# COMPUTATIONAL RNA ENZYMOLOGY

by

**COLIN S. GAINES**

**A dissertation submitted to the**

**School of Graduate Studies**

**Rutgers, The State University of New Jersey**

**In partial fulfillment of the requirements**

**For the degree of**

**Doctor of Philosophy**

**Graduate Program in Chemistry and Chemical Biology &**

**Computational Biology and Molecular Biophysics**

**Written under the direction of**

**Darrin M. York**

**and approved by**

_____

_____

_____

_____

**New Brunswick, New Jersey**

**October, 2020**

## ABSTRACT OF THE DISSERTATION

# Computational RNA Enzymology

**By Colin S. Gaines**

**Dissertation Director:**

**Darrin M. York**

With a recent surge in experimental work being conducted on catalytic RNA enzymes - or ribozymes - there was both a need and an opportunity to develop a comprehensive approach to the computational study of small nucleolytic ribozymes. In this dissertation, this approach to Computational RNA Enzymology is detailed (Chapter 2) and then applied to two recently discovered ribozymes. In Chapter 3, molecular dynamics simulations both in crystallo and in solution, along with molecular solvation theory, are leveraged to develop a model for the active state of the twister-sister (or TS) ribozyme in solution. An end-to-end study of the twister ribozyme, utilizing the full range of techniques involved in the Computational RNA Enzymology approach, is then presented in Chapters 4 and 5. Finally, Chapter 6 discusses how the insights gained from atomically detailed models of individual ribozyme systems resulted in the discovery of a catalytic core common to a majority of naturally occurring small-nucleolytic ribozymes as well as an artificial DNAzyme. The structure-function relationships gleaned from the computational work conducted on these ribozymes not only allows for the interpretation of experimental data, but also provides for experimentally testable predictions to guide future studies. Continuing to look forward, the characterization of a generalizable structural scaffold supporting catalysis may open the door to rational design of novel therapeutics built using catalytic RNA.

# Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Darrin M. York, for his wisdom, patience, and support throughout this journey. His lab is one where I have been encouraged to dive deep and work to understand every detail of my 'pet' ribozymes - Twister and TS. But, it has also been a place where I had the freedom to explore uncharted waters and discover new insights through close collaboration not just within the lab but with the experimental community working on these catalytic RNAs. Dr. York's joy in each accomplishment and the support and patience when doubt or anxiety got the better of me made it a true pleasure to work with him.

I would also like to thank, my wonderful collegues. Drs. Tim Giese and Tai-Sung Lee for lending their incomparable expertise. Ken Kostenbader, Emily Atieh, Drs. Şölen Ekesan and Abir Ganguly, I so greatly appreciate your friendship outside of the lab in addition to all of the great scientific discussions we had. All of the other members that have been part of the York lab during my time there, George Giambaşu, Haoyuan Chen, Ming Huang, Junjie Ouyang, Maria Panteva, Thakshila Dissanayake, Erich Kuechler, Pawel Janowski, Gerhard König, Brian Radak, Yujun Tao and Hsu-Chun Tsai, thank you all for the positive impact you have made.

Thank you to my committee, Drs. David A. Case, Wilma K. Olson, and Arek Kulczyk. And also, thank you to the experimental collaborators I have been privledged to work with - Drs. David Lilley, Tim Wilson, Joe Piccirilli, and Phil Bevilacqua.

# Dedication

*This dissertation is dedicated to my family, who have supported me throughout my long and winding path to get here. It is also dedicated to my incredible wife Rachel. I never could have done this without your love and sacrifice - together we really are the best satsuma.*

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1
# Introduction

The goal of this dissertation is to develop a comprehensive approach for applying molecular modeling methodologies to the study of small nucleolytic RNA enzymes, or ribozymes. The structure and catalytic mechanism of these ribozymes has been a topic of intense interest since the Nobel prize winning discovery of catalytic RNA over three decades ago. Since that initial discovery, a handful of unique classes of ribozymes have been identified and characterized - both experimentally and computationally. However, during the course of this research, there has been a flurry of new research both doubling the number of known small self-cleaving ribozyme classes and generating a wealth of experimental data. This boom in experimental work necessitated a corresponding effort in the development both of theoretical models through which the biochemical data could be interpreted and predictions for which future experiments could be designed to test.

In the following chapters, this work will describe efforts to construct models for two of the recently discovered ribozymes - twister and twister-sister (a.k.a TSrz), which then culminates in the proposal of the L-platform/L-anchor motif as a blueprint for the design of new classes of RNA cleaving enzymes. Chapter 2 will discuss background on RNA catalysis and an overview of how the molecular modeling methodologies were employed within this comprehensive approach to computational RNA enzymology. Additionally, an ontology for the mechanistic strategies leveraged by these ribozymes to catalyze 2′-O-transphosphorylation will be presented in Chapter 2 and will be used in describing both the experimental designs and interpretation of results throughout this work. Chapter 3 discusses the application of the computational RNA enzymology approach to the development of structural models for the active state in solution for the TS ribozyme. In Chapters 4 and 5, an end-to-end computational study of the twister ribozyme will be presented. Finally, Chapter 6 details the culmination of this work with the identification of a common catalytic core present in a majority of small nucleolytic ribozymes and a proposal of the L-platform/L-anchor framework as a blueprint for design.

# Chapter 2
# Background



Starting from their initial Nobel Prize winning discovery in the 1980's[21], the study of RNA enzymes - or ribozymes - has had significant impact across the fields of chemistry, biology, and medicine[22–27]. First, from the biological perspective, ribozymes are widespread throughout all domains of life, found in everything from bacterial to human genomes[28–31]. Within a cell, these ribozymes likely play complex roles in RNA processing, viral replication, and regulation of gene expression[22,27,32]. Notably, the catalytic core of both the ribosome[33–37] and spliceosome[38–41] were found to be comprised primarily of RNA - making them, at a fundamental level, RNA enzymes. The discovery of catalytic RNAs have had a transformative impact, spurring the idea of the RNA World hypothesis[42–46] and opening up the field of ribozyme engineering[47–55]. While the biological functions for many of the small nucleolytic ribozymes remains a mystery, the study of these RNAs has been invaluable from a biochemical perspective.

All of the currently known naturally occurring ribozymes are involved in the catalysis of the phosphoryl transfer reaction.[25,26,56,57] The small nucleolytic ribozymes that are the focus of this dissertation, more specifically, catalyze site specific cleavage via 2′-O-transphosphorylation. A deep understanding of how RNA, with its limited repertoire of chemical groups relative to protein enzymes, efficiently catalyzes this reaction can provide important chemical insights. These insights can then be leveraged for biotechnology and medical applications. In fact, progress has been made through utilizing the naturally occurring ribozymes, or even synthetic ones created via *in vitro* evolution, as a platform for bioswitches[50,58–62] and sensors[63–67]. However, continued work on atomically detailed models for known ribozymes and the patterns in active site architecture that can be gleaned from these models promises not only to advance the biological

and chemical understanding of RNA catalysis but also to unlock novel therapeutics.

## 2.1 Catalytic Strategies for 2′-O-transphosphorylation

In order to facilitate discussion of RNA strand cleavage by 2′-O-transphosphorylation, the reaction catalyzed by the small self-cleaving ribozymes studied in this work, an ontology for the catalytic strategies employed by these RNA enzymes will be presented. First and foremost, 2′-O-transphosphorylation involves an activated/deprotonated O2′ moiety making an in-line nucleophilic attack on the scissile phosphate to form a pentavalent dianionic transition state, followed by the departure of the O5′ leaving group (Figure 2.1). The framework for the strategies employed to catalyze this reaction, first presented by Breaker and Emilsson[1], will serve as the starting point for the following ontology.

The framework starts by describing the basic geometric constraints associated with 2′-O-transphosphorylation. The idealized transition state in this reaction is a dianionic pentavalent phosphorane in a trigonal bipyramidal geometry. The 2′ and 5′ oxygens (the nucleophile and leaving group, respectively in the forward reaction) are positioned at the apical positions, while the non-bridging phosphoryl oxygens (NPOs) and the 3′ oxygen are in the equatorial positions. This alignment of the nucleophile, phosphorus, and leaving group (O2′-P-O5′) at an angle of $\sim$180° at the transition state provides the basis for the first catalytic strategy - referred to as $\alpha$. If the structure of the ribozyme active site encouraged an in-line alignment of these atoms, that would lead to a enhancement of the observed rate of reaction. Next we consider the formal -2 charge delocalized among the O2′ nucleophile, O5′ leaving group, and the two NPOs at the transition state. The impact of the local chemical environment on the stabilization of the negative charge accumulating at the transition state is termed $\beta$. The final two catalytic strategies involve activation of the nucleophile, typically via deprotonation, and stabilization of the leaving group, typically through protonation. These two strategies are referred to as $\gamma$ and $\delta$, respectively. While this framework is quite useful in describing the general strategies employed by ribozymes to catalyze RNA strand cleavage, it needed expansion in order to characterize the reaction mechanisms with fine atomic detail. Towards this end, the ontology expands the framework of catalytic strategies by stratifying the $\beta$, $\gamma$, and $\delta$ strategies, with $\alpha$ being sufficient as previously defined[68].

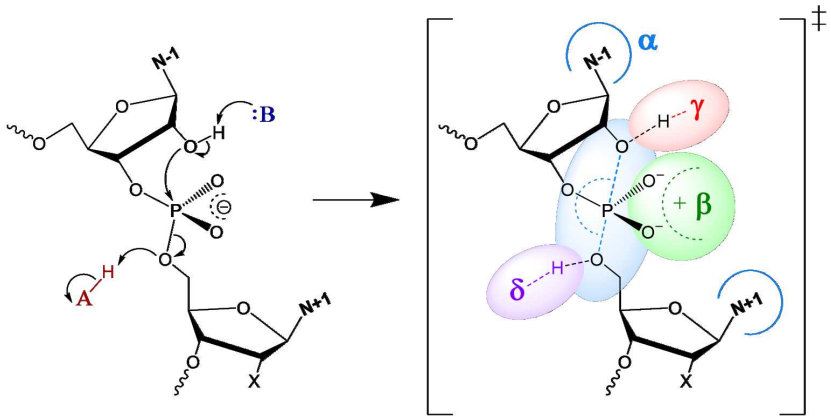



Figure 2.1: 2′-O-Transphosphorylation leading to cleavage of the RNA backbone (left of central arrow) and idealized transition state highlighting the general catalytic strategies[1] (right of central arrow). $\alpha$, Arrangement of the O2′ nucleophile, P (of scissile phosphodiester bond), and O5′ leaving group in an in-line attack geometry (facilitated by contacts, indicated by blue arcs, that splay the N1 and N+1 nucleobases). $\beta$, Stabilization (neutralization/protonation) of the negative charge accumulation on the nonbridging phosphoryl oxygens (NPOs). $\gamma$, Activation (deprotonation) of the O2′ nucleophile. $\delta$, Stabilization (neutralization/protonation) of the accumulating negative charge on the O5′ leaving group. Although this schematic uses a transition state model to illustrate the fundamental catalytic strategies, these strategies can impact any state along the reaction coordinate. Colored ovals highlight each strategy and encompass the primary atomic positions associated with the chemical space of bonds for each strategy.

For each catalytic strategy, we first identify the chemical space of bonds that are either broken or formed along the reaction coordinate and the *primary atomic positions* associated with these bonds. Due to the strong influence of divalent metal ions in RNA cleavage,[69,70] we include ionic bonding (direct inner-sphere coordination) as part of this chemical bonding space. Positions not associated with these bonds are *non-primary atomic positions.* The designations of the following primary and non-primary atomic positions are used to facilitate definitions of the primary, secondary, and tertiary catalytic effects.

The primary atomic positions are defined for each catalytic strategy as follows:

- **Primary $\beta$ atomic positions:** the NPO positions themselves, any atoms directly involved in protonation of the NPOs (*e.g.,* the proton itself and the heteroatom of the acid from which the proton was transferred), and any metal ion directly coordinated to the NPOs (atoms under the green oval in Figure 2.1); it does not include atoms on the acid

not directly involved in a bond with the proton, nor does it include atoms that hydrogen bond to the NPOs.

