, *41*, 2516-2525.

30. Bevington, P. R. McGraw-Hill, New York, **1969**.

31. Segel, I. H. Wiley, New York, **2003**.

32. Veeraraghavan, N., Bevilacqua, P. C., and Hammes-Schiffer, S. *J. Mol. Biol.* **2010**, *402*, 278-291.

33. Veeraraghavan, N., Ganguly, A., Chen, J.-H., Bevilacqua, P. C., Hammes-Schiffer, S., and Golden, B. L. *Biochemistry* **2011**, *50*, 2672-2682.

34. D. E. Shaw Research, N. Y. Desmond Molecular Dynamics System, **2008**.

35. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. *J. Amer. Chem. Soc.* **1995**, *117*, 5179-5197.

36. Qsite, *Schrodinger, LLC, New York, NY*, **2008**.

37. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372-1379.

38. Cramer, C. J. Computational chemistry, Chapter 5-8, **2002**.

39. Hehre, W. J. A guide to molecular mechanics and quantum mechanical calculations. Chapter 2, **2003**.

40. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474-6487.

41. Yoshida, A., Sun, S., and Piccirilli, J. A. *Nat. Struc. Biol.* **1999**, *6*, 318-321.

42. Hamm, M. L., Nikolic, D., Breemen, R. B. V., and Piccirilli, J. A. *J. Am. Chem. Soc.* **2000**, *122*, 12069-12078.

43. Shan, S., Kravchuk, A. V., Piccirilli, J. A., and Herschlag, D. *Biochemistry* **2001**, *40*, 5161-5171.

44. Veeraraghavan, N., Ganguly, A., Golden, B. L., Bevilacqua, P. C., and Hammes-Schiffer, S. *J. Phys. Chem. B* **2011**, *115*, 8346-8357.

45. Perrotta, A. T., and Been, M. D. *Biochemistry* **2006**, *45*, 11357-11365.

46. Wei, K., Liu, L., Cheng, Y.-H., Fu, Y., and Guo, Q.-X. *J. Phys. Chem. B.* **2007**, *111*, 1514-1516.

47. Ke, A., Zhou, K., Ding, F., Cate, J. H. D., Doudna, J. A., and A. Ke, K. Zhou, F. Ding, J. H. Cate, J. A. D. *Nature* **2004**, *429*, 201-205.

48. Wrzesinski, J., Wichłacz, A., Nijakowska, D., Rebowska, B., Nawrot, B., and Ciesiołka, J. *New J Chem.* **2010**, *34*, 1018-1026.

49. Rosenstein, S. P., and Been, M. D. *Nuc. Acids Res.* **1991**, *19*, 5409-5416.

50. Wadkins, T. S., Perrotta, A. T., Ferré-D'Amaré, A. R., Doudna, J. A., and Been, M. D. *RNA* **1999**, *5*, 720-727

51. Forconi, M., Lee, J., Lee, J. K., Piccirilli, J. A., and Herschlag, D. *Biochemistry* **2008**, *47*, 6883-6894.

52. Frey, P. A., and Sammons, D. *Science* **1985**, *228*, 541-545.

53. Kraut, D. A, Carroll, K. S., and Herschlag, D. *Annu. Rev. Biochem.* **2003**, *72*, 517-571.

54. Moody, E. M., Feerrar, J. C., and Bevilacqua, P. C. *Biochemistry* **2004**, *44*, 7992-7998.

55. Siegfried N. A., Bevilacqua P. C. *Methods Enzymol*. **2009**, *455*, 365-393.

56. Greenfeld, M., Solomatin, S. V., and Herschlag, D. *J. Biol. Chem.* **2011**, *286*, 19872-19879.

57. Bevilacqua, P. C. *Biochemistry* **2003**, *42*, 2259-2265.

58. Brown, T. S., Chadalavada, D. M., and Bevilacqua, P. C. *J. Mol. Biol.* **2004**, *341*, 695-712.

## FIGURES AND TABLES

**Table 6.1:** Thio effect and metal ion rescue for HDV ribozyme[a] in the presence of 10 mM $Mg^{2+}$ at pH 5.6 and 8.0[a]

pH 5.6 Conditions

| Ionic Conditions | Substrate | $k_{obs}$ (min$^{-1}$)[b] | $k_O/k_S$[c] | Metal Ion Rescue[d] |
|---|---|---|---|---|
| 10 mM $Mg^{2+}$ | Oxo | 4.10 ± 0.2 | N/A | N/A |
| | $S_P$ | 4.40 ± 0.2 | 0.93 ± 0.07 | |
| | $R_P$ fast | 1.46 ± 0.8 | 2.8 ± 1.4 | |
| | $R_P$ slow | 0.004 ± 0.001 | $(1.0 \pm 0.25) \times 10^3$ | |
| 10 mM $Mg^{2+}$/ 2 mM $Cd^{2+}$ | Oxo | 0.17 ± 0.07 | N/A | N/A |
| | $S_P$ | 0.09 ± 0.02 | 2 ± 0.9 | 0.5 ± 0.2 |
| | $R_P$ fast | 0.4 ± 0.2 | 0.4 ± 0.2 | 7.5 ± 4.3 |
| | $R_P$ slow | 0.003 ± 0.0007 | 57 ± 27 | 18 ± 9.6 |
| 10 mM $Mg^{2+}$/ 10 mM $Cd^{2+}$ | Oxo | 0.006 ± 0.0001 | N/A | N/A |
| | $S_P$ | 0.004 ± 0.0001 | 1.5 ± 0.06 | 0.5 ± 0.1 |
| | $R_P$ slow[e] | 0.002 | 3.0 ± 0.05 | $(3.33 \pm 1.0) \times 10^2$ |
| 10 mM $Mg^{2+}$/ 2 mM $Mn^{2+}$ | Oxo | 3.70 ± 0.1 | N/A | N/A |
| | $S_P$ | 4.01 ± 0.5 | 0.92 ± 0.12 | 1.0 ± 0.15 |
| | $R_P$ fast | 0.776 | 4.77 ± 0.15 | 0.6 ± 0.15 |
| | $R_P$ slow | 0.001 | $(3.70 \pm 0.15) \times 10^3$ | 0.3 ± 0.02 |

(continued on next page)

**Table 6.1 (cont.):**

pH 8.0 Conditions

| Ionic Conditions | Substrate | $k_{obs}$ (min$^{-1}$) | $k_O/k_S$ | Metal Ion Rescue |
|---|---|---|---|---|
| 10 mM $Mg^{2+}$ | Oxo | 7.0 ± 0.5 | N/A | N/A |
| | $S_P$ | 8.0 ± 1.6 | 0.87 ± 0.2 | |
| | $R_P$ fast | 1.45 ± 0.45 | 4.8 ± 1.5 | |
| | $R_P$ slow | 0.009 ± 0.001 | (7.7± 1.0) x $10^2$ | |
| 10 mM $Mg^{2+}$/ 0.5 mM $Cd^{2+}$ | Oxo | 0.1 ± 0.001 | N/A | N/A |
| | $S_P$ | 0.03 | 3.0 ± 0.3 | 0.3 ± 0.07 |
| | $R_P$ fast | 0.20±0.01 | 0.50 ± 0.06 | 10 ± 3.2 |
| | $R_P$ slow | 0.004±0.001 | 25 ± 6.5 | 31 ± 8.9 |

[a] $k_{obs}$ measured in the background of 10 mM $Mg^{2+}$ at 37° C at pH 5.6 (50 mM NaMES) or pH 8.0 (NaHEPES).

[b] Error in $k_{obs}$ is from the standard deviation of two or more experimental observations.

[c] Thio effect is the ratio $k_O/k_S$ at 10 mM $Mg^{2+}$, in the absence of thiophilic metal ions (eq 3). Error is propagated from the relative errors.[30]

[d] The metal ion rescue for each substrate is calculated as described in eq 4.

[e] At 10 mM $Cd^{2+}$, the $R_P$ substrate did not exhibit a biphasic behavior. We designated this $R_P$ substrate as '$R_P$ slow phase' because the observed rates for this substrate are close to the $R_P$ slow phase. Also with increasing concentrations of $Cd^{2+}$, the amplitude of the $R_P$ fast phase decreases.

**Table 6.2:** Thio effect in absence of divalent metal ion, AC double mutant, and C75Δ mutant

| Variant | pH | Ionic Conditions | Substrate | $k_{\text{obs}}$ (min$^{-1}$)$^{a}$ | $k_O/k_S$ $^{a}$ |
|---|---|---|---|---|---|
| WT no $Mg^{2+}$ | 6.0 | 1M NaCl/ 100 mM EDTA | Oxo | (0.5 ± 0.1) x $10^{-4}$ | N/A |
| | | | $S_P$ | (1.7 ± 0.1) x $10^{-4}$ | 0.3 ± 0.2 |
| | | | $R_P$ | (0.20 ± 0.04) x $10^{-4}$ | 2.5 ± 0.3 |
| AC DM | 5.6 | 10 mM $Mg^{2+}$ | Oxo | 0.06 ± 0.01 | N/A |
| | | | $R_P$ | 0.05 ± 0.01 | 1.1 ± 0.3 |
| | | 50 mM $Mg^{2+}$ | Oxo | 0.72 ± 0.01 | N/A |
| | | | $S_P$ | 0.57 ± 0.01 | 1.3 ± 0.03 |
| | | | $R_P$ | 0.16 ± 0.03 | 4.2 ± 0.8$^{b}$ |
| | 7.0 | 10 mM $Mg^{2+}$ | Oxo | 0.032 ± 0.005 | N/A |
| | | | $R_P$ | 0.022 ± 0.002 | 1.5 ± 0.3 |
| | | 50 mM $Mg^{2+}$ | Oxo | 0.026 ± 0.001 | N/A |
| | | | $R_P$ | 0.018 ± 0.003 | 1.4 ± 0.3 |
| C75Δ | 7.0 | 25 mM $Mg^{2+}$/ 400 mM imidazole | Oxo | (1.5 ± 0.01) x $10^{-4}$ | N/A |
| | | | $S_P$ | (2.1 ± 0.1) x $10^{-4}$ | 0.7 ± 0.03 |
| | | | $R_P$ | (0.44 ± 0.01) x $10^{-4}$ | 3.2 ± 0.08 |

[a] Errors in $k_{\text{obs}}$ and ratio of rate constants were determined as in Table 6.1. [b] The observed thio effect was partially (~2-fold) rescuable in the presence of 0.5 mM $Cd^{2+}$.

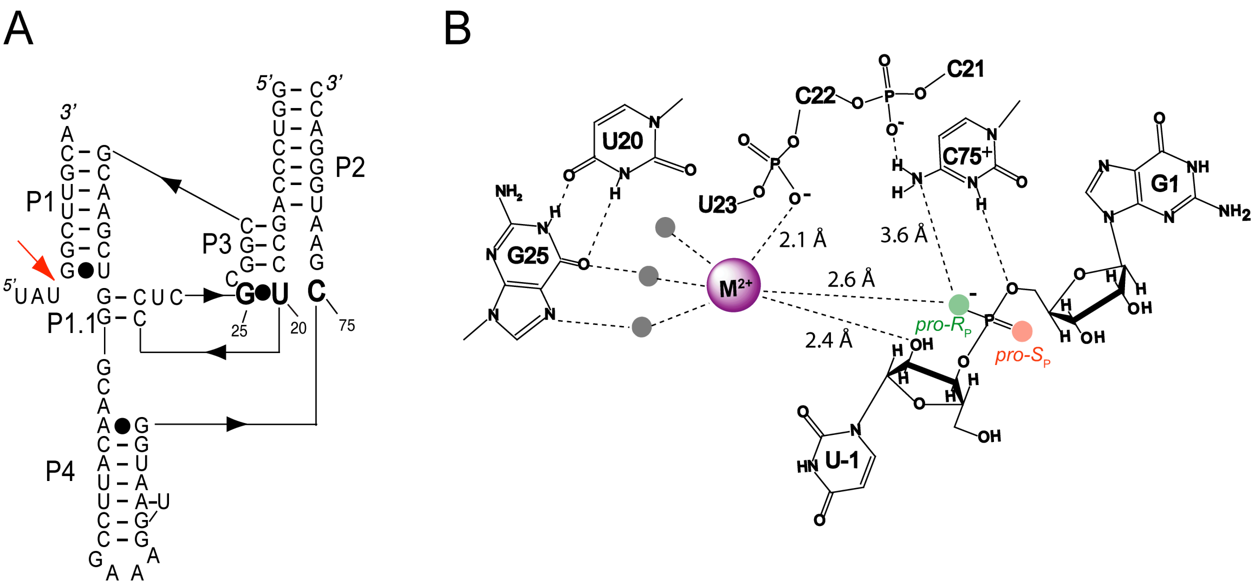

**Figure 6.1:** Secondary structure and active site interactions in the genomic HDV ribozyme. (A) Secondary structure of the *trans*-cleaving HDV ribozyme studied herein. The ribozyme consists of an enzyme strand and a substrate, which comprises the cleavage site shown by the red arrow. The catalytic nucleobase C75 and the G25•U20 reverse wobble, important in metal ion binding, are depicted in bold. In addition, there is a standard G1•U37 wobble at the cleavage site. The construct shown here is in the G11C/U27D background[58] and has a shortened P4, as used in crystallography;[5] original numbering of the 3'-portion is maintained. (B) Active site of the HDV ribozyme depicting catalytically important residues according to the crystal structure.[5] The *pro*-$R_P$ oxygen (green) is positioned near the catalytic metal ion (2.6 Å) and the amine of C75 (3.6 Å), which in turn interacts with the phosphate of C22. The active site $Mg^{2+}$ has three direct interactions: with the 2'OH of U-1, the *pro*-$R_P$ oxygen of the scissile phosphate, and the phosphate of U23. In addition, this $Mg^{2+}$ ion interacts with the reverse G•U wobble through its hydration shell. The distances are given according to reference.[5]

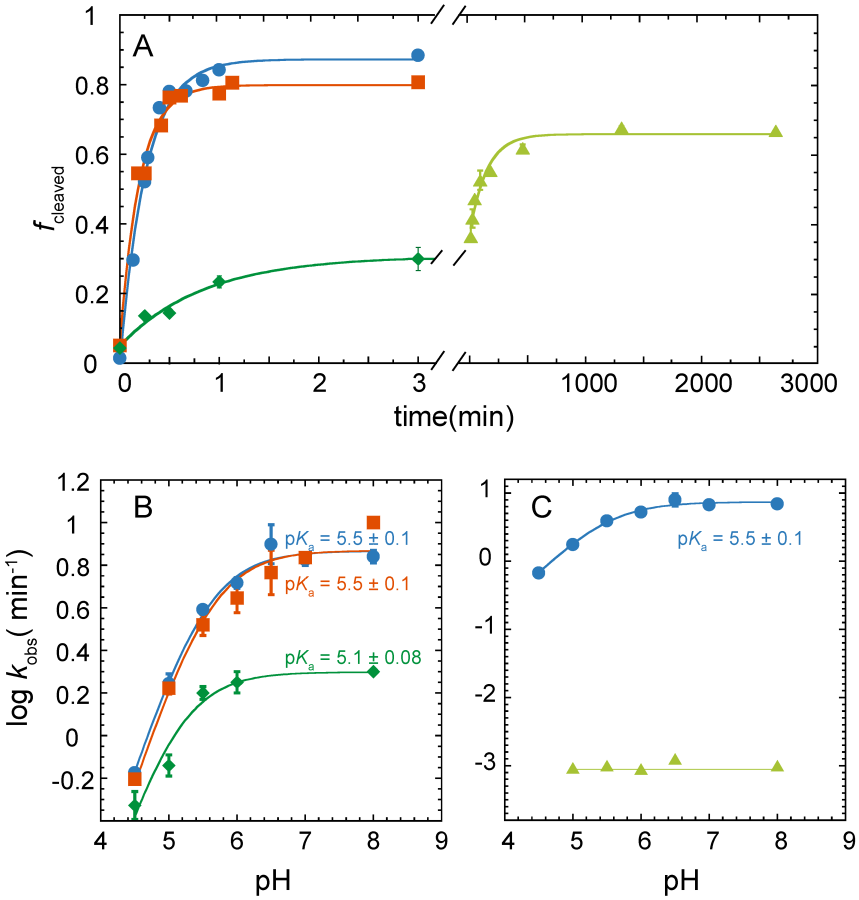

**Figure 6.2:** The $R_P$ substrate has substantial thio effects and a perturbed p$K_a$. (A) Fraction cleaved versus time (min) plots of oxo (blue circles), $S_P$ (red squares), and $R_P$ fast phase (dark green diamonds) and $R_P$ slow phase (light green triangles) in 10 mM $Mg^{2+}$ at pH 5.6 (same colors and shapes used in all graphs). Traces were fit to eq 1 for oxo and $S_P$ substrate and to eq 2 for $R_P$ substrate. The oxo and $S_P$ substrates react monophasically and close to completion, with $k_{obs}$ of 4.1 ± 0.2 and 4.4 ± 0.2 min$^{-1}$, respectively. The $R_P$ substrate, on the other hand, reacts biphasically, and both phases react slower than the oxo substrate, which constitutes thio effects. The fast and slow $R_P$ phases react ~3- ($k_{obs}$= 1.46 ± 0.8 min$^{-1}$) and ~1000-fold ($k_{obs}$=0.004 ± 0.001 min$^{-1}$) slower than the oxo substrate. (B, C) pH-rate profile of oxo, $R_P$, and $S_P$ substrate at 10 mM $Mg^{2+}$. (B) pH-rate profile of oxo, $S_P$, and $R_P$ fast-reacting substrate. The p$K_a$ of C75 when the substrate is oxo or $S_P$ was found to be ~5.5 ± 0.1 while the p$K_a$ of C75 when the substrate is $R_P$ fast substrate was 5.1 ± 0.08. (B) The pH-rate profile of the oxo substrate is shown in comparison to that of the slow-reacting $R_P$ substrate, which is independent of pH over this range. All data are in a background of 10 mM $Mg^{2+}$.

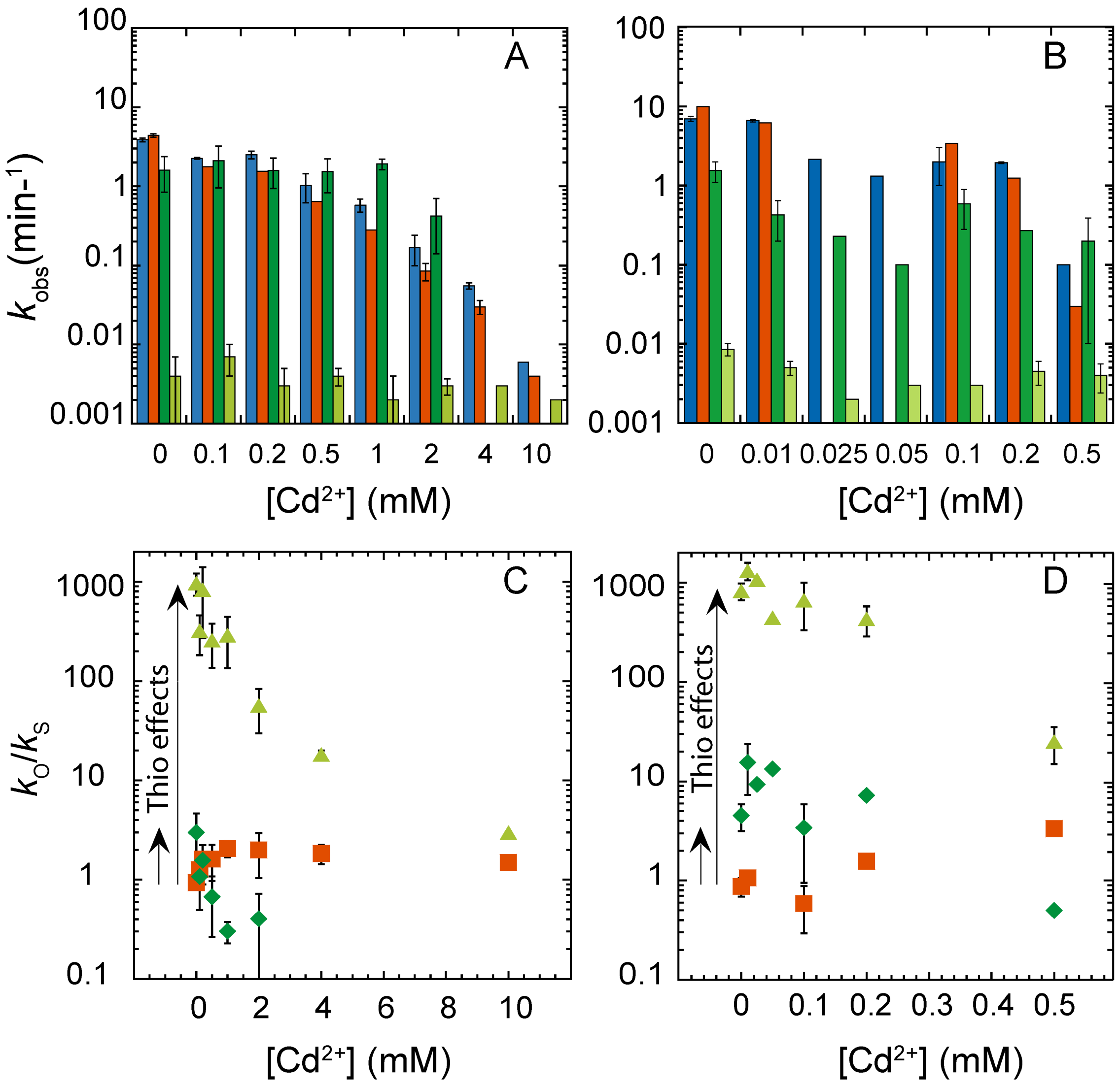

**Figure 6.3:** Dependence of $k_{obs}$ and $k_O/k_S$ on $Cd^{2+}$ concentration. (A and B) Histogram representing $k_{obs}$ values for the oxo (blue), $S_P$ (red), $R_P$ fast phase (dark green) and $R_P$ slow phase (light green) substrates as a function of $Cd^{2+}$ concentration at (A) pH 5.6 and (B) pH 8.0. (C and D) $k_O/k_S$ values for the $S_P$ (red squares), $R_P$ fast phase (green diamonds), and $R_P$ slow phase (light green triangles) substrates as a function of $Cd^{2+}$ concentration at (C) pH 5.6 and (D) pH 8.0. Note that there is only a single phase (light green triangles) for the $R_P$ substrate at $Cd^{2+}$ concentrations of >2 mM. Note that data for the $S_P$ substrate were not collected at 0.025 and 0.05 mM $Cd^{2+}$ at pH 8.0. All data are in the background of 10 mM $Mg^{2+}$. The thio effects in the absence of $Cd^{2+}$

for the $R_P$ substrate fast- and slow-phases are represented by the short and long arrows, respectively.

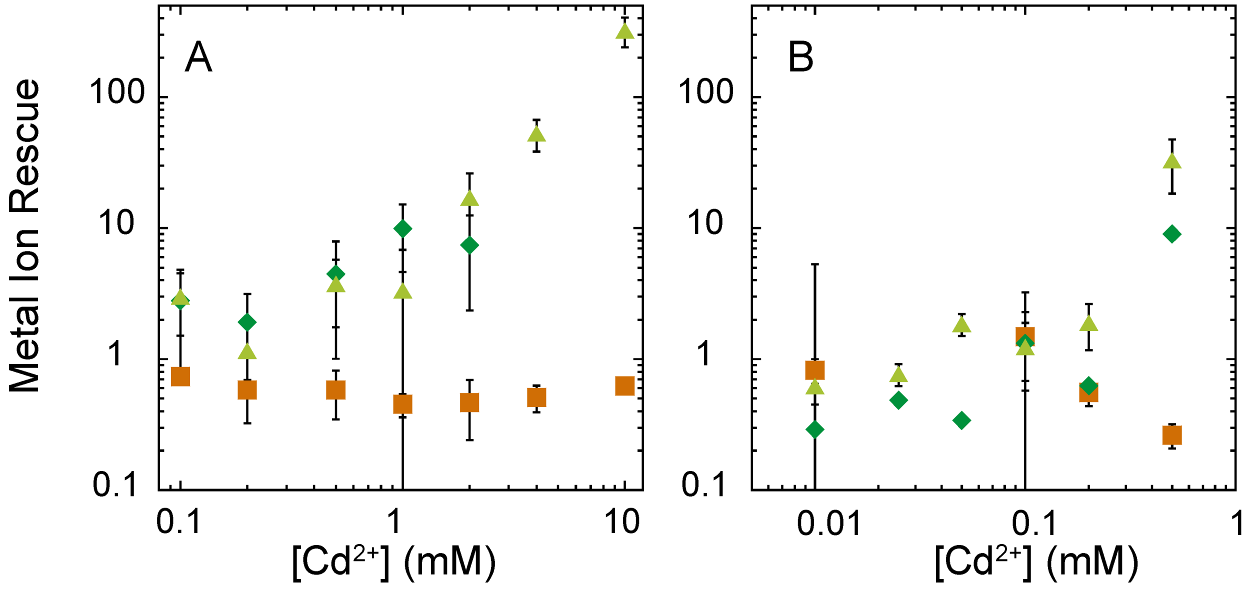

**Figure 6.4:** Metal ion rescue of the $S_P$ and $R_P$ substrate by Cd$^{2+}$ ions. Metal ion rescue for the $S_P$ (red squares), $R_P$ fast phase (dark green diamonds), and $R_P$ slow phase (light green triangles) substrates as a function of Cd$^{2+}$ concentration at (A) pH 5.6 and (B) pH 8.0. Note that there was only a single phase for the $R_P$ substrate at Cd$^{2+}$ concentrations above 2 mM. Data were fit to the following version of eq 4: metal ion rescue = $(k_O/k_S)^{Mg^{2+}} \big/ (k_O/k_S)^{Cd^{2+}}$ .

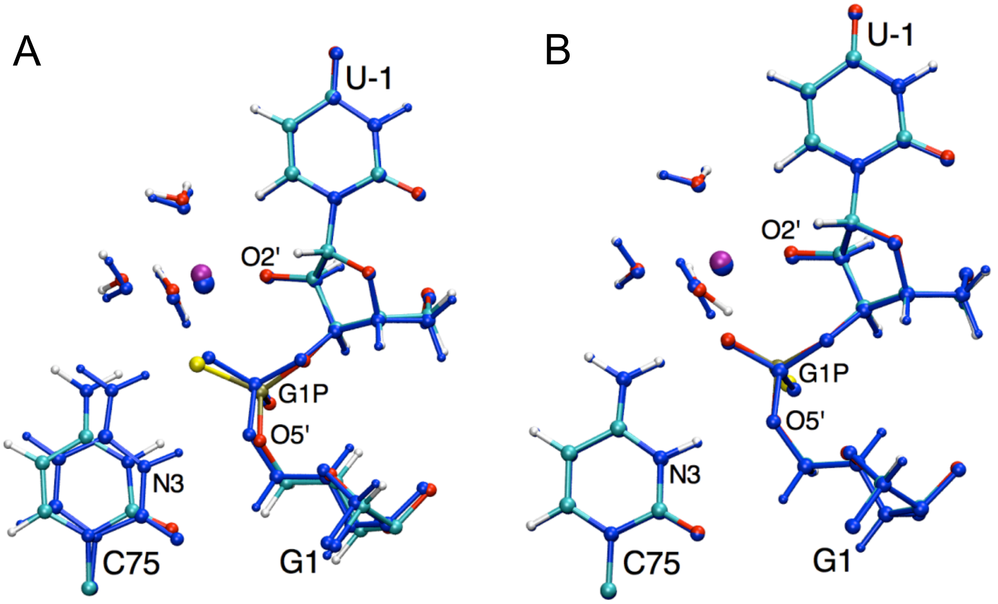

**Figure 6.5:** Overlay of the cleavage site of the QM/MM-optimized reactant states for the WT oxo substrate with the (A) $R_P$ and (B) $S_P$ substrates with $Mg^{2+}$. The phosphorothioate-containing structures are depicted in cyan and atom-specific colors with the catalytic $Mg^{2+}$ ion purple. The WT structure and its catalytic $Mg^{2+}$ ion are depicted entirely in blue. In all three structures, the N3 of C75 is protonated.

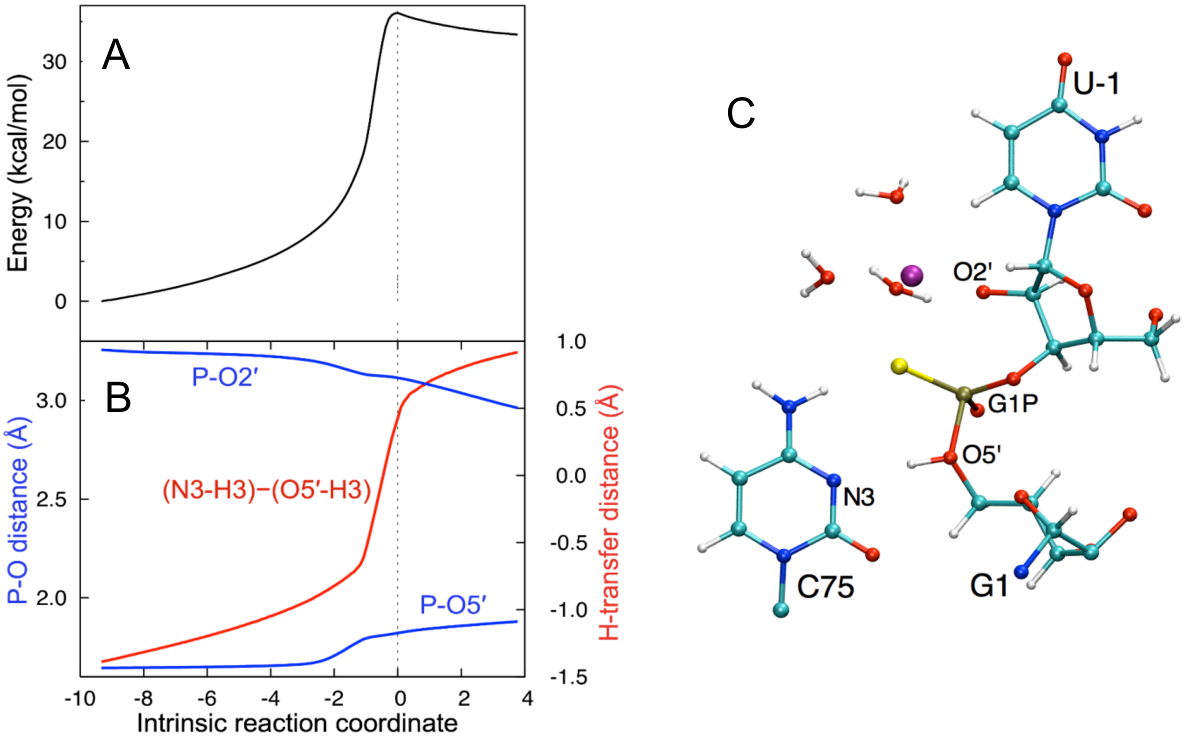

**Figure 6.6:** Minimum energy path (MEP) obtained from the QM/MM calculations for the initial proton transfer step of the self-cleavage reaction catalyzed by the $R_{\mathrm{P}}$ substrate with Mg$^{2+}$ ion at the catalytic site. This step corresponds to proton transfer from N3(C75) to O5'(G1) and would be followed by a step comprised of phosphorous-oxygen bond formation and cleavage. (A) Energy and (B) relevant distances shown along the intrinsic reaction coordinate. (C) Scissile phosphate region of the intermediate corresponding to a proton transferred from N3(C75) to O5'(G1), which is the product of the MEP for the step shown in panels A and B.

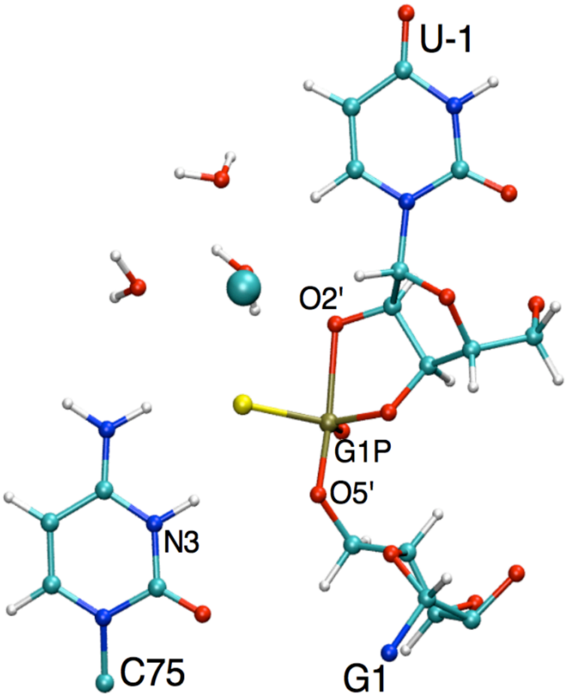

**Figure 6.7:** Cleavage site of the phosphorane intermediate structure obtained from the QM/MM calculations for the $R_P$ substrate with $Cd^{2+}$ ion at the catalytic site.

# CHAPTER 7: QUANTUM MECHANICAL/MOLECULAR MECHANICAL FREE ENERGY SIMULATIONS OF THE SELF-CLEAVAGE REACTION IN THE HEPATITIS DELTA VIRUS RIBOZYME[1]

## Abstract

The hepatitis delta virus (HDV) ribozyme catalyzes a self-cleavage reaction using a combination of nucleobase and metal ion catalysis. Both divalent and monovalent ions can catalyze this reaction, although the rate is slower with monovalent ions alone. Herein, we use quantum mechanical/molecular mechanical (QM/MM) free energy simulations to investigate the mechanism of this ribozyme and to elucidate the roles of the catalytic metal ion. With $Mg^{2+}$ at the catalytic site, the self-cleavage mechanism is observed to be concerted with a phosphorane-like transition state and a free energy barrier of ~13 kcal/mol, consistent with free energy barrier values extrapolated from experimental studies. With $Na^+$ at the catalytic site, the mechanism is observed to be sequential, passing through a phosphorane intermediate, with free energy barriers of 2-4 kcal/mol for both steps; moreover, proton transfer from the exocyclic amine of protonated C75 to the nonbridging oxygen of the scissile phosphate occurred to stabilize the phosphorane intermediate in the sequential mechanism. To explain the slower rate observed experimentally with monovalent ions, we hypothesize that the activation of the O2′ nucleophile by deprotonation and orientation is less favorable with $Na^+$ than with $Mg^{2+}$ ions. To explore this hypothesis, we experimentally measure the p$K_a$ of O2′ by kinetic and NMR methods and find it to be lower in the presence of divalent rather than only monovalent ions. The combined theoretical and

---

[1] This chapter was published in its entirety in the Journal of the American Chemical Society in 2014. The complete reference to the journal article is as follows: Ganguly, A.; Thaplyal, P.; Rosta, E.; Hammes-Schiffer, S.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2013**, 52, 6499-6514. A.G. performed the computational studies, P.T. performed the experiments, and E.R. provided valuable technical and scientific help.

experimental results indicate that the catalytic $Mg^{2+}$ ion may play three key roles: assisting in the activation of the O2′ nucleophile, acidifying the general acid C75, and stabilizing the nonbridging oxygen to prevent proton transfer to it.

## Introduction

The hepatitis delta virus (HDV) ribozyme belongs to a family of five distinct small ribozymes known as the nucleolytic ribozymes.[1,2] These ribozymes perform site-specific cleavage and ligation reactions in RNA and can utilize nucleobases in catalysis.[1,3,4] The HDV ribozyme is found in closely related genomic and antigenomic forms,[5] as well as an extended family of HDV-like ribozymes.[6-8] The $pK_a$'s of specific functional groups in ribozymes can be shifted toward neutrality to aid in general acid-base catalysis.[9-14] An active site cytosine, referred to as 'C75' in the genomic HDV ribozyme, has a $pK_a$ shifted toward neutrality and has been shown to be critical for its self-cleavage reaction, acting as the general acid by donating a proton to the O5′ oxyanion of the leaving group.[15,16] A role for C75 in proton transfer in the cleavage reaction of the genomic HDV ribozyme has been further supported by molecular dynamics (MD) and quantum mechanical/molecular mechanical (QM/MM) calculations.[17-20] In addition to nucleobase catalysis, the HDV ribozyme uses a metal ion in its cleavage reaction,[16,21,22] a strategy characteristic of the large ribozymes.[23,24] The catalytic divalent ion binds to a groove of extreme negative potential, where it interacts with a reverse G•U wobble[25] and the pro$R_P$ oxygen of the scissile phosphate.[22] In the case of the HDV ribozyme, both divalent and monovalent ions can catalyze the reaction, although the rate is ~25-fold slower in the case of monovalent ions alone.[21] The use of both metal ion and nucleobase catalysis in a single ribozyme is unusual[26] and thus has generated significant interest in understanding the mechanistic details.

In a previous study, we performed QM/MM calculations on the HDV ribozyme.[19] The active site of the HDV ribozyme and several possible mechanisms are depicted in Figure 7.1. Those calculations indicated that the mechanism of the self-cleavage reaction is concerted with a phosphorane-like transition state (TS) when a divalent metal ion, such as $Mg^{2+}$ or $Ca^{2+}$, is bound at the catalytic site (Figure 7.1A). In contrast, the self-cleavage reaction was found to be sequential with a phosphorane intermediate (PI) when a monovalent ion, such as $Na^{+}$, is at this site (Figure 7.1C with a $Na^{+}$ rather than $Mg^{2+}$ at the catalytic site). Electrostatic potential calculations suggested that the divalent metal ion at the active site lowers the p$K_a$ of protonated C75, aiding its proton donation and thus facilitating the concerted mechanism, in which the proton is partially transferred to the leaving group in the phosphorane-like TS. These observations are supported by several experimental studies, including measurements of the p$K_a$ in different $Mg^{2+}$ and $Na^{+}$ concentrations and proton inventories in the presence of $Mg^{2+}$ or $Na^{+}$ alone.[10,25,27] In this previous study, however, we emphasized that the results were only qualitatively meaningful because of the limitations of the methodology. Specifically, we optimized structures for the minima and transition states and calculated the minimum energy path (MEP) without including entropic contributions or conformational sampling. The potential energy surface for such large systems is expected to be complex, with numerous stationary points, and the reactant, product, and transition states identified by this methodology may be biased toward the starting configuration. Moreover, the inclusion of entropic contributions is essential for obtaining free energy barriers that are related to experimentally measured rate constants.