- **Primary $\gamma$ atomic positions:** the O2′ position itself, any atoms directly involved in nucleophile activation (*e.g.,* the O2′ proton itself and the heteroatom of the base to which the proton is transferred), and any metal ion directly coordinated to the O2′ position (atoms under the red oval in Figure 2.1); it does not include atoms on the base not directly involved in a bond with the proton, nor does it include nearby metal ions (not directly coordinated to the O2′) that electrostatically influence the p$K_a$ of the base.

- **Primary $\delta$ atomic positions:** the O5′ position itself, any atoms directly involved in leaving group stabilization (*e.g.,* the acid proton itself and the heteroatom of the acid from which the proton is transferred), and any metal ion directly coordinated to the O5′ position (atoms under the purple oval in Figure 2.1). It does not include atoms on the acid not directly involved in a bond with the proton, nor does it include nearby metal ions (not directly coordinated to the O5′) that electrostatically influence the p$K_a$ of the acid.

With these definitions, we now adopt conventions used in the discussion of isotope effects to categorize catalytic effects as *primary* or *secondary*[71] and concepts from structural biology to introduce *tertiary* catalytic effects.

### 2.1.1 Primary 1° Catalytic Effects.

A *primary catalytic effect* is one that results from changes in the identity of the primary atomic positions as defined above. Illustrative examples include (1) disruption of the direct coordination of a catalytic divalent metal ion at the NPO position (*e.g.,* thio substitution disrupting $Mg^{2+}$ coordination) would give rise to a primary $\beta$ effect, (2) a chemical modification that removes the general base heteroatom involved in activation of the nucleophile (*e.g.,* guanine general base N1C knockout) would give rise to a primary $\gamma$ effect, and (3) a chemical modification that removes the general acid heteroatom that donates a proton to the leaving group (*e.g.,* adenine (N1) general acid N1C knockout) would give rise to a primary $\delta$ effect.

### 2.1.2 Secondary 2° Catalytic Effects.

A *secondary catalytic effect*, on the other hand, is the change in the electronic environment of the primary atom resulting from changes in the identity of nonprimary atomic positions (and thus

is exclusively different from a primary catalytic effect). Modifications leading to a secondary catalytic effect have an indirect influence on the bonding environment of the primary atoms without involving any change to their identity. This can occur through electrostatic, inductive, or stereoelectronic effects perturbing the underlying electronic structure of the bond between primary atoms (*e.g.,* through either remote chemical modification or short-ranged nonbonded interactions).

Illustrative examples include the following: (1) Elimination of a stabilizing hydrogen bond to the NPO (while otherwise not changing the structure of the active site), such as deletion of a nucleobase exocyclic amine that hydrogen bonds to the NPO,[72] would give rise to a secondary $\beta$ effect. (2) Chemical modification at nonprimary atomic positions of the general base that changes the p$K_a$ of the primary position (*e.g.,* guanine general base N7C modification upshifting the p$K_a$ at the N1 position) would give rise to a secondary $\gamma$ effect. (3) Replacement of a divalent metal ion acting as a general acid through a coordinated water molecule with a different metal ion that has a shifted p$K_a$ would give rise to a secondary $\delta$ effect.

### 2.1.3 Tertiary 3° Catalytic Effects.

A *tertiary catalytic effect* reflects alteration of the position of the primary atoms resulting from modification of the *structural scaffold* or *hydrogen bond network* that organizes the enzyme active site. This alteration can lead to changes in (1) positions of key residues, functional groups, or protons in the active site, (2) interactions that support active conformations of the substrate itself, (3) binding modes or occupations of metal ions or other small molecules required for activity, or (4) orientation of solvent molecules that form hydrogen bond networks important for catalysis.

The end goal of experimental and computational work on a catalytic RNA system is to gain insight from the development of an atomically detailed model of mechanism that enables prediction. Thus, it is essential to have an ontology, such as this, that enables interpretation of data, including that from precision chemical modifications and molecular simulations, in terms of specific interactions within an active site and their contributions to catalysis.

## 2.2 Comprehensive Approach to Computational RNA Enzymology

In keeping with the goal of providing atomically detailed models of mechanism for catalytic RNAs, we have developed a comprehensive approach to the computational study of small nucleolytic ribozymes. This approach enables the end-to-end study of a ribozyme, integrating

the full body of available experimental data (structural and biochemical) into computational models that provide predictive insight into the active state, chemical mechanism, and the effects of modifications on catalysis.

The four broad categories of work involved in this approach are as follows:

- Modeling in crystallo
- Modeling of the active ribozyme in solution
- Exploration of reaction mechanism
- Generation of experimentally testable predictions

### 2.2.1 Modeling in crystallo

The computational RNA enzymology approach begins with modeling in crystallo as a means to establish a foundation from which a model of the active ribozyme in solution can be built. Structures from x-ray crystallography are used as a starting point, however, this structural data can present interesting challenges. While structures representing the cleaved, product state of a ribozyme can provide some valuable insights, typically these experiments seek to provide a structure that is representative of the pre-cleavage, reactant state. Since small nucleolytic ribozymes are self-cleaving enzymes, this necessitates some deactivating mutation - most commonly the substitution of the $2'$-hydroxyl nucleophile with a hydrogen or O-methyl group. Furthermore, crystal packing allows for interactions between adjacent monomers in the crystallographic unit cell as well as the stabilization of divalent metal binding sites. Both of which, in addition to the effect of a deactivating mutation, would need to be careful considered when constructing a model of the catalytically active state in solution. However, a comparative analysis of multiple crystal structures or the integration of biochemical or bioinformatics data can help to guide the interpretation of the crystallographic data. Finally, molecular dynamics simulation of the crystallographic unit cell allows for the construction of an atomically detailed dynamical model of the ribozyme in the crystal environment. This simulation not only serves to validate the computational methods when reproducing the experimentally observed structural averages and fluctuations but also allows for a foundational model to be constructed upon which a model in solution can be built.

### 2.2.2 Modeling of the active ribozyme in solution

In moving towards a model of the active state of a ribozyme in solution, a few initial assumptions must be made. The first is that the global fold of the ribozyme in crystallo is reasonable - meaning only local conformational rearrangements would be necessary for the ribozyme to adopt a catalytically active state. Second, it is assumed that the ribozyme employs all four catalytic strategies (detailed in Section 2.1) to varying degrees. The final assumption is that if a model of the active ribozyme in solution structurally provides for each of the catalytic strategies, it is likely to carry out the chemical reaction efficiently, but may represent an improbable state. These assumptions provide the framework for approaching the study of RNA catalysis using computational methods, that will integrate, interpret, and then guide experimental work.

A useful first step in transitioning from a model in crystallo to a model in solution, is to conduct long time-scale molecular dynamics simulation of the ribozyme in explicit solvent at various points along the reaction pathway. As an example, for a ribozyme whereby nucleobases (rather than metal ions) are utilized as the general acid and base, these simulations would include:

1. The "pre-reactive" state, where the presumptive general acid and base are in their standard neutral protonation states.

2. The "reactant" state, where the presumptive general acid and base are in their active, reverse protonation states.

3. The "activated precursor" state, where the nucleophile has been activated via proton transfer to the general base and the acid is still in its active, protonated state.

4. The "transition state mimic", where an artificial construct representing the pentavalent phosphorane intermediate is utilized. It may be valuable to simulate both the standard and reverse protonation states of the catalytic residues with this construct.

5. The "product" state, where the cleavage reaction has been completed yielding a $2'$,$3'$-cyclic phosphate and the $5'$ leaving group has accepted a proton from the general acid.

For each of these simulations the goal is to identify plausible active states in solution and the structural features (i.e. catalytic residues, metal ion binding modes, hydrogen bonding networks, etc.) that support each of the catalytic strategies. This atomically detailed, dynamical model would then be validated against - as well as utilized to interpret - any and all available experimental data (e.g. bioinformatics/sequence conservation data, mutational data, activity-pH profiles, metal ion titration data, etc.). Additionally, an estimate of the probability for the

catalytically active state is valuable in order to connect to experimentally observed rate data. This estimate for the probability of the active state may be derived from free energy simulations or theoretical calculations designed to probe the free energies associated with adopting an active conformation state, protonation state, or binding of catalytic metal ions, as well as explore the coupling between each of these degrees of freedom.

### 2.2.3 Determination of reaction pathways

By constructing a structural model of the catalytically active state under the assumption that its supports each of the catalytic strategies, mechanistic hypotheses naturally follow. These mechanistic hypotheses would then be explored by building up free energy surfaces for the reaction mechanism using QM/MM. Again using a ribozyme whereby nucleobases (rather than metal ions) are utilized as the general acid and base as an example, these free energy surfaces would explore the following chemical steps (many of which may be concerted):

1. Activation of the nucleophile via proton transfer to the general base
2. Nucleophilic attack of the phosphate
3. Formation of the pentavalent phosphorane transition states
4. Leaving group departure resulting in the 2′,3′-cyclic phosphate cleavage product
5. Stabilization of the leaving group via proton transfer from the general acid

In conjunction with the estimated probability of the active state, the free energy barrier for the intrinsic rate of reaction calculated here would provide a prediction for the overall kinetics that would be comparable to experimental rate data. This comprehensive model then provides insight into how a ribozyme balances a potentially improbable active state with efficient catalysis.

### 2.2.4 Providing experimentally testable predictions

The value of a comprehensive computational model for RNA catalysis comes first from its ability to aid in the interpretation of experimental data and then from its predictive capability, such that it can guide future experiments. As will be discussed in the following sections, the models built using this Computational RNA Enzymology approach seek to provide insight into a wide variety of experimental observables. First and foremost, the models explore the kinetics for both the wild-type ribozyme and sequences with a range of mutations. By providing a model

that accounts for both the probability of the active state and the free energy associated with the chemical steps, the effects of mutations (both base substitutions and chemical modifications of specific functional groups) can be analyzed in detail. Complex pH-dependent behavior or dynamic metal-ion binding can also be accounted for. Additionally, details about the nature of the chemical steps and specifically the rate limiting transition state can also be examined, for example by prediction of kinetic isotope effects. Looking beyond development of atomically detailed models for a specific ribozyme class, construction of these models in accordance with this Computational RNA Enzymology approach is an important step towards understanding patterns and trends across the various ribozyme classes.

# Chapter 3

# Model for the Functional Active State of the TS Ribozyme from Molecular Simulation



Recently, a crystal structure has been reported of a new catalytic RNA, the TS ribozyme, that has been identified through comparative genomics and is believed to be a metalloribozyme having novel mechanistic features. Although this data provides invaluable structural information, analysis suggests a conformational change is required to arrive at a catalytically relevant state. We report results of molecular simulations that predict a spontaneous local rearrangement of the active site, leading to solution structures consistent with available functional data and providing competing mechanistic hypotheses that can be experimentally tested. The two competing hypotheses differ in the proposed identity of the catalytic general acid: either a water molecule coordinating a $Mg^{2+}$ ion bound at the Watson-Crick edge of C7, or the N3 position of residue C7 itself.

## 3.1 Introduction

Small nucleolytic ribozymes are important model systems in the study of RNA catalysis[26]. Recently, there has been a surge of progress in the identification of new classes of self-cleaving nucleolytic ribozymes that have been revealed by comparative genomics analysis.[30,31] Following in the wake of this discovery have been intense efforts to determine crystal structures of these new ribozymes and gain insight into the origin of their function.[6,7,9,10,73–75] However, crystal structures often are not representative of the active states in solution. In the case of the recently discovered TS ribozyme, the crystal structure[7] does not provide a mechanistic explanation of available functional data, leading to the assertion that a local rearrangement of the active site must occur to form a catalytically active state.

Herein we use molecular dynamics (MD) simulations and molecular solvation theory to

develop two distinct models for the active state of the TSrz in solution that are consistent with the currently available body functional data[7,31]. The "Mg Acid" model (Appendix A.1), positions a divalent metal ion to play the role of the general acid, similar to what has been suggested as a plausible mechanism in the hammerhead ribozyme[76]. The "C7 Acid" model (Appendix A.2), points to a highly conserved cytosine residue that may play the role of the general acid, similar to that proposed for the HDV ribozyme[77–79]. Both models require local rearrangements of the active site from the crystal and, unlike the crystal structure, are able to both provide a plausible rationale for the current body of experimental functional data[7,31] and to suggest experimentally testable predictions that would distinguish pathways.

The TSrz crystal structure revealed several divalent metal ion binding sites (labeled M1-M7 in the original work[7]) that presumably stabilize tertiary contacts in the crystal. Initial simulations departing from the crystallographic structure that preserved these divalent metal ion binding modes did not lead to any significant rearrangement in solution (Figure 3.4). We then hypothesized that the crystal packing environment may have created divalent metal ion binding sites that differ from those present in solution. To explore this possibility, we used a 3D reference site interaction model (3D-RISM[80]) to predict cation occupation sites in TSrz both in a crystalline environment and in solution. Crystal packing models with 3D-RISM indicated cation binding sites consistent with the crystallographic data. However, calculations of the solution TSrz model revealed a shift in the predicted cation binding (Figure 3.1) from direct coordination of G5 *pro*R and C54 *pro*R (M4 site) in the crystal, to a distinct position at the nearby Watson-Crick edge of C7, referred to as the M4′ site.

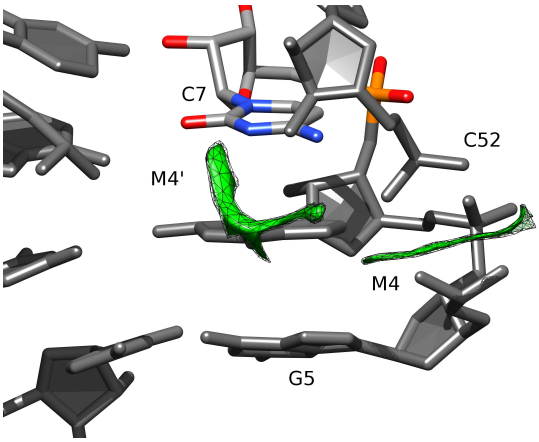



Figure 3.1: Difference map of cation number density from 3D-RISM calculations of the solution structure and crystallographic coordinates. The green mesh surfaces indicate that there is a predicted increase in cation density at both the crystallographic M4 site (where a $Mg^{2+}$ ion is observed to directly coordinate G5 *pro*R and C52 *pro*R), as well as the Watson-Crick edge of C7 (shown in a colored stick representation) labeled as the M4′ site.

## 3.2 Mg Acid Model

We therefore placed a $Mg^{2+}$ ion at this alternative position, and re-ran three independent simulations of a transition state mimic in solution. Within 2 ns, these simulations all produced a spontaneous rearrangement of the active site residues (C54, A55, U56, and U57) leaving the global fold, inferred from the original crystallographic work[7], largely unaffected (Figure 3.5 and Figure 3.6). In this model (Figure 3.2), a $Mg^{2+}$ ion at C7 is poised to act as a general acid, another interacting with A9 supports general base catalysis, and the rearrangement of A55 promotes in-line fitness and provides electrostatic stabilization of the transition state.

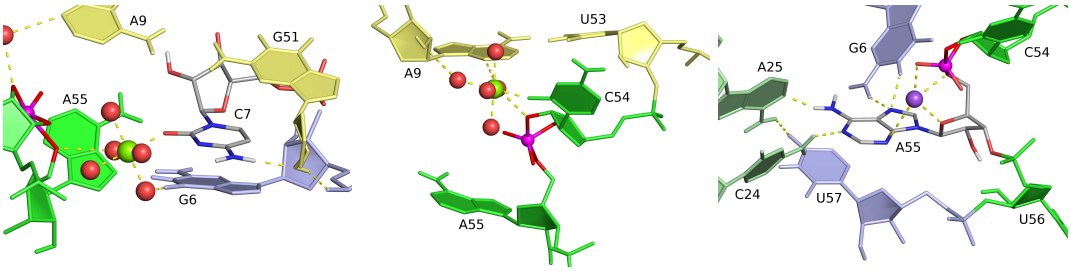



Figure 3.2: MD average structure for the "Mg Acid" model. Key nucleotides are shown in stick form, with hydrogen bonds indicated by broken lines. (a) Local environment of C7, highlighting hydrogen bonding between C7:N4 and G51:O2′ as well as the positioning of the $Mg^{2+}$ ion at the M4′ site, directly coordinating C7:O2. (b) View of the M1 site $Mg^{2+}$, outer sphere coordinating the sugar edge of A9 as well as the O2′ nucleophile. (c) Close up view of the hydrogen bonding scaffold that is proposed to anchor A55 in a splayed-out conformation relative to C54, in order to promote in-line fitness.

Table 3.1 summarizes the current body of available mutagenesis data and compares structural and functional interpretations to the experimental crystal structure and to the computational solution simulation models reported. The "Mg Acid" model predicts that the four most severe mutations involve disruptions to interactions that position C7 within the active site and to $Mg^{2+}$ binding at the O2 position. It should be noted that this $Mg^{2+}$ ion remained stably bound at C7:O2 throughout all MD simulations and 3D-RISM calculations of the average structure for the "Mg Acid" model predict a higher fractional occupancy at M4′ than at the M4 site (Figure 3.7). It follows that these mutations would be the most detrimental as they would directly impact the general acid step whereby a water molecule inner-sphere coordinated to the $Mg^{2+}$ bound at C7:O2 is poised to act as a proton donor to stabilize the negatively charged O5′ leaving group. Furthermore, this model suggests that the C7 5F mutation would be the least impactful of this group as it only weakens H-bonding and $Mg^{2+}$ binding rather than eliminating the key H-bonding interaction between C7:N4 and G51:O2′.

| Mutation | Fold Decr. | Crystal Structure (PDBID: 5T5A) | | Solution Simulation (Mg Acid) | | Solution Simulation (C7 Acid) | |
|---|---|---|---|---|---|---|---|
| | | Structural | Functional | Structural | Functional | Structural | Functional |
| C7U | N/A | C7:N4...G51:O2' | Folding | C7:N4...G51:O2' | General Acid $Mg^{2+}$ Binding at C7:O2 | C7:N4...A55:O5' | General Acid C7:N3H$^+$ |
| C7Z | N/A | | | | | | |
| G51 O2'H | 1600 | | | | | G51:O2'...G6:OP2 | Folding |
| C7 5F | 940 | | | | | C7:N4...A55:O5' | General Acid C7:N3H$^+$ |
| A9 N3CH O2'H | 890 | $Mg^{2+}$ ∘ A9:N3 | General Base $Mg^{2+}$ Binding | $Mg^{2+}$ ∘ A9:N3 | General Base $Mg^{2+}$ Binding | $Mg^{2+}$ ∘ A9:N3 | General Base $Mg^{2+}$ Binding |
| A9 O2'H | 3.2 | X | X | $Mg^{2+}$ ∘ A9:O2' | | $Mg^{2+}$ ∘ A9:O2' | |
| A55 N3CH O2'H | 270 | X | X | A55:N3 • $Na^+$ • A55:OP1 A55:O2'...U56:O3' | Positioning of A55 → Inline Fitness | A55:O2'...U56:O5' | Positioning of A55 → Inline Fitness |
| A55 N7CH | 59[a] | X | X | G6:N1/N2...A55:N7 | | G6:N1/N2...A55:N7 | |
| A55 I | 30[a] | A55:N6...C7:O2 | X | A55:N6...A25:N1 C24:N4...A55:N1 | | A55:N6...A25:N1 C24:N4...A55:N1 | |
| A55 O2'H | 9.5 | X | X | A55:O2'...U56:O3' | | A55:O2'...U56:O5' | |
| A55 P | 6.3[a] | A55:N6...C7:O2 | X | A55:N6...A25:N1 | | A55:N6...A25:N1 | |
| A25 I | 3.5 | X | X | A55:N6...A25:N1 | | A55:N6...A25:N1 | |

Table 3.1: Comparison of structural and functional interpretations of the available mutational data. Decrease in rate is shown as a fold decrease due to the mutation relative to the wild type sequence; N/A indicates no observable activity. The structural interpretation listed is the apparent interaction(s) disrupted by the mutation in the crystal or simulation model. A...B represents the disruption of a hydrogen bond, while A ∘ B indicates outer-sphere coordination of a metal and A • B indicates direct coordination. All rate measurements from Liu et al. 2017 or [a]personal communication (TJ Wilson and DMJ Lilley 2017). The listed functional interpretation for each mutation corresponds to its impact on one of the four widely accepted catalytic strategies[1–3] for RNA transesterification.

The crystallographic structure and the computational model are in close agreement for both the structural and functional explanations of the mutations to A9, with the sugar edge providing hydrogen bond acceptors for outer-sphere coordination of the $Mg^{2+}$ implicated in the general base step. The structural rearrangement in the computational model, however, allows for additional hydrogen bonding to A9:O2′ for which there is a small observed mutational effect.

Finally, the mutational data suggests that A55 is involved in numerous H-bonding interactions not observed in the crystal structure. Anomalously high B-values of A55 in the crystal ($> 60$ Å$^2$) suggest this residue may have a low barrier to local rearrangement[7]. The spontaneous structural rearrangement observed in our simulations positions A55 to form a constellation of hydrogen bonds that promotes an active in-line conformation of the nucleophile. The hydrogen bond between A55 and G6 along with the interaction of the Watson-Crick edge of A55 and residues C24 and A25 in the L4 loop provides a scaffold for positioning A55 in a splayed out conformation relative to C54, a common structural feature in ribozymes that favors in-line

attack[81].

## 3.3 C7 Acid Model

However, an alternative explanation for the mutagenesis data is that C7 may itself act as the general acid, similar to what has been proposed in the HDV ribozyme[77–79]. Starting from the previously discussed structural model with local rearrangement of A55, we explored simulations with C7 protonated while donating a hydrogen bond to the leaving group, as would be required to function as a general acid, and the $Mg^{2+}$ ion previously at the M4′ site is returned to the M4 site. 3D-RISM calculations predict significant cation density centered around the $Mg^{2+}$ ion bound at the M4 site, consistent with the position of metal ions in the crystal structure (Figure 3.8). The system was equilibrated and relaxed in this state, and simulated without restraints for 50 ns. The unrestrained simulation showed that this new structure was stable on that timescale, maintaining all of the key hydrogen bonding interactions representative of the model.

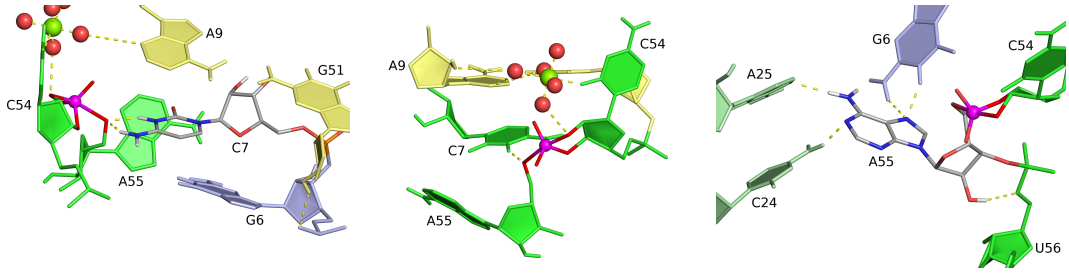



Figure 3.3: MD average structure for "C7 Acid" model. Key nucleotides are shown in stick form, with hydrogen bonds indicated by broken lines. (a) Close up view of C7:N3H$^+$ donating hydrogen bonds to the O5′ leaving group and thus poised to act as the general acid. (b) Interactions between the first solvation shell of the $Mg^{2+}$ ion at the M1 site and both the sugar edge of A9 and the O2′ nucleophile. (c) Hydrogen bonding scaffold supporting the positioning of A55 as splayed-out from C54, proposed to support in-line fitness.