An alternative approach that avoids these limitations is to combine QM/MM calculations with free energy simulations to generate the underlying free energy surface for the reaction.[28-35]

Exploring the entire free energy surface of the system is typically computationally prohibitive, however, because of the multidimensionality of these types of complex systems and the expense of the QM/MM calculations. A variety of methods have been developed to sample only the relevant portions of the free energy surface, thereby providing a computationally tractable approach for generating the minimum free energy path (MFEP). For example, Hummer and coworkers combined the finite temperature string method[36] with umbrella sampling simulations[37] to investigate the catalytic reaction in the ribonuclease H enzyme.[33] The key strategy of this method is to start with an initial guess reaction path described by an initial string in the multidimensional phase space and to repeatedly update the string based on umbrella sampling simulations for points along the string. In this manner, the regions of phase space relevant to the reaction are sampled, and the string evolves toward the MFEP.

In the present study, we use this finite temperature string umbrella sampling approach to investigate the mechanism of the HDV ribozyme self-cleavage reaction. We explore three possible reaction mechanisms for the self-cleavage reaction with a $Mg^{2+}$ ion bound at the catalytic site. As described below, all three initial reaction paths, which correspond to three different mechanisms, evolve toward the same MFEP, thereby illustrating the robustness of the approach in identifying the thermodynamically favorable mechanism. We also examine the reaction with a $Na^{+}$ ion rather than a $Mg^{2+}$ ion at the catalytic site and further illustrate that the mechanism is different in the presence of monovalent versus divalent ions. In contrast to previous calculations, these QM/MM free energy calculations provide estimates of the free energy barriers for comparison to experimental data and lead to different mechanistic predictions, such as an additional proton transfer reaction occurring to stabilize the phosphorane intermediate when $Na^{+}$ is at the catalytic site. Furthermore, these calculations lead to several hypotheses regarding the role of the catalytic metal ion. To test one of the hypotheses generated by these calculations, we perform kinetic and NMR experimental measurements to determine the p$K_a$ of the attacking O2′ in the presence of divalent or monovalent cations. The combined theoretical and experimental results provide a number of new key insights into the mechanism of the HDV ribozyme and, in particular, the roles of the catalytic metal ion.

## Methods

*QM/MM free energy simulations.*

The simulations were based on the precleaved crystal structure (PDB ID 3NKB)[16] solved with wild-type sequences and a $Mg^{2+}$ ion bound at the catalytic site, as in our previous QM/MM calculations.[19] The initial modeling performed on the crystal structure, particularly on the scissile phosphate and the upstream nucleotide, is discussed in our previous studies.[19,20,38] Several recent studies support the positioning of the active site nucleotides and metal ions in this model.[22,25] The nucleotides at positions 1 and 2 were modified from their deoxy form, which were present to prevent degradation during crystallography, by addition of 2′-hydroxyls with ideal bond lengths and bond angles. The residues C75 and C41 were kept protonated at their N3 positions to reflect the active form of the ribozyme, as discussed elsewhere.[39,40] The Accelrys Discover Studio Visualizer 2.0 program was used to add the hydrogen atoms. The ribozyme, along with its 10 crystallographically resolved $Mg^{2+}$ ions, was immersed in an orthorhombic box containing rigid TIP3P waters,[41] and $Na^+$ ions were added to neutralize the system. Additional $Na^+$ and $Cl^-$ ions were added to provide a physiological ionic strength of 0.15 M. The starting structures along the initial guess pathways for the QM/MM free energy simulations were obtained after classical molecular dynamics (MD) equilibration of the solvent and the monovalent $Na^+$ and $Cl^-$ ions, followed by deletion of all solvent molecules and ions further than

20 Å from the ribozyme. The MD simulations were performed using the DESMOND program[42] in conjunction with the AMBER99 forcefield,[43] periodic boundary conditions, and the Ewald treatment[44] of long-range electrostatics. The details of these classical MD simulations are provided in Ref. [20,38].

The QM/MM interface between Q-Chem[45] and CHARMM[46] was used to perform the QM/MM free energy simulations. The QM region consists of 87 atoms, as used in our previous study[19] and depicted in Figure 7.2. The O2′ of the U-1 residue remained deprotonated throughout all of our simulations to drive the nucleophilic attack on the scissile phosphate. The deprotonation of the U-1 2′OH was assumed to occur prior to the nucleophilic attack and was not considered explicitly in this study because the identity of the general base is unknown, although it may be a hydroxide anion ligated to the $Mg^{2+}$ ion. The QM atoms were treated with density functional theory (DFT) using the 6-31G** basis set and the B3LYP functional. The MM region was described by the AMBER99 forcefield,[43] and link atoms were utilized to treat the QM-MM boundary. The simulations were performed using Langevin dynamics with a 1 fs time step. During the simulations, all atoms lying outside a sphere of radius 20 Å centered at the phosphorus atom of the scissile phosphate remained fixed.

The multidimensional free energy surface and the minimum free energy path (MFEP) were determined by combining umbrella sampling simulations with a finite temperature string method.[33] In this method, an *M*-dimensional curve corresponding to the initial string, $s_0(\xi)=\left(f_0^1(\xi),\ \ldots,f_0^M(\xi)\right)$, is constructed along an initial guess reaction pathway from the reactant to the product state. Here *M* is the number of reaction coordinates, $\xi$ is the reduced length along the curve, $0\le\xi\le1$, and each function $f_0^i(\xi)$ represents the value of the *i*th reaction coordinate at $\xi$. The reactant and product states correspond to $f_0^i(0)$ and $f_0^i(1)$, respectively, for

$i = 1$ to $M$. The initial string along $\xi$ is then divided equally into $N$ images at positions $\xi_j$ with $j = 1$ to $N$, where each image corresponds to certain values of the $M$ reaction coordinates. Umbrella sampling is performed for each image, with harmonic restraints centered at $s_0(\xi_j)$ for each of the $M$ reaction coordinates. After 100 fs of MD, the string is updated by fitting the average reaction coordinates of each image along $\xi$ to a new curve. The $N$ images are redistributed along this updated string with equal spacings to obtain the new centers of the restraining potentials, $s_1(\xi_j)$, $j = 1$ to $N$, for the umbrella sampling. This process of updating the string based on umbrella sampling simulations is repeated until the string is converged. We considered the string to be converged when the root-mean-squared deviation (RMSD) of all coordinates of the latest string from the mean value of the previous ten iterations, summed over all images, fell below a threshold of 0.1 Å. Finally, the results are unbiased using the multidimensional weighted histogram analysis method (WHAM)[47] with a convergence threshold of 0.001 kcal/mol.

We utilized this method to explore the possibility of three different reaction mechanisms of the HDV ribozyme self-cleavage reaction. Independent sets of simulations were performed for each case, starting with an initial guess string associated with each proposed reaction mechanism. The following three mechanisms were investigated: (A) a concerted mechanism, in which the reaction passes through a single phosphorane-like transition state (Figure 7.1A), (B) a sequential mechanism, in which the reaction passes through a bridging O5′ proton-transferred intermediate (Figure 7.1B), and (C) another sequential mechanism, in which the reaction passes through a phosphorane intermediate (Figure 7.1C). The initial string in simulation set A was constructed from configurations along the minimum energy path obtained in our previous QM/MM calculations.[19] The initial strings for simulation sets B and C were generated from interpolation between representative structures along the respective pathways. For these three simulation sets, A, B, and C, a $Mg^{2+}$ ion was bound at the catalytic site, as in the crystal structure. We also investigated the impact of the valency of the catalytic metal ion on the reaction mechanism by replacing the $Mg^{2+}$ ion at the catalytic site with a $Na^{+}$ ion in certain calculations. On the basis of our previous QM/MM calculations,[19] the initial string for this simulation set, denoted D, was chosen to be similar to that for simulation set C, passing through a phosphorane intermediate.

For each of these four sets, A, B, C, and D, independent simulations were performed using the finite temperature string method with umbrella sampling described above. Twelve atom-to-atom active site distances were used as reaction coordinates, as illustrated in Figure 7.2. Note that many of these distances did not change significantly along the initial strings but were included in case they were found to be important during the finite temperature string calculations. The inclusion of reaction coordinates that do not change substantially during the reaction does not increase the computational cost significantly. Restraining potentials with a force constant of 100 kcal/mol were used for the umbrella sampling simulations. For set A, initially 15 images were used along the reaction pathway, and later the number of images was increased to 30 to ensure better resolution of the MFEP. In all other sets, 21 images were used along the reaction pathway. To obtain the starting configurations for the production simulations, the configurations representing the images along the initial strings for each of the sets were subjected to 150 fs of QM/MM equilibration with harmonic restraints applied to the reaction coordinates. The total simulation times for sets A, B, C, and D were 52 ps, 66 ps, 37 ps, and 56 ps, respectively. A limitation of this approach is that such relatively short simulation times do not enable comprehensive sampling of conformational space. As will be shown below, however, the qualitatively different initial strings for simulation sets A, B, and C resulted in similar MFEPs and overall mechanisms, providing a degree of validation for the convergence of these simulations in terms of the qualitative behavior. Set D corresponds to a different ion at the catalytic site (i.e., $Na^+$ instead of $Mg^{2+}$) and results in a different MFEP and mechanism, illustrating the flexibility of the approach in terms of identifying different mechanisms.

### *Measurements of O2′ pKa.*

The p$K_a$ of O2′ was determined using two different experimental methods. The first method measured the dependence of cleavage rate of a chimeric oligonucleotide with a single ribose linkage on pH using kinetic assays, and the second method measured the dependence of the $^1H$ chemical shift of 3′ adenosine monophosphate (3’AMP) on pH using proton NMR.

In the first method, the chimeric DNA/RNA was obtained from IDT (Coralville, IA), with the same sequence 5’ d(CGACTCACTAT)rU*d(GGAAGAGATG), as used in a previous study.[48] The “*”denotes the site of the only RNA linkage in the chimeric oligonucleotide. The sequence U*G was chosen for the ribose linkage in order to correspond with the -1 and +1 nucleotides at the cleavage site of the HDV ribozyme. The chimeric oligonucleotide was 5′ $^{32}P$-labeled using T4 polynucleotide kinase (New England Biolabs) and [$\gamma$-$^{32}P$] ATP, and subsequently purified on 10% polyacrylamide-7 M urea gels before being used for the kinetic assays.

The cleavage assays in the presence of $Na^+$ and $K^+$ were performed similarly to those previously described.[48] Reactions were conducted at 23 °C in the presence of 3.16 M monovalent ion. The reaction was maintained at a particular pH by using the appropriate concentration of the corresponding metal hydroxides (KOH or NaOH). The total $M^+$ ion concentration was maintained at 3.16 M at various pH values by addition of appropriate amounts of the corresponding metal chloride (KCl or NaCl).

Cleavage assays in the presence of $Ca^{2+}$ were performed under an inert argon atmosphere to avoid the precipitation of $Ca^{2+}$ as calcium carbonate. Solid $Ca(OH)_2$ and $CaCl_2$ (Sigma Aldrich, 99.99% purity) were used to make stock solutions of 10 mM and 20 mM, respectively, using degassed water under an argon atmosphere. The solutions were made fresh for each new set of experiments and stored under argon. For each pH tested, the total $Ca^{2+}$ concentration was maintained at 10 mM, which is relevant to typical HDV ribozyme experiments.[10,49] Higher $Ca^{2+}$ concentrations were not attempted due to the precipitation of $Ca(OH)_2$.

In each metal ion, the reaction was initiated by the addition of a premixed solution of metal hydroxide and metal chloride to the $5'^{32}$P-labeled chimeric substrate. Subsequently, 3 uL aliquots were removed at regular intervals and neutralized by addition to an equal volume of 100 mM Tris-HCl buffer (pH 7.0). The solution was then diluted with an equal volume of 20 mM EDTA and 90% formamide and placed on dry ice to prevent further reaction. The reactant and product were fractionated on a 10% polyacrylamide-7 M urea gel, and the fraction of the cleaved substrate with respect to time was quantified using a Typhoon PhosphorImager (Molecular Dynamics). Data from cleavage of the chimeric substrate were fit to a single exponential equation:

$$f_{\text{cleaved}} = A\left(1 - e^{-k_{\text{obs}}t}\right) \tag{7.1}$$

where $f_{\text{cleaved}}$ is the fraction of substrate cleaved at time $t$, $k_{\text{obs}}$ is the observed first-order rate constant, and $A$ is the fraction of the substrate cleaved at completion. For reactions at lower pH, which had slower observable rates ($< 10^{-3}$ min$^{-1}$), early time points corresponding to the initial reaction rate were taken and fitted to a linearized form of Eq. (1):

$$f_{\text{cleaved}} = Ak_{\text{obs}}t \tag{7.2}$$

where $A$ is the amplitude of the reaction, as in Equation 7.1, and $Ak_{\text{obs}}$ is the slope of the line. For reactions in the presence of monovalent ions, the reactions went to completion leading to A~1. In the presence of $Ca^{2+}$, the reaction showed complete cleavage of substrate (A ~ 1) at high pH (pH ≥ 11.5) in an Ar atmosphere. Reactions carried out at the same pH but in the presence of atmospheric $CO_2$ showed similar rates, but gave only partial cleavage of substrate (amplitude~0.5-0.6). Thus, $CO_2$ present in air was found to affect the amplitude but not the rate of the reaction. Reaction performed under argon atmosphere at pH 11.0 in the presence of $Ca^{2+}$ went to 70% completion and so a value of 0.7 was used for $A$ in eq (2) for slow reactions (pH < 11.0).

The rate-pH profiles obtained were fitted to the following equation:

$$k_{\text{obs}} = \frac{k_{\text{max}}}{1+10^{n(\text{p}K_{\text{a}}-\text{pH})}} \tag{7.3}$$

where $k_{\text{max}}$ is the maximal observed rate constant, $n$ is the Hill coefficient, and the p$K_{\text{a}}$ corresponds to the value of pH at which $k_{\text{obs}}$ is equal to $k_{\text{max}}/2$. Fits were with non-linear least squares using a Marquadt algorithm Kaleidagraph (Synergy Software).

To test whether the p$K_{\text{a}}$ measured from the kinetic method was associated, at least in part, with the deprotonation of a calcium-bound water, we directly measured the p$K_{\text{a}}$ of the 2′OH of 3′AMP using NMR. We chose to use 3′AMP because a ribose linkage such as that in the above chimeric oligonucleotide would be cleaved at high pH during the NMR experiment. Similar studies on 3′AMP have been carried out previously by Chattopadhyaya and co-workers[50] but under different ionic conditions. NMR data were collected on a Bruker AV-3-600 spectrophotometer at 25° C. Even though it required water suppression, we chose to determine p$K_a$'s in 90% $H_2O$ solutions rather than 100% $D_2O$ because pure $D_2O$ solutions generally shift p$K_a$ values of N-linked protons higher by ~0.4-0.6 units.[10,51] The 3′AMP was from Sigma Aldrich (St. Louis, MO). Solutions of 3′AMP (~1-2 mM) were prepared in 10% $D_2O$ and contained 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) as an internal standard. Water suppression was achieved through presaturation using the noesygppr1d pulse sequence. Each NMR-detected pH titration consisted of 12-15 points, each of which was prepared independently. Appropriate volumes of sodium or calcium hydroxide were used to bring the solution to the required pH in a total volume of 500 mL. Throughout the titration, the total final concentration of sodium ions was kept constant at 0.5 M and that of calcium ions was kept constant at 10 mM by adding appropriate volumes of sodium chloride and calcium chloride. Sodium ion concentrations higher than 0.5 M, corresponding to pH 13.7, were avoided due to NMR tuning issues associated with high ionic strength solutions, while pH was limited to 12.2 in the presence of calcium ions to avoid precipitation of calcium hydroxide. The calcium hydroxide-containing samples were degassed under inert argon to avoid precipitation of calcium carbonate. Degassing by bubbling argon into the NMR tube or degassing the buffers-only gave similar results. The pH of each solution was measured after each NMR spectrum was acquired and was found to agree with calculated values. All solutions were prepared freshly on the day that the NMR experiments were performed. The chemical shift of the H1′ peak was monitored as a function of pH because this peak is relatively well isolated from the water peak.

Plots of observed chemical shift versus pH were used to determine the p$K_a$ according to the following equation

$$d_{\text{obs}} = d_{\text{A}} + \frac{d_{\text{AH}} - d_{\text{A}}}{1 + 10^{(\text{pH} - \text{p}K_a)}} \tag{7.4}$$

where $d_A$ is the chemical shift at high pH, $d_{AH}$ is the chemical shift at low pH, and $d_{obs}$ is the observed chemical shift at a given pH. As stated above, the highest pH attainable in the presence of 0.5 M $Na^+$ was 13.7, and that in the presence of 10 mM $Ca^{2+}$ was 12.2. Due to incomplete ionization of the 2'OH of AMP at these pH values, the high pH baseline was not well defined for either $Na^+$ or $Ca^{2+}$ experiments. To determine the value of $d_A$, we used a linearized form of Eq. 4 as previously described[50]

$$d_{\mathrm{obs}} = d_{\mathrm{A}} + \left(d_{\mathrm{AH}} - d_{\mathrm{obs}}\right)\left(\mathrm{a}_{\mathrm{H}^+}/K_{\mathrm{a}}\right) \tag{7.5}$$

where $a_{H+}$ is the activity of the proton and $K_a$ is the acid dissociation constant. A plot of $d_{obs}$ versus $(d_{AH}\text{-}d_{obs})*a_{H+}$ gives a straight line with a y-intercept of $d_A$. This value of $d_A$ was then used in Eq. 4 to obtain the p$K_a$.

To compare the change in protonation as a function of pH for sodium and calcium ions on the same scale, the fraction of protonated species was plotted as a function of pH according to the following equation:

$$f_{\mathrm{AH}} = \frac{d_{\mathrm{obs}} - d_{\mathrm{A}}}{d_{\mathrm{AH}} - d_{\mathrm{A}}} \tag{7.6}$$

where $f_{AH}$ is the fraction of protonated species and the other parameters were determined from Eq. 4.

## Results and Discussion

As discussed in the Introduction, our previous QM/MM calculations[19] implicated a concerted reaction pathway for the HDV ribozyme when a $Mg^{2+}$ ion is at the catalytic site. Figure 7.3 depicts the initial (dashed) and final (solid) strings from the three sets of simulations with $Mg^{2+}$ at the catalytic site: A (green), B (purple), and C (red). The strings are projected in the two-dimensional (2D) space of the collective reaction coordinates (r1-r2) and (r3-r4). As shown in Figure 7.2, (r1-r2) is associated with proton transfer from N3(C75) to O5′(G1), and (r3-r4) is associated with the P(G1)-O2′(U-1)/P(G1)-O5′(G1) bond forming/breaking. Figure 7.3 indicates that the strings from all three sets converged to a similar reaction pathway, corresponding to a concerted mechanism (i.e., a predominantly straight line along the diagonal in the 2D space of these two collective reaction coordinates). The initial string for set B (dashed purple curve) is in a very different region of configurational space than the initial strings for sets A and C (dashed green and red curves) because the mechanism associated with the initial string for set B has an intermediate with the proton transferred to the 5′-bridging oxygen. In contrast, the mechanisms associated with the initial strings for sets A and C have a phosphorane-like TS and phosphorane intermediate, respectively, which are located in similar regions of configurational space. The converged strings for sets A and C (solid green and red curves) are nearly indistinguishable. To conserve computational resources, the string calculations for set B were terminated after the reaction pathway, indicated by the latest string (solid purple curve), was determined to be qualitatively similar to that obtained from sets A and C and continuing to evolve in that direction. Details regarding the convergence of the strings are provided in the SI.

The 2D free energy profile of the self-cleavage reaction obtained from set A is depicted in Figure 7.4A. The 2D surface is constructed by projecting the free energy along the (r1-r2) and (r3-r4) collective reaction coordinates. The black dashed and solid lines indicate the initial and converged strings, respectively. The free energy surface implicates a concerted mechanism for the reaction without stable intermediates. Specifically, the MFEP (solid black line) is a predominantly straight line along the diagonal with a single maximum in free energy. The 1D free energy profile along the converged string, which corresponds to the MFEP, is depicted in

Figure 7.4B. The free energy barrier along this string is ~13 kcal/mol. This barrier is consistent with the experimental observation that the intrinsic rate constant of the HDV ribozyme catalytic reaction is $10^2$-$10^4$ $s^{-1}$ when the general base is extrapolated to the fully functional deprotonated O2′ form as used in the reactant state for the calculations.[10] This range of experimental rate constants corresponds to a free energy barrier of 11-14 kcal/mol using transition state theory (TST), assuming a prefactor of ~1 $ps^{-1}$ in the TST rate constant expression. In this qualitative analysis, we are neglecting the corrections associated with dynamical recrossings and Jacobian factors related to the choice of reaction coordinates, but these effects are not expected to significantly alter the qualitative rate constant. The free energy profiles along the converged strings for sets B and C are similar and are provided in the SI.

Figure 7.4C depicts the changes in the most important reaction coordinates involved in the self-cleavage reaction along the MFEP. The reaction coordinates shown here are as follows: r1 (the N3-H3 distance), r2 (the O5′-H3 distance), r3 (the P-O5′ distance), and r4 (the P-O2′ distance), as depicted in Figure 7.2. The other eight reaction coordinates do not change significantly along the MFEP, as depicted in Figure F.7 in Supporting Information. The mechanism of the reaction is further illustrated in terms of representative structures along the MFEP in Figure 7.5. According to Figure 7.4B, the cleavage reaction in the HDV ribozyme is concerted, evolving from the reactant state to the product state through a single TS. Furthermore, Figures 7.4C and 5 provide detailed mechanistic information along the MFEP. In the early stages of the reaction (i.e., over the first three points in Figure 7.4C), the r2 coordinate (blue curve in Figure 7.4C) decreases, mainly due to a decrease in the proton donor-acceptor distance (i.e., decrease of the distance between O5′ and N3). Simultaneously, the r4 coordinate decreases as the O2′ begins to attack the scissile phosphate. In Figure 7.4A, these two movements are manifested as an initial increase in (r1-r2), as well as an increase in (r3-r4), in the early stages of the MFEP to prime the system prior to the overall concerted mechanism. Following this initial preparation, the O2′ continues to move toward the phosphorus, as indicated by the further decrease of the r4 coordinate in Figure 7.4C. This attack, as depicted in structure 2 of Figure 7.5, primes the local geometry for the subsequent proton transfer from N3(C75) to O5′(G1). Subsequently, movements along (r3-r4) and (r2-r1) are concerted, as manifested by the diagonal, nearly linear MFEP with slope of ~1 between -0.5 and +0.5 Å for the two collective coordinates (Figure 7.4A, solid line). Moreover, the scissile phosphate evolves into a phosphorane-like geometry in the TS region, as depicted in structure 3 of Figure 7.5. In the TS region, r4 ≈ r3, suggesting that the P-O2′ and P-O5′ bonds are half formed and half broken, and r1 ≈ r2, suggesting that the proton H3 is midway between the donor N3(C75) and the acceptor O5′(G1). In other words, the two collective reaction coordinates, (r1-r2) and (r3-r4), are both ≈ 0 in the TS region. Finally, the proton fully transfers to the O5′ of G1, as shown in structure 4 of Figure 7.5, and subsequently the P-O2′ and P-O5′ bonds are completely formed and broken, respectively, as shown in structure 5 of Figure 7.5. We emphasize that the structures in Figure 7.5 do not represent stable intermediates along the reaction pathway but rather represent selected structures along a concerted pathway comprised of a single TS between two stable minima.

The MFEP described above is consistent with the results from our previous theoretical study[19] with $Mg^{2+}$ bound at the catalytic site. In that study, we used QM/MM methods to generate a minimum energy path (MEP) for the HDV ribozyme self-cleavage reaction with a $Mg^{2+}$ ion at the catalytic site. The MEP was generated by identifying a TS (i.e., a first-order saddle point) on the potential energy surface and following the paths of steepest descent to the reactant and product minima. When $Mg^{2+}$ was bound at the catalytic site, the mechanism was found to be concerted with a phosphorane-like TS. We searched for stable intermediates but were unsuccessful for this case. One key limitation associated with this previous study was the lack of conformational sampling and hence the neglect of entropic contributions. In the present study, the MFEP was also determined to correspond to a concerted mechanism with a phosphorane-like TS.

Although both the MEP and the MFEP correspond to an overall concerted mechanism, we observed certain differences in the specific pathways followed. The most important difference is that the previously determined MEP (dashed black line in Figure 7.4A) was characterized as concerted but asynchronous. Specifically, the proton transfer coordinate in the MEP remained nearly constant during the initial O2′ attack on the scissile phosphate, which was followed by proton movement with a nearly constant P(G1)-O2′(U-1) distance to generate a phosphorane-like TS with the proton partially transferred. This reaction pathway was still considered to be concerted, however, because there was a single TS. In contrast, the MFEP is more synchronous, as indicated by the solid black line in Figure 7.4A and by Figures 7.4C and 5. Specifically, in the TS region of Figure 7.4A, the MFEP (solid line) is nearly linear with a slope of ~1, whereas the MEP (dashed line) is nearly linear with a slope of ~0. Moreover, the distance between the proton donor, C75(N3), and the proton acceptor, G1(O5′), decreases from ~3.0 Å in structure 1 to ~2.7 Å in structure 2 of Figure 7.5. The previous MEP calculations did not exhibit this decrease in the proton donor-acceptor distance during the initial O2′ attack on the scissile phosphate. This decrease in the proton donor-acceptor distance is likely to facilitate the proton transfer, leading to earlier proton motion and a more synchronous mechanism.

Furthermore, the MFEP indicates that the reaction is modestly exergonic, whereas the previously determined MEP indicated that the reaction is highly endothermic. Although a direct comparison of free energy and potential energy profiles is not meaningful, the entropic differences between the reactant and product states are unlikely to be large enough to account for this difference. The procedure used to generate the MEP probably resulted in a local minimum for the product state, whereas the finite temperature string umbrella sampling method used to generate the MFEP included conformational sampling and therefore was able to identify a more stable product structure, as shown in Figures 7.4A and 7.4B. Note that these calculations do not include the separation of the U-1 product fragment from the rest of the ribozyme following self-cleavage, which would further impact the entropic contributions to the overall process.

In contrast to the concerted mechanism observed when a $Mg^{2+}$ ion is bound at the catalytic site, the previous QM/MM MEP calculations indicated that the mechanism is sequential with a phosphorane intermediate when a monovalent ion is at the catalytic site. The free energy simulations from set D are consistent with this hypothesis but, as discussed below, differ from the previous MEP calculations in the protonation state of the intermediate. The initial string for set D ($Na^+$ ion at the catalytic site) corresponded to a sequential mechanism with a phosphorane intermediate, similar to the initial string used for set C ($Mg^{2+}$ ion at the catalytic site). Figure 7.6 compares the initial string (dashed lines) and the converged string (solid lines) from set D in the presence of $Na^+$ (blue lines) to those from set C in the presence of $Mg^{2+}$ (red lines). Again the strings are projected in the 2D space of the collective coordinates (r1-r2) and (r3-r4). Although the initial strings are similar for set C ($Mg^{2+}$ ion at the catalytic site) and set D ($Na^+$ ion at the catalytic site), the converged MFEP for set C corresponds to a concerted mechanism with a phosphorane-like TS, as discussed above, whereas the converged MFEP for set D corresponds to a sequential mechanism with a phosphorane intermediate indicated by an additional minimum on the free energy surface, as discussed below. In Figure 7.6, the red and blue circles indicate the positions of the phosphorane-like TS structure and the phosphorane intermediate structure, respectively.

The 2D free energy surface obtained from set D is depicted in Figure 7.7A. The black dashed and solid lines indicate the initial and converged strings, respectively. In this case, the free energy surface exhibits a minimum corresponding to a phosphorane intermediate structure. Figure 7.7B depicts the 1D free energy profile along the MFEP and also illustrates that the self-cleavage reaction in the presence of a $Na^+$ ion at the catalytic site occurs by a sequential mechanism with a phosphorane intermediate. The first and second free energy barriers obtained from this free energy profile are ~3.5 kcal/mol and ~2 kcal/mol, respectively.

The detailed mechanism of the self-cleavage reaction with a $Na^+$ ion at the catalytic site is depicted in Figures 7.7C and 7.8. Initially the O2′ attacks the scissile phosphate to form a geometry in which the P-O2′ and P-O5′ bonds are half formed and half broken, respectively, as indicated by the decrease of the r4 coordinate (i.e., the P-O2′ distance) at the beginning of the MFEP in Figure 7.7C. Structure 2 of Figure 7.8 corresponds approximately to the TS for the first step. This part of the mechanism is similar to that observed in the MFEP obtained with a $Mg^{2+}$ ion at the catalytic site. Next the P-O2′ distance decreases further, and a proton is transferred from the exocyclic amine of C75 (N4 in Figure 7.8) to the nonbridging oxygen of the scissile phosphate, denoted the pro$R_P$ oxygen (O1P in Figure 7.8), to make a stable penta-coordinated monoanionic phosphorane structure (structure 3 of Figure 7.8). This structure corresponds to the intermediate minimum observed in the free energy profiles shown in Figures 7.7A and 7.7B. This part of the mechanism is different from that observed with a $Mg^{2+}$ ion at the active site and is also different from that observed in our previous MEP calculations with a $Na^+$ ion at the active site. Following this step, the P-O2′ distance decreases a bit further and the P-

O5′ distance increases substantially, leading to the transfer of the proton from the pro$R_P$ oxygen (O1P) back to N4 of C75, as well as initiating the proton transfer from N3(C75) to O5′(G1). Structure 4 of Figure 7.8 corresponds approximately to the TS for the second step (i.e., the proton transfer reaction from N3(C75) to O5′(G1). Finally, once the proton (H3) is transferred to O5′(G1), the P-O2′ and P-O5′ bonds are completely formed and broken, respectively.

These free energy calculations indicate that the mechanism is concerted when $Mg^{2+}$ is at the catalytic site and sequential when $Na^+$ is at the catalytic site. An explanation for these different mechanisms is that the proton H3 of $C75^+$ is less acidic with a monovalent ion than with a divalent ion at the catalytic site. Indeed, solution and crystallographic experiments have measured lower p$K_a$'s for C75 in the presence of $Mg^{2+}$ ions than in the presence of $Na^+$ ions.[10,11,27] In particular, kinetics experiment indicate that the p$K_a$ value of C75 decreases from 7.25 to 5.9 upon the binding of a $Mg^{2+}$ ion at the catalytic site[10,52] and to be ~7.5 in the presence of up to 1M $Na^+$ concentration without $Mg^{2+}$.[27] As a result, the proton H3 of $C75^+$ is more difficult to transfer from N3(C75) to O5′(G1) when $Na^+$ is at the catalytic site, and a protonated phosphorane monoanion intermediate is formed prior to this proton transfer reaction.

A new aspect of the mechanism illustrated by these free energy calculations is the transfer of a proton from the exocyclic amine of C75 to the pro-$R_P$ oxygen to generate a protonated phosphorane monoanion intermediate. There is experimental support for such proton transfer occurring to generate a monoanionic phosphorane in the monovalent ion mechanism of the genomic HDV ribozyme. While proton inventory measurements on this ribozyme in the presence of $Mg^{2+}$ gave an inventory of 2, consistent with a concerted reaction mechanism,[53] those in the presence of 1 M $Na^+$ gave an inventory of 1,[27] supporting a stepwise pathway such as that shown in Figure 7.7. In addition, we observed a normal thio effect at the pro-$R_P$ oxygen in the presence of 1 M $Na^+$, where a thio effect is the ratio of rate constants with oxo and thio at this oxygen, $k_O/k_S$.[22] Observation of a thio effect supports the importance of the role of the pro-$R_P$ oxygen in the mechanism in the absence of divalent ion coordination and thus is consistent with the stepwise mechanism. We also note that studies by Anslyn and Breslow on dinucleotide cleavage in the presence of imidazole buffer and no divalent ions support a similar stepwise mechanism with a protonated, monoanionic phosphorane intermediate.[54] These observations suggest that divalent ions, in addition to acidifying the general acid, interact with the pro-$R_P$ oxygen and stabilize charge development at the scissile phosphate, preventing proton transfer to a non-bridging oxygen in the HDV ribozyme mechanism. This idea is supported by recent metal ion rescue experiments that establish an interaction between the pro-$R_P$ oxygen and divalent metal ion in the transition state.[22] Moreover, it is notable that the $Mg^{2+}$ remains within 2.1 Å of the pro-$R_P$ oxygen throughout the MFEP.

The proton transfer from C75 N4 to the pro-$R_P$ oxygen (O1P) observed during formation of the monoanionic phosphorane intermediate with $Na^+$ at the catalytic site is reasonable on the basis of energetic considerations. According to high level gas phase quantum calculations,[55,56] the neutral imino tautomer of cytosine shown in structure 3 of Figure 7.8 has been estimated to be only slightly higher in energy (0.8 kcal/mol) than the standard amino tautomer of cytosine, suggesting that this species is energetically accessible. Secondly, the $pK_a$ of the amino group (N4) of cytosine is significantly acidified, from a $pK_a$ of 18 to 9, upon protonation at the imino nitrogen (N3) as determined by proton exchange NMR measurements.[57-59] Estimates of the $pK_a$ of the nonbridging oxygens of the dianionic phosphorane species are near 14.[60,61] Thus, protonation at the N3 of C75 could acidify the N4 enough to allow proton transfer to the pro-$R_P$

oxygen, which is basic in the phosphorane intermediate. Protonation at one atom facilitating transfer at another could be a general catalytic strategy for enhancing ribozyme mechanism.[62]

The rate of the HDV ribozyme catalytic reaction has been experimentally measured to be ~25-fold higher in the presence of $Mg^{2+}$ ions than in the presence of only $Na^{+}$ ions.[21] However, the QM/MM free energy simulations indicate that the free energy barrier for the self-cleavage reaction is significantly higher with a $Mg^{2+}$ ion at the catalytic site than with a $Na^{+}$ ion at the catalytic site. These free energy barriers are only qualitatively meaningful, but nevertheless this substantial difference in free energy barriers suggests that the rate constant for self-cleavage should be smaller in the presence of $Mg^{2+}$ ions. This result from the simulations can be reconciled with the experimental data by considering the initial deprotonation of the 2′OH nucleophile. On the basis of simple electrostatic arguments concerning stabilization of the negatively charged O2′, the initial deprotonation of 2′OH is expected to be more thermodynamically favorable with $Mg^{2+}$ than with $Na^{+}$.

To test this hypothesis, we measured the $pK_a$ of the O2′ in the presence of monovalent and divalent ions using kinetic and NMR methods. First, we used the kinetic method of Li and Breaker,[48] in which we utilized a chimeric 22mer with a single internal ribose linkage. This oligonucleotide was 5′-end labeled and its self-cleavage monitored as a function of pH in the presence of various monovalent and divalent ions. Li and Breaker reported $pK_a$ values in the presence of 0.5 to 3.16 M $K^{+}$ but not in the presence of $Mg^{2+}$ owing to precipitation of magnesium hydroxide above pH~9. We repeated their studies in 3.16 M $K^{+}$ and then measured the $pK_a$ in 3.16 M $Na^{+}$, which is the monovalent ion we used in prior HDV ribozyme experiments. We then conducted measurements in the presence of $Ca^{2+}$ as the divalent ion, as its hydroxide has a higher $K_{sp}$ than magnesium hydroxide, allowing studies up to a pH of 12.3

equivalent to 10.0 mM $Ca(OH)_2$. We previously observed that HDV ribozyme kinetics are similar in $Mg^{2+}$ and $Ca^{2+}$, with nearly indistinguishable rate-pH profiles and slightly faster rates in the presence of identical concentrations of $Ca^{2+}$, suggesting that this is a reasonable approach.[10,49] Molar amounts of monovalent ion and millimolar amounts of divalent ion correspond to the concentrations of these ions used in our HDV ribozyme experiments.[10,21]

Sample PAGE gels and kinetic traces in the presence of $Na^+$ and $Ca^{2+}$ are provided in Figure F.8, and rate-pH profiles are provided in Figure 7.9. Cleavage experiments yielded a single product by PAGE analysis and reactions went to completion (Figure F.8). As shown in Figure 7.9A, a plot of rate versus pH in the presence of 3.16 M $K^+$ has a sigmoidal shape. Fitting of this plot to a standard rate-pH profile equation (Eq. (7.3)) yielded a p$K_a$ of 13.3 in 3.16 M $K^+$, in good agreement with Li and Breaker,[48] and measurements in the presence of 3.16 M $Na^+$ provided a p$K_a$ of 13.6, indicating little dependence of the p$K_a$ of the O2′ on monovalent ion identity as expected. Our experiments and simulations focus on $Na^+$. Extrapolation of the p$K_a$ value in $Na^+$ from 3.16 to 1 M monovalent ion conditions,[48] which were used in our earlier experiments on the genomic HDV ribozyme,[10,21] leads to a p$K_a$ for the O2′ ~0.5 units higher at ~14.1. The rate versus pH profiles in the presence of $Ca^{2+}$ yielded a much lower p$K_a$ of 11.4 (Figure 7.9). Thus, using kinetic measurements we found that the p$K_a$ of the O2′ is shifted ~2.7 units lower in the presence of millimolar amounts of divalent ions as compared to molar amounts of sodium ions.

The above kinetic method is not necessarily specific to the p$K_a$ of O2′ because deprotonation of a calcium-bound water that could act as a general base might also be detected. We thus utilized the approach used by Chattopadhyaya and co-workers[50] to directly measure the p$K_a$ of the O2' in 3′AMP using $^1H$ NMR. Although we used similar methods to theirs, we conducted our experiments in 90% $H_2O$ rather than 100% $D_2O$, kept ionic strength constant throughout the titration, and explored different cations and cation concentrations (see Materials and Methods). We first conducted the experiments in the presence of 0.5 M sodium ions. As reported previously,[50] the chemical shift versus pH profile did not yield a sigmoidal curve owing to incomplete ionization of the 2′OH at the highest pH attainable (Figure F.9A). We used Eq. 5 to estimate the chemical shift at high pH (Figure F.9C) and then fit the data in Figure F.9A to Eq. 4. This method yielded a $pK_a$ of 13.2 in the presence of 0.5 M $Na^+$, which is similar to that previously reported for 3′AMP in $Na^+$ by $^1H$ NMR , $pK_a$ 13.38[50].