From these simulations we present an alternative "C7 Acid" model that can be used as an alternative to the "Mg Acid" model to interpret the experimental functional data in Table 3.1. The only notable differences between the two computational models involve C7 (Figure 3.3). In the "C7 Acid" model, C7:N4 is hydrogen bonding to the leaving group instead of G51:O2′ in the crystal and the "Mg Acid" model. The interpretation of the complete loss of activity with C7U and C7Z mutants is then both the shifting of the p$K_a$ of C7:N3 away from neutrality and the loss of the H-bond from C7:N4. The C7 5F mutation is now explained by the shift

in the C7:N3 p$K_a$ down 2 units (cytosine: p$K_a$ = 4.2, 5-fluorocytidine: p$K_a$ = 2.3), lowering the probability of C7 being protonated and the ribozyme being in an active protonation state at neutral pH, in addition to weakening of the hydrogen bonding between C7:N4 and the O5′ leaving group. A shift in p$K_a$ down by 2 units would presumably disfavor the protonated state by 2.6 kcal/mol corresponding to an approximately 80-fold decrease in the observed rate. This is an order of magnitude less than the observed decrease (940 fold) in rate, and it could be the weakening of the hydrogen bonding between C7:N4 and the leaving group, due to the 5F substitution, that partially accounts for the remainder of the rate decrease. This supposition of the catalytic importance of hydrogen bonding between C7 and the leaving group is supported by the elimination of activity observed with the C7Z mutation, since the p$K_a$ of zebularine and 5-fluorocytidine are approximately the same, but zebularine lacks the exocyclic amine entirely. Finally, the proposed "C7 Acid" model presents a hypothesis that the G51:O2′H mutation is detrimental due to the importance of G51 in folding, rather than a direct interaction with C7.

## 3.4 Conclusion

To test these competing mechanistic hypotheses, the activity of a TSrz ribozyme mutant with C7 replaced by N3-methylcytosine or N3-deazacytosine (both eliminate the ability of C7 to donate a proton from the N3 position) could be measured experimentally. If these mutants had only a moderate effect on catalysis, the most straightforward interpretation would be that C7:N3 most likely does not act as the general acid. However, if this mutation eliminated catalytic activity, the result would be consistent with the role of C7 as a general acid, and could be further validated by testing whether introduction of an enhanced leaving group, such as a 5′ thio substitution, would have a rescue effect. It is clear that while a crystal structure is an essential starting point for the study of the TS ribozyme, computational modeling can provide important insights that not only aid in the interpretation of the available experimental data, but also drive future studies. Additional computational work must be done to explore the relevant chemical steps along the reaction pathway. Likewise, additional biochemical experiments are needed to refine the working hypothesis for the mechanism of this catalytic RNA.

## 3.5 Supporting Information for: Model for the Functional Active State of the TS Ribozyme from Molecular Simulation

### 3.5.1 Supporting Methods

The initial coordinates used in these computational studies were taken from the 2.0Å crystal structure of the TS ribozyme[7] (PDB 5T5A). All simulations were performed with the AMBER14 package[82]; using rism1d and rism3d.snglpnt for the 3D-RISM[80] calculations and pmemd.cuda[83] for the molecular dynamics simulations. For each of the simulations, the ribozyme was solvated in a truncated octahedral box with a 15.0Å buffer of TIP4P-Ew[84] waters and 140mM NaCl. The AMBER ff14SB force field[85] was used, with monovalent and divalent ion parameters designed for used with the TIP4P-Ew water model[86,87]. Prior to production runs of the molecular dynamics simulations all structures were subjected to 10ns of solvent annealing and solute equilibration, with the details of this procedure previously described in the ribozyme literature[88].

For the 3D-RISM calculation of the crystallographic coordinates, the missing hydrogen atoms were added in using LEAP and the metal ions were removed. The solution structure was taken from the end of the aforementioned equilibration procedure and was stripped of all solvent and metal ions prior to the 3D-RISM calculations. The initial 1D solvent susceptibility was calculated using SPC/E water[89] with $Na^+$ and $Cl^-$ ions with concentrations of 55.5M, 140mM, and 140mM respectively. The 3D-RISM calculations were solved with both the crystallographic and solution structure aligned to the same 142 by 156 by 110Å grid with gridpoints spaced every 0.5Å, such that a difference map between the two calculations could be constructed from the cation number density.

In order to explore potential rearrangements from the crystallographic structure to an active conformation in solution, the O2′ nucleophile was modeled in and covalently bonded to the scissile phosphate making a pentacoordinate phosphorane transition state mimic. First, a 100 ns simulation of a transition state mimic starting from the crystallographic coordinates for the RNA and $Mg^{2+}$ ions was completed. Guided by the results from the 3D-RISM calculations, the $Mg^{2+}$ ion at the M4 site in the both crystal structure and previous simulation was moved to the Watson-Crick edge of C7 (M4′ site) and restrained to directly coordinate the N3 of that residue. An additional restraint was applied to bring the O5′ leaving group within outer-sphere coordination distance (5Å) of the $Mg^{2+}$, consistent with the hypothesized model that a $Mg^{2+}$ bound water could act as the general acid. Three independent trajectories (different starting velocities) were propagated for 20 ns with statistics being calculated from the final 10 ns. From

these simulations where the spontaneous rearrangement of the active site was observed, the two proposed computational models were developed. For the "Mg Acid" model, an additional 25 ns of molecular dynamics was carried out with the sole restraint keeping the $Mg^{2+}$ (presumed to act as the general base) outer-sphere coordinated to the N3 of A9.

The "C7 Acid" model was developed by starting with the structure observed in the simulation where a $Mg^{2+}$ at C7 induced the rearrangement of the active site. This $Mg^{2+}$ was returned to the crystallographic M4 site, while C7 was protonated at the N3 position. Additionally, U57 was deleted (a mutation shown to have minimal effect on the rate)[7] as a means to reduce the conformational sample space. Following roughly 10 ns of equilibration where C7:N3H$^{+}$ was restrained to hydrogen bond with the O5′ leaving group, all restraints were slowly removed and 50 ns of unrestrained dynamics were performed. A second 25 ns long trajectory with a restraint keeping the general base $Mg^{2+}$ within 5Å of A9:N3 was completed in order to focus in on conformations that would be representative of the ribozyme at the transition state.

The "Mg Acid" and "C7 Acid" models provided here as supplementary datasets 1 and 2, respectively, are the average structures from the two 25 ns trajectories with the presumptive general base $Mg^{2+}$ restrained to within 5Å of A9:N3. These two average structures were then used in the 3D-RISM calculations confirming the positioning of the active site $Mg^{2+}$ ions seen in the MD trajectories (Figures S4 and S5).

### 3.5.2 Supporting Figures

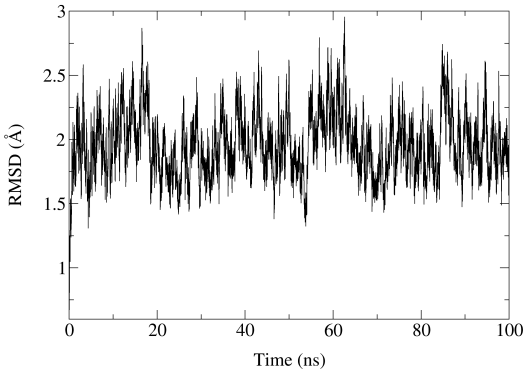



Figure 3.4: Time series of the best-fit heavy atom RMSD for an unrestrained 100 ns of the TSrz starting from the crystallographic coordinates for all RNA and $Mg^{2+}$ ions. Additionally, C54:O2′ was modeled in as part of pentacoordinate phosphorane transition state mimic. Snapshots were taken every 10 ps and aligned to the crystallographic coordinates deposited in the Protein Data Bank as PDB ID: 5T5A[7]

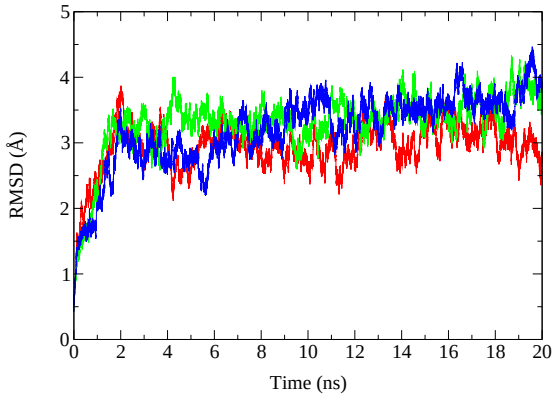



Figure 3.5: Time series of the best-fit heavy atom RMSD from the three independent trajectories (same initial coordinates, but different starting velocities) where a $Mg^{2+}$ ion at the M4′ site induced a spontaneous rearrangement of the TSrz active site.

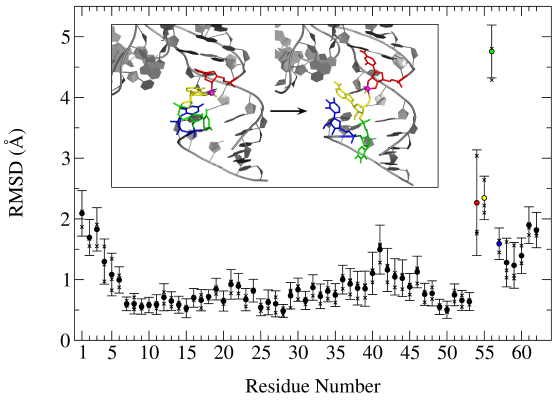



Figure 3.6: Average per residue RMSD (and standard deviation) from the final 10 ns of the simulation trajectory where a $Mg^{2+}$ ion at the M4′ site induced a spontaneous rearrangement of the TSrz active site (Figure S2). The reference structure, to which all residues were aligned and the RMSD was calculated, is the structure corresponding to the end of an equilibration run where the crystallographic coordinates relax to the inclusion of the transition state mimic phosphorane. Residues 54 - 57 are highlighted with the colors used for the corresponding residues in the inset figure. Inset: Rearrangement of the TSrz active site following the placement of a $Mg^{2+}$ ion at the M4′ site. Residues C54, A55, U56, and U57 (the residues that experience the significant rearrangement) are represented as sticks and highlighted in red, yellow, green, and blue, respectively. The scissile phosphate is highlighted as a magenta sphere. Inset left: Crystallographic structure PDB ID: 5T5A.[7] Inset right: Average structure from the final 10 ns of the three "rearrangement" trajectories combined.

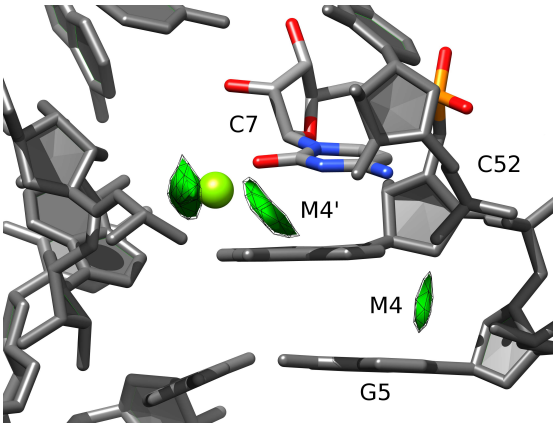



Figure 3.7: Cation density around C7 predicted by 3D-RISM for the "Mg Acid" model overlayed on the average structure from a 25 ns MD trajectory with the presumptive general base restrained. There is significant density at both the M4 and M4′ sites, with the higher predicted fractional occupancy at the M4′ site. The C7 residue is highlighted in color and a stick representation, while the $Mg^{2+}$ ion is shown as a green sphere.

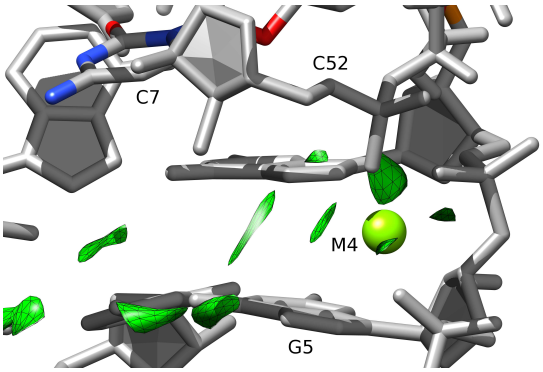



Figure 3.8: Predicted cation density from the 3D-RISM, using the average structure from a 25 ns simulation of the "C7 acid" model. Cation density is centered around the crystallographically observed M4 site. While no density was predicted directly at the M4 $Mg^{2+}$ ion, this is likely due to 3D-RISM not being able to place cations in the small space available between the non-bridging phosphoryl oxygens of C52 and G5 tightly bound to the $Mg^{2+}$ ion (Note: The $Mg^{2+}$ ions are stripped from the coordinate file, and the 3D-RISM calculation is run without additional relaxtion or rearrangement of the RNA structure). The C7 residue is highlighted in color and a stick representation, while the $Mg^{2+}$ ion is shown as a green sphere.