Subsequently, we carried out NMR experiments in the presence of calcium ions. As in our kinetic assays, we kept the concentration of $Ca^{2+}$ constant at 10 mM throughout the titration. As in the presence of $Na^+$, the chemical shift value at high pH was estimated from Eq. 5 (Figure F.9D). In the presence of $Ca^{2+}$, the $pK_a$ was found to be shifted to 11.9 (Figure F.9B), which is in good agreement with the value of 11.4 from our kinetic assay under identical divalent ion conditions. Thus, using NMR we found that the $pK_a$ of the O2′ is shifted ~1.3 units lower in the presence of millimolar amounts of divalent ions as compared to molar amounts of sodium ions. This result is clearly visualized in the difference in the flex points of the NMR species plots provided in Figure 7.10. Additionally, the difference in $pK_a$ values in the presence of $Ca^{2+}$ and $Na^+$ determined by NMR is in reasonable agreement with the difference obtained by kinetics, especially considering that lower baselines had to be estimated for the NMR method. Both methods clearly show that deprotonation of the 2′OH is considerably more favored in $Ca^{2+}$ than $Na^+$. Lastly, we note that the $pK_a$ in the presence of $Ca^{2+}$ ions measured using the kinetic method may be a result of convolution due to the deprotonation of the 2′OH and a calcium-bound water both having similar $pK_a$ values. Indeed, the Hill coefficient from the fit of the kinetic data is close to 2 (Figure 7.9), suggesting that more than one deprotonation event is being measured, as opposed to the NMR experiment, which is specific to the nucleotide.

Overall, the experimental results are consistent with the hypothesis generated by the simulations, namely that 2′OH deprotonation is less favorable in the presence of only monovalent ions than when a divalent ion is bound at the catalytic site. For both types of ions, the shapes of the rate-pH profiles of the HDV ribozyme reaction reflect participation of C75 in the rate-limiting step; thus 2′OH deprotonation does not appear to be rate-limiting in either case, but rather is in rapid equilibrium. Thus, we can estimate the observed rate constant, $k_{\text{obs}}$, as the product of an equilibrium constant, $K_{\text{eq}}$, associated with O2′ activation and the rate constant, $k_{\text{cleavage}}$, associated with self-cleavage: $k_{\text{obs}} = K_{\text{eq}} k_{\text{cleavage}}$. We can use the following three observations to qualitatively estimate the impact of this pre-equilibrium on the observed rate constant: (1) the experimental observation of a ~25-fold slower overall rate in the case of monovalent ions alone;[21] (2) the experimentally measured O2′ p$K_a$ that is ~1.3-2.7 units lower for divalent ions than for sodium ions under typical reaction concentrations; and (3) the calculated free energy barriers for the self-cleavage reaction with $Mg^{2+}$ or $Na^+$ at the catalytic site. Given the calculated free energy barriers for the self-cleavage reaction (~13 kcal/mol and ~3.5 kcal/mol for divalent and monovalent ions, respectively), in conjunction with the p$K_a$ differences that favor deprotonation in the presence of $Mg^{2+}$ by ~1.8-3.8 kcal/mol, the 2′OH activation in the presence of $Na^+$ is expected to be disfavored by an additional ~8-10 kcal/mol to render the observed rate ~25-fold slower with $Na^+$. Such an effect could arise from an inability of $Na^+$ to effectively orient the O2′ for attack on the adjacent phosphorus. In contrast, $Mg^{2+}$ at the active site appears to act as a Lewis acid and directly coordinate the nucleophilic O2′ of U-1, which may provide not only enhanced acidity of the O2′, as indicated in the experiments herein, but also proper orientation for attack.[16,19] Along these lines, we note that the monovalent and divalent ions are within 2.3 and 2.1 Å respectively, of the deprotonated O2′ throughout the MFEPs, indicating closer approach of the divalent ion and greater stabilization throughout. Moreover, these distances represent *direct* interactions of the $Na^+$ and $Mg^{2+}$ with the O2' (i.e., without a water bridge); thus the differential energetics for dehydrating these ions and the O2′ within the active site may also influence the pre-equilibrium process.

Another effect that could contribute to the pre-equilibrium constant in the overall observed rate constant is the relative probability of a $Mg^{2+}$ ion versus a $Na^+$ ion occupying a binding site that allows interaction with the O2′. To investigate this effect, we analyzed a series of 25 nanosecond classical MD trajectories from our previous work.[20,38] Specifically, we analyzed two independent trajectories of the HDV ribozyme with $Mg^{2+}$ ion at the catalytic site and two independent trajectories in which the catalytic $Mg^{2+}$ ion was replaced by two $Na^+$ ions in the bulk. In the latter case, we observed that a $Na^+$ ion moved into the active site region during the equilibration stage. The charge isodensity plots illustrating positive charge in the active site region and the associated cumulative distribution functions for metal ion occupancy are depicted in Figures F.10 and F.11, respectively. These plots indicate that the $Mg^{2+}$ ion is highly localized with 100% occupancy of a binding site that allows strong interaction with the O2′ of U-1. In contrast, the $Na^+$ ion is more delocalized with multiple binding sites that often do not allow the ion to interact with the O2’. These observations suggest that the pre-equilibrium constant is significantly decreased for the $Na^+$ ion relative to the $Mg^{2+}$ ion due to the lower probability of the $Na^+$ ion occupying a binding site that allows it to interact with the O2′. It is interesting to note that Ke and co-workers solved a structure of the HDV ribozyme in the pre-cleaved state with the heavy monovalent ion $Tl^+$ and found that the ion interacts in a non-productive way with the O2′

and pro$S_P$ oxygen, which would be consistent with an unfavorable conformation in the presence of monovalent ions.[63]

To complete this study, we performed additional QM/MM molecular dynamics simulations to investigate the plausibility of a hydroxide ion bound to the catalytic $Mg^{2+}$ ion acting as a general base and activating the O2′ of U-1. The starting structures for these simulations were obtained from the solvated pre-cleaved crystal structure following equilibration of the solvent and ions. In these starting structures, the O2′ of U−1 was protonated, the $Mg^{2+}$ ion was directly coordinated to the O2′, and one of the $Mg^{2+}$-coordinated crystallographic water molecules was replaced by a hydroxide ion. The QM region consisted of the U−1 sugar, the phosphate, the atoms C5′ and O5′ of G1, the catalytic $Mg^{2+}$ ion, the two $Mg^{2+}$-coordinating water molecules, and the hydroxide ion. Two independent molecular dynamics trajectories, in which different water molecules were replaced by a hydroxide ion, were propagated for 300 fs and 900 fs, respectively. In both trajectories, we observed that the hydroxide ion moved to a position that enabled it to interact strongly with the O2′, either by moving spatially or by a Grötthuss-type mechanism,[64,65] and subsequently abstracted the proton from the O2′. Movies of these two trajectories are available. These QM/MM molecular dynamics trajectories suggest that a hydroxide ion ligated to the catalytic $Mg^{2+}$ can potentially act as a general base, thereby activating the O2′. A dual role of a partially hydrated $Mg^{2+}$ ion as a Lewis acid and a Brønsted base, such as that observed here, is consistent with earlier experiments and models,[10,25] and is also consistent with the kinetic-based p$K_a$ studies conducted herein, where a Hill coefficient of approximately two was observed.

## Concluding Remarks

We carried out QM/MM free energy simulations to investigate the mechanism of the self-cleavage reaction in the HDV ribozyme. We explored the possibility of three different reaction mechanisms with a $Mg^{2+}$ ion bound at the catalytic site. Our simulations indicated that the self-cleavage reaction for this case is concerted with a phosphorane-like TS. The free energy barrier of the reaction along the concerted pathway is ~13 kcal/mol, which is consistent with values extrapolated from experimental studies. In contrast to our previous MEP calculations, the inclusion of conformational sampling and entropic effects resulted in a modestly exergonic reaction and a concerted mechanism that is synchronous, with the proton transfer and oxygen-phosphorus bond breaking/forming occurring simultaneously. Furthermore, although these free energy simulations were initiated with the O2′ already deprotonated, separate QM/MM molecular dynamics trajectories beginning with a protonated O2′ provided evidence that a $Mg^{2+}$-bound hydroxide ion could potentially activate the O2′ via deprotonation, thereby suggesting a Brønsted base as well as a Lewis acid role for the catalytic partially hydraged $Mg^{2+}$.

We also investigated the reaction with a $Na^{+}$ ion instead of a $Mg^{2+}$ ion at the catalytic site. Our results indicated that the self-cleavage reaction is sequential in this case, passing through a phosphorane intermediate. The free energy barriers along this sequential path are much lower, ~3.5 kcal/mol and 2 kcal/mol for the first and second steps, respectively. In contrast to our previous MEP calculations, we observed a proton transfer from the N4-exocyclic amine of N3-protonated C75 to the nonbridging pro-$R_P$ oxygen of the scissile phosphate to stabilize the phosphorane intermediate. This new observation is consistent with previous experimental data.

To explain the experimental observation that the rate is ~25-fold slower with $Na^{+}$ than with $Mg^{2+}$, we hypothesized that the activation of the O2′ nucleophile of the U-1 nucleobase by deprotonation and orientation is more disfavored when $Na^+$ is at the catalytic site than when $Mg^{2+}$ is bound at the catalytic site. In addition to providing greater electrostatic stabilization of the negatively charged O2′, $Mg^{2+}$ directly coordinates to the O2′ as a Lewis acid, thereby increasing its acidity and potentially facilitating an in-line attack. This hypothesis is supported by our experimental measurements of the $pK_a$ of the O2′ nucleophile in the presence of monovalent and divalent ions ions. In particular, the $pK_a$ of the O2′ was measured to be ~1.3-2.7 units lower in the presence of divalent rather than monovalent ions. In addition, analysis of the metal ion distribution functions in the active site region for classical MD simulations indicates that $Na^+$ populates the site near the O2′ much less frequently than does $Mg^{2+}$. Overall, the combined theoretical and experimental data suggest that the initial 2′OH deprotonation step is in rapid equilibrium, with a significantly lower equilibrium constant in the presence of monovalent ions compared to divalent ions.

Thus, the catalytic partially hydrated $Mg^{2+}$ ion may play multiple key roles in the mechanism: assisting in the activation of the O2′ nucleophile of the U-1 nucleobase by Lewis acid and Brønsted base catalysis, acidifying the general acid C75, and stabilizing the nonbridging oxygen so as to prevent proton transfer to it. The studies herein indicate that the proton may transfer to a nonbridging oxygen in the phosphorane intermediate in the absence of divalent ions, most likely because the above roles are altered without divalent ions present. Further experiments relating rates to specific atomic substitutions at active site residues will be performed to examine the significance of these functional roles.

## Acknowledgements

This work was supported by NIH Grant GM056207 to S.H.-S. and NIH Grant R01GM095923 to P.C.B. We thank Squire Booker and Brad Landgraf for assistance with Ar-atmosphere experiments, and Chad Lawrence and Emmanuel Hatzakis for help with the NMR experiments. We thank Alexander Soudackov, Gerhard Hummer, Yihan Shao and H. Lee Woodcock for useful discussions. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575.

## Supporting Information

Information provided in Appendix F include two-dimensional free energy surfaces and mechanistic analysis for Simulation Sets B and C; convergence analysis for Simulation Sets A, B, C, and D; average values of all reaction coordinates for the final iteration of Simulation Sets A, B, C, and D; autoradiograms and plots of the chimeric substrate cleavage with respect to time in the presence of $Na^+$ and $Ca^{2+}$; determination of the p$K_a$ of the 2’OH of 3’AMP by $^1$H NMR; charge-isodensity plots of metal ions at the active site; cumulative distribution functions of the occupancy of metal ions at the active site.

## References

(1) Fedor, M. J.; Williamson, J. R. *Nat. Rev. Mol. Cell. Biol.* **2005**, *6*, 399-412.

(2) Golden, B. L.; Hammes-Schiffer, S.; Carey, P. R.; Bevilacqua, P. C. An Integrated Picture of HDV Ribozyme Catalysis. In *Biophysics of RNA Folding*; Russell, R., Ed.; Springer New York, 2013; Vol. 3; pp 135-67.

(3) Bevilacqua, P. C.; Yajima, R. *Curr. Opin. Chem. Biol.* **2006**, *10*, 455-64.

(4) Lilley, David M. J. *Biochem. Soc. T* **2011**, *39*, 641-46.

(5) Been, M. D. HDV Ribozymes. In *Curr. Top. Microbiol. Immunol.*; Casey, J. L., Ed.; Springer Berlin Heidelberg, 2006; Vol. 307; pp 47-65.

(6) Salehi-Ashtiani, K.; Lupták, A.; Litovchick, A.; Szostak, J. W. *Science* **2006**, *313*, 1788-92.

(7) Webb, C.-H. T.; Riccitelli, N. J.; Ruminski, D. J.; Lupták, A. *Science* **2009**, *326*, 953-53.

(8) Webb, C.-H. T.; Luptak, A. *RNA Biol.* **2011**, *8*, 719-27.

(9) Perrotta, A. T.; Shih, I. h.; Been, M. D. *Science* **1999**, *286*, 123-26.

(10) Nakano, S.; Chadalavada, D. M.; Bevilacqua, P. C. *Science* **2000**, *287*, 1493-97.

(11) Gong, B.; Chen, J.-H.; Chase, E.; Chadalavada, D. M.; Yajima, R.; Golden, B. L.; Bevilacqua, P. C.; Carey, P. R. *J. Am. Chem. Soc.* **2007**, *129*, 13335-42.

(12) Guo, M.; Spitale, R. C.; Volpini, R.; Krucinska, J.; Cristalli, G.; Carey, P. R.; Wedekind, J. E. *J. Am. Chem. Soc.* **2009**, *131*, 12908-09.

(13) Gong, B.; Klein, D. J.; Ferre-D'Amare, A. R.; Carey, P. R. *J. Am. Chem. Soc.* **2011**, *133*, 14188-91.

(14) Liberman, J. A.; Guo, M.; Jenkins, J. L.; Krucinska, J.; Chen, Y.; Carey, P. R.; Wedekind, J. E. *J. Am. Chem. Soc.* **2012**, *134*, 16933-36.

(15) Das, S. R.; Piccirilli, J. A. *Nat. Chem. Biol.* **2005**, *1*, 45-52.

(16) Chen, J.-H.; Yajima, R.; Chadalavada, D. M.; Chase, E.; Bevilacqua, P. C.; Golden, B. L. *Biochemistry* **2010**, *49*, 6508-18.

(17) Banas, P.; Rulisek, L.; Hanosova, V.; Svozil, D.; Walter, N. G.; Sponer, J.; Otyepka, M. *J. Phys. Chem. B* **2008**, *112*, 11177-87.

(18) Lee, T.-S.; Giambasu, G. M.; Harris, M. E.; York, D. M. *J. Phys. Chem. Lett.* **2011**, *2*, 2538-43.

(19) Ganguly, A.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Phys. Chem. Lett.* **2011**, *2*, 2906-11.

(20) Veeraraghavan, N.; Ganguly, A.; Golden, B. L.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Phys. Chem. B* **2011**, *115*, 8346-57.

(21) Nakano, S.; Proctor, D. J.; Bevilacqua, P. C. *Biochemistry* **2001**, *40*, 12022-38.

(22) Thaplyal, P.; Ganguly, A.; Golden, B. L.; Hammes-Schiffer, S.; Bevilacqua, P. C. *Biochemistry* **2013**, *52*, 6499-514.

(23) DeRose, V. J. *Curr. Opin. Struc. Biol.* **2003**, *13*, 317-24.

(24) Hougland, J.; Kravchuk, A.; Herschlag, D.; Piccirilli, J. *PLoS Biol.* **2005**, *3*, e277-1548.

(25) Chen, J.; Ganguly, A.; Miswan, Z.; Hammes-Schiffer, S.; Bevilacqua, P. C. *Biochemistry* **2013**, *52*, 557-67.

(26) Golden, B. L. *Biochemistry* **2011**, *50*, 9424-33.

(27) Cerrone-Szakal, A. L.; Siegfried, N. A.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2008**, *130*, 14504-20.

(28) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2000**, *105*, 569-85.

(29) Faradjian, A. K.; Elber, R. *J. Chem. Phys.* **2004**, *120*, 10880-89.

(30) Zheng, H.; Zhang, Y. *J. Chem. Phys.* **2008**, *128*, 204106-8.

(31) Pan, A. C.; Sezer, D.; Roux, B. *J. Phys. Chem. B* **2008**, *112*, 3432-40.

(32) Zhu, F.; Hummer, G. *Proc. Nat. Acad. Sci.* **2010**, *107*, 19814-19.

(33) Rosta, E.; Nowotny, M.; Yang, W.; Hummer, G. *J. Am. Chem. Soc.* **2011**, *133*, 8934-41.

(34) Lans, I.; Medina, M.; Rosta, E.; Hummer, G.; Garcia-Viloca, M.; Lluch, J. M.; González-Lafont, À. *J. Am. Chem. Soc.* **2012**, *134*, 20544-53.

(35) Wojtas-Niziurski, W.; Meng, Y.; Roux, B.; Bernèche, S. *J. Chem. Theory. Comput.* **2013**, *9*, 1885-95.

(36) Ren, W. Q.; Vanden-Eijnden, E. *J. Phys. Chem. B* **2005**, *109*, 6688-93.

(37) Torrie, G. M.; Valleau, J. P. *J. Comp. Phys.* **1977**, *23*, 187-99.

(38) Veeraraghavan, N.; Ganguly, A.; Chen, J.-H.; Bevilacqua, P. C.; Hammes-Schiffer, S.; Golden, B. L. *Biochemistry* **2011**, *50*, 2672-82.

(39) Veeraraghavan, N.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Mol. Biol.* **2010**, *402*, 278-91.

(40) Veeraraghavan, N.; Ganguly, A.; Chen, J.; Bevilacqua, P. C.; Hammes-Schiffer, S.; Golden, B. L. *Biochemistry* **2011**, *50*, 2672-82.

(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-35.

(42) *D. E. Shaw Research, New York, NY* **2009**.

(43) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-97.

(44) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089-92.

(45) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J. r.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O?Neill, D. P.; DiStasio Jr, R. A.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Yeh Lin, C.; Van Voorhis, T.; Hung Chien, S.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Min Rhee, Y.; Ritchie, J.; Rosta, E.; David Sherrill, C.; Simmonett, A. C.; Subotnik, J. E.; Lee Woodcock Iii, H.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer Iii, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172-91.

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

(47) Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011-21.

(48) Li, Y.; Breaker, R. R. *J. Am. Chem. Soc.* **1999**, *121*, 5364-72.

(49) Nakano, S.; Cerrone, A.; Bevilacqua, P. C. *Biochemistry* **2003**, *42*, 2982-94.

(50) Velikyan, I.; Acharya, S.; Trifonova, A.; Földesi, A.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **2001**, *123*, 2893-94.

(51) Schowen, B. K.; Schowen, R. L. *Method enzymol* **1982**, *87*, 551-606.

(52) Nakano, S.; Bevilacqua, P. C. *Biochemistry* **2007**, *46*, 3001-12.

(53) Nakano, S.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2001**, *123*, 11333-34.

(54) Anslyn, E.; Breslow, R. *J. Am. Chem. Soc.* **1989**, *111*, 4473-82.

(55) Kobayashi, R. *J. Phys. Chem. A* **1998**, *102*, 10813-17.

(56) Šponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2001**, *61*, 3-31.

(57) McConnell, B. *Biochemistry* **1978**, *17*, 3168-76.

(58) McConnel, B.; Politowski, D. *Biophys. Chem.* **1984**, *20*, 135-48.

(59) Guéron, M.; Leroy, J.-L. *Method enzymol* **1995**, *261*, 383-413.

(60) Davies, J. E.; Doltsinis, N. L.; Kirby, A. J.; Roussev, C. D.; Sprik, M. *J. Am. Chem. Soc.* **2002**, *124*, 6594-99.

(61) Lopez, X.; Schaefer, M.; Dejaegere, A.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 5010-18.

(62) Wilkox, J. L.; Ahluwalia, A. K.; Bevilacqua, P. C. *Acc. Chem. Res.* **2011**, *44*, 1270-79.

(63) Ke, A.; Ding, F.; Batchelor, J. D.; Doudna, J. A. *Structure* **2007**, *15*, 281-87.

(64) Grotthus, C. *Ann. Chim.* **1806**, *58*, 54.

(65) Agmon, N. *Chem. Phys. Lett.* **1995**, *244*, 456-62.

## FIGURES

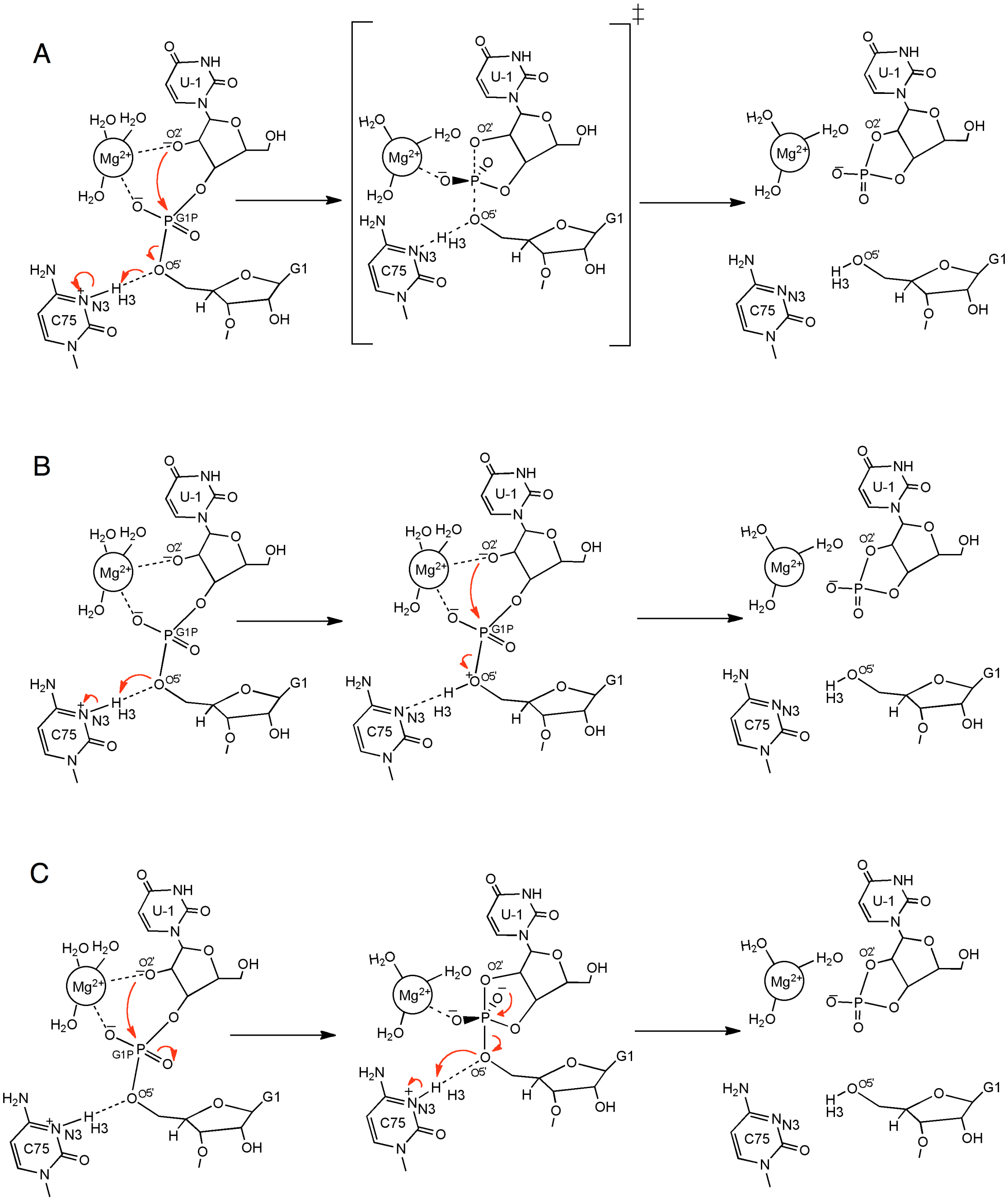

**Figure 7.1.** Plausible reaction pathways explored in this study. (A) The concerted mechanism, in which the reaction passes through a phosphorane-like TS, which is depicted as the middle structure. The minimum energy pathway obtained in our previous study[19] corresponds to this pathway. (B) A sequential pathway that passes through a bridging O5′ proton-transferred intermediate, which is depicted as the middle structure. (C) A sequential pathway that passes

through a non-protonated phosphorane intermediate, which is depicted as the middle structure. Note that all of the pathways are reversible but are shown only in one direction for clarity.

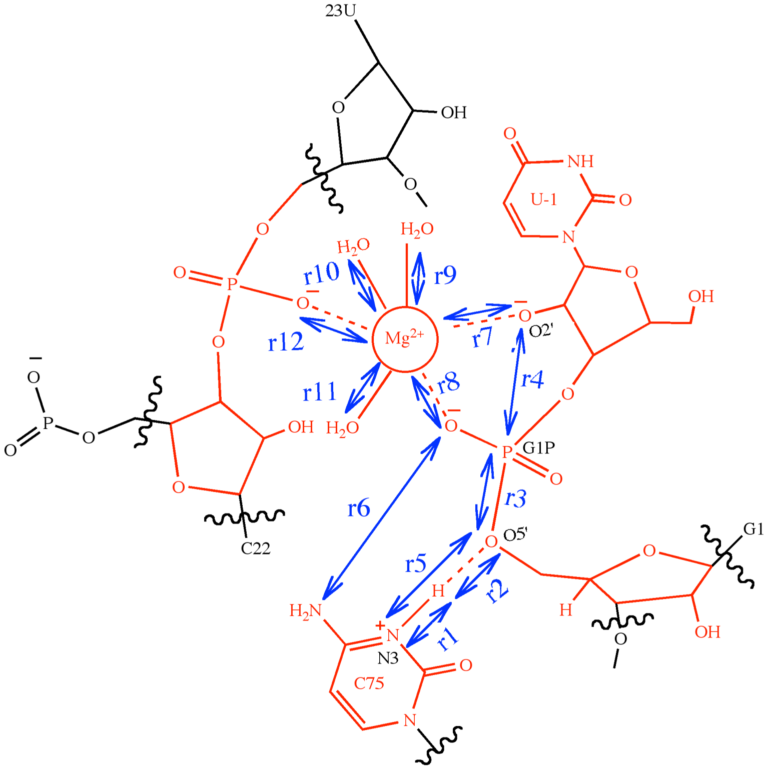

**Figure 7.2.** Schematic illustration of the active site of the genomic HDV ribozyme and the reaction coordinates included in the string calculations. The QM region for the QM/MM calculations consists of the 87 atoms shown in red, including the U-1 base and sugar, the G1 phosphate (the scissile phosphate) and sugar, the protonated C75 base, the U23 phosphate, the sugar of C22, the active site $Mg^{2+}$ ion, and the three crystallographic waters coordinating to the $Mg^{2+}$ ion. The reaction coordinates included in the string calculations are indicated by the blue double-headed arrows. In total, 12 active site distances were considered, although r1, r2, r3, and r4 were found to be the most important. The reaction paths are typically projected onto collective coordinates, the difference of r1 and r2 (r1 – r2) and r3 and r4 (r3 – r4).

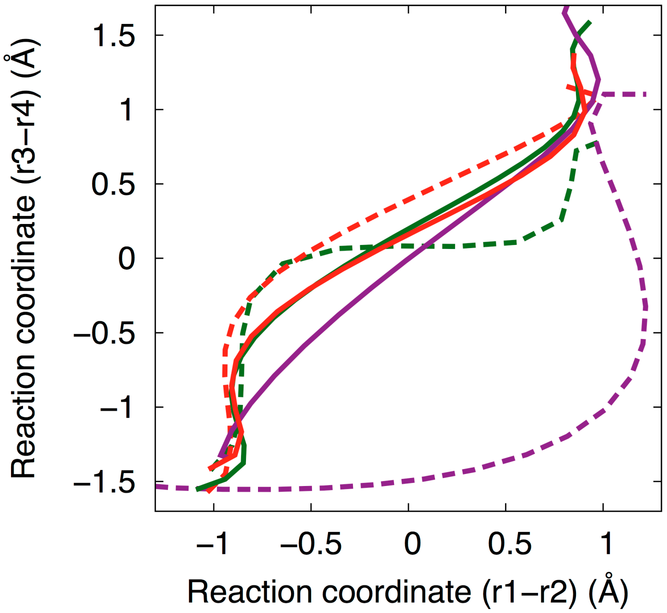

**Figure 7.3.** The initial (dashed lines) and final (solid lines) strings from the three sets of simulations with $Mg^{2+}$ bound at the catalytic site: A (green, initial path is concerted mechanism), B (purple, initial path is sequential with proton transferred intermediate), and C (red, initial path is sequential with phosphorane intermediate). The three initial paths are depicted in Figure 7.1, but the final paths all correspond qualitatively to pathway A in Figure 7.1. The strings are projected in the space of the collective coordinates (r1-r2) and (r3-r4), corresponding to the proton transfer reaction to the O5′ and the oxygen-phosphorus bond breaking/forming, respectively. The strings from sets A and C have converged to a similar path. The string from set B is still evolving, as described in the text, but is already close to the converged strings from A and C.

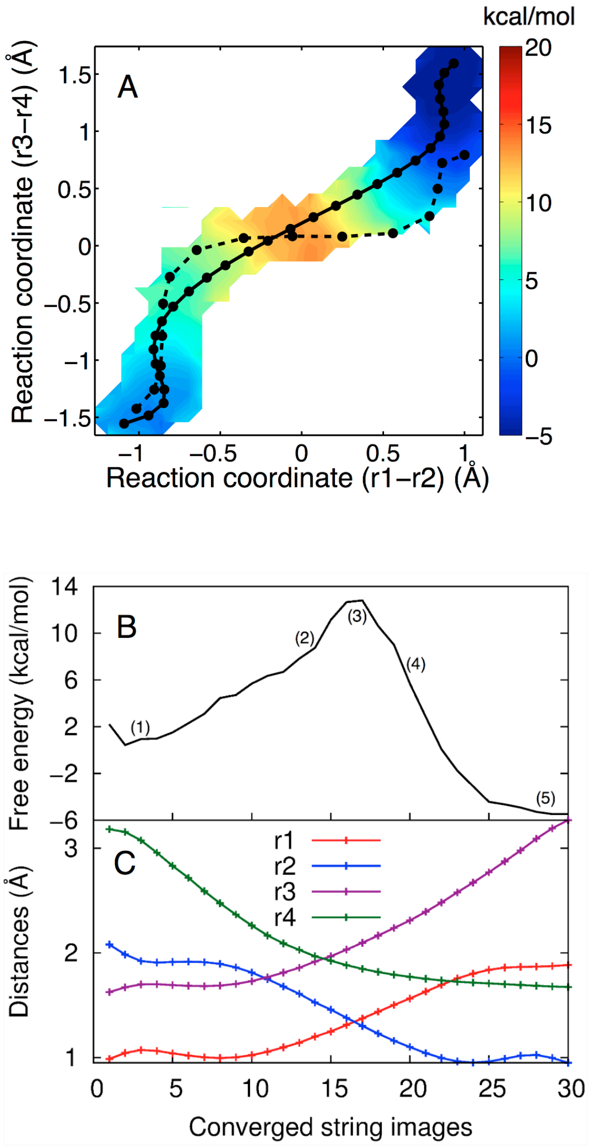

**Figure 7.4.** (A) The 2D free energy surface obtained from set A, where $Mg^{2+}$ is bound at the catalytic site, projected in the (r1-r2) and (r3-r4) space. The initial string (dashed black line) corresponds to the MEP obtained in our previous study.[19] The converged string (solid black line) corresponds to the MFEP obtained from set A. Both the MEP and MFEP correspond to a concerted mechanism with a phosphorane-like TS, but the MFEP is more synchronous than the previously obtained MEP. Each circle corresponds to an image along the string. The color scale denotes free energy in units of kcal/mol. (B) The 1D free energy profile along the MFEP obtained from set A, where $Mg^{2+}$ is bound at the catalytic site. (C) Values of the most important reaction coordinates, r1, r2, r3, and r4, along the MFEP. Each circle corresponds to an image along the string.

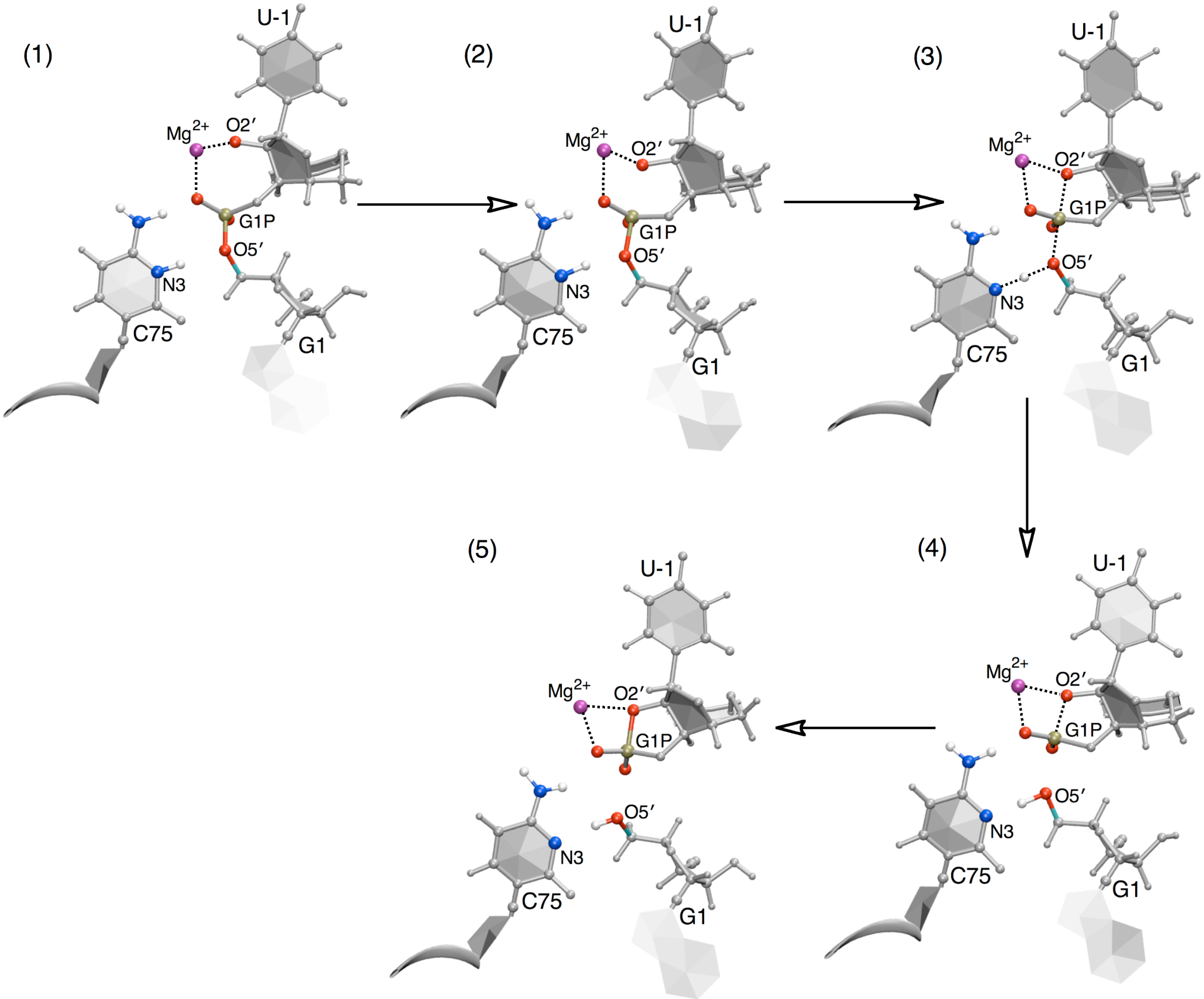

**Figure 7.5.** Representative structures obtained along the MFEP obtained from set A, where Mg$^{2+}$ is bound at the catalytic site. The structures are numbered according to the locations along the MFEP identified in Figure 7.4B. The MFEP passes through a single transition state, which is phosphorane-like, as depicted in the reaction pathway in Figure 7.1A and represented here in structure (3). Note that these structures do not represent stable intermediates along the reaction pathway but rather represent selected structures along a concerted pathway.

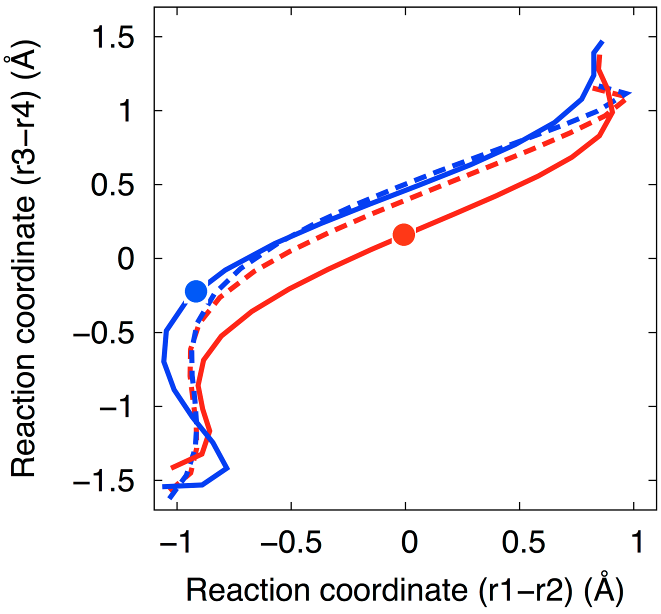

**Figure 7.6.** The initial (dashed lines) and final (solid lines) strings from sets C (red, $Mg^{2+}$ at catalytic site), and D (blue, $Na^{+}$ at catalytic site). Set C, both solid and dashed red lines, is reproduced from Figure 7.3. The strings are projected in the space of the collective coordinates (r1-r2) and (r3-r4). The converged string from set D is distinct from the converged string from set C. The red circle indicates the position of the phosphorane-like transition state structure (free energy maximum for the concerted mechanism with $Mg^{2+}$), while the blue circle indicates the position of the phosphorane intermediate structure (free energy minimum for the sequential mechanism with $Na^{+}$).