# Chapter 4

# Ribozyme Catalysis with a Twist: The Active State of the Twister Ribozyme in Solution Predicted from Molecular Simulation

Reprinted with permission from Colin S. Gaines and Darrin M. York. Ribozyme Catalysis with a Twist: Active State of the Twister Ribozyme in Solution Predicted from Molecular Simulation. *J. Am. Chem. Soc.*, 138(9):3058–3065, 2016. 

We present results from molecular dynamics simulations and free energy calculations of the twister ribozyme at different stages along the reaction path in order to gain insight into its mechanism. The results, together with recent biochemical experiments, provide support for a mechanism involving general acid catalysis by a conserved adenine residue in the active site. Although adenine has been previously implicated as a general acid acting through the N1 position in other ribozymes such as the hairpin and VS ribozymes, in the twister ribozyme there may be a twist. Biochemical experiments suggest that general acid catalysis may occur through the N3 position, which has never before been implicated in this role; however, currently, there is a lack of a detailed structural model for the active state of the twister ribozyme in solution that is consistent with these and other experiments. Simulations in a crystalline environment reported here are consistent with X-ray crystallographic data, and suggest that crystal packing contacts trap the RNA in an inactive conformation with U-1 in an extruded state that is incompatible with an in-line attack to the scissile phosphate. Simulations in solution, on the other hand, reveal this region to be dynamic and able to adopt a conformation where U-1 is stacked with G33. In this state, the nucleophile is in line with the scissile phosphate, and the N1 position of G33 and N3 position of A1 are poised to act as general base and acid, respectively, as supported by mutational experiments. Free energy calculations further predict the electrostatic environment causes a shift of the microscopic p$K_a$ at the N3 position of A1 toward neutrality by approximately 5 p$K_a$ units. These results offer a unified interpretation of a broad range of currently available experimental data that points to a novel mode of general acid catalysis

through the N3 position of an adenine nucleobase, thus expanding the repertoire of known mechanistic strategies employed by small nucleolytic ribozymes.

## 4.1 Introduction

The twister ribozyme is a recently discovered member of the group of small nucleolytic ribozymes that catalyze the site-specific cleavage of the RNA phosphodiester backbone. This group of self-cleaving ribozymes serves as an important testbed in the study of the mechanisms of RNA catalysis[56,70,90–93], and has impacted in the design of new biomedical technology[65,94–98] and theories of evolution[44,45].

The twister ribozyme was originally identified from bioinformatics, and subsequently has been studied extensively by X-ray crystallography,[6,9,10] mutagenesis,[6,10,30] and other biochemical experiments[6,10]. The twister ribozyme is of special interest in that experiments suggest that metal ions, although important for folding, do not play a direct role in catalysis[30]. This implies that catalysis derives mainly from the electrostatic engineering of the active site and participation of nucleobase functional groups, which are generally considered as fairly chemically inert when compared to amino acids[91,99]. In the case of the twister ribozyme, mutational studies have implicated that the N3 position of a conserved adenine residue, A1, is critical for activity, and may serve, in protonated form, as a general acid in the reaction. If this supposition is true, it would represent an intriguing new twist on general acid catalysis, and therefore would be important in furthering our understanding of the mechanistic strategies employed by RNA enzymes.

Crystallographic data[6,9,10] suggest that the N3 position of A1 is in reasonably close proximity to the O5′ leaving group to potentially act as a general acid. However, none of the currently available crystal structures[6,9,10] are likely representative of the active state of the twister ribozyme, obfuscating their functional interpretation. Thus, there remain significant gaps in our understanding of the twister ribozyme based on currently available experimental data that prevent definitive conclusions to be drawn about its mechanism. Specifically, there currently does not exist a structural (or dynamical) model of the active state of the twister ribozyme in solution that is consistent with the current body of available experimental data. As a result, there remain several key questions: (1) What is the origin of the inactive state observed crystallographically[6,9]? (2) What conformational events are required to form an active in-line conformation in solution, and what is the probability of finding the ribozyme in this conformation? (3) What residues likely act as the general base and acid? (4) How do electrostatics shift

the microscopic p$K_a$ values of implicated general acid and base residues, and help to stabilize the transition state?

In this paper, we utilize molecular dynamics simulations in a crystalline environment and in solution, along with free energy calculations, to gain insight into these key questions. Our results provide a unified molecular-level interpretation of a broad range of experimental data[6,9,10,30] that sheds new light onto the mechanism of the twister ribozyme.

## 4.2 Results and Discussion

### 4.2.1 Crystal packing causes extrusion of U-1 from the active site leading to a conformation that is not catalytically relevant.

Crystallographic data provide valuable information about the ribozyme structure and function; however, it remains an open question as to the degree to which a static picture of a deactivated ribozyme in a crystalline environment may have direct catalytic relevance for the active ribozyme in solution[100]. In the case of the twister ribozyme, with each of the three available crystal structures (PDB ID: 40JI[6], 4RGE[10], and 4QJD/4QJH[9]), U-1 both lacks the O2′ nucleophile and is seen to be extruded from the active site. As a consequence, these structures do not reflect a conformation where the nucleophile is both in line (so as to be able to attack the scissile phosphate) and directly interacting, through hydrogen bonding, with a residue with the potential to act as the general base. These features are generally considered to be a requirement for catalysis[101]. This observation then begs the question: Is the structure in the crystal catalytically relevant? To address these questions, we performed molecular dynamics simulations of the twister ribozyme, departing from the high resolution structure[6] (PDBID: 4OJI), both in a crystalline environment and in solution (see 4.4). We first set out to ascertain the consistency of our crystal simulations with available crystallographic data, and second, to characterize the dynamics in the region of the active site and, in particular, the local geometry associated with the orientation of U-1.

Overall, the crystal simulations were in very close agreement with available experimental data. The average structure from the crystal simulation was almost identical to the crystal structure (Figure 4.1), particularly in the active site where the heavy-atom root-mean-square deviation (RMSD) was 0.44 Å. Furthermore, the B-values determined from the simulated mean-square positional fluctuations (Figure 4.2) represent the degree of structural variation, and agree closely with those determined from crystallographic data (linear correlation coefficient of 0.90). Additional details on the observed structural fluctuations within the crystal simulation can be

found in Figure 4.11. Taken together, this provides credence for the reliability of the simulations, and also underscores the supposition that the region of residues 17 and 18 are highly disordered as interpreted by the lack of electron density in the $2F_o$-$F_c$ map.[6]. The crystal simulations predict that U-1 is stabilized by a crystal packing contact with G14 in a symmetry related monomer, where it remains trapped for the duration of the simulation and does not sample configurations that allow the 2′-oxygen to come in line with the scissile phosphate (Figure 4.12). Together with the relatively modest experimental and predicted fluctuations in this region, this suggests U-1 does not sample catalytically relevant conformations in the crystal.

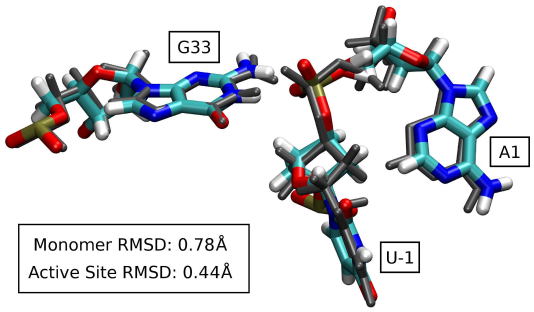



Figure 4.1: Overlay of experimental (PDB: 4OJI, gray) and average crystal simulation (colored) active sites. The positional, heavy atom RMSD is 0.78 Å for the full monomer and 0.44 Å for the active site (U-1, A1 and G33) alone.

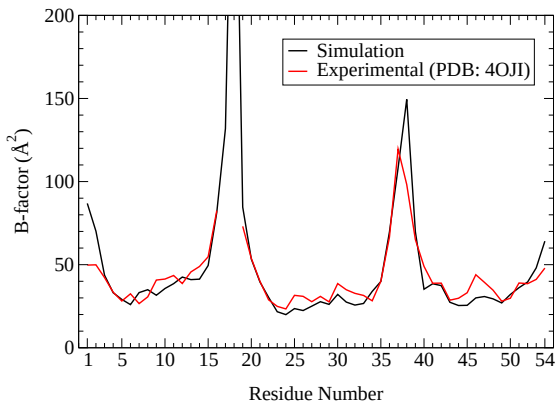
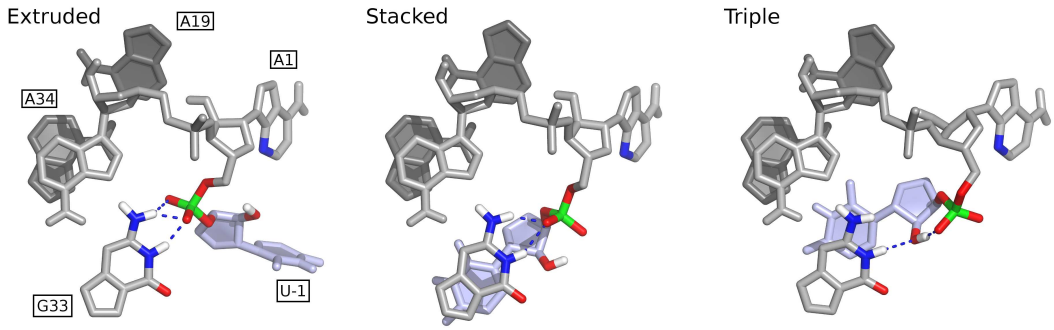



Figure 4.2: Comparison of simulation and experimental B-factors calculated for each residue (linear correlation coefficient: 0.90; standard error: 10.68 Å$^2$). Experimental B-factors for residues 17 and 18 were not reported as this region was interpreted to be highly disordered. Residue numbers along the x-axis correspond to the order of residues found in the PDB (Figure 4.10a) and map to the general numbering scheme for twister (Figure 4.10b).

### 4.2.2 Simulations in solution suggest that U-1 is able to transition between three distinct conformational states.

Over the course of a 120 ns unrestrained simulation of the twister ribozyme in solution, we identified three distinct conformational states of U-1 (Figure 4.13) that we denote as "extruded", "stacked" and "triple". In the extruded state, U-1 is completely solvent exposed in an orientation similar to that of the crystal (but with greater dynamical fluctuations since it is not stabilized by crystal packing contacts). In the stacked state, U-1 stacks under G33 while forming a hydrogen bond with a non-bridging oxygen of A34, and in the triple state, the additional hydrogen bonds characteristic of a WC/H/WC (U-1/A34/A19) base triple are observed. We found these different states could be distinguished by a single coordinate, the distance between U-1:N3 and A34:N7, henceforth denoted as $D_{stack}$: $D_{stack} > 7.0$ Å(extruded), 3.75 Å $< D_{stack} < 7.0$ Å(stacked), and $D_{stack} < 3.75$ Å(triple). The average structures from a set of three 75 ns simulations restrained to each of the aforementioned states are shown in Figure 4.3 (with a similar set of structures for G33 deprotonated shown in Figure 4.14) Additionally, this coordinate can be used to assess the occupancy of each state in solution, and estimate the kinetics of transitions between these states. Figure 4.4 shows the free energy profile along the stacking coordinate for simulations with G33, which has been implicated as a general

base, in its neutral form and deprotonated at the N1 position as would be required for general base catalysis.




(a) G33N1H - Extruded (b) G33N1H - Stacked (c) G33N1H - Triple

Figure 4.3: Average structures from the restrained simulations exploring the conformational states of U-1 (light blue) are labeled as follows: (a) extruded, (b) stacked, and (c) triple. The characteristic hydrogen bonding between G33 in its neutral, protonated form (G33N1H) and either the scissile phosphate or the nucleophile is shown in dark blue.