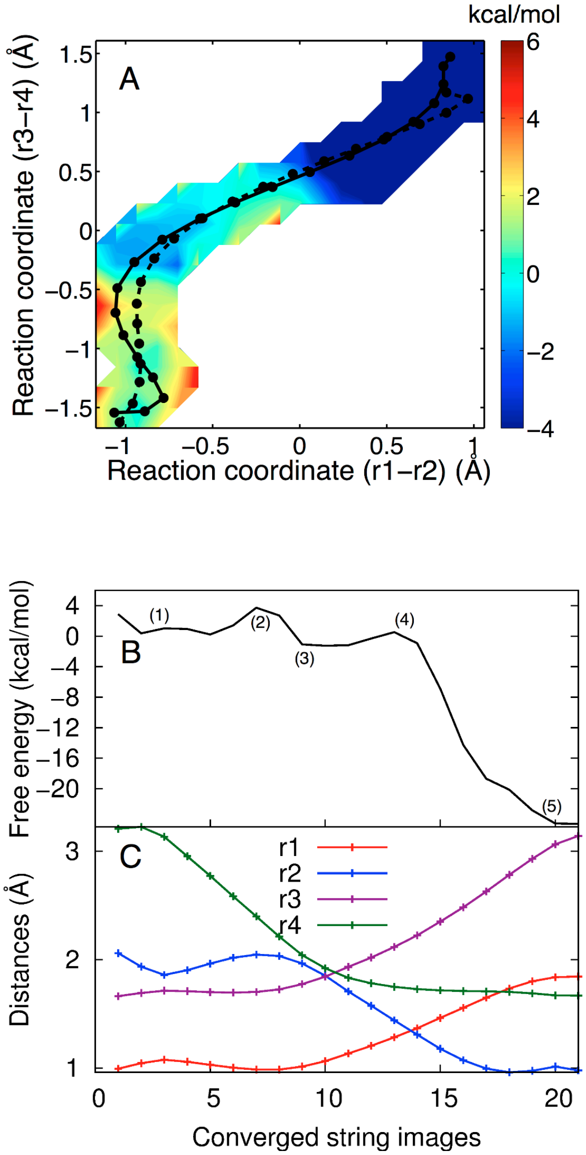

**Figure 7.7.** (A) The 2D free energy surface obtained from simulation set D, where a $Na^+$ ion is at the catalytic site, projected in the (r1-r2) and (r3-r4) space. The initial and converged strings are illustrated on the surface as the black dotted and solid lines, respectively. In this case, the MFEP (solid black line) is sequential, passing through a phosphorane intermediate. Each circle corresponds to an image along the string. The color scale denotes free energy in units of kcal/mol. (B) The 1D free energy profile along the MFEP obtained from set D, where a $Na^+$ ion is at the catalytic site. (C) Values of the most important reaction coordinates, r1, r2, r3, and r4, along the MFEP. Each circle corresponds to an image along the string.

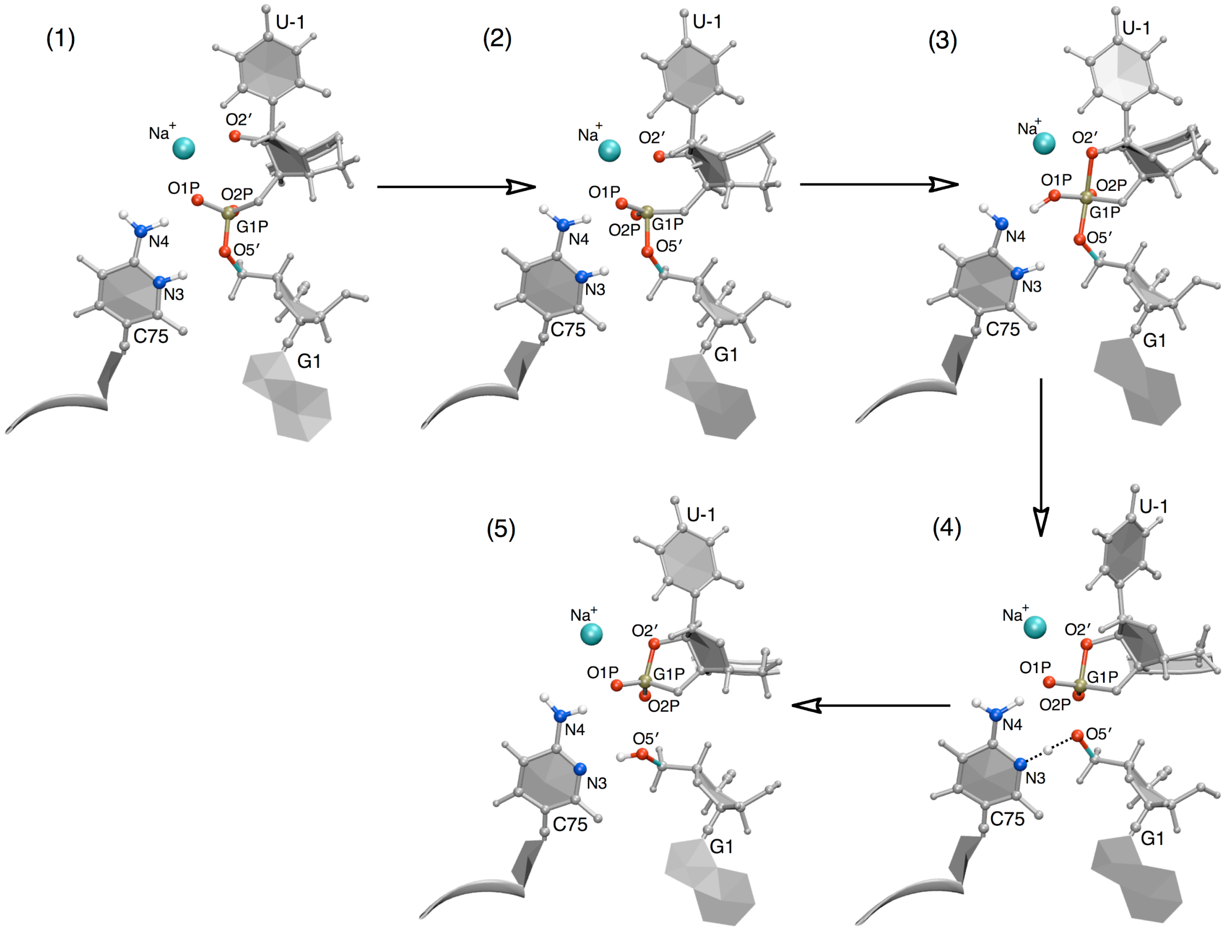

**Figure 7.8.** Representative structures obtained along the MFEP obtained from simulation set D, where $Na^+$ is at the catalytic site. The structures are numbered according to the locations along the MFEP identified in Figure 7.7B. The MFEP passes through a phosphorane intermediate, as represented here in structure (3), which is similar to the intermediate depicted in Figure 7.1C but with protonation at the pro$R_P$ oxygen (O1P) rather than at N4 of C75. Note that these structures do not represent stable intermediates along the reaction pathway but rather simply represent selected structures along a sequential pathway.

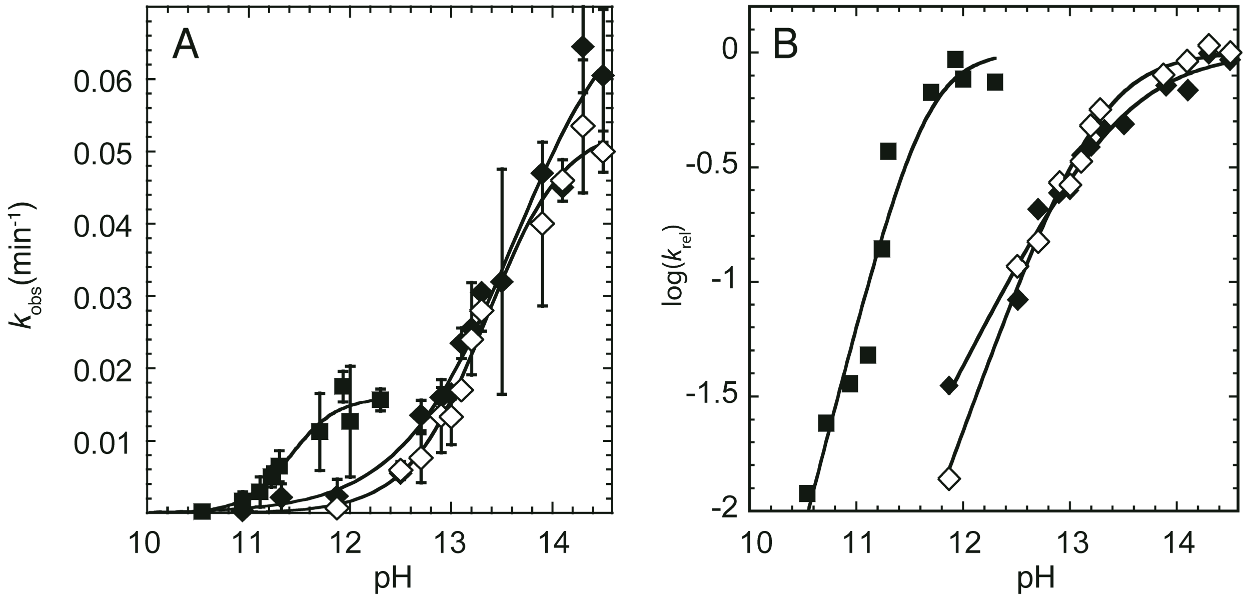

**Figure 7.9:** Determination of the p$K_a$ of the 2′OH by the kinetic method. A) Plot of $k_{obs}$ as a function of pH in the presence of 3.16 M Na$^+$ (filled diamonds), 3.16 M K$^+$ (open diamonds), and 10 mM Ca$^{2+}$ (filled squares), B) Plot of log ($k_{rel}$) as a function of pH using the same symbols as in panel A. $k_{rel}$ was obtained by normalizing the data of panel A to the $k_{max}$ value from the fit. The p$K_a$ values were determined from fitting the rate versus pH plots in panel A to Eq. 3 and were found to be 13.6, 13.3, and 11.4 in the presence of 3.16 M Na$^+$, 3.16 M K$^+$, and 10 mM Ca$^{2+}$, respectively. (Note that for panel A, the p$K_a$ is the pH where $k_{obs}$ is ½ $k_{max}$, while for panel B the p$K_a$ is the pH near the flex point.) The Hill coefficients determined from the fit of the data in panel A to Eq. 3 were found to be 0.8, 1.1, and 1.9 for Na$^+$, K$^+$, and Ca$^{2+}$, respectively. These Hill coefficients are apparent in the relative slopes in panel B. The value of $k_{max}$ is ~3-fold higher in monovalent than divalent ions, as revealed in panel A, but p$K_a$ values are unaffected by $k_{max}$ differences as they are determined entirely from the shape of the curves. The origin of the $k_{max}$ difference is unclear, but could reflect different interactions of divalent and monovalent ions with this 22mer.

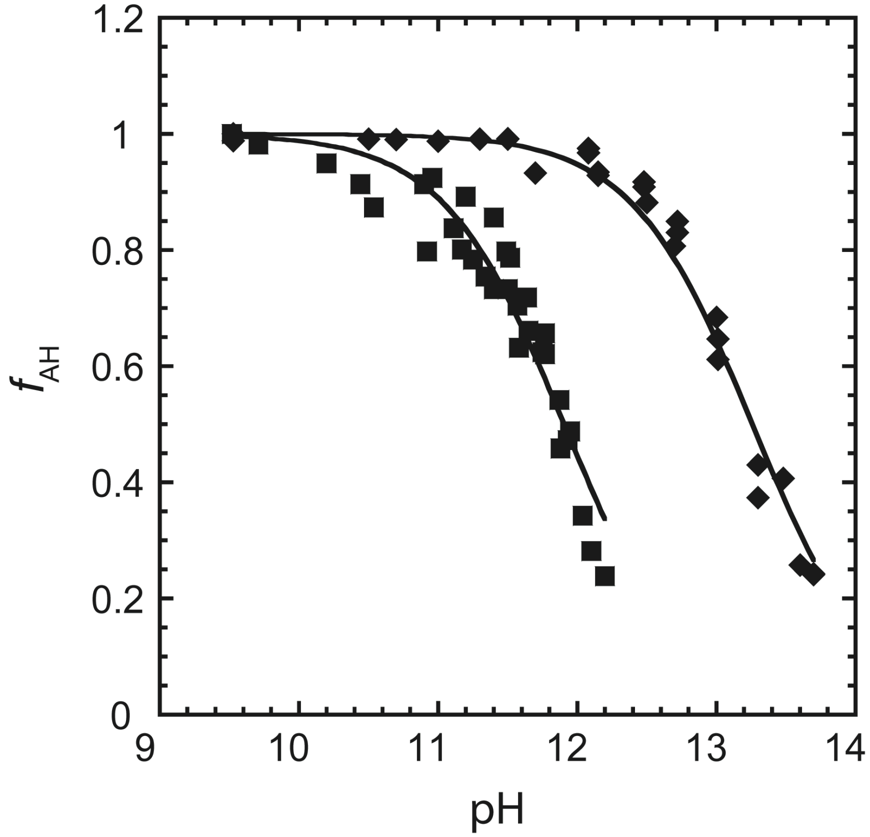

**Figure 7.10:** Fraction of protonated species as a function of pH in the presence of 0.5 M $Na^+$ (filled diamonds) and 10 mM $Ca^{2+}$ (filled squares) as determined by the NMR methods. The fraction of protonated species at a given pH was generated by Eq. 6. The $pK_a$ values were determined from fitting $d_{obs}$ versus pH plots to Eq. 4, as shown in Figure F.9, and were found to be 13.2 and 11.9 in the presence of 0.5 M $Na^+$ and 10 mM $Ca^{2+}$, respectively.

# CHAPTER 8: CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation, the hepatitis delta virus (HDV) ribozyme was investigated using a variety of computational approaches. From the methodological perspective, classical molecular dynamics (MD) simulations were used to study the structure and dynamics of the ribozyme, characterize metal ion interactions at specific regions, and explore the importance of electrostatics in the ribozyme's biological function. MD simulations on mutant ribozymes were performed to study the impact of key residues on the structure and motion of the ribozyme. Quantum mechanical/molecular mechanical (QM/MM) methods were employed to study the ribozyme self-cleavage reaction. The favorable mechanisms of the catalytic reaction were explored and the roles of the catalytic metal ion were elucidated.

These studies addressed a wide range of important questions related to the structure and function of the HDV ribozyme. The significance of the rare reverse G•U wobble found close to the active site of the ribozyme was explored. The 'chelated' metal ion interaction at this base pair was found to be distinct from the 'diffused' metal ion interaction at a nearby, geometrically similar standard G•U wobble. Non-linear Poisson-Boltzmann (NLPB) calculations revealed a negative pocket close to the reverse wobble that led to its strong interaction with metal ions. The reverse wobble was found to interact with the putative catalytic ion and play a role in positioning the ion for catalysis. Mutation of this reverse wobble to an isosteric G•C base pair resulted in the weakening of metal ion binding at the catalytic position. This was reflected in the lowering of the experimentally measured rate of reaction of the mutant ribozyme. The impact of protonating nucleobase C75 on the stability and motions of the ribozyme was also studied. The protonated state of C75 was found to be essential in keeping the active site of the ribozyme organized for catalysis.

The HDV ribozyme self-cleavage reaction was found to follow a concerted pathway in presence of a divalent ion at the catalytic site. The reaction, however, adopted a sequential mechanism in presence of a monovalent ion at the catalytic site, passing through a phosphorane intermediate. The divalent ion shifted the pKa of C75 towards neutrality, making its proton donation more facile and thereby favoring the concerted mechanism. Several key roles of the catalytic metal ion were illustrated, such as, assisting in the activation of the O2′ nucleophile, acidifying the general acid C75, and stabilizing the non-bridging oxygens of the scissile phosphate at the cleavage site. These findings were largely consistent with experimental measurements.

These studies not only contribute significantly towards the understanding of the functioning of the HDV ribozymes, but also provide insights into the chemistry employed by catalytic RNAs in general, and open up several interesting future directions in the field of ribozymes. Some of these future directions are discussed below.

*Identification of the fast and slow reacting phases of the $R_P$ substrate:* The phosphorothioate substitution studies on the HDV ribozyme indicated biphasic kinetics for the $R_P$ substrate. The identities of the two reacting phases, however, were ambiguous. Characterization of these two phases with very different reaction rates are of interest because they can provide insights into the different channels by which the HDV ribozyme can perform catalysis. One explanation of the slow phase kinetics is that the reaction passes through a channel that involves a 2′-5′ pseudorotation at the scissile phosphate. QM/MM simulations of the phosphorothioate substituted ribozymes can help distinguish between these two phases.

*Investigation of the inverse thio-effect of the $S_P$ substrate:* The phosphorothioate substitution studies also indicated that the $S_P$ substrate reacted at a higher rate than the oxo substrate in presence of $Na^+$ ions. More recent unpublished data suggests that at higher concentrations of $Na^+$ ions this inverse thio-effect diminishes, and the $S_P$ substrate cleaves almost at the same rate as the oxo substrate, similar to the behavior observed in presence of $Mg^{2+}$ ions. QM/MM simulations of the cleavage reaction in presence of a $Na^+$ ion at the active site revealed a proton transfer from the exocyclic amine of C75 to the *pro*-$R_P$ oxygen of the scissile phosphate. This proton transfer suggests the involvement of the non-bridging oxygens of the scissile phosphate in the catalytic reaction and hints the possibility of different reaction mechanisms of the phosphorothioate substituted substrates in presence of $Na^+$ ions. Further studies are required to explore these new mechanisms.

*Estimation of $pK_a$s of nucleobases in the ribozyme environment:* The $pK_a$s of specific nucleobases in ribozymes are often shifted far from their solution values allowing them to participate in catalysis in form of general acids or bases. In the HDV ribozyme the $pK_a$ of the nucleobase C75 is shifted by more than 2 units, enabling the base to act as a general acid. Other nucleolytic ribozymes are also known to use similar strategies in their catalytic reactions. Thus, the accurate estimation of the nucleobase $pK_a$s in the ribozymes is an important part of understanding the chemistry of the catalytic RNAs. There are existing strategies in literature by which $pK_a$s of isolated bases in solution can be calculated accurately. However, as most of these methods are based on quantum mechanical calculations, their application to ribozymes is not feasible. One proposed route to calculate the $pK_a$s of nucleobases in the ribozyme environment is as follows: the shift in the $pK_a$ of a particular nucleobase in the ribozyme from its solution value can be assumed to be an effect of replacing the aqueous environment of the nucleobase by the ribozyme environment. Under this assumption, the difference in the free energies of solvation of the nucleobase deprotonation process in the ribozyme and in water can be directly related to its p$K_a$ shift. A method on the lines of free energy perturbation can be used to estimate the above free energies of solvation.

*Extending the mechanistic studies to other nucleolytic ribozymes:* The nucleolytic ribozymes: hairpin, hammerhead, Varkud satellite (VS), GlmS, and HDV ribozyme, catalyze the same phosphodiester cleavage reaction using different mechanistic approaches. All of these ribozymes employ nucleobase catalysis, while only the hammerhead and HDV ribozymes have been shown to use metal ion catalysis. Several aspects of their catalysis are still unknown. While numerous quantum mechanical studies have been performed to study the mechanisms of these ribozymes, most of these studies were based on cutoff models of their respective active sites. Hybrid QM/MM methods such as those discussed in this dissertation will provide a more rigorous framework for studying the catalytic mechanisms in the other nucleolytic ribozymes.

# APPENDIX A: SUPPORTING INFORMATION FOR CHAPTER 2[1]

**Contents:**

I. Supporting Methods
  1) Molecular Dynamics (MD) equilibration procedure.
  2) Updated protonated cytosine partial charges. See Figure A.1 and Tables A.1, A.2.

II. Other Supporting Figures
  Figure A.2: Additional independent trajectories of the reverse G•U wobble.
  Figure A.3: Metal ion movements during MD simulations of product C75 state.
  Figure A.4: Heavy-atom RMSD plots.
  Figure A.5: Additional NLPB results.

III. Other Supporting Tables
  Table S3: Crystallographic data collection statistics.
  Table S4: Crystallographic $Na^+$ ion coordinates for precleaved ribozyme.
  Table S5: Average distances between reverse G•U wobble and metal ions in product C75 and precleaved $C75^+$

IV. Supporting References

---

[1] This appendix was published in its entirety in the journal *Biochemistry* in 2011. The complete reference to the journal article is as follows: Veeraraghavan, N.; Ganguly, A.; Chen, J.-H.; Bevilacqua, P. C.; Hammes-Schiffer, S.; Golden, B. L., *Biochemistry*, **2011**, *50*, 2762-2682. N.V. performed the computational analysis, A. G. performed the simulations, and J. C. H. performed the experiments.

# I. Supporting Methods

1) Molecular Dynamics (MD) equilibration procedure.
For both the product and precleaved ribozyme, initially an energy minimization of the solvent and ions was conducted with the ribozyme fixed, followed by an energy minimization of the entire system (see below). Subsequently, we followed a comprehensive simulated annealing equilibration procedure:

1. Energy optimization of the solvent (waters, $Na^+$ and $Cl^-$ ions) with the solute (RNA and crystallographic $Mg^{2+}$) fixed.
2. Simulated annealing (NVT ensemble, Nose-Hoover thermostat) of solvent keeping solute fixed:
    - i. Increase temperature from 0K to 298K in 10 ps, MD @ 298 K for 50 ps
    - ii. Increase temperature from 298K to 498K in 10 ps, MD @ 498K for 50 ps
    - iii. Increase temperature from 498K to 698K in 10 ps, MD @698K for 50 ps
    - iv. Decrease temperature from 698K to 498K in 10 ps, MD @ 498K for 50 ps
    - v. Decrease temperature from 498K to 298K in 10 ps, MD @ 298 K for 160 ps

    The above protocol was performed two times.

3. MD at 298K with NPT ensemble using Nosé-Hoover barostat (*1, 2*) for 100 ps with solute fixed.
4. Energy optimization of the solute with the solvent fixed.
5. Simulated annealing (NPT ensemble, Nosé-Hoover barostat) of both solute and solvent:
    - i. Increase temperature form 0 to 100K in 10 ps, MD @ 100K for 100 ps
    - ii. Increase temperature from 100K to 200K in 10 ps, MD @ 200K for 100 ps
    - iii. Increase temperature from 200K to 298K in 10 ps, MD @ 298 K for 500 ps
6. MD at 298K with NVT ensemble for 1-3 ns using Nosé-Hoover thermostat

Data were collected at 298 K in the NVT ensemble. A Nosé-Hoover thermostat was used to maintain the temperature and pressure, and the time step was 1 fs for all MD trajectories. During data collection, snapshots were output every 5 ps. Distance plots were generated with a running average over 50 snapshots. Heavy-atom RMSDs for all trajectories are given in Figure A.4.

2) Updated protonated cytosine partial charges

Given the availability of the new crystal structure of the precleaved form of HDV ribozyme, we performed a thorough investigation of the protonated cytosine charges. Coordinates of C75, C41, and C48 were extracted from the precleaved HDV ribozyme crystal structure (PDB: 3NKB) after hydrogens were added to the entire ribozyme using Maestro; the same software was used to protonate each of these three cytosines at the N3 position. C75 and C41 were chosen because they are in a rare puckered form (2'-endo) but have pivotal roles in the catalytic activity of HDV ribozyme (*3, 4*), and C48 was chosen as a control because it is in the more typical puckered form (3' endo). We performed the RESP calculations starting from each of these three cytosine structures. We used the charges generated from C75 for protonated cytosine in our MD simulations.

The RED-II program was used to calculate the partial charges of the base atoms of the cytosines. The calculation of RESP charges involved three steps.

1. Optimization of the structure with Gaussian 03 at the HF/6-31G* level.
2. Generation of the Molecular Electrostatic Potential (MEP) for the target structure on a three dimensional grid.
3. Exportation of the grid into the Restrained Electrostatic Potential (RESP) program, which is used to fit atom-centered charges to the MEP.

During the third step, the charges on the atoms of the sugar (excluding C1' and H1') were fixed to the standard Amber charges (*5, 6*).

Figure A.1 depicts the neutral and protonated cytosine molecules that were used in these calculations. Table S1 compares the Amber partial charges of neutral cytosine with those calculated with the starting structures C75 and C48. The level of agreement confirms that our procedure is consistent with the Amber forcefield. Table S2 compares the partial charges of protonated cytosine calculated with the starting structures C75, C41, and C48.

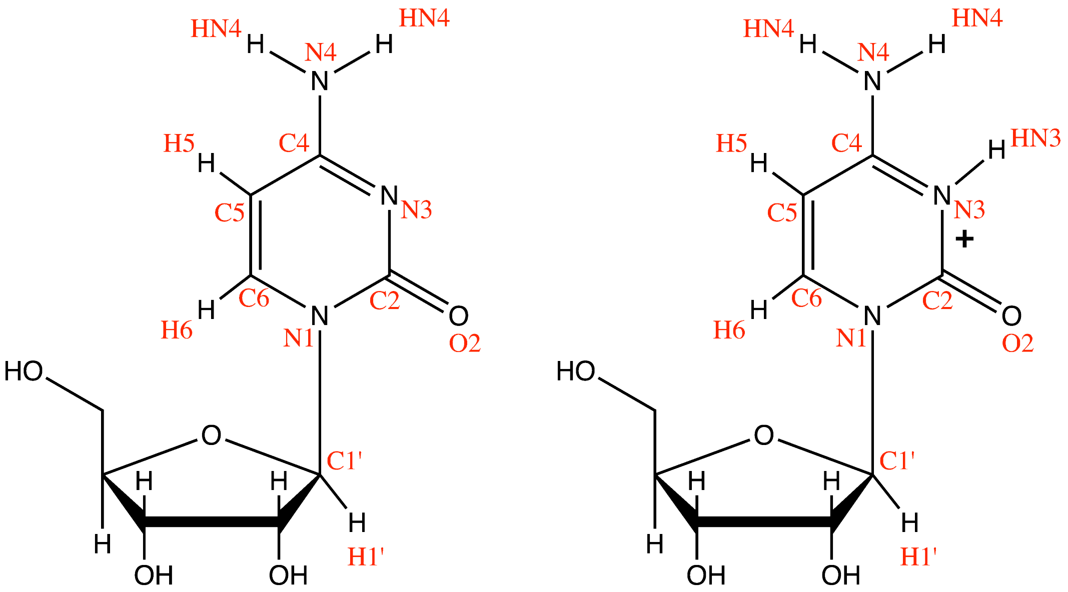

**Figure A.1.** Neutral (left) and protonated (right) cytosine.

| | Amber | C75 | C48 |
|---|---|---|---|
| **C1'** | 0.0066 | -0.0464 | 0.0323 |
| **H1'** | 0.2029 | 0.124 | 0.1787 |
| **N1** | -0.0484 | 0.0877 | -0.0325 |
| **C6** | 0.0053 | -0.032 | -0.0276 |
| **H6** | 0.1958 | 0.1837 | 0.2005 |
| **C5** | -0.5215 | -0.5442 | -0.5086 |
| **H5** | 0.1928 | 0.1993 | 0.1896 |
| **C4** | 0.8185 | 0.939 | 0.8161 |
| **N4** | -0.953 | -1.0711 | -0.9552 |
| **HN4** | 0.4234 | 0.4496 | 0.4116 |
| **HN4** | 0.4234 | 0.4496 | 0.4116 |
| **N3** | -0.7584 | -0.8243 | -0.767 |
| **C2** | 0.7538 | 0.7644 | 0.772 |
| **O2** | -0.6252 | -0.5633 | -0.6055 |

**Table A.1.** Partial charges calculated for neutral cytosine with two different starting structures and compared to standard Amber charges.

| | C75H$^+$ | C41H$^+$ | C48H$^+$ |
|---|---|---|---|
| **C1'** | 0.5035 | 0.4557 | 0.4659 |
| **H1'** | -0.014 | 0.0061 | 0.0524 |
| **N1** | -0.149 | -0.1421 | -0.0742 |
| **C6** | 0.0118 | -0.0456 | -0.1747 |
| **H6** | 0.1257 | 0.1695 | 0.2815 |
| **C5** | -0.2521 | -0.223 | -0.235 |
| **H5** | 0.196 | 0.1897 | 0.1891 |
| **C4** | 0.462 | 0.4553 | 0.4618 |
| **N4** | -0.8545 | -0.8758 | -0.8829 |
| **HN4** | 0.4641 | 0.4736 | 0.4691 |
| **HN4** | 0.4641 | 0.4736 | 0.4691 |
| **N3** | -0.2119 | -0.1764 | -0.1732 |
| **C2** | 0.4925 | 0.4837 | 0.4189 |
| **O2** | -0.4218 | -0.4183 | -0.4599 |
| **HN3** | 0.2994 | 0.2899 | 0.3081 |

**Table A.2.** Partial charges calculated for protonated cytosine with three different starting structures

## II. Other Supporting Figures

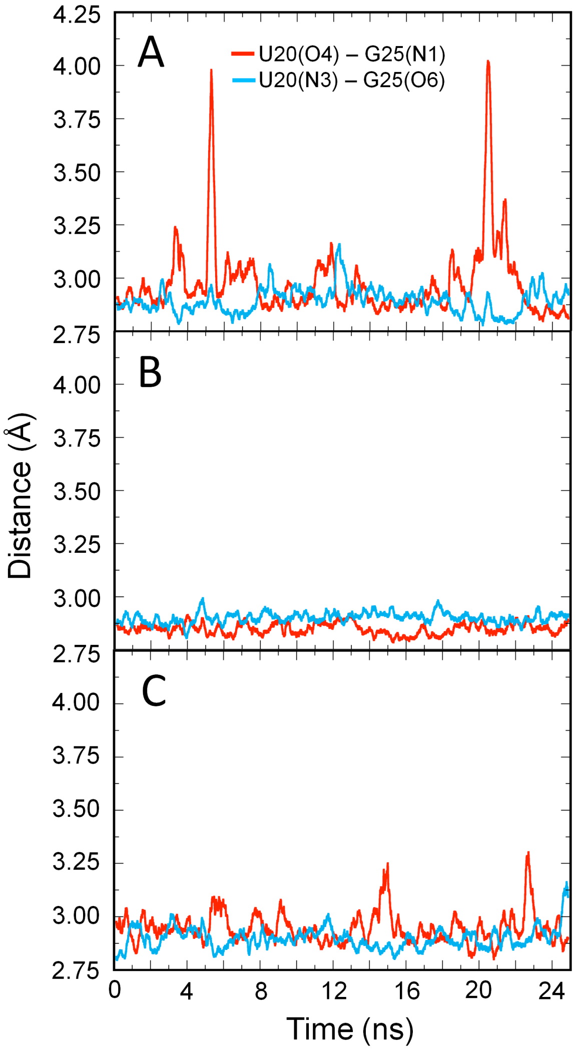

**Figure A.2.** Additional trajectories of the reverse G•U wobble. (A) Second independent trajectory of product structure, as per Figure 3B. (B) Second independent trajectory of precleaved structure, as per Figure 3C. (C) Independent trajectory of product structure with U27 removed.

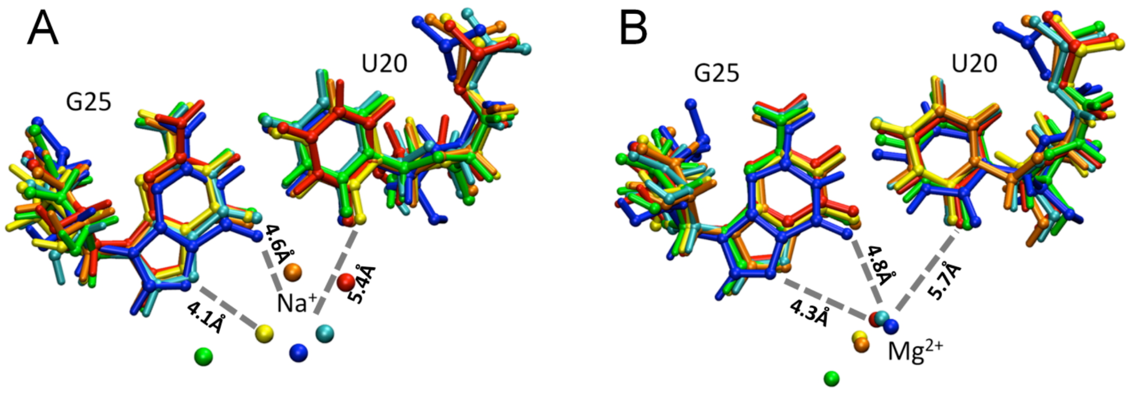

**Figure A.3.** Snapshots of movement of metal ion during MD trajectories of the product C75 state of the HDV ribozyme. (A) Coordination of $Na^+$ for a trajectory without $Mg^{2+}$ bound to the reverse G•U wobble. (B) Coordination of $Mg^{2+}$ for a trajectory with a $Mg^{2+}$ ion placed in the active site and two $Na^+$ ions removed from the bulk prior to equilibration. Colors show snapshots at various time steps: blue – 0 ns, cyan - 5ns, green – 10 ns, yellow – 15 ns, orange – 20 ns, red – 25 ns. These results were obtained for a single independent trajectory of each type. The snapshots at these time steps were aligned to minimize the RMSD over residues G25 and U20. Only heavy atoms were considered for the alignment. Average distances between key atoms of the reverse G•U wobble and these metal ions are indicated by dashed lines. These distances, as well as their standard deviations, are provided in Table S5 (first trajectory). As a comparison, in the crystal structure of the HDV ribozyme (PDBID-3NKB) the distance between the $Mg^{2+}$ ion to its second shell ligands, N7 and O6 of G25, is 4.1 Å.

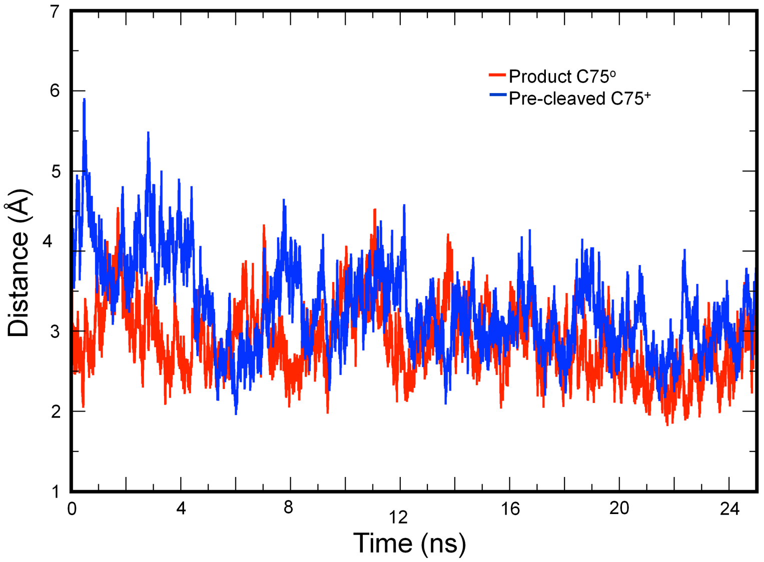

**Figure A.4.** Root mean square deviation (RMSD) plots for the heavy atoms of the product and precleaved structures corresponding to the trajectories in Figure 3. RMSDs were calculated with reference to the crystal structures.

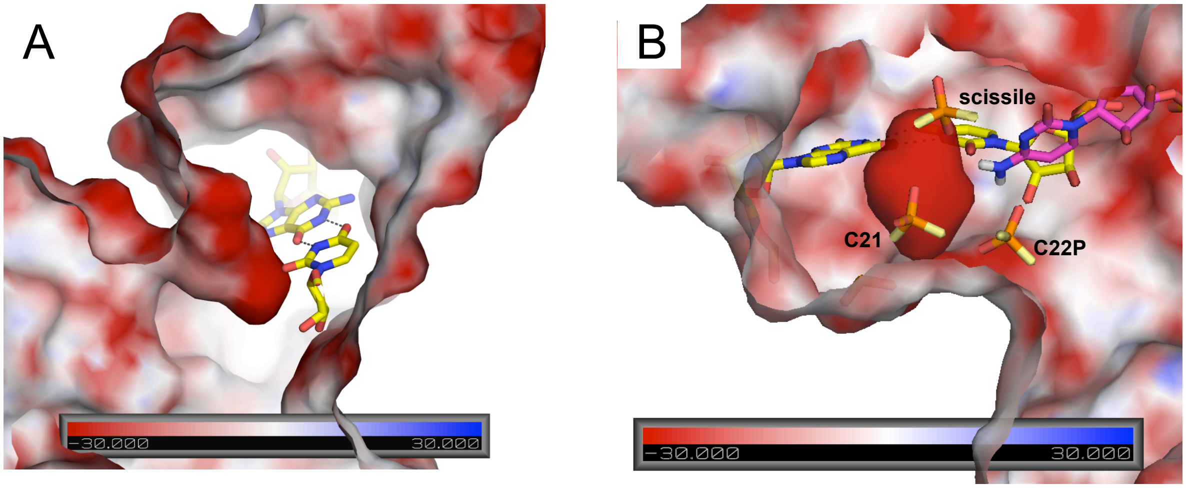

**Figure A.5.** Surface electrostatic potential of product and precleaved HDV ribozyme. The potential is colored according to the scales provided in the base of each panel. Views shown here are near the reverse G•U wobble, which is depicted in yellow sticks with hydrogen bonding in black. (A) Product state after 20 ns of MD simulation, which is from the trajectory provided in Figure 3B; similar potentials are found at other times (not shown).(B) Pprecleaved HDV ribozyme with specific molecular interactions displayed. C75 is in magenta with N4 hydrogens shown explicitly. C21, C22, and scissile phosphates are annotated and contribute to the negative potential; non-bridging oxygens, which bear the negative charge, are in red, while bridging oxygens are in pale yellow. Note that the C75 is oriented toward the highly negatively charged pocket generated near the reverse G•U wobble. The amine of C75 is a second shell ligand to the catalytic $Mg^{2+}$ (not shown here) and is hydrogen bonded to the scissile phosphate. View is rotated counterclockwise by ~90° from panels (A) in order to facilitate viewing of these phosphates.

## III. Other Supporting Tables

**Table A.3.** Crystallographic data collection statistics

| | |
|---|---|
| Space group | $C222_1$ |
| Cell dimensions | |
| a, b, c (Å) | 64.65, 84.23, 102.06 |
| $\alpha, \beta, \gamma$ | 90, 90, 90 |
| Wavelength (Å) | 0.9794 |
| Resolution (Å) | 50 – 1.9 (1.97 – 1.9) |
| $R_{sym}$ (%)[a] | 3.8 (53.9) |
| $I/\sigma(I)$[a] | 64.8 (1.5) |
| Data completeness (%)[a] | 91.4 (42.9) |
| Redundancy[a] | 12.6 (3.1) |

---

[a]The high-resolution shell is listed in parentheses.