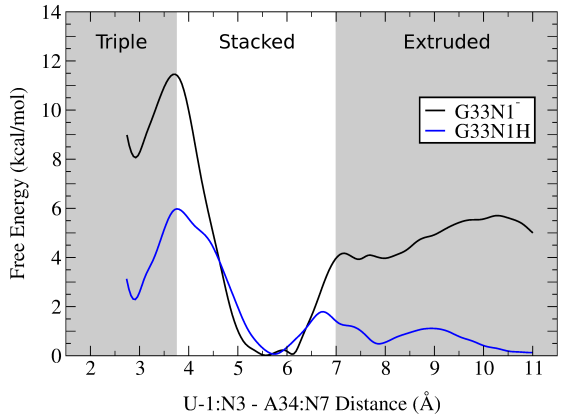



Figure 4.4: Free energy profile along the U-1 stacking coordinate defined by the distance between U-1:N3 and A34:N7 ($D_{stack}$), for G33 both deprotonated at the N1 position ($G33N1^{-}$ in black) and neutral (G33N1H in blue). The shaded region with $D_{stack} < 3.75$ Å corresponds to the "triple" state, while the shaded region with $D_{stack} > 7$ Å defines the "extruded" state. The "stacked" state for U-1 is left unshaded with $D_{stack}$ between 3.75 and 7.0 Å.

In both cases, simulations predict that U-1 rotation from the extruded state to the stacked

state (the global free energy minimum) is essentially unhindered in solution. While the triple state is higher in free energy than the stacked state by less than 2.5 kcal/mol with G33 in its neutral form, it is considerably higher (greater than 8 kcal/mol) for G33 in the deprotonated form. Further, there is a considerable forward (∼3.5 kcal/mol) and reverse (6-11.5 kcal/mol) barriers to transitions between the triple and stacked states. Overall, the stacked state is predicted to be long-lived and the dominant state in solution. As will be discussed below, this has important implications for catalysis.

### 4.2.3 The stacked state of U-1 promotes active in-line conformations and supports the role of G33 as a general base.

To explore the impact of the different stacking states on the active site structure where the nucleophile adopts a catalytically active in-line conformation, we performed sets of simulations restrained to each of these states, considering G33 both in the neutral (protonated at N1) and anionic (deprotonated at N1) forms (see 4.4).

Figure 4.5 shows a profile for the probability of finding the nucleophile in an active in-line attack conformation for each of the stacking states with G33 in its ground-state neutral form. Whereas active in-line attack conformations are not observed in the extruded state, and are partially observed (51.0% occupancy) in the triple state, they are dominantly clustered in the stacked state. This clustering is preserved (96.4% occupancy) in simulations with G33 activated (deprotonated) for base catalysis (Figure 4.5: G33N1$^-$ Stacked), and further analysis reveals this state is marked by preservation of key hydrogen bonds between the endocyclic and exocyclic nitrogens of G33 with the nucleophile 2′-OH and non-bridging oxygen of the scissile phosphate, respectively (Figure 4.5). This provides strong computational support that the twister ribozyme forms in-line attack conformations while in the dominant stacked state in solution, and upon deprotonation of G33, forms a hydrogen bond network that facilitates nucleophile activation with the N1 position of G33 acting as a general base catalyst.

With G33 poised to act as a base (G33N1$^-$), the sampling of in-line conformations significantly increases for both the triple and extruded states, with occupancies of 93.1% and 57.5%, respectively (Figure 4.15). However, the sampling of in-line conformations is only one of the bases for developing a model of the twister ribozyme's mechanism. With this in mind, the triple state is disfavored as the most relevant active state, first, because of the higher free energy associated with sampling that state relative to the stacked or extruded states (Figure 4.4)

and, second, because of the loss of the hydrogen bond between G33:N2 and the non-bridging phosphoryl oxygen of A1, which aids in neutralizing the negative charge on the phosphate (Figure 4.3c and Figure 4.14c). In contrast, with G33 deprotonated, the extruded state does exhibit the important hydrogen bonding interactions between G33 and A1. However, the stacked state is both lower in free energy and has a significantly higher probability of the nucleophile being positioned in line, when compared to the extruded state.

In addition to the six restrained simulations exploring the impact of U-1 stacking on the sampling of in-line conformations and the potential role of G33 as both a general base and hydrogen bond donor to the scissile phosphate, one additional restrained simulation was run to examine the interaction between the proposed general acid (A1:N3) and the 5′-oxygen leaving group. As with the other simulations of U-1 in the stacked state, the probability of finding the ribozyme in an in-line conformation (Figure 4.6a) is extremely high (99.1% occupancy). Additionally, the hydrogen bonding between G33 and either the nucleophile and the pro-$R_{\mathrm{P}}$ non-bridging oxygen are observed in a majority of simulation snapshots. As seen in Figure 4.6b, the distribution of distances between A1:N3 and A1:O5′ is bimodal, indicating that there are two distinct states being sampled. The first state (Figure 4.6c), is strikingly similar to the G33N1$^{-}$ stacked state (Figure 4.14b) where the general acid is neutral, and there is no direct interaction between the presumed acid and the leaving group. However, a small fluctuation in the position of A1 (Figure 4.6d) allows for N3 to donate a hydrogen bond to the 5′-oxygen, enhancing it as a leaving group. These data are consistent with the proposed active state model, where U-1 stacks beneath the general base G33 and A1:N3 is poised to act as the general acid.

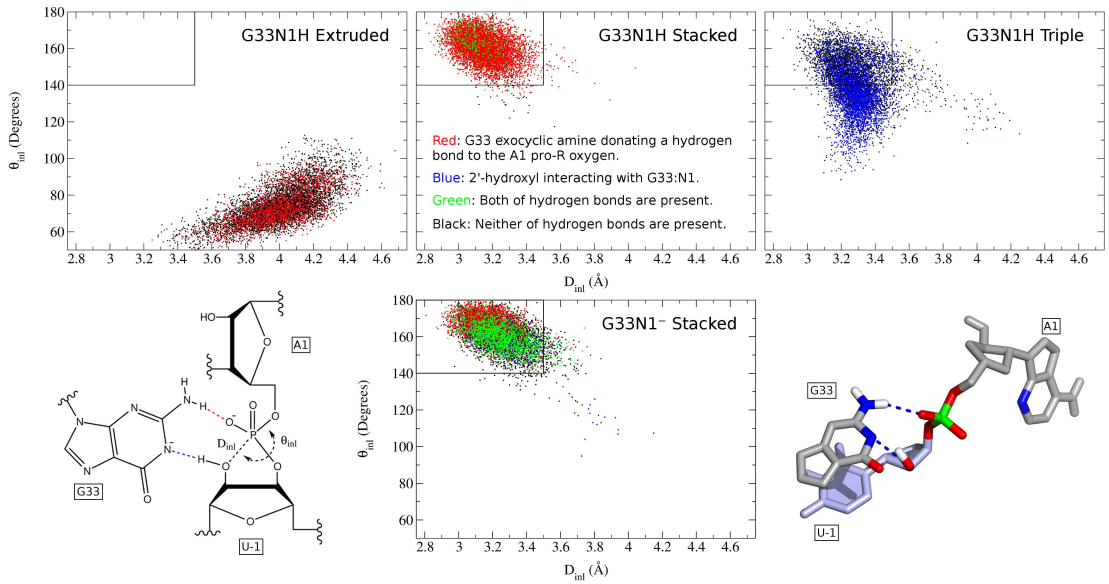



Figure 4.5: First row: nucleophile attack angle versus nucleophile-phosphate distance with G33 neutral (G33N1H) and U-1 in each of the three conformational states (extruded, stacked, triple). Second row: schematic highlighting hydrogen bonding used for clustering analysis, G33 deprotonated (G33N1$^-$) with U-1 in the stacked state, and the average structure of the active site for the G33N1$^-$ "stacked" simulation. The black outlined region ($\theta_{inl} > 140°$ and $D_{inl} < 3.5$ Å) indicates active "in-line" conformations that favor catalysis.

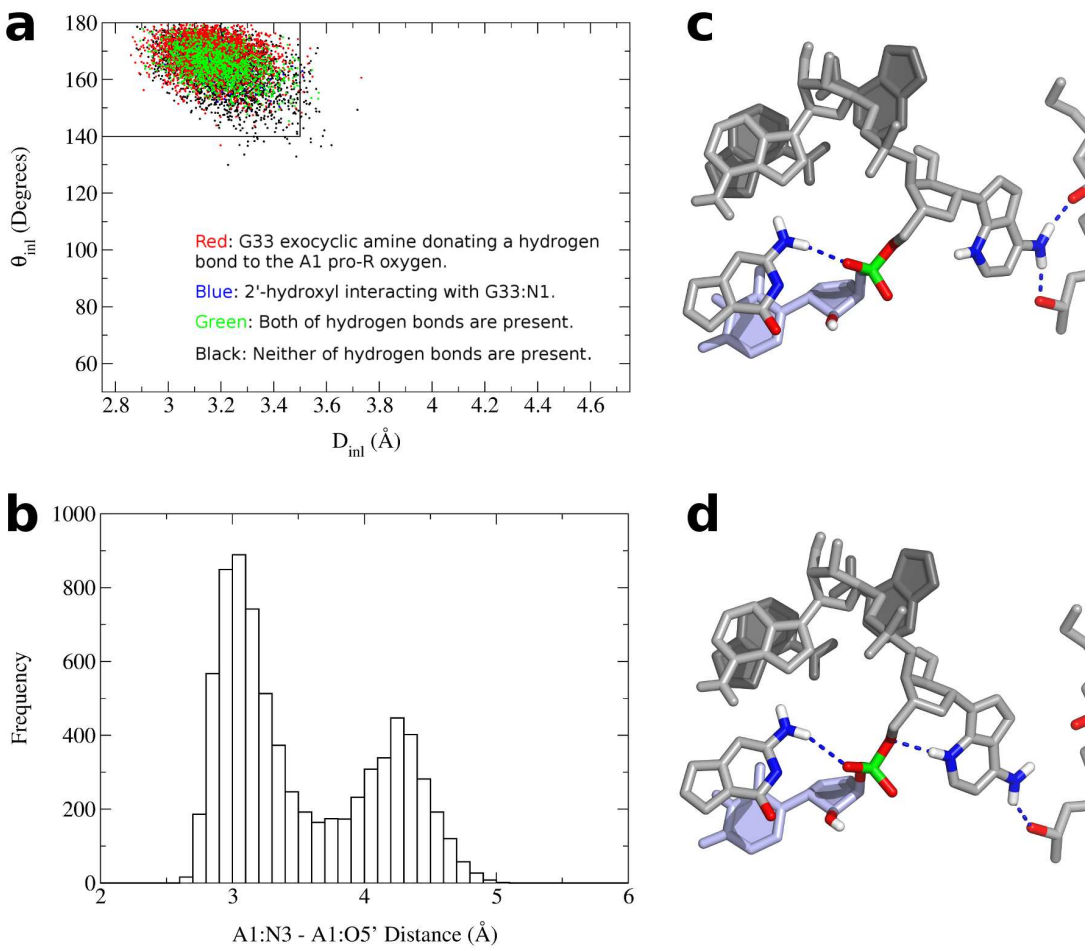



Figure 4.6: Restrained simulation of presumed active state. (a) Nucleophile attack angle versus nucleophile-phosphate distance clustered based on hydrogen bonding interactions with G33. The black outlined region ($\theta_{inl} > 140°$ and $D_{inl} < 3.5$ Å) indicates active "in-line" conformations that favor catalysis. (b) Distribution of distances between A1:N3 and A1:O5′, (c) Average structure of the frames where the A1:N3 - A1:O5′ distance is greater than 3.5 Å. (d) Average structure for the cluster where the presumptive general acid is directly hydrogen bonding to the leaving group. The two clusters were generated using the criteria that a heavy atom distance of than 3.5 Å is considered as the cutoff for a hydrogen bond.