[b]The $R_{sym}$ is the linear R-factor of $\Sigma|I-\langle I\rangle|/\Sigma I$, where the $I$ is the observed intensity and $\langle I\rangle$ is the average intensity of the multiple observations of symmetry-related reflections.

**Table A.4.** Crystallographic $Na^+$ ion coordinates for precleaved ribozyme

| Atom | x (Å) | y (Å) | z (Å) | Distance to $Na^+$ (Å) |
|---|---|---|---|---|
| $Na^+$ | -21.798 | -10.123 | -9.158 | - |
| Solvent1 | -22.919 | -12.084 | -10.098 | 2.4 |
| Solvent2 | -20.806 | -9.677 | -11.333 | 2.4 |
| Solvent3 | -23.223 | -8.409 | -9.969 | 2.4 |
| Solvent4 | -21.033 | -8.492 | -7.745 | 2.3 |

Values are from the $2F_o$-$F_c$ map calculated with the crystal soaked in $Na^+$-containing buffer as described in Materials and Methods and Table S3.

**Table A.5.** Average distances between reverse G•U wobble atoms and bound metal ion in product and precleaved forms.

| Structure | Trajectory | $Mg^{2+}$ present initially | G25(N7) – metal ion (Å) | G25(O6) – metal ion (Å) | U20(O2) – metal ion (Å) |
|---|---|---|---|---|---|
| Product ($C75^o$) | | | | | |
| | 1 | No | 4.1 (1.5) | 4.6 (1.1) | 5.4 (1.8) |
| | 2 | No | 4.8 (2.7) | 4.7 (1.6) | 4.8 (1.7) |
| | 1 | Yes | 4.3 (0.3) | 4.8 (0.8) | 5.7 (1.2) |
| | 2 | Yes | 4.3 (0.3) | 4.3 (0.4) | 5.0 (0.5) |
| Precleaved ($C75^+$) | | | | | |
| | 1 | Yes | 4.7 (0.2) | 4.1 (0.2) | 4.0 (0.2) |
| | 2 | Yes | 4.7 (0.2) | 4.1 (0.2) | 4.0 (0.2) |
| | 1 | No | 5.4 (1.5) | 3.8 (1.1) | 3.0 (0.7) |
| | 2 | No | 7.2 (0.4) | 4.9 (0.4) | 2.7 (0.5) |

Average distances between key atoms of the reverse G•U wobble and the metal ion observed in this region for the product $C75^o$ and the precleaved $C75^+$ states. For the trajectories with a bound $Mg^{2+}$ in the starting structure, these distances are from the reverse G•U wobble atoms to the $Mg^{2+}$ ion. For the trajectories without a bound $Mg^{2+}$ ion in the starting structure, these distances are from the reverse G•U wobble to the $Na^+$ ion that moves into this region. Only those snapshots exhibiting metal ion residency are considered for calculating average distances and standard deviations. Metal ion residency is defined in the footnote of Table 1. Results are obtained from 25 ns trajectories with data points taken every 5 ps. Standard deviations are given in parenthesis.

**Table A.6.** Refinement of the HDV ribozyme product structure

| | $R_{free}$ | $R_{work}$ | RMSD bonds (Å) | RMSD angles (°) |
|---|---|---|---|---|
| (A) 1CX0 | 27.9 | 24.6 | 0.011 | 1.400 |
| B. product with wobble | 27.1 | 24.9 | 0.010 | 1.106 |
| C. 1CX0-Phenix | 27.1 | 24.9 | 0.009 | 1.055 |

Crystallographic refinement statistics for (A) the HDV ribozyme product crystal structure refined using CNS {Brunger, 2007 #5055;Brunger, 1998 #5056} as reported previously, (B) the HDV ribozyme product crystal structure, modified to introduce a G25•U20 reverse wobble as described in the text and refined in Phenix {Adams, 2010 #5053}, and (C) the HDV ribozyme product structure, PDB ID 1CX0, refined in Phenix without any further adjustment.

## IV. Supporting References

(1) Nosé, S. *Mol. Phys.* **1984**, *52*, 255-268.
(2) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695-1697.
(3) Ferre-D'Amare, A. R., and Doudna, J. A. *J. Mol. Biol.* **2000**, *295*, 541-556.
(4) Chen, J. H., Yajima, R., Chadalavada, D. M., Chase, E., Bevilacqua, P. C., and Golden, B. L. *Biochemistry* **2010**, *49*, 6508-6518.
(5) Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. J., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(6) Wang, J. M., Cieplak, P., and Kollman, P. A. *J. Comp. Chem.* **2000**, *21*, 1049-1074.

# APPENDIX B: SUPPORTING INFORMATION FOR CHAPTER 3[1]

**Contents:**

---

[1] This appendix was published in its entirety in the Journal of Physical Chemistry B in 2011. The complete reference to the journal article is as follows: Veeraraghavan, N.; Ganguly, A.; Golden, B. L.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Phys. Chem. B* **2011**, *115*, 8346-8357. N. V. performed the analysis and A. G. performed the simulations.

## I. Supporting Methods

1.) Molecular Dynamics (MD) equilibration procedure.
The equilibration procedure for all MD trajectories was as follows:

1. Energy optimization of the solvent (waters, $Na^+$ and $Cl^-$ ions) with the solute (RNA and crystallographic $Mg^{2+}$) fixed.

2. Simulated annealing (NVT ensemble, Nosé-Hoover thermostat) of solvent keeping solute fixed:
   - i. Increase temperature from 0K to 298K in 10 ps, MD @ 298 K for 50 ps
   - ii. Increase temperature from 298K to 498K in 10 ps, MD @ 498K for 50 ps
   - iii. Increase temperature from 498K to 698K in 10 ps, MD @698K for 50 ps
   - iv. Decrease temperature from 698K to 498K in 10 ps, MD @ 498K for 50 ps
   - v. Decrease temperature from 498K to 298K in 10 ps, MD @ 298 K for 150 ps

   The above protocol was performed two times.

3. MD at 298K with NPT ensemble using Nosé-Hoover barostat [1,2] for 100 ps with solute fixed.

4. Energy optimization of the solute with the solvent fixed.

5. Simulated annealing (NPT ensemble) of both solute and solvent:
   - i. Increase temperature form 0 to 100K in 10 ps, MD @ 100K for 100 ps
   - ii. Increase temperature from 100K to 200K in 10 ps, MD @ 200K for 100 ps
   - iii. Increase temperature from 200K to 298K in 10 ps, MD @ 298 K for 500 ps

6. MD at 298K with NVT ensemble for 1 ns using a Nosé-Hoover thermostat.

For the U37C mutant, U37 was changed to C in the starting structure, prior to equilibration, and this GC base pair was optimized with everything else frozen. Then the same equilibration as outlined above was followed.

## II. Supporting Tables

**Table B.1.** RMSD of thermally averaged structures of C75$^+$/Mg$_{rev}$ and C75$^o$/Mg$_{rev}$.

| Region | Trajectory 1 (Å) | Trajectory 2 (Å) |
|---|---|---|
| **Whole RNA** | 3.28 | 2.97 |
| **Active site** (G1, U20, C21, C22, G25, U37, G38, G39 and C75) | 1.47 | 1.32 |
| **Reverse GU wobble** | 0.83 | 0.23 |
| **Normal GU wobble** | 0.89 | 0.90 |
| **P1** | 1.01 | 0.62 |
| **P2** | 0.91 | 0.66 |
| **P3** | 0.69 | 0.26 |
| **P4** | 0.61 | 0.43 |
| **L3** | 2.25 | 0.63 |
| **J1.1/4** (C41 and G40 base triple) | 0.87 | 1.01 |
| **J4/2** (G76, A77, A78) | 0.80 | 1.09 |

Values are calculated by aligning the residues from thermally averaged structures of C75$^+$/Mg$_{rev}$ and C75/Mg$_{rev}$ so as to minimize their RMSD, considering only regions shown in column 1.

**Table B.2.** RMSD of thermally averaged structures of C75$^+$/Mg$_{rev}$ and C75$^+$/noMg$_{rev}$.

| Region | Trajectory 1 (Å) | Trajectory 2 (Å) |
|---|---|---|
| **Whole RNA** | 1.48 | 2.11 |
| **Active site** (G1, U20, C21, C22, G25, U37, G38, G39 and C75) | 0.46 | 0.72 |
| **Reverse GU wobble** | 0.15 | 0.14 |
| **Normal GU wobble** | 0.22 | 0.70 |
| **P1** | 0.57 | 0.61 |
| **P2** | 1.05 | 0.57 |
| **P3** | 0.58 | 0.67 |
| **P4** | 0.57 | 0.45 |
| **L3** | 0.96 | 0.91 |
| **J1.1/4** (C41 and G40 base triple) | 0.55 | 0.90 |
| **J4/2** (G76, A77, A78) | 0.58 | 0.67 |

Values are calculated by aligning the residues from thermally averaged structures of C75$^+$/Mg$_{rev}$ and C75$^+$/noMg$_{rev}$ so as to minimize their RMSD, considering only regions shown in column 1.

## II. Supporting Figures

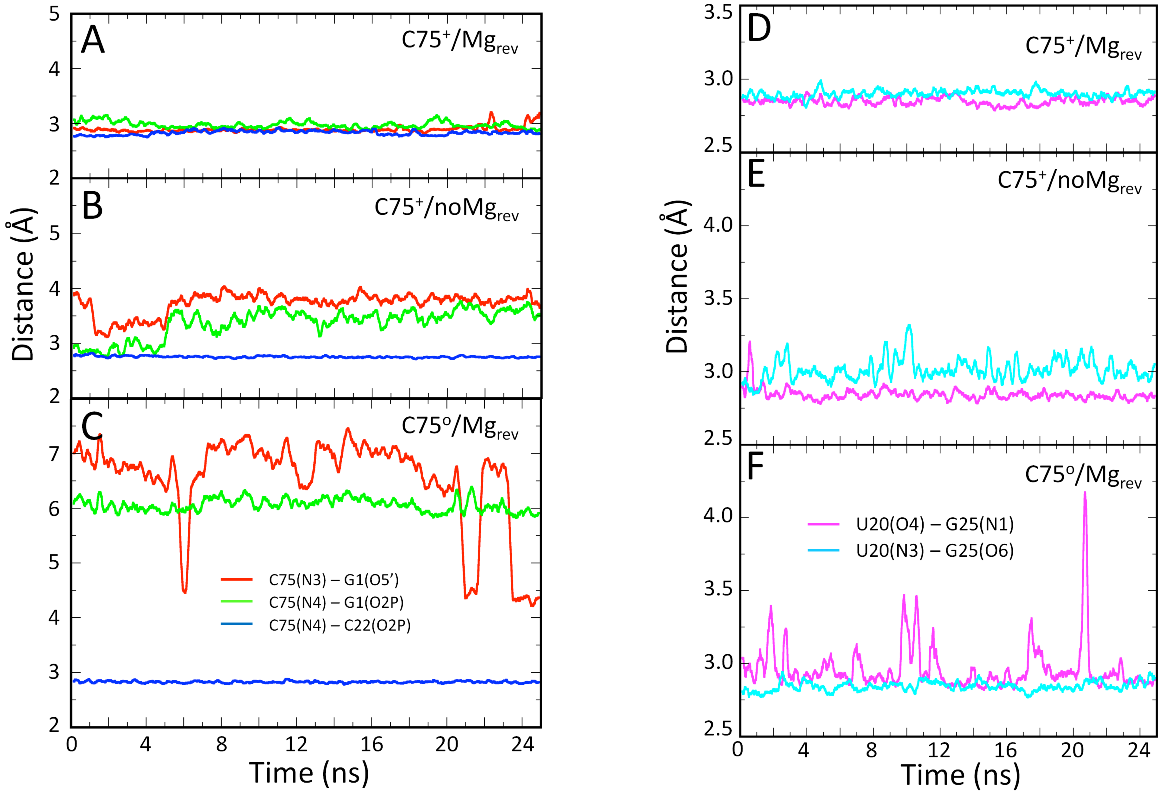

**Figure B.1.** Hydrogen bonding interactions in the active site of the HDV ribozyme. Second independent trajectory of the precleaved structure, as per Figure 3.3.

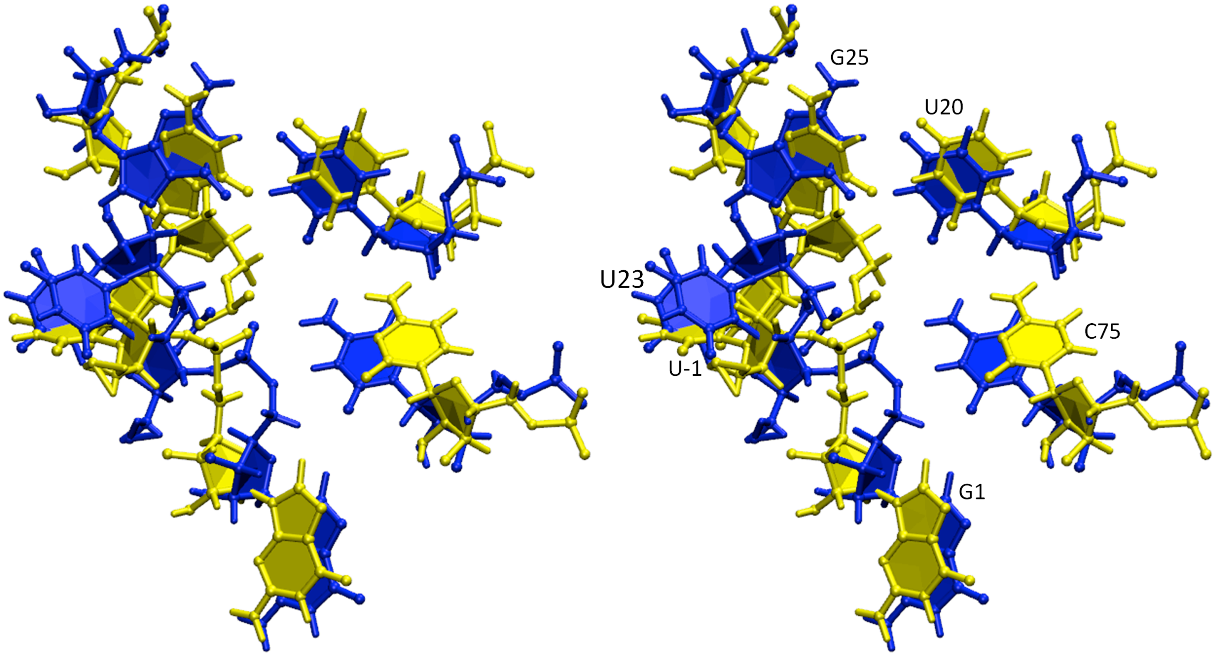

**Figure B.2.** Overlay of thermally averaged structures from the C75$^{+}$/Mg$_{rev}$ (blue) and C75$^{o}$/Mg$_{rev}$ (yellow) trajectories in stereoview. Note the absence of N3H on C75 in yellow. The distance between C75(N3) and G1(O5') has lengthened from ~3 to ~7 Å upon deprotonating C75, as per Figures 3.3A and 3.3C.

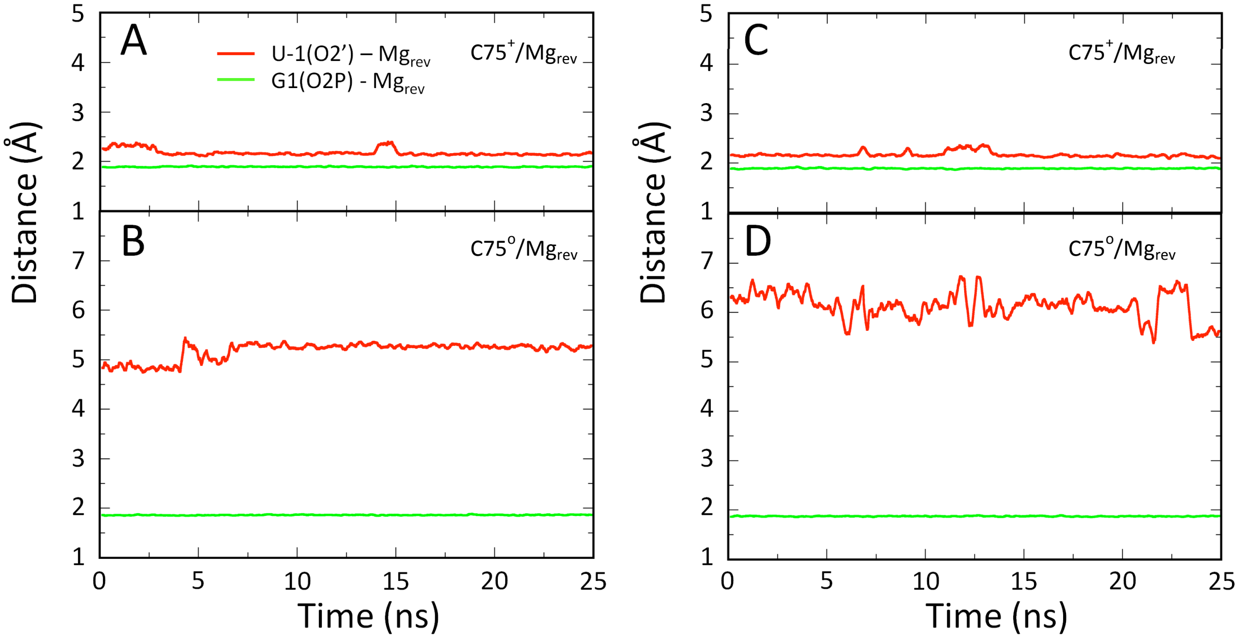

**Figure B.3.** Interactions of Mg$_{rev}$ to the ribozyme. The distances shown are between species provided in the key in panel A above. Results are provided for the two different types of trajectories: C75$^+$/Mg$_{rev}$ (A and C), and C75$^o$/Mg$_{rev}$ (B and D). The color scheme is the same as in Figure 3.1B.

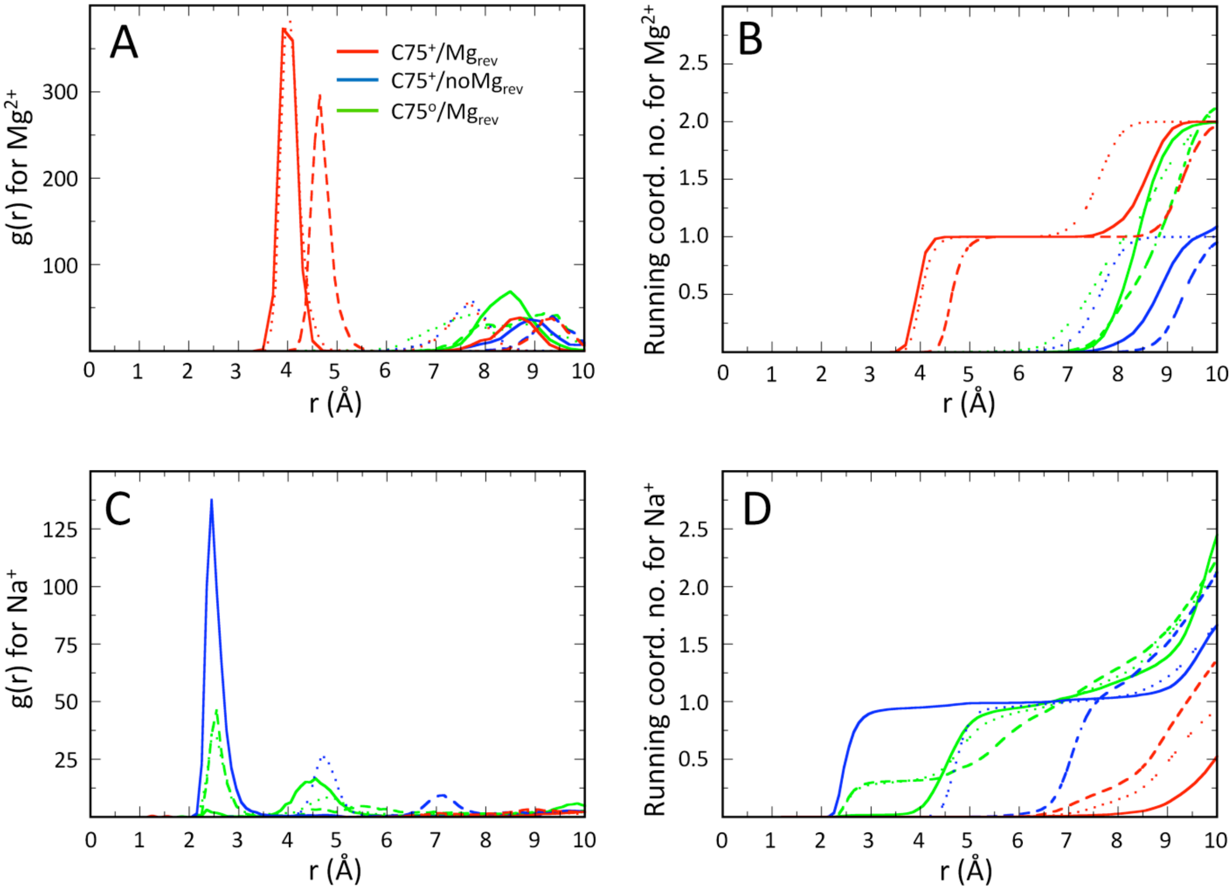

**Figure B.4.** Radial distribution function of metal ions for G25•U20 reverse wobble. Second independent trajectory of the pre-cleaved structure for C75$^+$/Mg$_{\text{rev}}$ (red), C75$^+$/noMg$_{\text{rev}}$ (blue) and C75$^o$/Mg$_{\text{rev}}$ (green), as per Figure 3.4. Panels A and C show the radial distribution functional of the Mg$^{2+}$ and Na$^+$ ions respectively, around each of atoms G25(N7) (dashed line), G25(O6) (dotted line) and U20(O2) (solid line). Panels B and D show the running coordination numbers of the Mg$^{2+}$ and Na$^+$ ions respectively, around each of atoms G25(N7), G25(O6) and U20(O2).

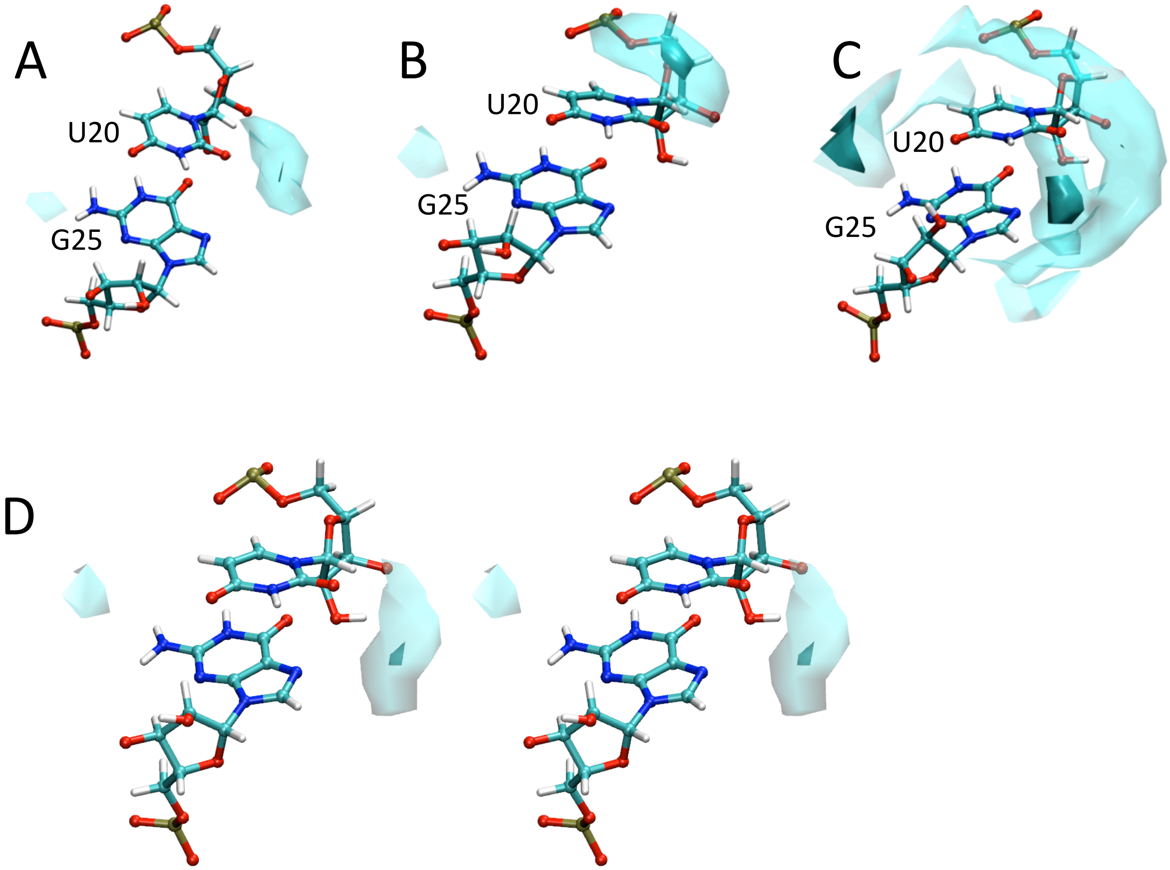

**Figure B.5.** Charge iso-density plots of metal ions for G25•U20 reverse wobble. Second independent trajectory for (A) $C75^+/Mg_{rev}$, (B) $C75^+/noMg_{rev}$, and (C) $C75^o/Mg_{rev}$. The darker cyan volumes represent the top 30% occupancy of a unit positive charge and the transparent cyan volumes represent the top 80% occupancy, as per Figure 3.5. (D) Stereoview of Figure 3.5A.

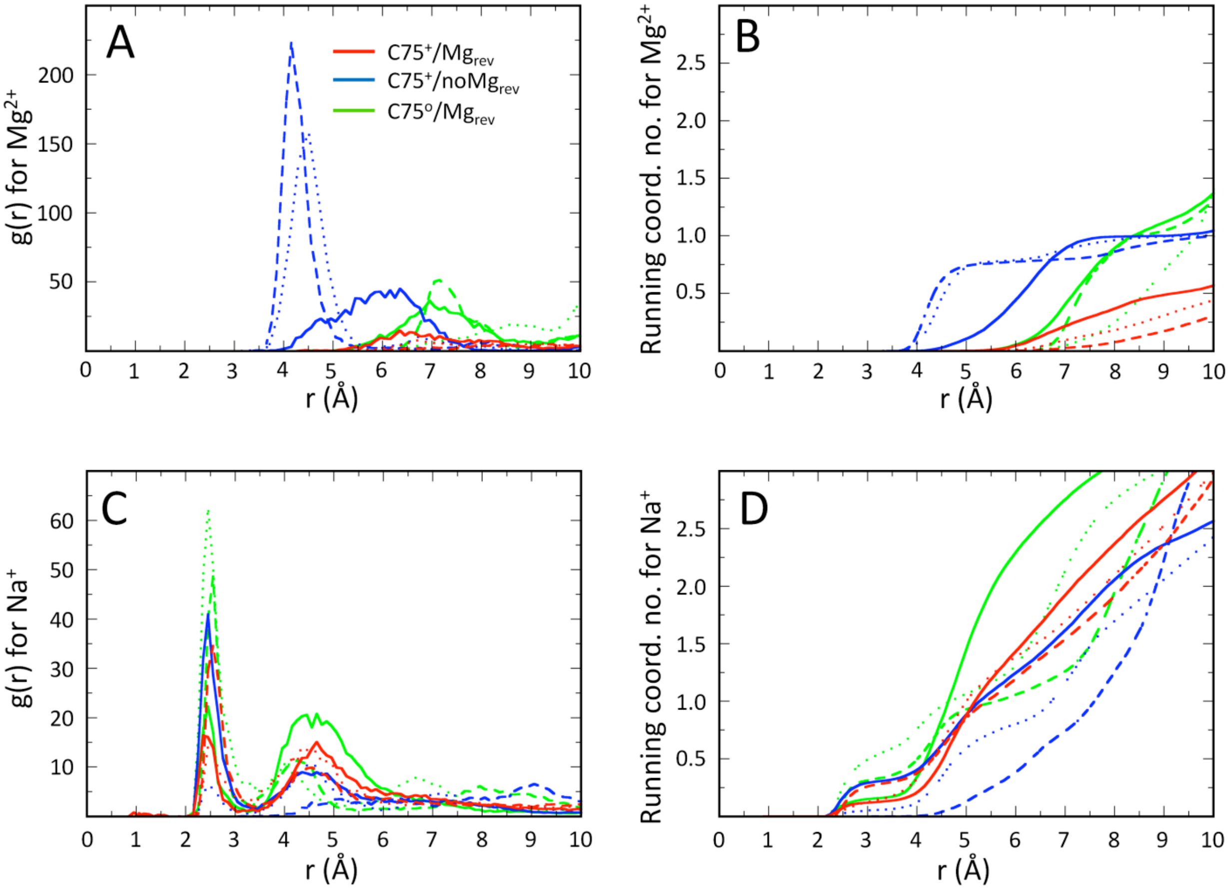

**Figure B.6.** Radial distribution function of metal ions for G1•U37 wobble. Second independent trajectory of the pre-cleaved structure for C75$^+$/Mg$_{rev}$ (red), C75$^+$/noMg$_{rev}$ (blue) and C75$^o$/Mg$_{rev}$ (green), as per Figure 3.6. Panels A and C show the radial distribution functional of the Mg$^{2+}$ and Na$^+$ ions respectively, around each of atoms G1(N7) (dashed line), G1(O6) (dotted line) and U37(O4) (solid line). Panels B and D show the running coordination number of the Mg$^{2+}$ and Na$^+$ ions respectively, around each of atoms G1(N7), G1(O6) and U37(O4).

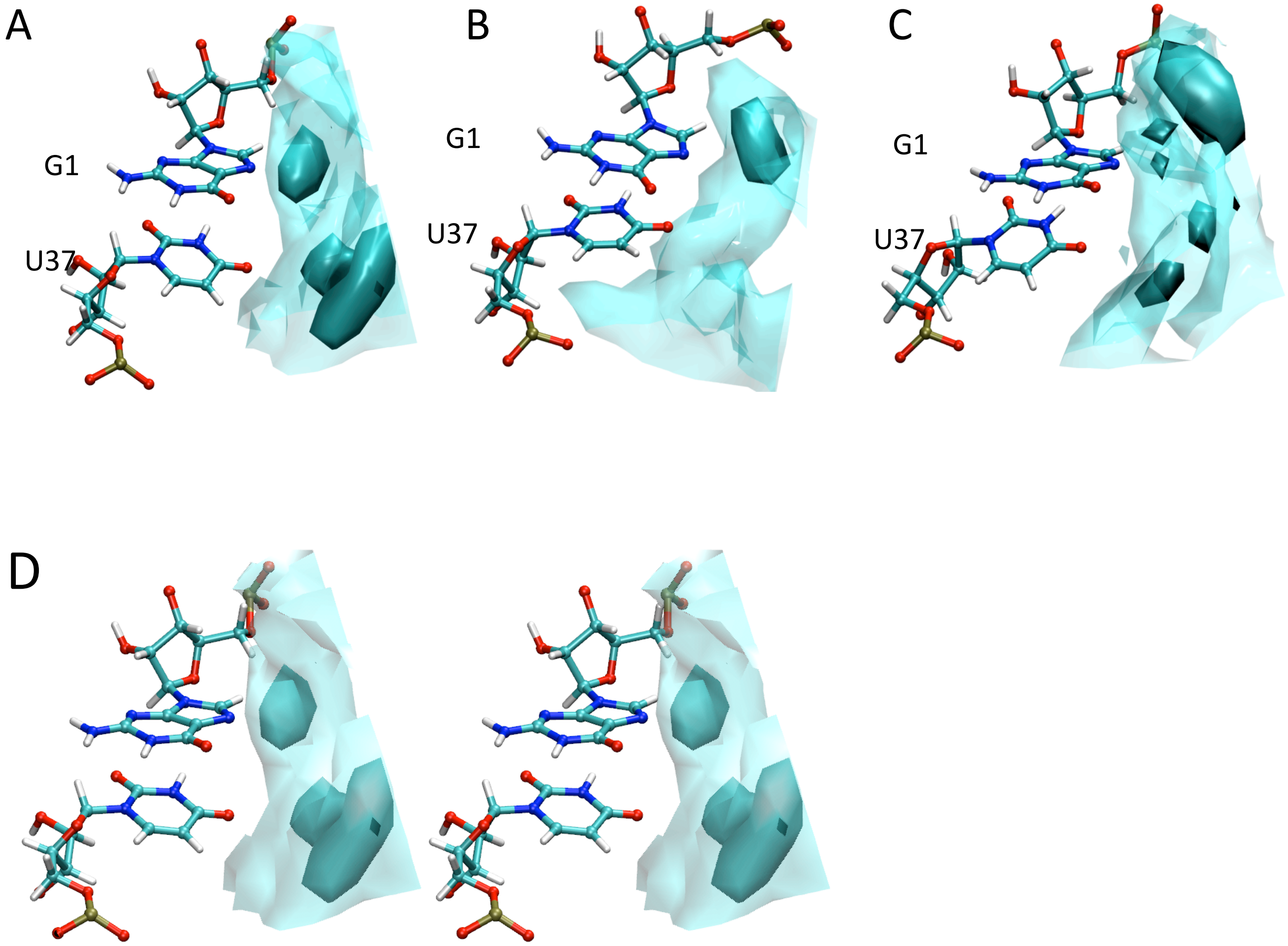

**Figure B.7.** Charge iso-density plots of metal ions for G1•U37 normal wobble. Spatial localization of a unit positive charge around the reverse G•U wobble for (A) C75$^+$/Mg$_{rev}$, (B) C75$^+$/noMg$_{rev}$, and (C) C75$^o$/Mg$_{rev}$. The darker cyan volumes represent the top 30% occupancy of a unit positive charge and the transparent cyan volumes represent the top 80% occupancy. (D) Stereoview of Figure 3.7A. Panel B.4A is the same as Figure 3.7A.

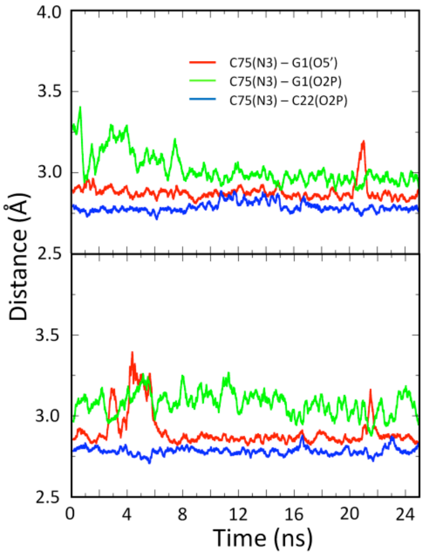

**Figure B.8.** Hydrogen bonding interactions in the active-site of C75$^+$/Mg$_{rev}$ U37C mutant. Figure shows two independent trajectories.

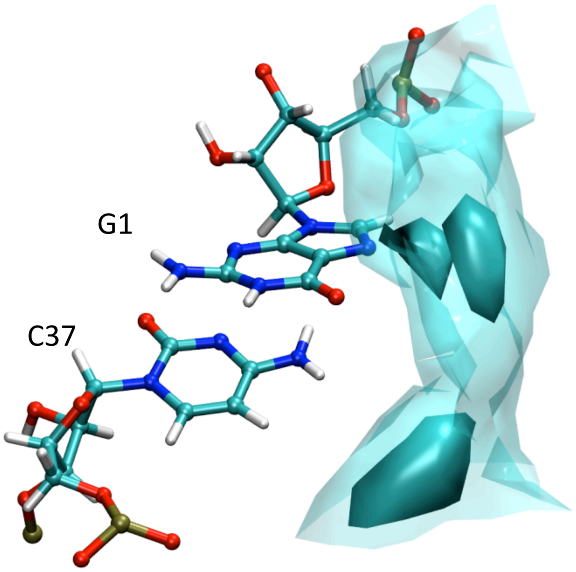

**Figure B.9.** Charge iso-density plots of metal ions for U37C mutant. Second independent trajectory of C75$^+$/Mg$_{rev}$ when U37 is mutated to C37, as per Figure 3.7B. The darker cyan volumes represent the top 30% occupancy of a unit positive charge and the transparent cyan volumes represent the top 80% occupancy.

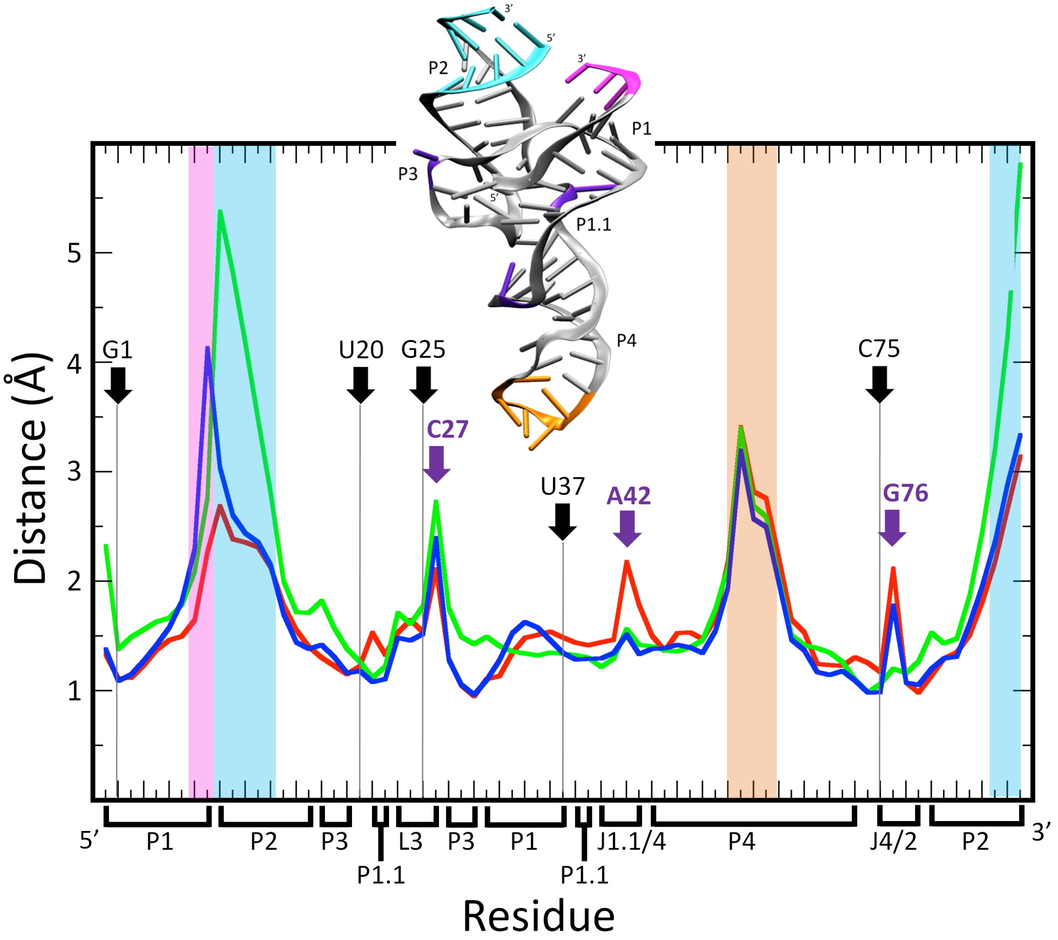

**Figure B.10.** Root Mean Square Fluctuation (RMSF) of heavy atoms. Second independent trajectory similar to Figure 3.9.