### 4.2.4 Transition state mimic simulations further suggest that A1:N3 acts as a general acid.

Simulations of a transition state mimic (TSM) allow the prediction of the active site architecture and hydrogen bonding network at a critical point along the reaction pathway. Simulations of the TSM were performed with the twister ribozyme in a protonation state that would support general acid/base catalysis in accordance with implicated residues identified through mutagenesis and biochemical studies[6]. Specifically, both the N3 position of A1 (the presumed general acid) and the N1 position of G33 (the presumed general base) were protonated, representing a transition state that follows from the active state (designated as ${}_{AH^+}\mathrm{E}_{B^-}$) where G33 was negatively charged and has accepted a proton from the 2′-oxygen nucleophile. Under these assumptions, examination of the TSM simulations serves as a predictive test; if the integrity of the active site and catalytic requirements of the hydrogen bond network is preserved, this provides support for the roles of A1 and G33 as general acid and base, respectively. On the other hand, if the TSM simulations are not stable, this creates strong evidence against the working hypothesis.

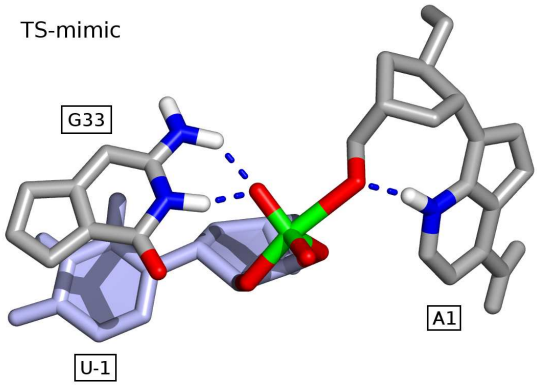



Figure 4.7: Average structure of the proposed dianionic phosphorane transition state mimic along with the characteristic hydrogen bonding network. With U-1 in the stacked state, the presumptive base (G33) donates dual hydrogen bonds to the pro-$R_\mathrm{P}$ non-bridging oxygen, while N3 of A1 (dark blue) is seen hydrogen bonding with the leaving group.

Overall, simulations of the TSM in the stacked state preserve the key hydrogen bonds necessary for catalysis, while providing important electrostatic stabilization of the negatively charged

transition state. Figure 4.16 shows that G33 maintains hydrogen bonding with the pro-$R_P$ oxygen of the scissile phosphate, while at the same time, A1 donates a hydrogen bond via the protonated N3 position to the O5′ leaving group. The results of the TSM simulations in the triple state (Figure 4.16b) are generally similar with those of the stacked state. Whereas, comparison of the TSM simulations in the extruded state (Figure 4.16a) are markedly different. In particular, there is a notable rotation of the backbone alpha and gamma torsion angles for the scissile phosphate. The $\alpha/\gamma$ torsion angles flip from $g^-/g^-$ in the extruded state to $t/g^+$ and $g^+/t$ for the triple and stacked states, respectively. The rearrangement of the backbone to $t/g^+$ or $g^+/t$ allows for the 5′ leaving group to be positioned in close proximity to A1. These results suggest U-1 must stack underneath G33 in order to preserve the necessary hydrogen bond network to promote general acid/base catalysis by A1/G33, consistent with the stacked state as the preferred working model of catalysis for the twister ribozyme.

### 4.2.5 The local electrostatic environment shifts the p$K_a$ of A1:N3 toward neutrality.

Although the present simulations, along with crystallographic data[6], indicate that A1:N3 is positioned to donate a hydrogen bond to the O5′ leaving group, the microscopic p$K_a$ of the N3 site for adenosine in solution has been predicted to be quite low (1.5 ±0.3)[4], raising possible objections to its effectiveness as an acid catalyst near neutral pH[9]. This poses a question as to whether the electrostatic environment in the active site might cause the p$K_a$ of A1:N3 to shift toward neutrality, and if so, to what extent. The observation of such a large shift in the p$K_a$ of a nucleic acid fuctional group would not be unprecedented in either experimental[102,103] or computational studies[104]. Therefore, we have performed free energy calculations to predict the apparent p$K_a$ values for the proposed general acid (A1:N3) and base (G33:N1) in order to address this key question.

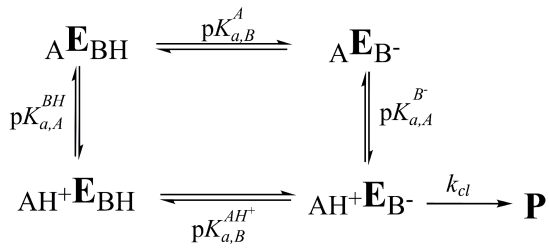



Figure 4.8: Microscopic general acid-base protonation state model[8]: The four microstates for the ribozyme (labeled E) are ${}_{AH^+}\mathrm{E}_{B^-}$ (the presumed active state), ${}_{AH^+}\mathrm{E}_{BH}$, ${}_{A}\mathrm{E}_{BH}$, and ${}_{A}\mathrm{E}_{B^-}$. The subscripts A and B indicate the p$K_a$ was calculated for the acid or base, respectively, given the protonation state of the residue noted by the superscript.

In keeping with the microscopic kinetic model for general-acid base catalysis[8] (Figure 4.8), the free energy (and thus the p$K_a$) for each leg of the cycle was calculated using a series of thermodynamic integration calculations (see 4.4). From these free energy calculations, predicted fractions of each microstate could be determined from the partition function $Q$,

$$Q = 1 + 10^{(\mathrm{p}K_{\mathrm{a,B}}^{\mathrm{AH}^+} - \mathrm{p}H)} + 10^{(\mathrm{p}K_{\mathrm{a,B}}^{\mathrm{AH}^+} - \mathrm{p}K_{\mathrm{a,A}}^{\mathrm{BH}})} + 10^{(\mathrm{p}H - \mathrm{p}K_{\mathrm{a,A}}^{\mathrm{B}^-})} \tag{4.1}$$

with the fraction of the active state, ${}_{AH^+}\mathrm{E}_{B^-}$, determined as

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

The observed rate, which is pH-dependent, is presumed to be directly proportional to the active fraction according to the relation:

$$k_{\mathrm{obs}} = k_{\mathrm{cl}} \ \times \ f_{(\mathrm{AH}^+/\mathrm{B}^-)} \tag{4.3}$$

where $k_{cl}$ is a proportionality constant, representing the pH independent rate of cleavage. Furthermore, the simulated active fraction (${}_{AH^+}\mathrm{E}_{B^-}$) can be fit to a two state apparent p$K_a$ model[105] that assumes the p$K_a$ values of the general acid and base are uncorrelated. This model has the three free parameters $k_{cl}$, p$K_{a,A}$, and p$K_{a,B}$, with the latter two parameters representing the apparent p$K_a$ of the acid and base participating in catalysis, respectively:

$$k_{\mathrm{obs}} = k_{\mathrm{cl}} \ \times \ 1/[1 + 10^{(pK_{\mathrm{a,B}} - pH)} + 10^{(pK_{\mathrm{a,B}} - pK_{\mathrm{a,A}})} + 10^{(pH - pK_{\mathrm{a,A}})}] \tag{4.4}$$

By fitting the function as described in Equation 4.4 to the simulated active fraction (i.e., in the same manner that experimental activity-pH curves are analyzed and interpreted), one

obtains a predicted pH-activity curve and apparent p$K_a$ values that can be compared with experiment (Figure 4.9). The predicted apparent p$K_a$ values of 6.5±0.4 and 9.0±0.4 for A1:N3 and G33:N1 acting as the general acid and base respectively, are very close to the experimentally observed values of 6.9 and 9.5, respectively, and the overall shape of the simulated and experimental pH-activity profiles are very similar[6]. These calculations support the hypothesis that the local environment within the active site of the twister ribozyme significantly shifts the p$K_a$ of A1:N3 (by ≈5 p$K_a$ units) towards neutrality, to provide a new mode of general acid-base catalysis.

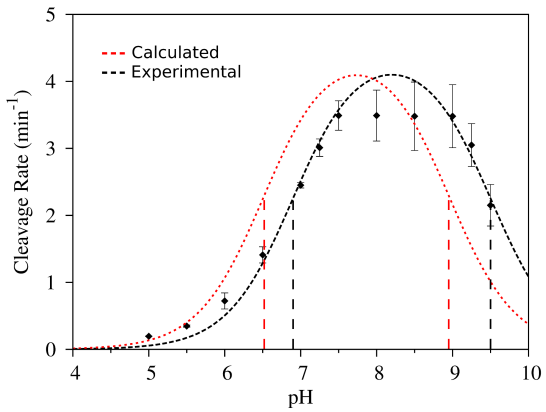



Figure 4.9: Comparison of the simulated and experimental pH-activity profiles and apparent p$K_a$ values. The predicted pH-activity profile (red) is scaled such that the peak maximum corresponds to that of a two state apparent p$K_a$ model fit to the experimental data[6] (black). The derived apparent p$K_a$ values for the acid and base are highlighted by the dotted lines for both the calculated (6.5 ± 0.4 and 9.0 ± 0.4) and experimentally derived (6.9 and 9.5) curves. The experimental rates shown here are the mean of at least three independent experiments, while the error bars are 2 times the standard deviation. For the calculated curve, the error associated with the predicted p$K_a$ values (± 0.4 p$K_a$ units) is derived primarily from the error in the experimental reference data used rather than from the free energy calculations or fitting procedure.

## 4.3 Conclusion

The simulations presented here provide a uniform interpretation of a large body of experimental data in terms of a molecular-level description of catalytic mechanism for the twister ribozyme.

The results depict a mechanism whereby U-1, which is trapped in an extruded state in the crystal due to crystal packing contacts, adopts a stable stacked state with G33 in solution. In this stacked state, the active site is able both to adopt an active in-line attack conformation of the nucleophile with the scissile phosphate and to position nucleotide residues to act as general acid and base catalysts. The transition state mimic simulations support a model whereby G33:N1 maintains hydrogen bonding with the O2′ position (which has a partial covalent bond to phosphorus) and A1:N3 donates a hydrogen bond to the O5′ position of the leaving group. Finally, free energy calculations suggest that the active site provides electrostatic stabilization of the transition state, and shifts the p$K_a$ of A1:N3 toward neutrality by almost 5 p$K_a$ units. Together, these results support the hypothesis that the twister ribozyme employs a new mode of general acid-base catalysis involving the N3 position of a protonated adenine, adding to the growing body of knowledge about the diversity of strategies employed by catalytic RNAs.

## 4.4 Methods

### 4.4.1 Crystal Simulation

In keeping with the methodology set forth in the literature[15,106–108], we completed a 120 ns simulation of the ribozyme in the crystalline environment in an effort to both validate our computational models and provide the foundation for investigating twister in solution. The simulated crystal was generated by applying the appropriate symmetry operations to the A conformation of the X-ray structure deposited as PDB ID: 4OJI[6], such that the unit cell contained 12 monomers arranged in the space group $P6_522$. The system was then solvated with TIP4P-Ew[84] water, and equilibrated over 20 ns at constant temperature and pressure conditions to ensure that the system volume was within 0.3% of the experimental value. In a continued effort to replicate the experimental conditions, ammonium acetate and magnesium chloride were added to the system to match the experimental bulk concentrations of 0.1M and 0.02M, respectively. Following approximately 20 ns of equilibration, the simulation was carried out using the GPU accelerated version of PMEMD14[82,83] at 280K and 1 atm using the AMBER FF14SB force field[85] with ion parameters[86] for use with the TIP4P-Ew water model. The average monomer structure from this simulation was generated by first reversing the symmetry operations for each of the 12 monomers in the unit cell in each 10 ps snapshot and then averaging over the coordinates of all monomers as well as over all snapshots in the last 100 ns of the 120 ns trajectory. Analysis of the data was conducted using both the XtalAnalyze[108] module in AMBER14 and in-house scripts.

### 4.4.2 General solution phase simulation parameters

For each of the solution phase simulations, the ribozyme was solvated in a truncated octahedral box of TIP4P-Ew water with 12 Å between the solute and box, additionally including the crystallographically observed $Mg^{2+}$ as well as 140 mM NaCl. All of the following solution phase simulations were carried out at 300K and 1 atm, using the AMBER FF14SB force field. The simulated annealing and solute equilibration procedure applied here has been used for other ribozymes and is detailed elsewhere[88]. Details on restraints applied and data collection specific to the various simulations completed is discussed below.