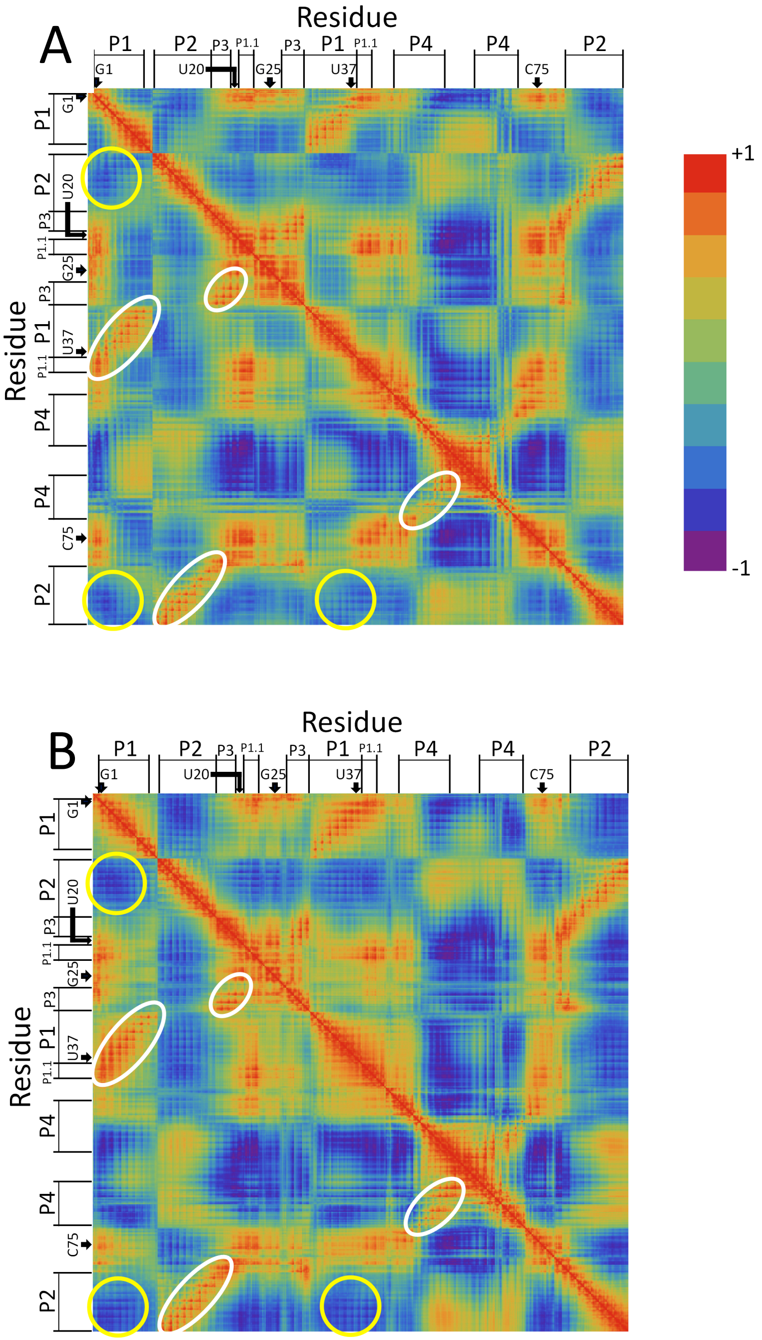

**Figure B.11.** Cross-correlation plot of heavy atoms. Heavy atom based cross-correlation plot for C75$^{+}$/noMg$_{rev}$ and C75$^{o}$/Mg$_{rev}$ over 25ns of MD trajectory is shown in panels A and B respectively, similar to Figure 3.10.

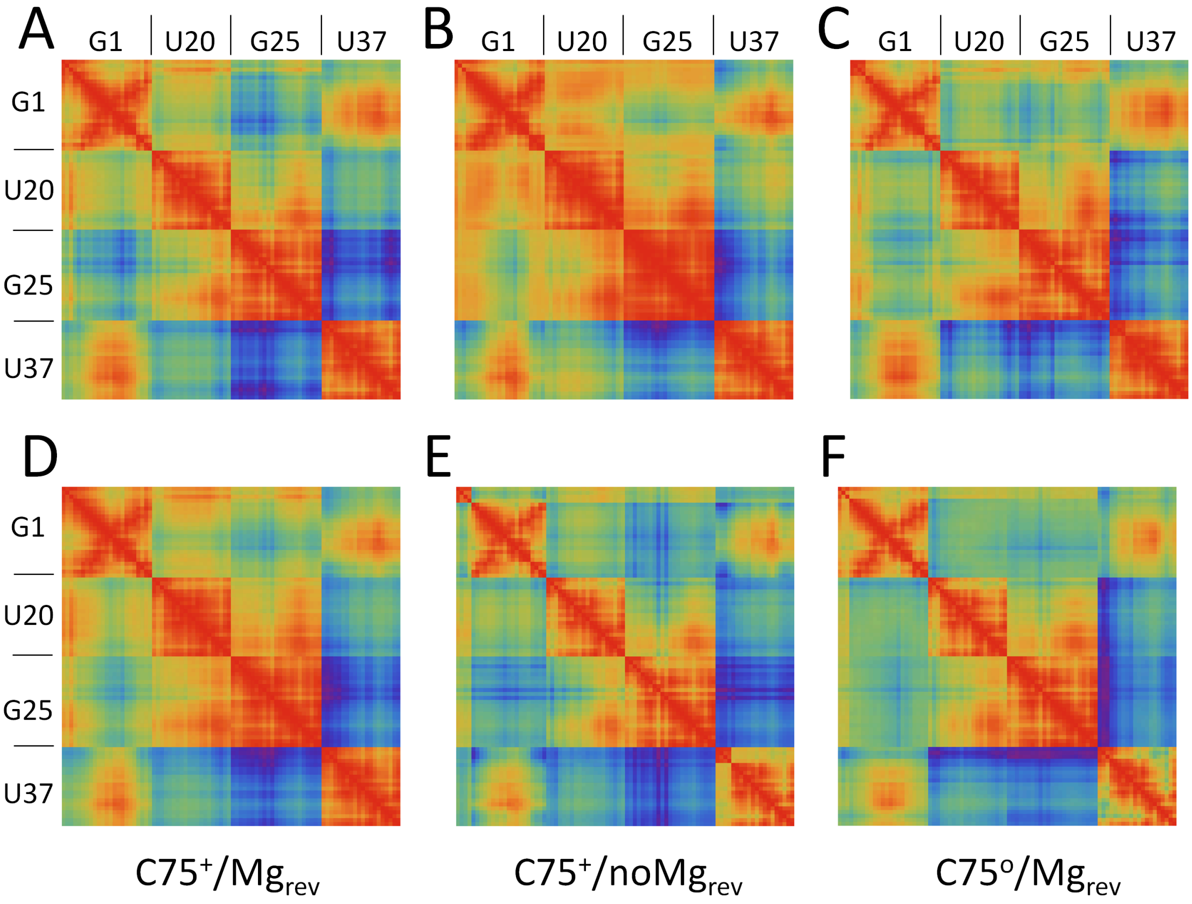

**Figure B.12.** Cross-correlation plot of heavy atoms in the reverse (G25•U20) and normal (G1•U37) wobbles. Heavy atom based cross-correlation plot for $C75^{+}/Mg_{rev}$, $C75^{+}/noMg_{rev}$ and $C75^{o}/Mg_{rev}$ over 25ns of MD trajectory is shown in panels A, B and C respectively. Corresponding plots for a second independent trajectory is shown in panels D, E and F. Residues corresponding to the heavy atoms are shown on both axes. The coloring scheme is as per Figure 3.10. Panel A is the same as Figure 3.11.

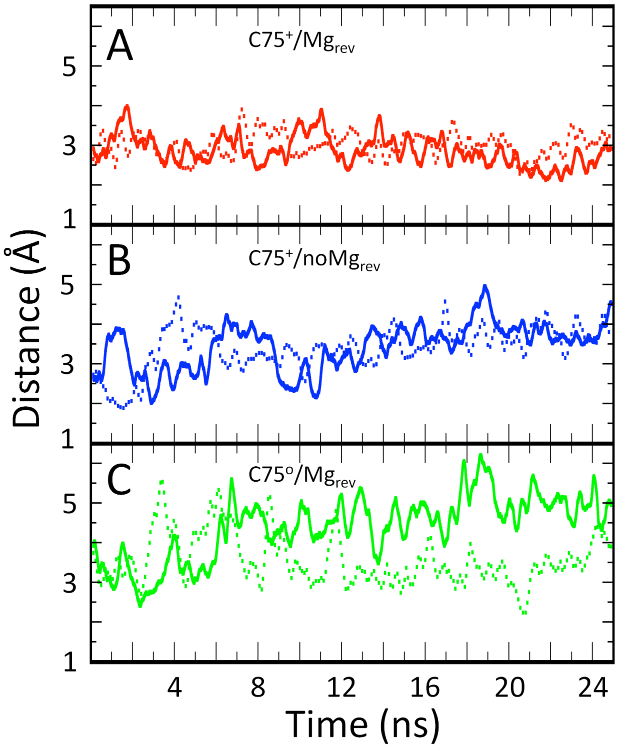

**Figure B.13.** Root mean square deviation (RMSD) plots for the heavy atoms precleaved structures. Solid and dotted lines indicate two independent trajectories. RMSDs were calculated with reference to the crystal structures.

## IV. Supporting References

(1) Nosé, S. *Mol. Phys.* **1984**, *52*, 255-268.
(2) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695-1697.

# APPENDIX C: SUPPORTING INFORMATION FOR CHAPTER 4[1]

**Contents:**

---

[1] This appendix was published in its entirety in the journal Biochemistry in 2013. The complete reference to the journal article is as follows: Chen, J.; Ganguly, A.; Miswan, Z.; Hammes-Schiffer, S.; Bevilacqua, P. C.; Golden, B. L. *Biochemistry*, **2013**, *52*, 557-567. C. J. and M. Z. performed the experimental studies and A. G. performed the computational studies.

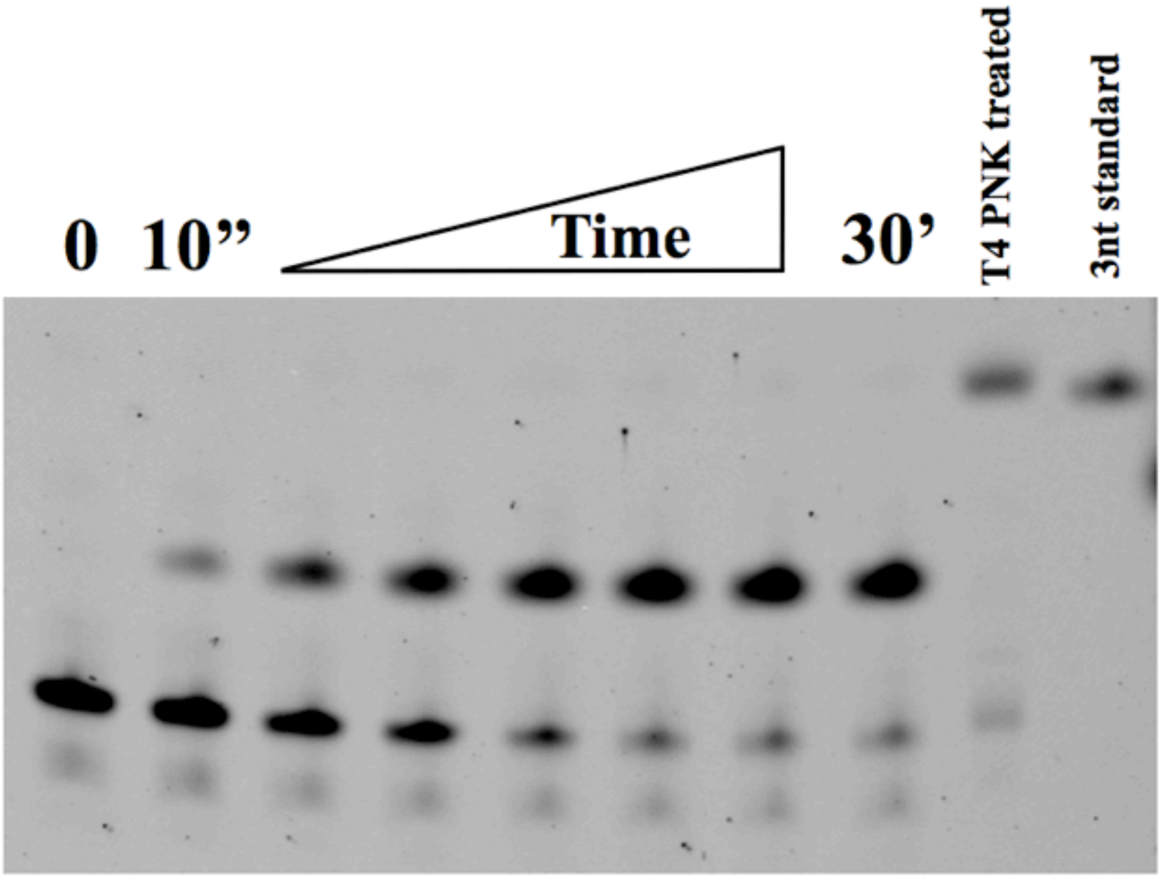

**Figure C.1.** Identification of the HDV ribozyme cleavage product. The WT HDV ribozyme was mixed with a fluorophore-labeled substrate DY547-UAU*GGCUUGCA and 0.5 mM $MgCl_2$ in MOPS pH 7.0. Ribozyme cleavage resulted in a product with reduced mobility as predicted for the reaction product, DY547-UAU with a 2’, 3’-cyclic phosphate terminus, as the fluorophore is bulky and carries a positive charge. To verify the nature of this product, it was treated with T4 polynucleotide kinase and calf intestinal phosphatase to remove the 2’,3’-cyclic phosphate[1] and generate DY547-UAU with free 2’- and 3’-hydroxyl groups. This product co-migrates with an authentic DY547-UAU standard (see last two right-hand lanes).

Materials and Methods for Figure C.2

*pH-rate profile of the G25A•U20C double mutant in the absence of Mg2+.* To characterize the pH dependence of ribozyme reaction under $Mg^{2+}$-free conditions, the *cis*-cleaving one-piece -30/99 HDV ribozyme was transcribed as described above for the trans-cleaving ribozyme. This RNA is based on the sequence of the genomic HDV ribozyme and incorporates the G11C mutation that enhances the folding properties of the RNA. It has a 5'-end that is 30 nucleotides upstream of the cleavage site and a 3'-end that is 99 nucleotides downstream of the cleavage site. This is referred to as the wild-type ribozyme for the purpose of this study. The A25•C20 mutation was introduced into this construct using QuikChange to generate an G25A•U20C double mutant self-cleaving RNA. Preparation of the one-piece ribozyme was necessary because the two-piece HDV ribozyme does not react in the conditions needed to monitor the reaction in the absence of $Mg^{2+}$ ion.

The 5'-end of the RNA was radiolabeled using polynucleotide kinase and $[\gamma\text{-}^{32}P]$-ATP and gel-purified as described for the two-piece ribozyme. To perform the cleavage assays, ~2 nM of labeled HDV ribozyme was renatured at 55°C for 10 min in TE buffer (0.5 mM Tris-HCl pH7.5 and 0.05 mM EDTA pH 8.0) and allowed to cool down at room temperature for 10 min. The folded RNA was then pre-incubated at 37°C for 2 min with saturating 10 μM DNA oligomer (AS1) that is complementary to the single-stranded RNA upstream of the cleavage site (nucleotides -30 through -3) in 50 mM (final concentration) of one of the following buffers: potassium acetate (pH5.0 and 5.5), potassium MES (pH6.0 and pH6.5) or Tris-HCl (pH7.0-pH8.5). AS1 DNA oligomer sequesters the inhibitory strand upstream of the cleavage site and helps to prevent misfolding.{Chadalavada, 2000 #1386} All of the buffers were supplemented with 100 mM EDTA (final concentration) prior to adjusting the pH to the final values. No attempt was made to correct the meter reading for effects on ionic strength. This may lead to an underestimate of $pK_a$ values by ~ 0.5 units, but does not affect comparative values.{Cerrone-Szakal, 2008 #665} A zero time point was removed followed by adding pre-warmed 1 M NaCl to initiate the reaction. The time course experiment was incubated at 37°C for up to ~200 h. Samples at appropriate time points were removed, quenched and analyzed on a 6% denaturing PAGE. Quantitation was performed by a PhosphorImager (Molecular Dynamics). Reaction progress was analyzed by KaleidaGraph 4.1 (Synergy Software). Data fitting was described in Materials and Methods, with an exception of the G25A•U20C double mutant reacting at pH8.0, where $f = k_{obs}t$ was used to determine the $k_{obs}$ value (~60% of completion was reached at 144h, only the initial ~30% was used for data fitting).

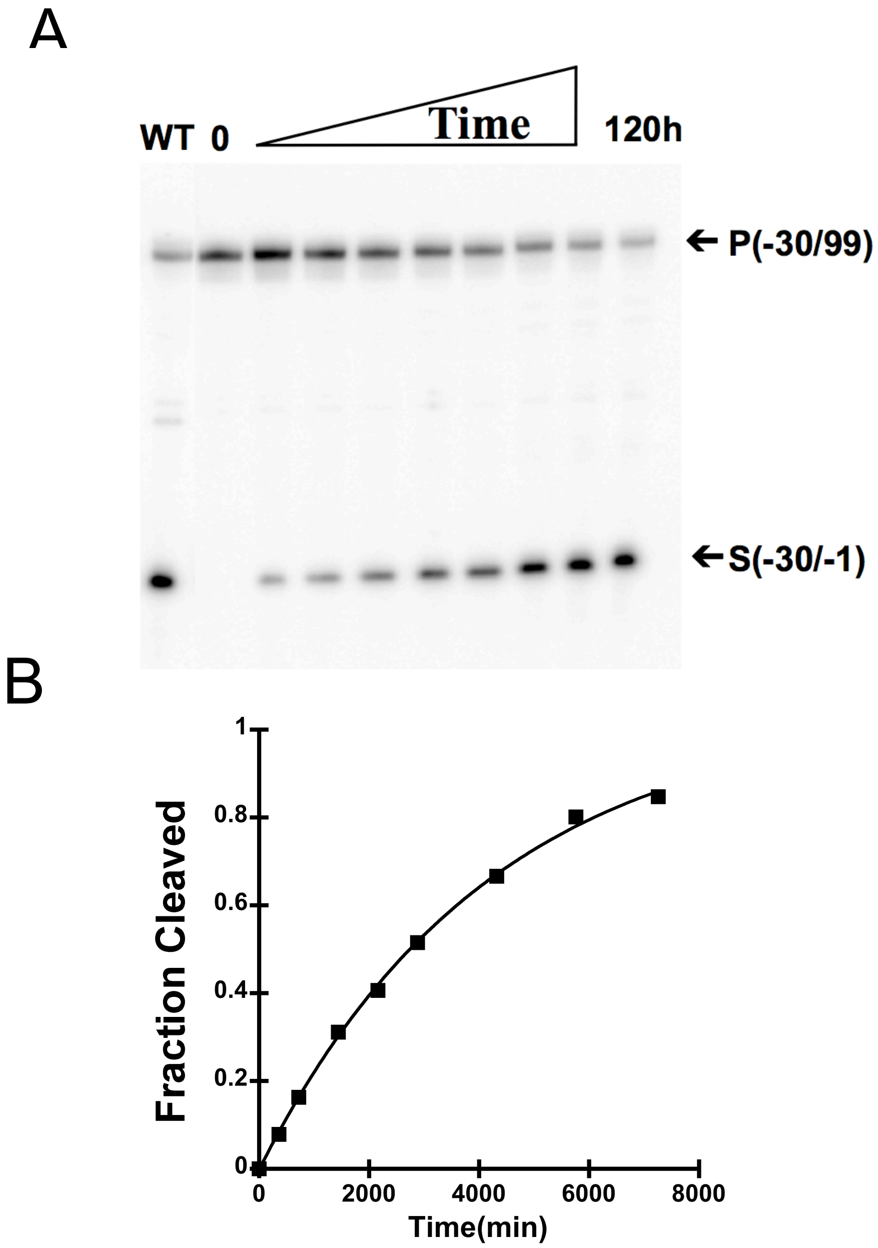

**Figure C.2.** The self-cleavage reaction of the G25A•U20C double mutant. **A.** The denaturing PAGE shows a cleavage reaction of the 5'-end-labeled G25A•U20C double mutant in 1M NaCl, 50 mM Tris, and 100mM EDTA at pH7.0. The -30/99 precursor ribozyme and the -30/-1 cleavage substrate are indicated as P and S, respectively. A wild type ribozyme (WT) cleaved at 24h under the same condition is shown for comparison. **B.** Panel A was quantitated and fraction of cleavage was plotted against time to yield the graph of reaction progress. The G25A•U20C double mutant cleaves to more than 80% of completion at 120h in a monophasic fashion. The data were fit to equation (1).

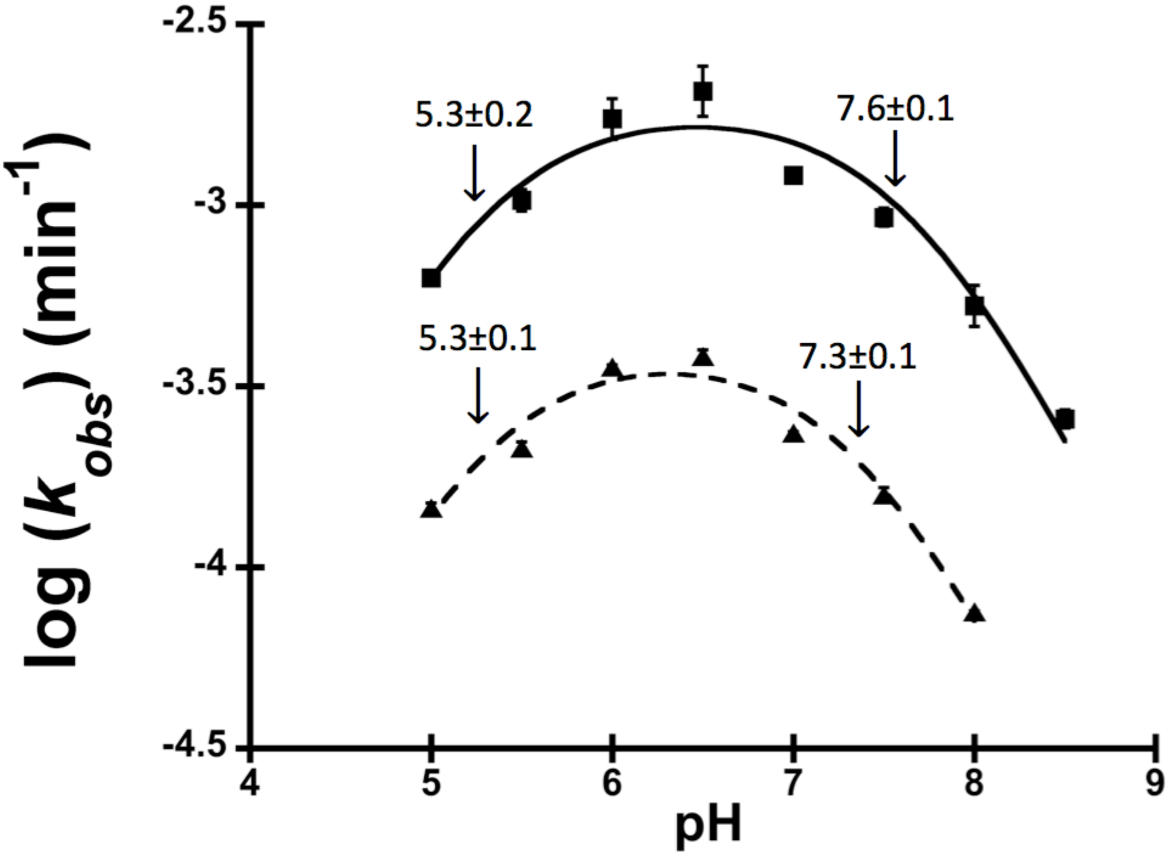

**Figure C.3.** The pH-rate profiles for the self-cleavage reactions of the G25A•U20C double mutant and WT in the absence of $Mg^{2+}$ ion. The reactions were performed in 1 M NaCl, 50 mM of buffer, and 100 mM EDTA for the WT (squares) and the G25A•U20C double mutant (triangles). The data were fit to equation (5) to account for the two ionizable groups (See Discussion), which yielded apparent p*K*a's of 5.3 ± 0.2 and 7.6 ± 0.1 for the WT and apparent p*K*a's of 5.3 ± 0.1 and 7.3 ± 0.1 for the the G25A•U20C double mutant, as illustrated in the figure.

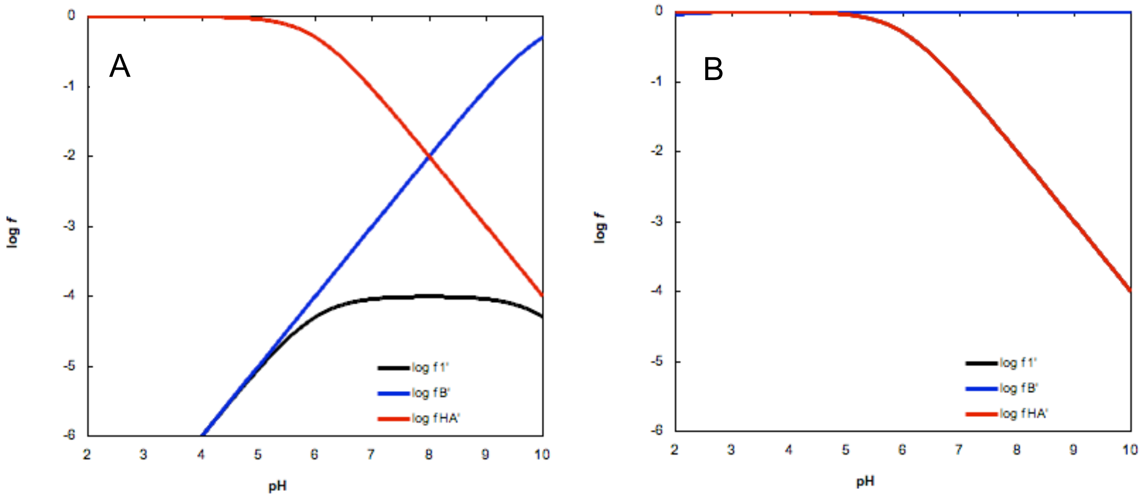

**Figure C.4.** Sum of rate-pH profiles to generate observed rate-pH profiles. Simulations are according to Bevilacqua.[2] **A.** Origin of rate-pH profile in the *presence* of catalytic $Mg^{2+}$. The functional form of the base—either magnesium hydroxide as a Brønsted (general) base or magnesium as a Lewis acid and OH- as a specific base (see Figure C.5A and B for these and related, dual mechanisms)—is depicted in blue, and the functional form of the acid, presumably $C75^{+}$, is depicted in red. Their logarithmic sum, which represents the fraction of the ribozyme in the fully functional form, is shown in black. The shape of the black curve is similar to the rate-pH profiles seen for WT in the presence of $Mg^{2+}$ (Figure 4.4, squares). **B.** Origin of rate-pH profile in the *absence* of catalytic $Mg^{2+}$. The functional form of the base—either water or a non-ionizing functional group on the RNA, and/or some other base possibly in a non-rate limiting step (see Figure C.5C)—is depicted in blue, and the functional form of the acid, presumably $C75^{+}$, is depicted in red. Their logarithmic sum, which represents the fraction of the ribozyme in the fully functional form, is black (hidden under red curve). The shape of the black curve is similar to profiles seen for the G25A•U20C double mutant in the presence of $Mg^{2+}$, which does not participate in the reaction (Figure 4.4, triangles). In both simulations (panels A and B), the $pK_a$ of the general acid is set at 6; the $pK_a$ of the general base is set at 10 in panel A and at -1.7 (the $pK_a$ of water) in panel B. Experiments can only be conducted up to pH~8 owing to alkaline denaturation of the ribozyme, which comes from hydroxide competing for hydrogen bonding.[3] Another simulation could be made for panel B with a $pK_a$ of 5.4 for the general base that would produce the slight low pH arm in Figure 4.4 (triangles); this was not done here because of the possibility that this effect arises from acid denaturation or contribution of a catalytic metal ion only at low pH (see main text).

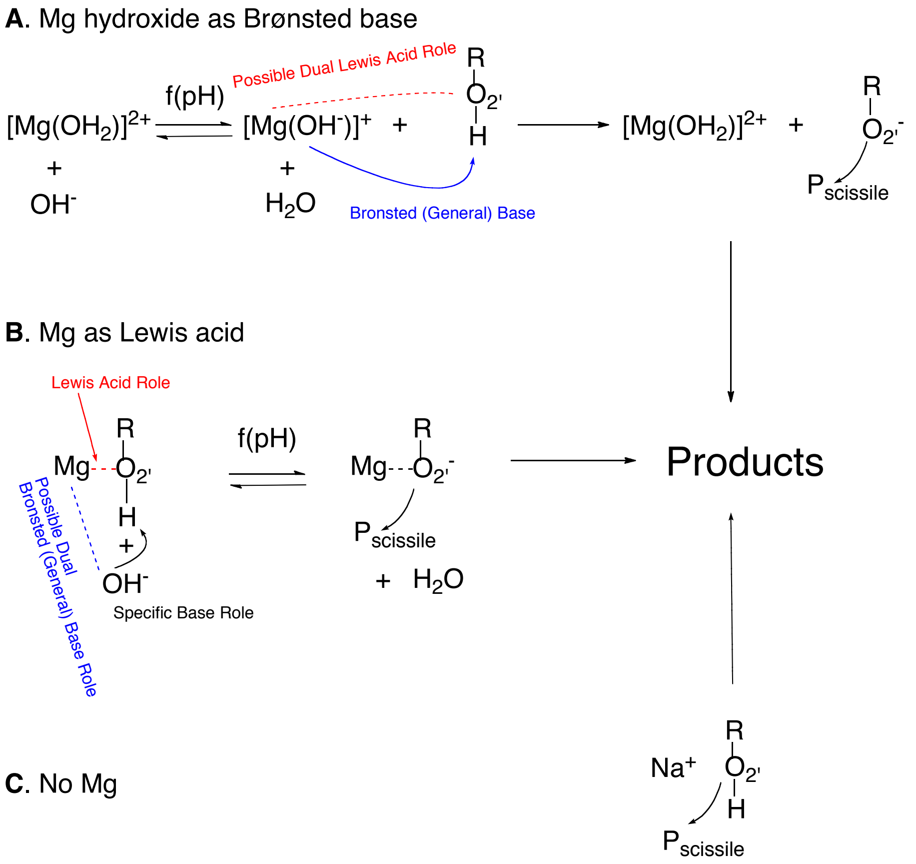

**Figure C.5.** Possible mechanisms for deprotonation of the 2'-hydroxyl. Panels **A** and **B** depict mechanisms in the *presence* of $Mg^{2+}$, which are for WT, while panel **C** depicts a mechanism in the *absence* of $Mg^{2+}$, which is for G25A•U20C double mutant. Steps that are a function of pH, which is important in interpreting rate-pH profiles such as in Figure C.4, are noted. **A.** Magnesium hydroxide serves as a Brønsted (general) base (blue), with a possible dual role as a Lewis acid at the 2'O (red). **B.** Magnesium serves as a Lewis acid (red), with a possible dual role as a general base through a $Mg^{2+}$-bound hydroxide (blue), otherwise some other species needs to act as the base in a pH-dependent fashion, such as hydroxide as a specific base. Alternately, the $pK_a$ reflects the $pK_a$ of the 2'OH of U(-1) itself. **C.** Reaction without catalytic magnesium. Here population of the general base is independent of pH (Figure C.4B, blue line), so either water ($pK_a$ of -1.7) or some other low $pK_a$ group ($pK_a$ should be <~4, the lower end of the experimental pH range) perhaps on the RNA (not shown) acts as the base, and/or proton transfer from the 2'OH occurs in a non rate-determining step. $Na^+$ bound nearby could have a general electrostatic effect (see text). See Emilsson *et al.* for related mechanisms.[4] A reason hydroxide could deprotonate the 2'OH in panel B but not C is that the $Mg^{2+}$ could lower the $pK_a$ of the 2'OH in panel B.

**Table C.1.** Statistics from MD simulation of the G25A•U20C double mutant of the HDV ribozyme[a]

| Trajectory | $Mg^{2+}$ ion present initially | A25(N1)-C20(N4) (Å) | A25(N6)-C20(N3) (Å) | A25(N6)-metal ion (Å) | A25(N7)-metal ion (Å) | C20(O2)-metal ion (Å) |
|---|---|---|---|---|---|---|
| | | | | | | |
| 1 | Yes | 3.10 (0.19) | 2.91 (0.09) | 4.32 (0.23) | 4.76 (0.17) | 4.26 (0.22) |
| 2 | Yes | 3.00 (0.12) | 2.90 (0.09) | 4.25 (0.17) | 4.90 (0.20) | 4.00 (0.16) |
| | | | | | | |
| 1 | No | 3.03 (0.15) | 2.97 (0.16) | 3.41 (0.35) | 5.20 (0.29) | 2.53 (0.53) |
| 2[b] | No | 3.05 (0.15) | 2.97 (0.12) | 4.23 (0.32) | 4.95 (0.26) | 4.14 (0.66) |

[a] Values are distances with standard deviations shown in parentheses.

[b] In this trajectory, a sodium ion did not move into the catalytic site until the first 1.5 ns. The data shown here was collected after the sodium ion moved in, over the next 23.5 ns.

## Supporting References

(1) Morse, D. P., and Bass, B. L. *Biochemistry* **1997**, *36*, 8429-8434.
(2) Bevilacqua, P. C. *Biochemistry* **2003**, *42*, 2259-2265.
(3) Moody, E. M., Lecomte, J. T., and Bevilacqua, P. C. *RNA* **2005**, *11*, 157-172.
(4) Emilsson, G. M., Nakamura, S., Roth, A., and Breaker, R. R. *RNA* **2003**, *9*, 907-918.

# APPENDIX D: SUPPORTING INFORMATION FOR CHAPTER 5[1]

**Contents:**

---

[1] This chapter was published in its entirety in the Journal of Physical Chemistry Letters in 2013. The complete reference to the journal article is as follows: Ganguly, A.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Phys. Chem. Lett.* **2011**, *2*, 2906-2911.

## I. Supporting Methods

1) Molecular Dynamics (MD) equilibration procedure

The equilibration procedure for the starting structure of the QM/MM calculations was as follows:

1. Energy optimization of the solvent (waters, $Na^+$ and $Cl^-$ ions) with the solute (RNA and crystallographic $Mg^{2+}$) fixed.

2. Simulated annealing (NVT ensemble, Nosé-Hoover thermostat[1,2]) of solvent keeping solute fixed:

i. Increase temperature from 0K to 298K in 10 ps, MD @ 298 K for 50 ps

ii. Increase temperature from 298K to 498K in 10 ps, MD @ 498K for 50 ps

iii. Increase temperature from 498K to 698K in 10 ps, MD @ 698K for 50 ps

iv. Decrease temperature from 698K to 498K in 10 ps, MD @ 498K for 50 ps

v. Decrease temperature from 498K to 298K in 10 ps, MD @ 298 K for 150 ps

The above protocol was performed two times.

3. MD at 298K with NPT ensemble using Nosé-Hoover barostat[1,2] for 100 ps with solute fixed.

## II. Supporting Figures

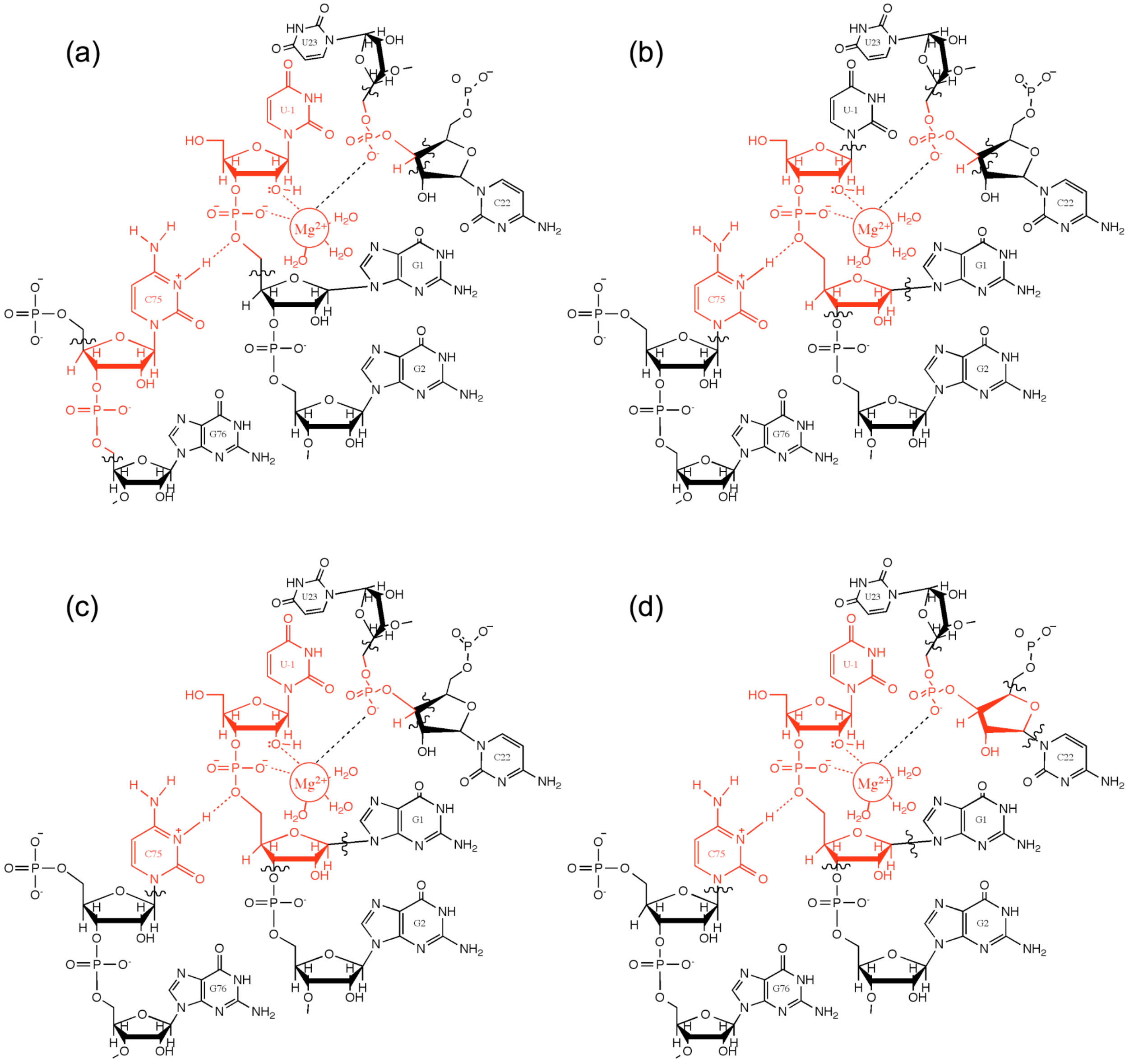

**Figure D.1.** QM regions considered for the QM/MM calculations. Atoms in red are included in the QM region. Wavy lines indicate the QM-MM boundary. Table D.1 provides the key distances after a QM/MM optimization with each of these QM regions. Region (d), which contains 87 QM atoms, was chosen for the subsequent calculations used to generate the results in the paper and is provided in Figure 5.1.

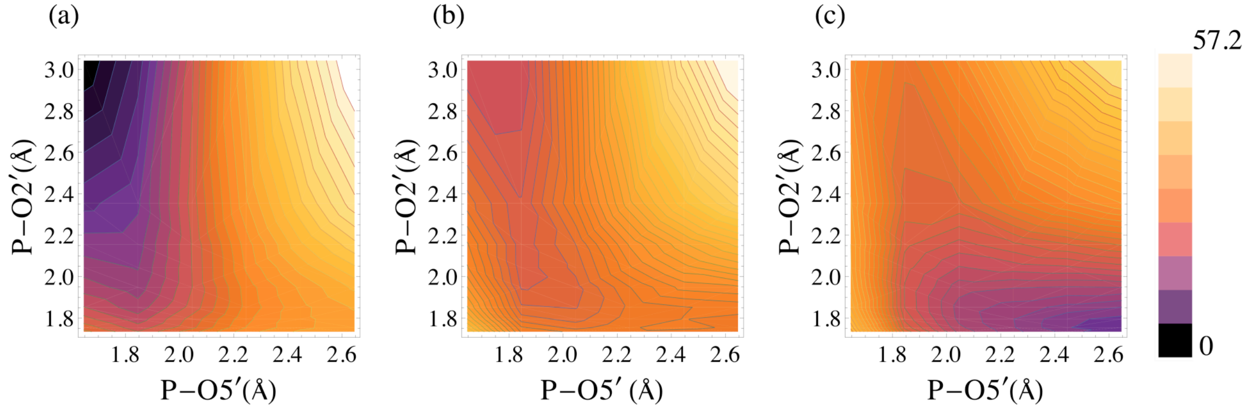

**Figure D.2**. Potential energy surfaces obtained from constrained optimizations along P(G1)-O5′(G1) and P(G1)-O2′(U-1) coordinates for (a) H3 bonded to N3 of C75, (b) H3 in the middle of N3(C75) and O5′(G1) at the TS position, (c) H3 bonded to O5′(G1). The legend for the color-coding is shown on the extreme right in units of kcal/mol. In part (a), to mimic the proton transfer reactant state, the H3(C75)-N3(C75) distance was constrained to its value in the optimized reactant structure. In part (b), to mimic the proton transfer transition state, the H3(C75)-N3(C75), H3(C75)-O5′(G1), and N3(C75)-O5′(G1) distances were constrained to the distances in the TS. In part (c), to mimic the proton transfer product state, the H3(C75)-O5′(G1) distance was constrained to its value in the optimized product structure. To generate each point on the potential energy surfaces, the P-O5′ and P-O2′ interatomic distances, as well as the relevant proton transfer interface distances, were constrained while all other coordinates were optimized using the QM/MM methods described in the paper. The minimum observed in panel (a) (top left of the surface) corresponds to the reactant state, and the minimum observed in panel (c) (bottom right of the surface) corresponds to the product state. No minima corresponding to intermediate states were observed.

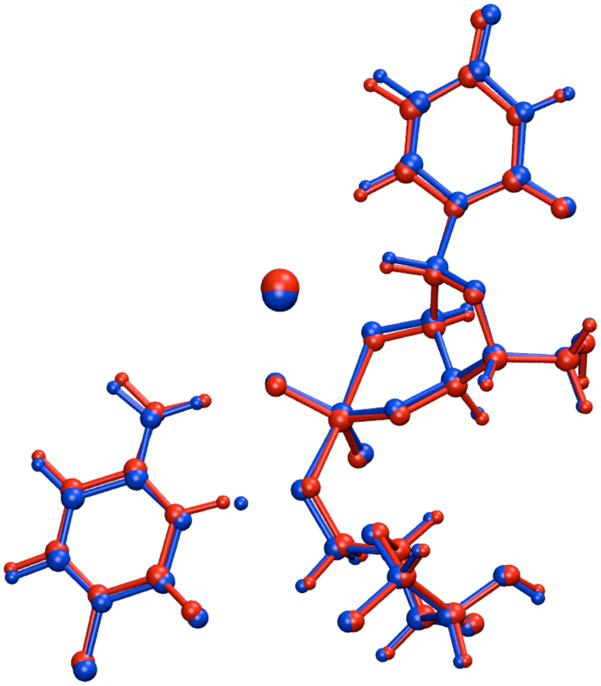

**Figure D.3.** Overlay of the phosphorane-like transition state (blue) with the phosphorane intermediate (red). The prominent difference is in the proton H3 position, which is midway between N3(C75) and O5′(G1) in the transition state but bonded to N3(C75) in the phosphorane intermediate. C75 is shown in the lower left of the figure.

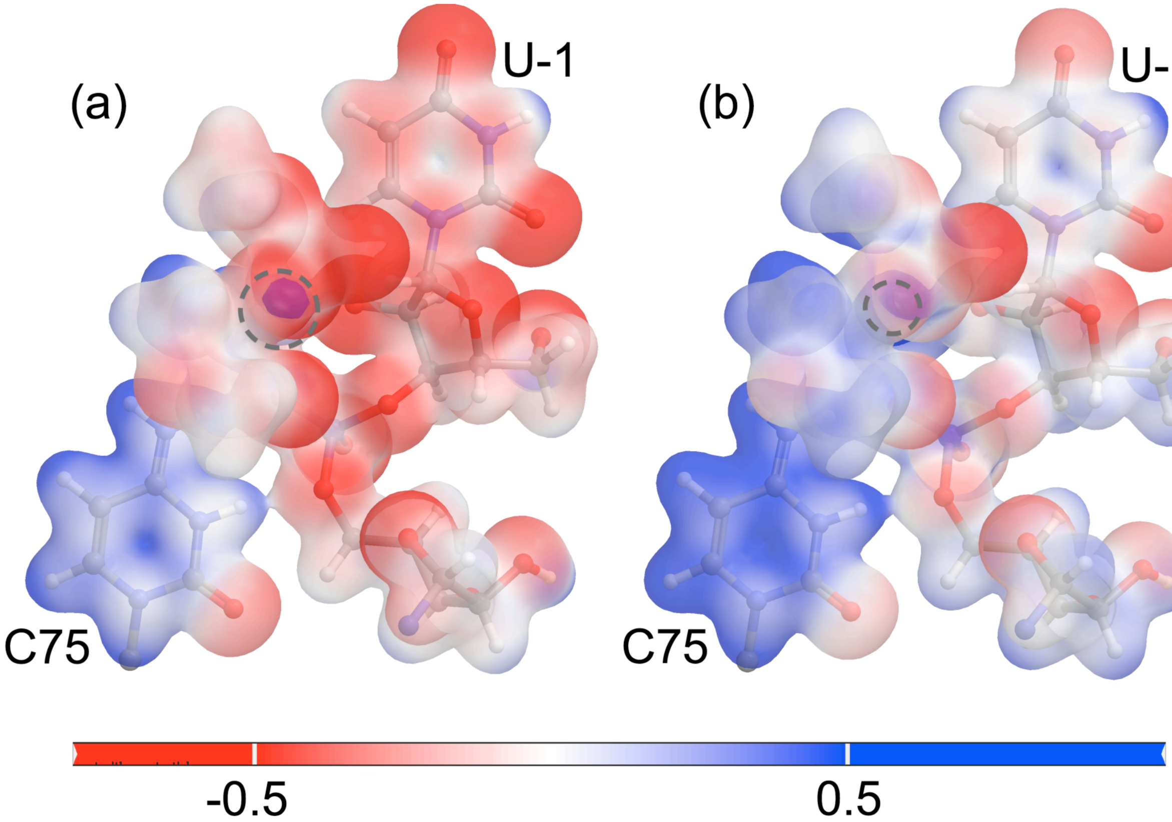

**Figure D.4**. Molecular electrostatic potential mapped onto the electron density isosurface. The QM region is shown here for the (a) optimized reactant state with the $Mg^{2+}$ at the catalytic position, with the $Mg^{2+}$ ion replaced by a $Na^{+}$, and (b) optimized reactant state with a $Na^{+}$ ion at the catalytic position, with the $Na^{+}$ ion replaced by a $Mg^{2+}$. In both panels, the scale is from -0.5 to 0.5 kT/e at 298 K. The dotted gray circle indicates the position of the metal ion in each case. The similarity of panel (a) with Figure 5.4(b) and of panel (b) with Figure 5.4(a) suggests that the difference in the electrostatic potential of the residue C75 is governed by the presence of a divalent or a monovalent ion at the catalytic site, rather than structural differences.

## I. Supporting Tables

| | crystal structure | Region (a) | Region (b) | Region (c) | Region (d) |
|---|---|---|---|---|---|
| N3(C75)-O5′(G1) (Å) | 3.58 | 3.51 | 3.24 | 3.25 | 3.22 |
| N3(C75)-H-O5′(G1) (°) | 154.43 | 162.75 | 163.25 | 160.93 | 157.0 |
| $Mg^{2+}$-HOH (1) (Å) | 1.97 | 2.20 | 2.26 | 2.19 | 2.19 |
| $Mg^{2+}$-HOH (2) (Å) | 1.98 | 2.04 | 2.05 | 2.04 | 2.05 |
| $Mg^{2+}$-HOH (3) (Å) | 1.91 | 2.13 | 2.14 | 2.13 | 2.13 |
| $Mg^{2+}$-O2′(U-1) (Å) | 2.37 | 2.30 | 2.29 | 2.28 | 2.28 |
| $Mg^{2+}$-O1P(U23) (Å) | 2.24 | 1.94 | 1.93 | 1.94 | 1.95 |
| O2′(U-1)-P(G1) (Å) | 3.30 | 3.12 | 3.10 | 3.08 | 3.09 |
| N4(C75)-O2P(G1) (Å) | 3.58 | 3.41 | 3.13 | 3.16 | 3.14 |

**Table D.1.** Comparison of key distances and angles in the active site of the HDV ribozyme after QM/MM optimization with the QM regions depicted in Figure D.1. Based on these values, region (d) was chosen for the subsequent calculations used to generate the results in the paper. Region (a) was eliminated because the H-cap on C5′(G1) was too close to the reaction center and affected the N3(C75)-O5′(G1) distance. Region (b) was eliminated because the exclusion of the U-1 residue affected the $Mg^{2+}$-HOH (1) distance. Region (c) was eliminated because H-capping inside the C22 sugar led to artifacts in frequency calculations. Note that the optimized reactant structure for region (d) differs slightly from that reported in the paper because the optimization of the reactant reported in this table was performed directly from the starting structure, while the optimized reactant reported in the paper was obtained from the minimum energy path starting at the transition state.

| | TS with $Mg^{2+}$ | TS with $Ca^{2+}$ | PI with $Li^{+}$ | PI with $Na^{+}$ | PI with $K^{+}$ |
|---|---|---|---|---|---|
| N3(C75)-O5′(G1) (Å) | 2.63 | 2.64 | 2.94 | 2.91 | 2.89 |
| N3(C75)-H3(C75) (Å) | 1.32 | 1.45 | 1.05 | 1.06 | 1.06 |
| O5′(G1)-H3(C75) (Å) | 1.31 | 1.19 | 1.90 | 1.86 | 1.84 |
| P(G1)-O5′(G1) (Å) | 2.05 | 1.87 | 1.81 | 1.85 | 1.86 |
| P(G1)-O2′(U-1) (Å) | 1.95 | 2.41 | 1.95 | 1.91 | 1.89 |
| P(G1)-O2P(G1) (Å) | 1.56 | 1.53 | 1.57 | 1.56 | 1.56 |
| O1P(G1)-O2P(G1)-P(G1)-O3′(U-1) (improper dihedral) (°) | 178.6 | 160.4 | 175.2 | 179.5 | 178.9 |

**Table D.2.** Comparison of the active site distances and angle of the phosphorane-like transition state (TS) with a $Mg^{2+}$ or $Ca^{2+}$ ion at the catalytic position and the phosphorane intermediate (PI) with monovalent ions at the catalytic position.

## I. Supporting References

(1) Nosé, S. *Mol. Phys.* 1984, 52, 255-268.
(2) Hoover, W. G. *Phys. Rev. A* 1985, 31, 1695-1697.

## APPENDIX E: SUPPORTING INFORMATION FOR CHAPTER 6[1]

**Contents:**

---

[1] This chapter was published in its entirety in the Journal of Biochemistry in 2013. The complete reference to the journal article is as follows: Thaplyal, P.; Ganguly, A.; Golden, B. L.; Hammes-Schiffer, S.; Bevilacqua, P. C. *Biochemistry* **2013**, 52, 6499-6514. P.T. performed the experiments and A.G. performed the computational studies.

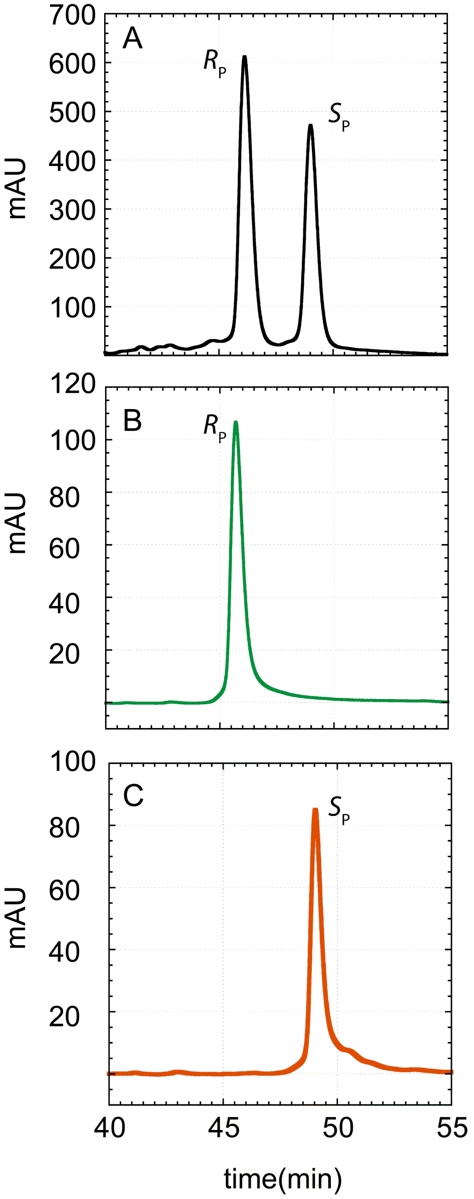

**Figure E.1.** Separation of $R_P$ and $S_P$ diastereomers using reverse phase HPLC. (A) Absorbance versus time (min) for a mixture of $R_P$ and $S_P$ diastereomers containing 300 pmoles of oligonucleotide. The $R_P$ diastereomer elutes first followed by the $S_P$ diastereomer, as confirmed by snake venom phosphodiesterase cleavage (see main text). (B) Re-injection of the purified $R_P$ diastereomer after several weeks of storage at -20 ºC in sterile water. (C) As per panel B, except for re-injection of the purified $S_P$ diastereomer. Oxo elutes at ~43 min (not shown). Note that there is no detectable contamination of oxo or the other diastereomer substrates in panels (B) and (C).

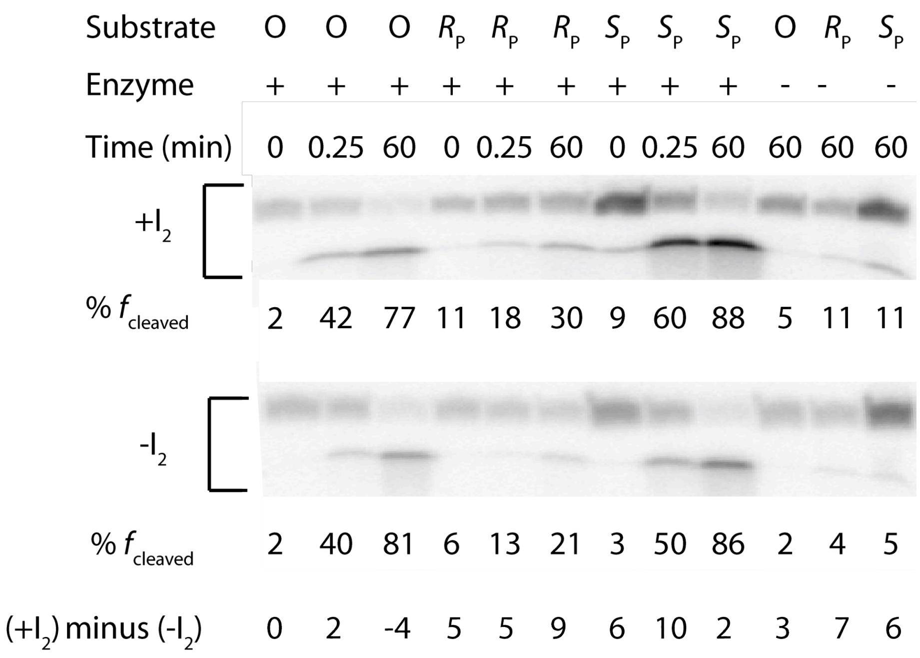

**Figure E.2.** Iodine cleavage of oxo, *R*$_P$, and *S*$_P$ substrates. Samples were treated with 1 mM $I_2$ at 90 ºC for 2 min *after* any enzyme reaction. In the absence of enzyme (last three lanes, top and bottom panels), oxo substrate shows minimal cleavage (~3%) upon treatment with $I_2$ (bottom row) whereas both *S*$_P$ and *R*$_P$ substrates show ~7% cleavage for the same treatment (bottom row). In the presence of enzyme, differential cleavage (+$I_2$ minus –$I_2$) is noted (see values across bottom of figure): oxo substrate still shows minimal cleavage (-4 to 2% differential cleavage) whereas both *R*$_P$ and *S*$_P$ substrates again show higher cleavage in the presence of iodine (5 to 10% differential cleavage). Note that the low extent of backbone cleavage overall for both of the phosphorothioate substrates is intrinsic to the $I_2$ reaction and is due to the majority of the $I_2$-specific reaction proceeding by another channel, as described by Eckstein.[1,2] Nonetheless, it is clear that the *R*$_P$ and *S*$_P$ substrates are equally reactive with $I_2$ and the oxo substrate is unreactive. Also, the contribution of $I_2$ to overall cleavage of the substrate does not change appreciably after the ribozyme is introduced. (The low differential cleavage for *S*$_P$ substrate at 1 h is because the reaction has already proceeded nearly to completion before $I_2$ treatment.)

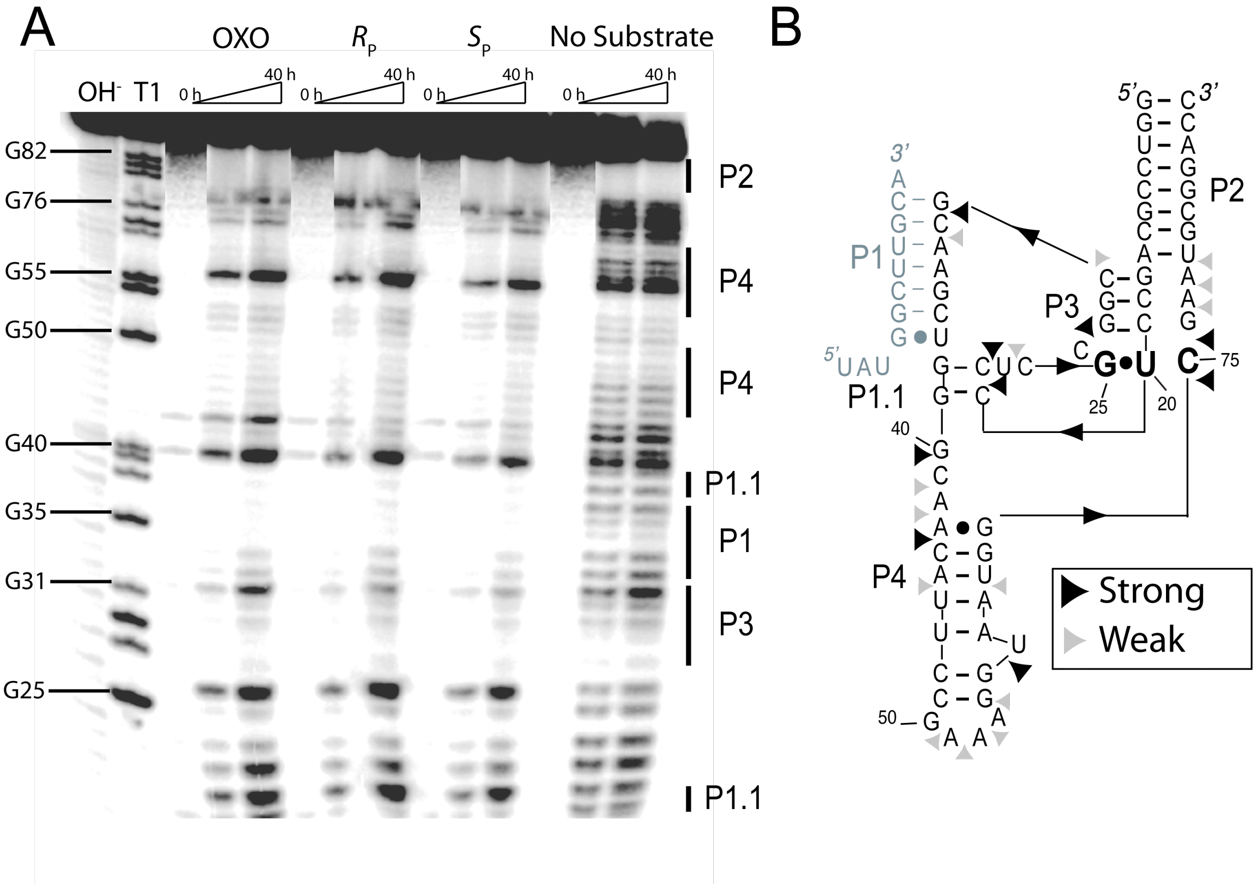

**Figure E.3:** In-line probing (ILP) of the enzyme strand of the HDV ribozyme. Conditions were pH 8.3, 20 mM Mg$^{2+}$, 100 mM K$^{+}$, 37 °C at 0, 20, and 40 h, similar to typical in-line probing conditions.[3] The enzyme strand is 5'-end labeled with $^{32}$P with unlabeled oxo, $R_P$, and $S_p$ substrates present in saturating amounts (see Materials and Methods). A.) Denaturing 8% polyacrylamide gel. 'OH$^{-}$' is a denaturing alkaline cleavage treatment, which reveals position of all residues, and 'T1' is a denaturing RNase T1 treatment, which reveals positions of all G residues. The oxo, $R_P$, and $S_P$ lanes exhibit similar structural changes in the enzyme strand during the course of 40 h, including intermediate times, suggesting no major conformational differences. The 'No Substrate' lane has very different reactivity than the other three conditions, supporting that the enzyme is saturated with substrate in the other conditions. B) Secondary structure mapping of the in-line probing gel. Most of the nucleotides undergoing in-line probing attack are in non-base paired regions or near the end of a helix. The intensity of the bands has been denoted by different colors, with strong and weak intensity bands colored black and grey, respectively. Substrate is colored blue to emphasize that it is not directly detected in this experiment.

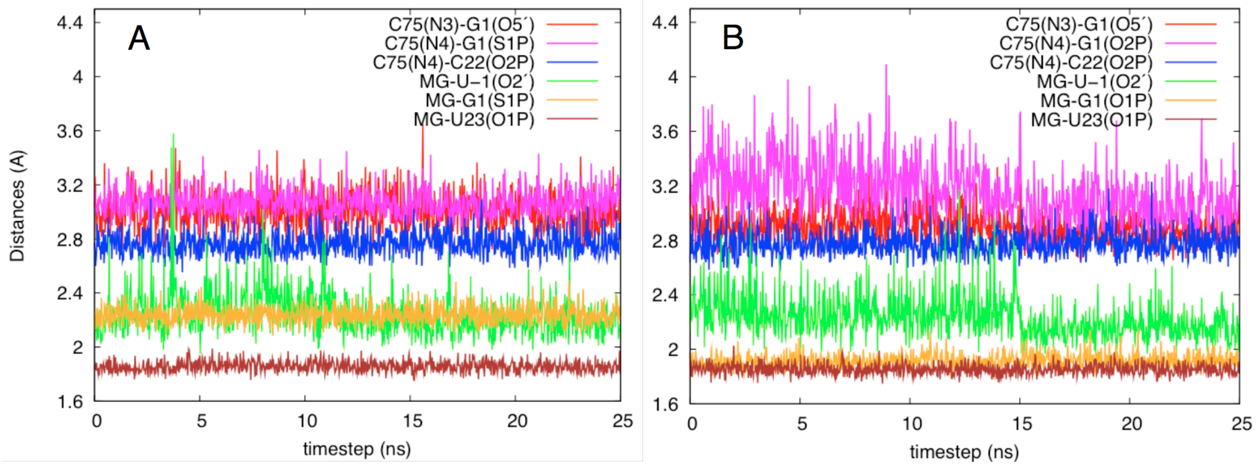

**Figure E.4.** Molecular dynamics simulation of the phosphorothioate substrates. The key active site hydrogen bonding distances for A) $R_{\mathrm{P}}$ and B) $S_{\mathrm{P}}$ substrates bound to the HDV ribozyme. The active site remained stable throughout the course of the 25 ns trajectory for both substrates. The key hydrogen bonds involving C75 remained intact throughout the course of the trajectory, showing that C75 was not being ejected out from the active site for either substrate. Two independent trajectories were carried out for both $S_{\mathrm{P}}$ and $R_{\mathrm{P}}$ substrates, both of which gave similar results.

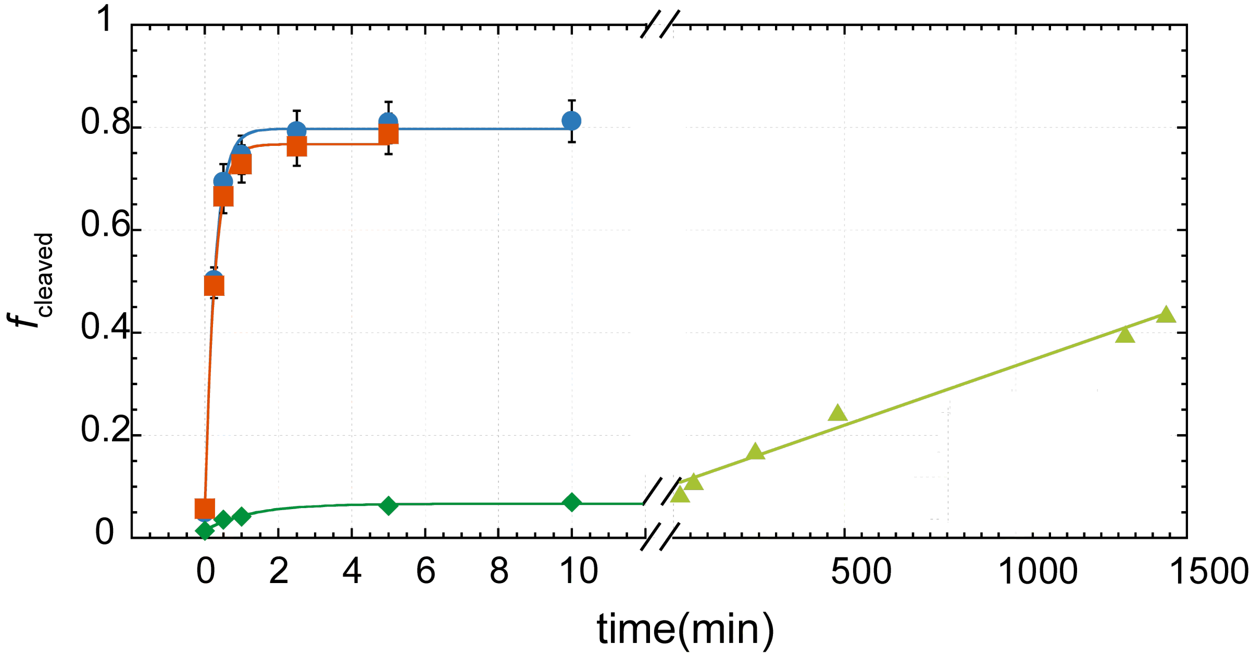

**Figure E.5.** Absence of metal ion rescue in the presence of $Mn^{2+}$. Fraction cleaved vs time plots of the oxo (blue circles), $S_P$ (red squares), $R_P$ fast phase (dark green diamonds), and $R_P$ slow phase (light green triangles) in the presence of 2 mM $Mn^{2+}$ at pH 5.6. No rescue is observed with $Mn^{2+}$ (Table 1).

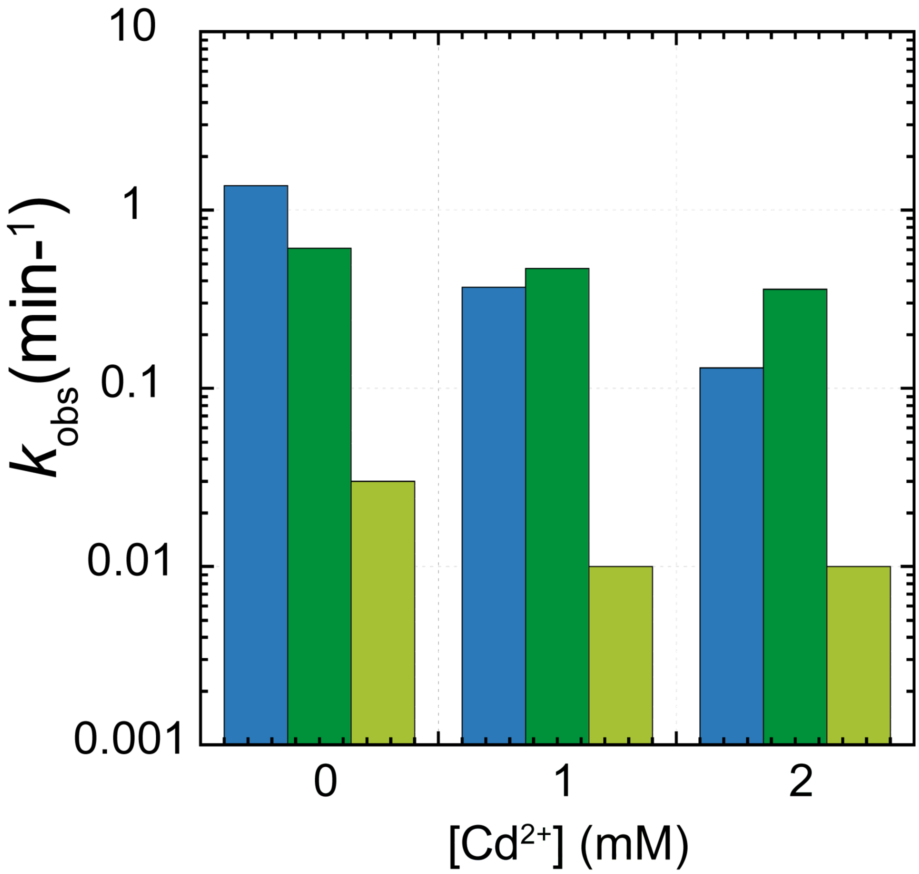

**Figure E.6:** Reduced inhibition by $Cd^{2+}$ in the presence of a GC base pair at the cleavage site. Histogram representing the inhibition in the presence of $Cd^{2+}$ for oxo (blue), $R_P$ fast phase (dark green), and $R_P$ slow phase (light green) substrates. All reactions were in the presence of 10 mM $Mg^{2+}$ and pH 5.6. The rate constants for the oxo substrate were as follows, GC Watson-Crick: 1.29, 0.40, and 0.13 min$^{-1}$ in 0, 1, and 2 mM $Cd^{2+}$ (from this Figure), and G•U wobble pair: 4.1, 0.60, and 0.17 min$^{-1}$ in 0, 1, and 2 mM $Cd^{2+}$ (from Figure 3, Main Text), respectively. Thus, the inhibition by $Cd^{2+}$ for the GC Watson-Crick base pair at the active site was reduced relative to wild-type, although it was not removed.

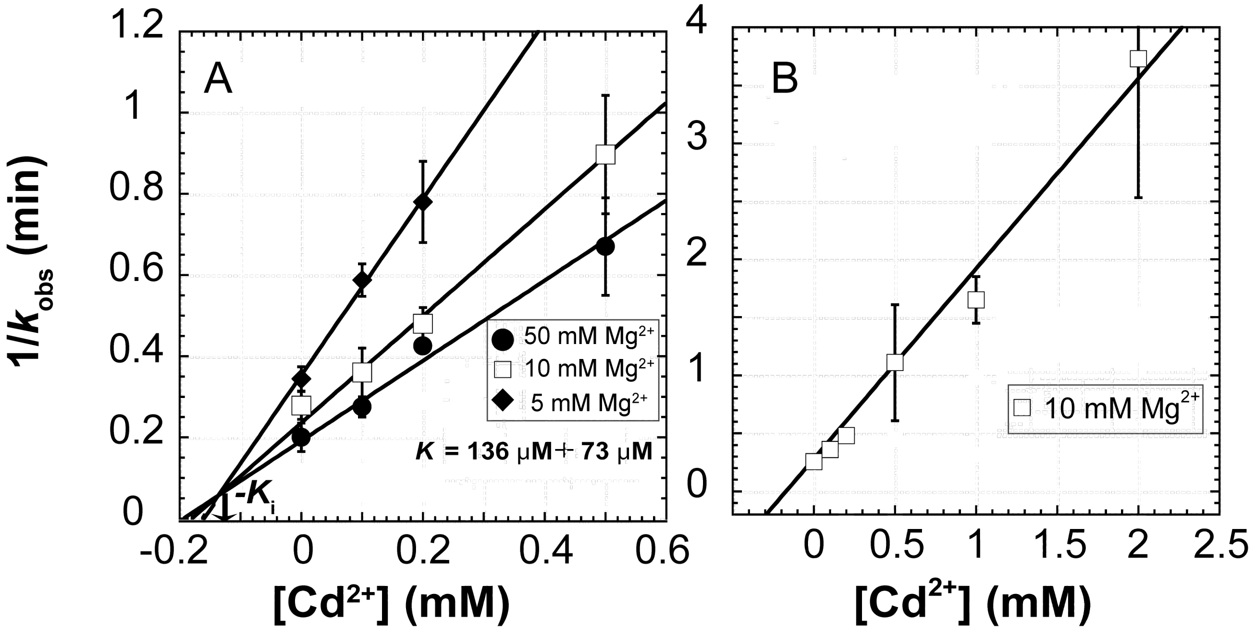

**Figure E.7:** Competitive inhibition by $Cd^{2+}$. A.) Dixon plot for the inhibition of the reaction rate, with increasing concentration of $Cd^{2+}$ for the oxo substrate at pH 5.6 in 5, 10, and 50 mM $Mg^{2+}$. The lines corresponding to the three different $Mg^{2+}$ concentrations intersect in the second quadrant, suggesting that $Cd^{2+}$ acts as a competitive inhibitor, with $K_\mathrm{i}$ of 136 ± 73 $\mu$M depicted graphically in the figure. Note that these data were globally fit forcing them to converge at a single point. The equations for the global fit were obtained using gnuplot and then the figure was made on Kaleidagraph. B.) Inhibition of the reaction rate in the presence of 10 mM $Mg^{2+}$ including data out to 2 mM $Cd^{2+}$. This plot shows that the linear trend of the inhibition of reaction rate by $Cd^{2+}$ shown in panel A (the exact same line is plotted in both panels) is maintained even at these higher concentrations of $Cd^{2+}$ where the inhibitor $Cd^{2+}$ is in large excess of the inhibitor enzyme complex. This provides evidence that the free concentrations of $Cd^{2+}$ equal the total concentrations of $Cd^{2+}$ even at lower concentrations of $Cd^{2+}$, owing to the linear extrapolation. Apparently the concentration of free $Cd^{2+}$ is not affected appreciably by binding diffusely to the ribozyme, perhaps because diffuse sites are filled by the excess divalent species, which is $Mg^{2+}$. Increasing error bars with increasing $Cd^{2+}$ concentration are due to the ordinate being a reciprocal axis, not to the percentage error systematically increasing.

Supplementary References:

(1) Eckstein, F., and Gish, G. *Trends in Biochem. Sci.* **1989**, *14*, 97–100.

(2) Gish, G., and Eckstein, F. *Science* **1988**, *240*, 1520–1522.

(3) Soukup, G. A., and Breaker, R. R. *RNA* **1999**, *5*, 1308–1325.

# APPENDIX F: SUPPORTING INFORMATION FOR CHAPTER 7[1]

## Contents

---

[1] This chapter was published in its entirety in the Journal of the American Chemical Society in 2014. The complete reference to the journal article is as follows: Ganguly, A.; Thaplyal, P.; Rosta, E.; Hammes-Schiffer, S.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2013**, 5*2*, 6499-6514. A.G. performed the computational studies, P.T. performed the experiments, and E.R. provided valuable technical and scientific help.

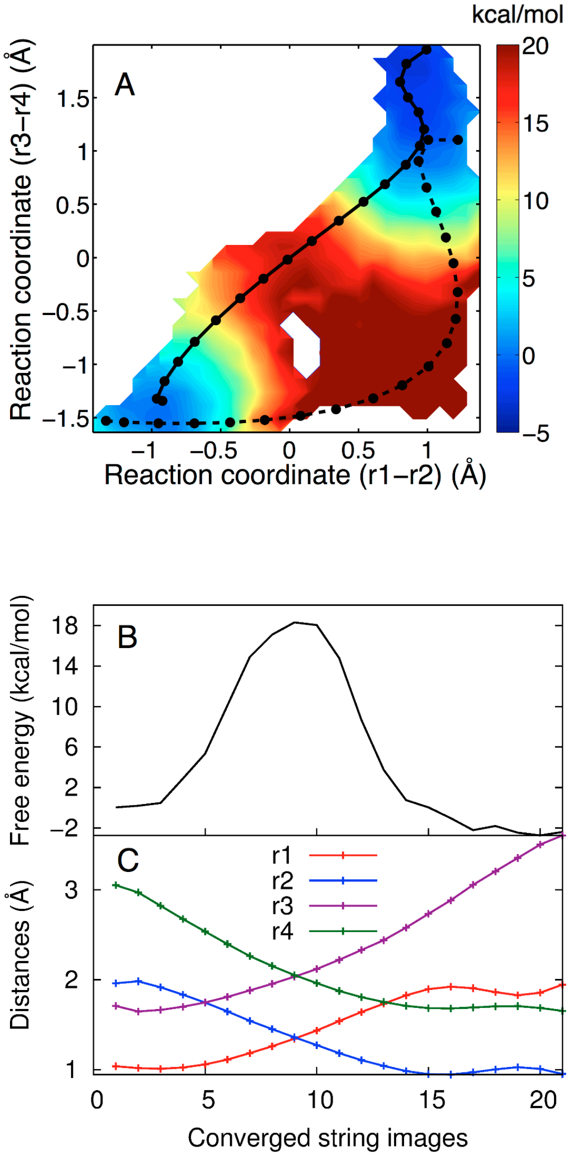

**Figure F.1.** (A) The 2D free energy surface obtained from set B, where Mg$^{2+}$ is bound at the catalytic site, projected in the (r1−r2) and (r3−r4) space. The initial string (dashed black line) corresponds to an initial guess pathway in which the proton is transferred before the oxygen-phosphorus bond breaking/forming. To conserve computational resources, the string calculations in this set were terminated after 30 string iterations, when the reaction pathway associated with the final string (solid black line) was determined to be qualitatively similar to that obtained from sets A and C. The MFEP corresponds to a concerted mechanism with a phosphorane-like TS. Each circle is associated with an image along the string. The color scale denotes free energy in units of kcal/mol. (B) The 1D free energy profile along the MFEP obtained from set B, where $Mg^{2+}$ is bound at the catalytic site. The modestly higher free energy barrier (~5 kcal/mol) along this MFEP, as compared to the barrier along the MFEP obtained from set A, suggests that the string is still evolving but is nearing convergence. (C) Values of the most important reaction coordinates, r1, r2, r3, and r4, along the MFEP. Each circle corresponds to an image along the string.

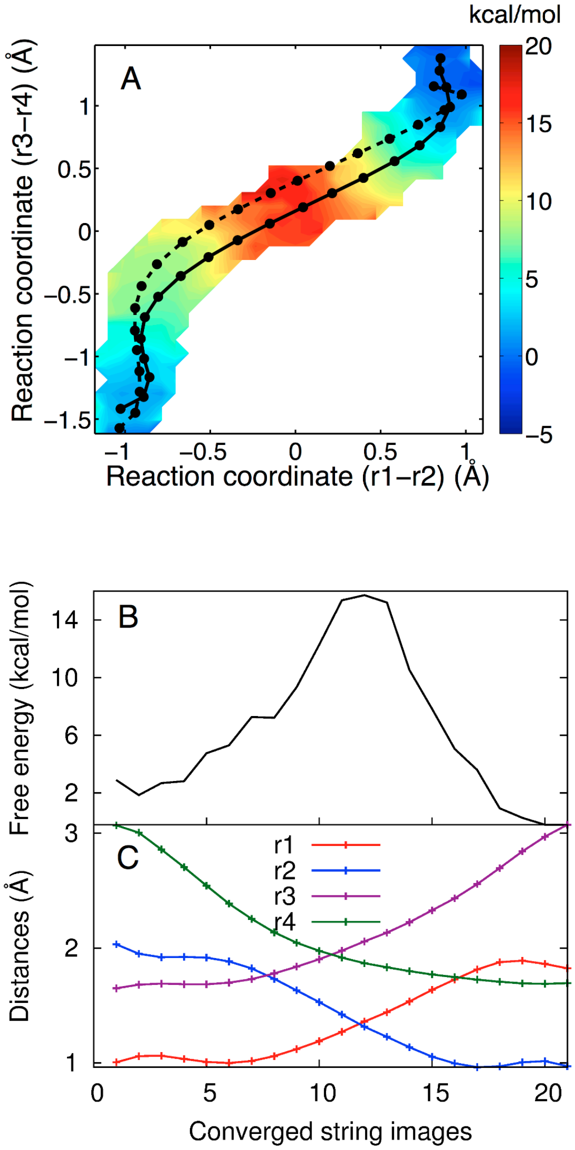

**Figure F.2.** (A) The 2D free energy surface obtained from set C, where $Mg^{2+}$ is bound at the catalytic site, projected in the (r1–r2) and (r3–r4) space. The initial string (dashed black line) corresponds to an initial guess pathway passing through a phosphorane intermediate. The converged string (solid black line) represents the MFEP obtained from set C. The MFEP corresponds to a concerted mechanism with a phosphorane-like TS. Each circle is associated with an image along the string. The color scale denotes free energy in units of kcal/mol. (B) The 1D free energy profile along the MFEP obtained from set B, where $Mg^{2+}$ is bound at the catalytic site. The slightly higher free energy barrier (~2 kcal/mol) along this MFEP, as compared to the barrier along the MFEP obtained from set A, suggests that the string is nearly converged within the accuracy of the method. (C) Values of the most important reaction coordinates, r1, r2, r3, and r4, along the MFEP. Each circle corresponds to an image along the string.

## STRING CONVERGENCE

### Description of string convergence criterion

The string in a given simulation set was considered to be converged when the root-mean-squared deviation (RMSD) of all coordinates of the latest string in that particular set from the mean value of the previous ten iterations, summed over all images, fell below a given threshold, which was chosen to be 0.1 Å. Mathematically, the RMSD of the reaction coordinate *k* for the *j*th string update is calculated according to the equation

$$\mathrm{RMSD}_{j,k} = \sqrt{\frac{1}{N}\sum_{i}^{N}\left(f_j^k(\xi_i) - \overline{f_j^k(\xi_i)}\right)^2} \qquad \text{(F.1)}$$

where

$$\overline{f_j^k(\xi_i)} = \frac{1}{10}\sum_{j'=j-10}^{j-1} f_{j'}^k(\xi_i). \qquad \text{(F.2)}$$

Here *N* is the number of images considered along the string, and a specific image is represented by the index *i*. String convergence is plotted for several sets in the figures F.3-F.6.

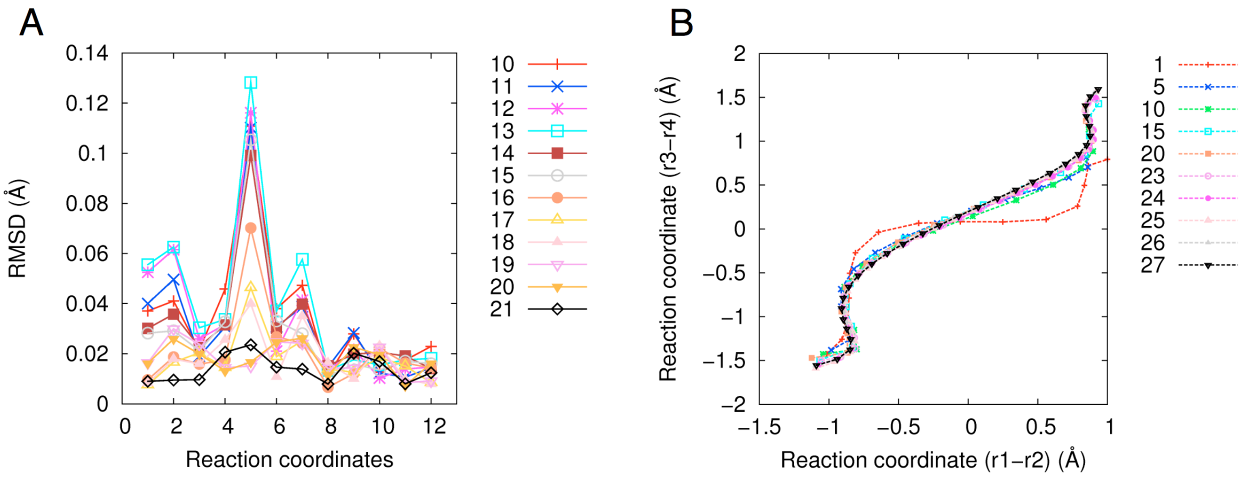

**Figure F.3.** String convergence for set A, where $Mg^{2+}$ is bound at the active site. In total, 27 string iterations were performed. For iterations 1-21, 15 images were considered along the string. For iterations 22-27, the number of images along the string was increased to 30. (A) RMSD of all of the reaction coordinates is shown for iterations 10-21. Iterations 22-27 are excluded from the figure because the number of images in these iterations is not consistent with the rest of the iterations. Moreover, these iterations are essentially superimposable on one another. This figure indicates that the RMSD of the string for all reaction coordinates is well below 0.1 Å by iteration number 21. The additional iterations were performed to get a higher resolution of the MFEP and a smoother free energy profile. (B) The strings are plotted in the 2D space of the (r1−r2) and (r3−r4) coordinates for the selected iterations, as indicated in the figure key. This figure suggests that the string has stopped evolving in the space of the (r1−r2) and (r3−r4) coordinates.

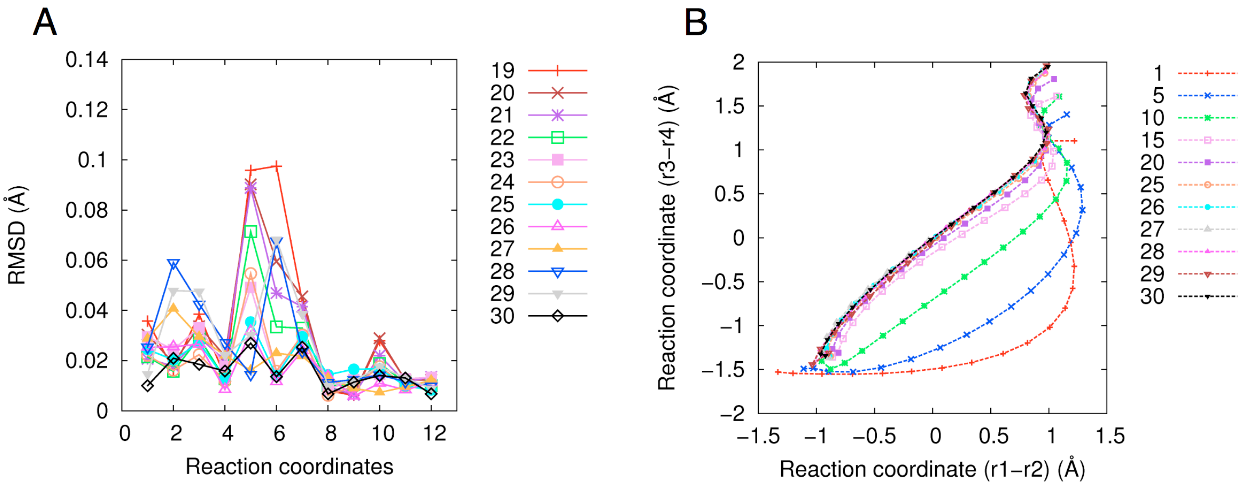

**Figure F.4.** String convergence for set B, where $Mg^{2+}$ is bound at the active site. In total, 30 string iterations were performed. For all iterations, 21 images were considered along the string. (A) RMSD of all of the reaction coordinates is shown for the last 12 iterations. (B) The strings are plotted in the 2D space of the (r1–r2) and (r3–r4) coordinates for the selected iterations, as indicated in the figure key. As shown in the figure, the strings evolve significantly for the early iterations. The evolution of the strings for the later iterations is slow but is in the direction of the final strings obtained from simulation sets A and C. The string calculations for this set were terminated after iteration 30, when the reaction pathway, as indicated by the latest string (black curve), was determined to be qualitatively similar to that obtained from sets A and C.

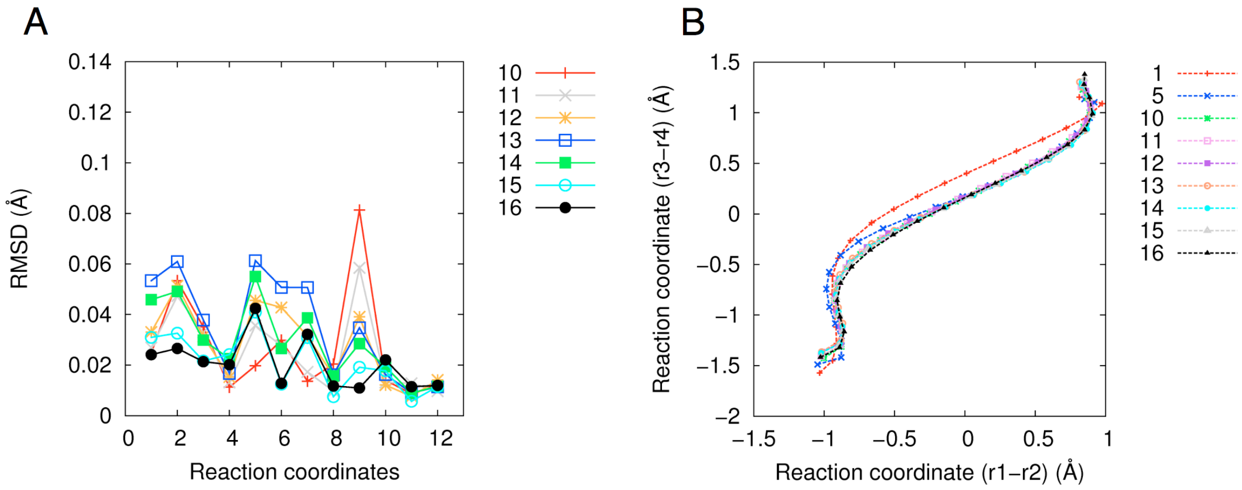

**Figure F.5.** String convergence for set C, where $Mg^{2+}$ is bound at the active site. In total, 16 string iterations were performed. For all iterations, 21 images were considered along the string. (A) RMSD of all of the reaction coordinates is shown for the last 7 iterations. (B) The strings are plotted in the 2D space of the (r1–r2) and (r3–r4) coordinates for the selected iterations, as indicated in the figure key. The string calculations for this set were terminated after iteration 16, when the reaction pathway, as indicated by the latest string (black curve), was determined to be very similar to the converged string from simulation set A.

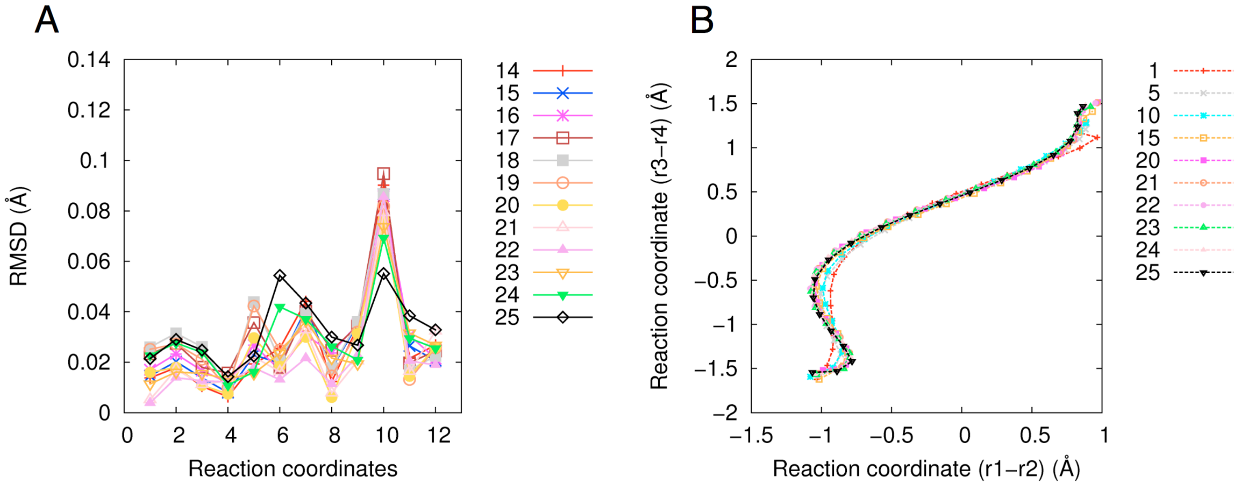

**Figure F.6.** String convergence for set D, where Na$^+$ is bound at the active site. In total, 25 string iterations were performed. For all iterations, 21 images were considered along the string. (A) RMSD of all of the reaction coordinates is shown for the last 12 iterations. (B) The strings are plotted in the 2D space of the (r1–r2) and (r3–r4) coordinates for the selected iterations, as indicated in the figure key.

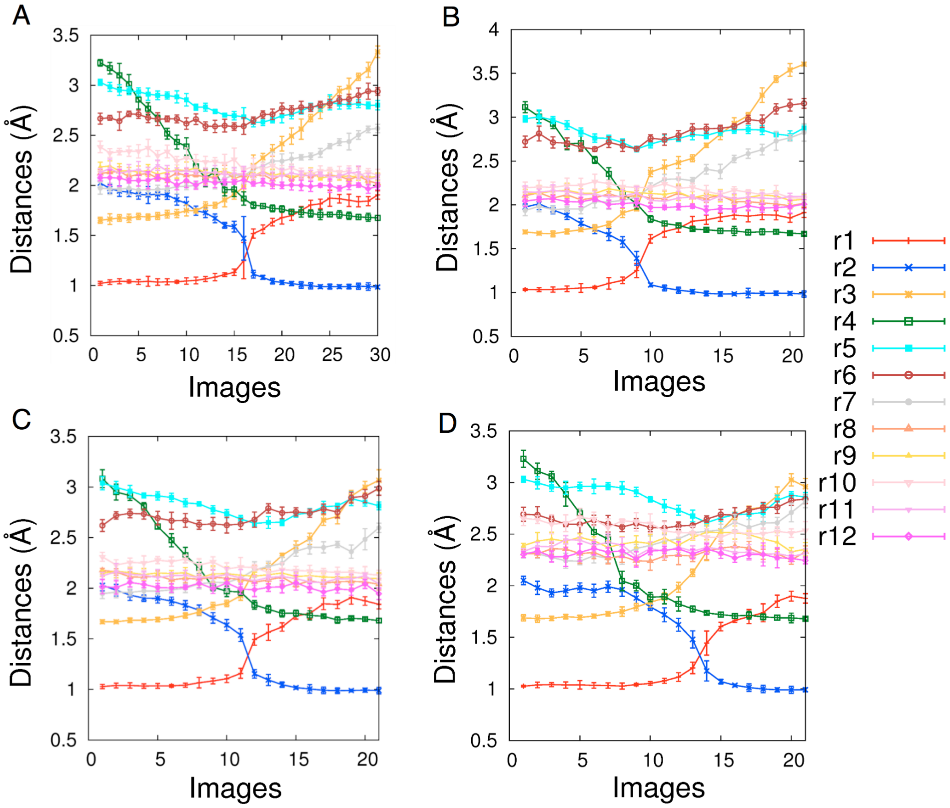

**Figure F.7.** Average values of all reaction coordinates considered in the string simulations for the final iteration of simulation set A (Panel A), set B (Panel B), set C (Panel C), and set D (Panel D). The figure illustrates that in each of the simulation sets, the most important reaction coordinates are r1, r2, r3, and r4 because these coordinates change significantly along the MFEP.

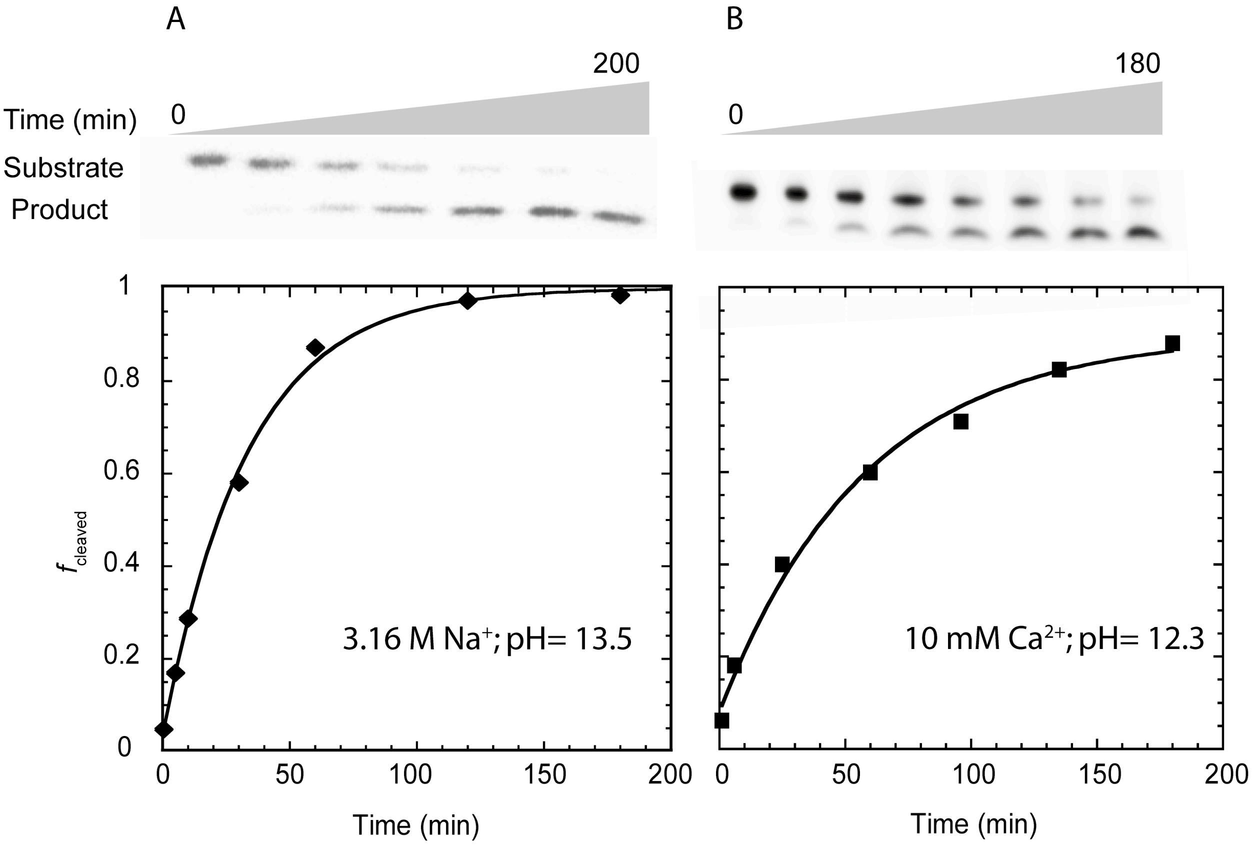

**Figure F.8.** Autoradiograms of the chimeric substrate (upper row) and plots of fraction cleaved versus time (lower row) in the presence of A) 3.16 M $Na^+$ at pH 13.5, and B) 10 mM $Ca^{2+}$ at pH 12.3. The chimeric substrate undergoes complete cleavage yielding a single product under both of these conditions.

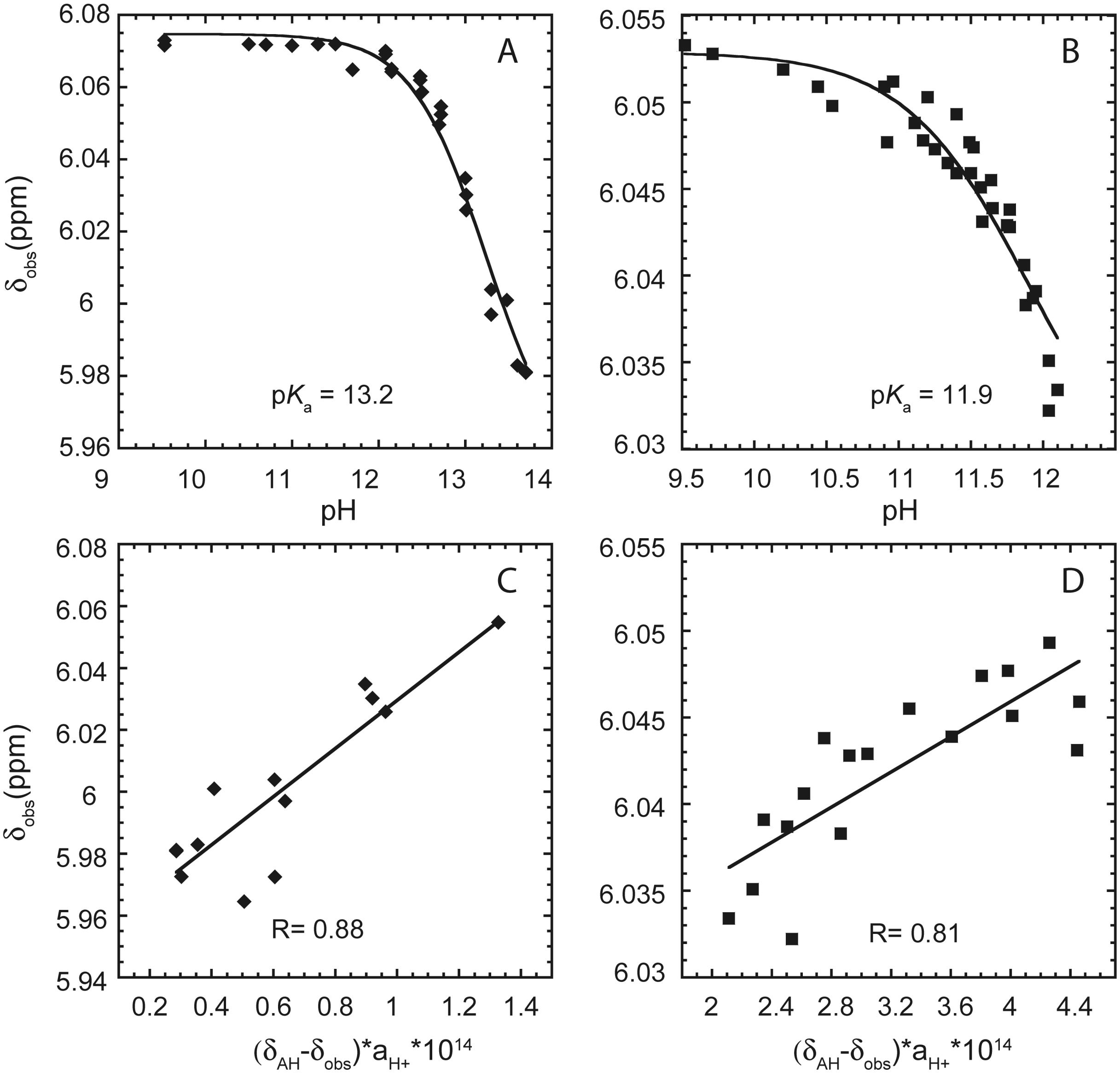

**Figure F.9:** Determination of the p$K_a$ of the 2′OH of 3′AMP by $^1$H NMR. A) B) NMR titration monitoring H1′ chemical shift as a function of pH in the presence of A) 0.5 M $Na^+$ and B) 10 mM $Ca^{2+}$. The data were fit to Eq. 4. C) D) The value of $d_A$ used in Eq 4 was first determined by linearizing the sloping portion of the panel A and B data using Eq. 5 as shown in C) 0.5 M $Na^+$ and D) 10 mM $Ca^{2+}$. The p$K_a$ values were determined from fitting to panels A and B and found to be 13.2 and 11.9 in the presence of 0.5 M $Na^+$ and 10 mM $Ca^{2+}$, respectively.

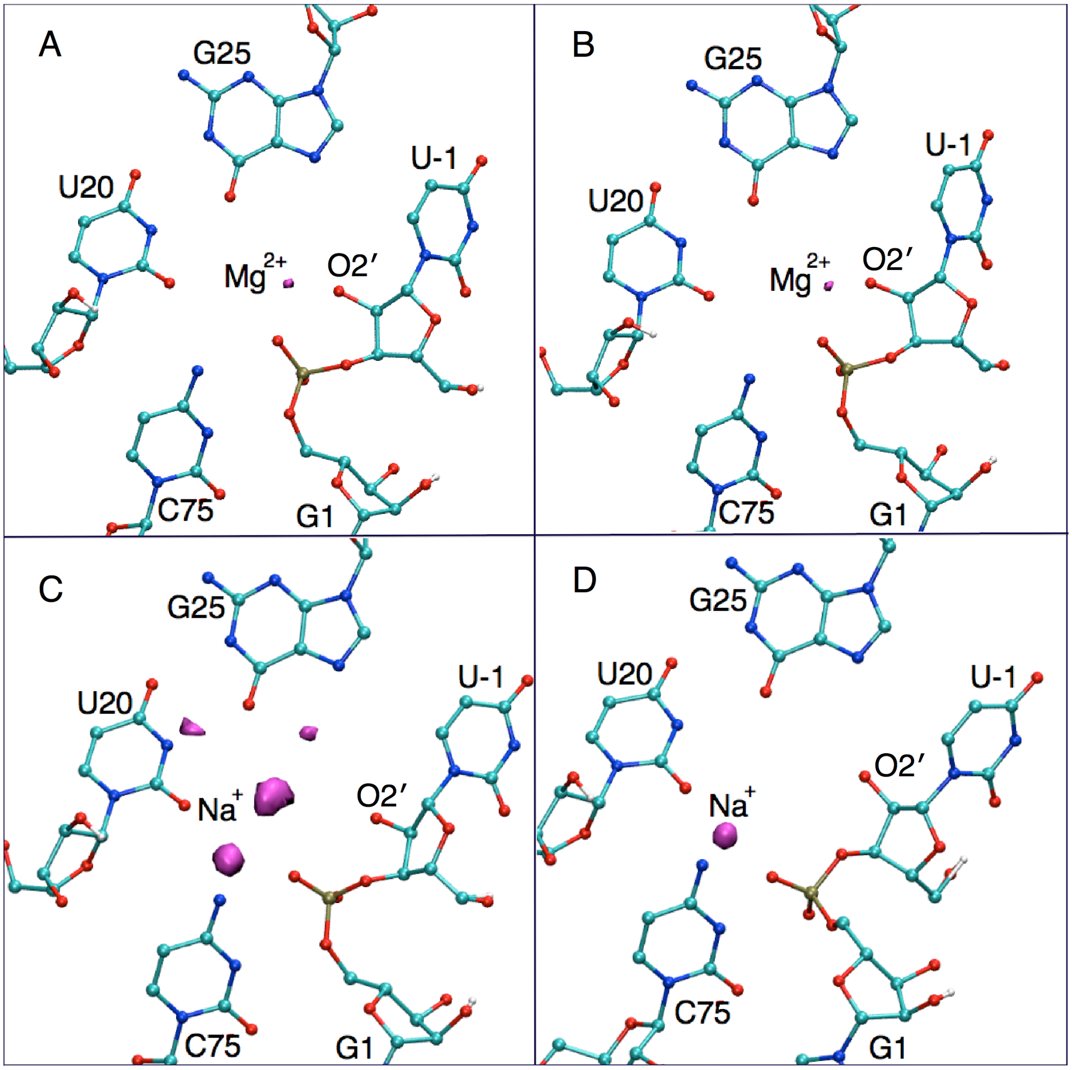

**Figure F.10.** Charge isodensity plots illustrating positive charge in the active site region of the HDV ribozyme calculated from a series of 25 nanosecond classical MD trajectories. Panels A and B correspond to two independent trajectories of the WT ribozyme with a $Mg^{2+}$ ion at the catlytic site. Panels C and D correspond to two independent trajectories of the ribozyme in which the catalytic $Mg^{2+}$ ion was replaced by two $Na^+$ ions in the bulk, and $Na^+$ moves into the region of the catalytic site. The details pertaining to the MD simulations are provided in refs. 1 and 2. In each of the panels, the purple surface represents the greatest 30% of the positive charge density. In panels A and B, the positive charge is due to the $Mg^{2+}$ ion, while in panels C and D, the positive charge is due to a $Na^+$ ion. Panels A and B indicate that the $Mg^{2+}$ ion is highly localized and is positioned at a site where it interacts strongly with the O2′ of U-1. Panel C suggests a more delocalized nature of the $Na^+$ ion at the active site with multiple binding sites, only one of which allows the ion to interact with the O2′. Panel D suggests that the favored binding site of the $Na^+$ ion does not allow the ion to interact with O2′. All of these simulations were performed with the O2′ protonated, but the hydrogen atoms are not shown for clarity. The differences between Panels C and D arise from differences in conformational sampling for two independent MD trajectories starting at different initial configurations.

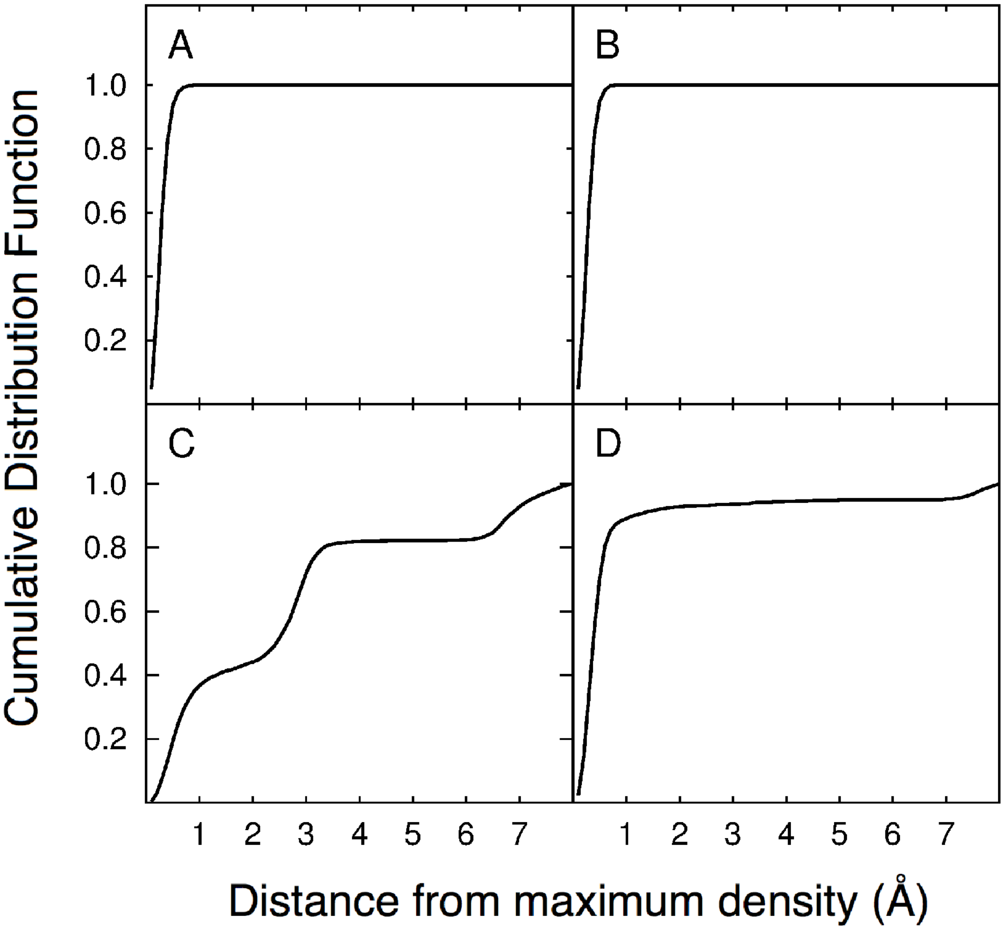

**Figure F.11.** Cumulative distribution functions (CDFs) for metal ion occupancy at the active site region of the HDV ribozyme calculated from MD simulations. The CDF for a metal ion at a given distance *r* is the probability of finding that ion within a sphere of radius *r*. Panels A and B correspond to two independent trajectories of the WT ribozyme with a $Mg^{2+}$ ion at the catlytic site. Panels C and D correspond to two independent trajectories of the ribozyme in which the catalytic $Mg^{2+}$ ion was replaced by two $Na^+$ ions in the bulk, and $Na^+$ moves into the region of the catalytic site. These panels are associated with the respective panels depicted in Figure F.10. In each case, the site of maximum positive charge density from the isodensity surfaces is chosen as the origin of the CDF, and only ions lying within a distance of 8 Å in all directions from the origin are considered. Thus, the CDFs measure the occupancy of the metal ions at the various binding sites that are revealed in Figure F.10. In panels A and B, the CDF reaches unity within 1 Å, confirming that the binding sites of $Mg^{2+}$ revealed in the charge isodensity plots depicted in Figures F.10A and F.10B are highly localized and have 100% occupancy. Panel C indicates that the two prominent binding sites of $Na^+$ observed in Figure F.9C each have ~40% occupancy. Panel D indicates that the occupancy of the binding site observed in Figure F.9D has ~90% occupancy. The differences between Panels C and D arise from differences in conformational sampling for two independent MD trajectories starting at different initial configurations.

## Supporting references

(1) Veeraraghavan, N.; Ganguly, A.; Chen, J.-H.; Bevilacqua, P. C.; Hammes-Schiffer, S.; Golden, B. L. *Biochemistry* **2011**, *50*, 2672-82.

(2) Veeraraghavan, N.; Ganguly, A.; Golden, B. L.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Phys. Chem. B* **2011**, *115*, 8346-57.