### 4.4.3 Free Energy of stacking U-1 beneath G33

Two free energy profiles were generated via umbrella sampling, by first minimizing representative structures of the stacked state (Figure 4.3b or Figure 4.14b) while restraining the U-1:N3 - A34:N7 distance to 5.5 Å. This distance restraint was applied using a harmonic potential with a force constant of 50 kcal/mol-Å$^2$. A short equilibration step of 1 ns was performed under constant temperature and pressure conditions, while continuing to enforce the above restraint. The structures used for each of the umbrella sampling windows were spaced every 0.5 Å from 3.0 to 11.0 Å and were equilibrated in serial from the previously equilibrated, adjacent window (starting from the initial window at 5.5 Å). The distance distribution data for each of the 17 windows, used for generating the free energy profile, was gathered every picosecond from the last 10 ns of 15 ns trajectories. For the data collection stage of the simulations, the restraint weight was reduced to 10 kcal/mol-Å$^2$ to provide adequate overlap between adjacent sampling windows. One additional window at 3.75 Å was needed for the G33N1$^-$ profile to improve sampling at the top of the barrier between the stacked and triple states. The free energy profiles were built using vFEP[109] and are presented in Figure 4.4. To estimate the statistical error of these calculations, 100 random subsamples of 10% of the data were collected. The 95% confidence intervals calculated from this distribution of subsamples are essentially the width of the lines in the plots presented in Figure 4.4.

### 4.4.4 Restrained MD exploring U-1 stacking states

During the equilibration stage (20 ns) restraints enforcing both the stacking state of U-1 and an active in-line conformation were maintained. The additional restraints supporting the active in-line conformation were slowly relaxed in the same fashion as described in the literature[88], and then removed for the data production stage. Keeping the restraints on $D_{stack}$ ensured that throughout the course of the simulations statistics could be gathered on a single stacking state of U-1, since the barriers are sufficiently small to allow for transitions between the states. The nucleophile attack angle ($\theta_{inl}$) is defined as the angle between U-1:O2′, A1:P, and A1:O5′, and distance between the nucleophile and the phosphate is denoted as ($D_{inl}$)[101]. These data were collected every 10 ps over the course of each of seven 75 ns restrained simulations, and clustered on the basis of the hydrogen bonding patterning involving G33, the 2′-hydroxyl nucleophile, and the pro-$R_P$ phosphate oxygen as seen in Figure 4.5 and Figure 4.6a.

### 4.4.5 Transition state mimic simulations

Initially, restraints enforcing the characteristic hydrogen bonding observed for each of the U-1 stacking states were applied. Once minimized, these structures were equilibrated, removing the restraints, using the same procedure as referenced previously and then simulated at constant temperature and pressure for 120 ns. Data were collected from the last 100 ns of the simulations.

### 4.4.6 Estimates of pH-activity profiles and apparent p$K_a$ values

To determine the pH dependent probabilities of each of the four microstates in Figure 4.8, a series of thermodynamic integration calculations were completed. First, as a reference, the free energy of deprotonating the N3 position of adenosine and the N1 position of guanosine was calculated and assigned p$K_a$ values of 1.5±0.3[4] and 9.4±0.2[5], respectively. The model nucleotides were simulated as methyl capped trinucleotides, including the residues flanking the proposed general acid and base in twister (e.g. A1 was modeled by the trinucleotide: $CH_3$-pU-**pA**-pA-$CH_3$). Each model was simulated for a total of 5 ns (1 ns per value of lambda), with the data being collected over the final 750 ps of each trajectory. The free energy for each leg of the thermodynamic cycle presented as Figure 4.8 was calculated for the full ribozyme in the U-1 stacked conformation, from 500 ps simulations at each of nine values for $\lambda$. A set of restraints reinforcing the hydrogen bonding interactions between the A2-G33 base pair and

from residues A1 to C16:OP1 and C17:OP2 were applied to maintain the integrity of the active site throughout the simulation, in addition to restraining $D_{stack}$ to between 3.5 and 7.0 Å. These restraints allowed for a more stable calculation by restricting the sampling to just the conformational basin (U-1 Stacked) predicted to be catalytically active. The difference in free energy between the models and the ribozyme microstates was then used to calculate the shifted (at 300 K) microscopic p$K_a$ values relative to the experimentally determined reference values for the nucleotides in solution:

$$pK_{\mathrm{a,ribo}} = pK_{\mathrm{a,model}} + (\Delta G_{\mathrm{ribo}} - \Delta G_{\mathrm{model}})/(RT * ln(10)) \tag{4.5}$$

From these data, the macroscopic "apparent p$K_a$" values could then be estimated, by fitting a two state apparent p$K_a$ model[105] to the active fraction, ${}_{AH^+}\mathrm{E}_{B^-}$, as described in the text.

It should be noted that the thermodynamic cycle imposes the contraint that the legs of the cycle sum to zero; however, the free energy calculations performed here were done so independent of one another. As a result the calculated cycle did not close and instead summed to -0.53 ± 0.19 kcal/mol. To close the cycle, the residual energy (-0.53 kcal/mol) was distributed evenly across the four microstates and the apparent pKa model was fit to the active fraction generated from the "closed cycle" pKa values. The data for these free energy calculations and the resulting pKa values used in the fitting procedure are presented in Table 4.1.

## 4.5 Supporting Information for: Ribozyme Catalysis with a Twist: The Active State of the Twister Ribozyme in Solution Predicted from Molecular Simulation

### 4.5.1 Supporting Figures

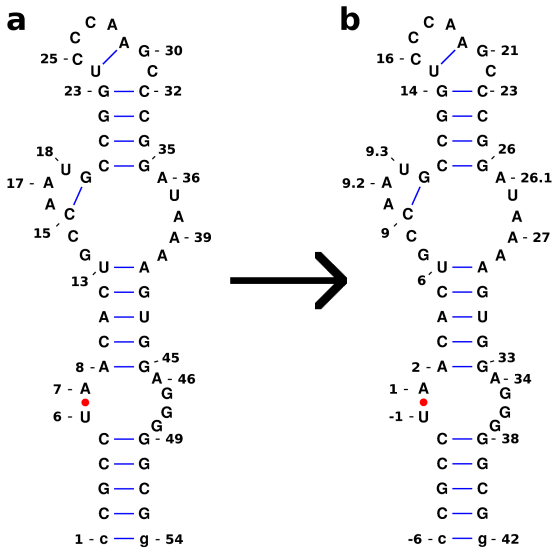



Figure 4.10: Mapping of residues numbered in order of appearance (a) in PDBID: 4OJI to numbering scheme (b) starting from the scissile phosphate marked by the red dot, in order to be generally independent of sequence length for the twister ribozyme. NB: base pairs invloved in forming the two pseudoknots not shown for clarity.

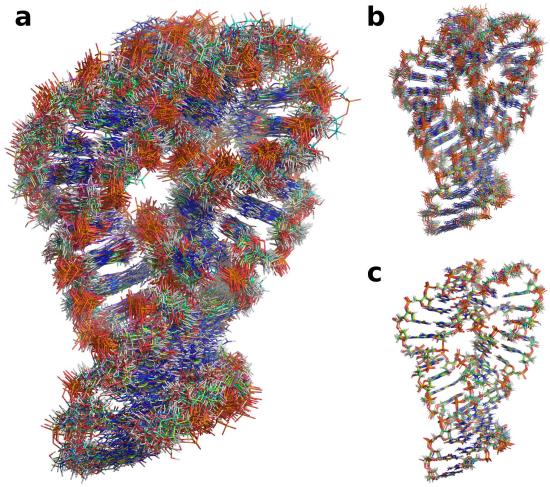



Figure 4.11: Overlays of representative structures from simulation of the twister ribozyme in the crystal environment. (a) Snapshots of each of the 12 monomers in the crystallographic unit cell taken every 20 ns. (b) Overlay of monomer structures averaged over the length of the 100 ns crystal simulation. (c) Snapshots taken every 20 ns of the trajectory generated by averaging over the coordinates of the 12 monomers within the unit cell. The overlay discussed in the main text (Figure 4.1) is an average over both time and monomer coordinates as that provides the more directly comparable structure to experiment.

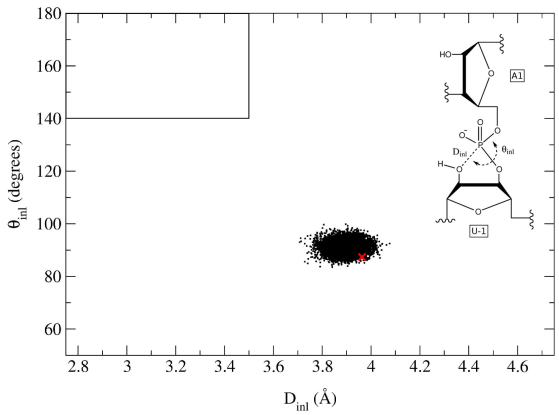



Figure 4.12: Scatter plot of in-line conformations for the crystal simulation. The black outlined region ($\theta_{inl} > 140°$ and $D_{inl} < 3.5$ Å) indicates active "in-line" conformations that favor catalysis. The red X indicates the average values of the nucleophile attack angle (91.6°) and distance (4.05 Å) for the double conformation of A1 in the deposited experimental structure (PDB: 4OJI[6]) with U-1:O2′ modeled in.

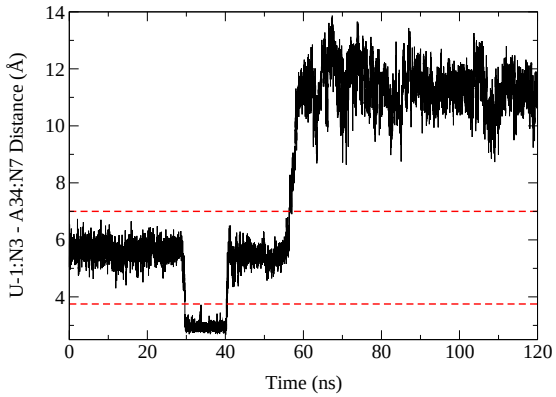
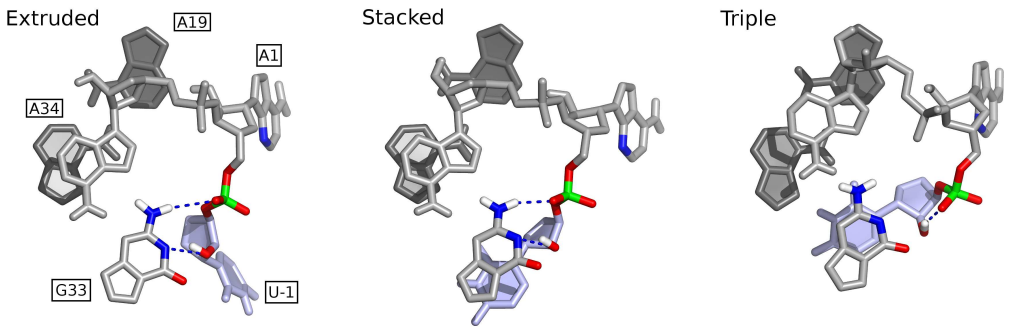



Figure 4.13: Time evolution of the U-1:N3 - A34:N7 distance, $D_{stack}$, showing the three stacking states of U-1. As indicated by the red dashed lines, the states are defined as follows: "Extruded" ($D_{stack} > 7.0$ Å), "Stacked" (3.75 Å $< D_{stack} <$ 7.0 Å) and "Triple" ($D_{stack} < 3.75$ Å).




(a) G33N1$^-$ - Extruded (b) G33N1$^-$ - Stacked (c) G33N1$^-$ - Triple

Figure 4.14: Average structures from the restrained simulations exploring the conformational states of U-1 (light blue) are labeled as follows: (a) Extruded, (b) Stacked, and (c) Triple. The characteristic hydrogen bonding between G33 in its deprotonated form (G33N1$^-$) and either the scissile phosphate or the nucleophile is shown in dark blue.

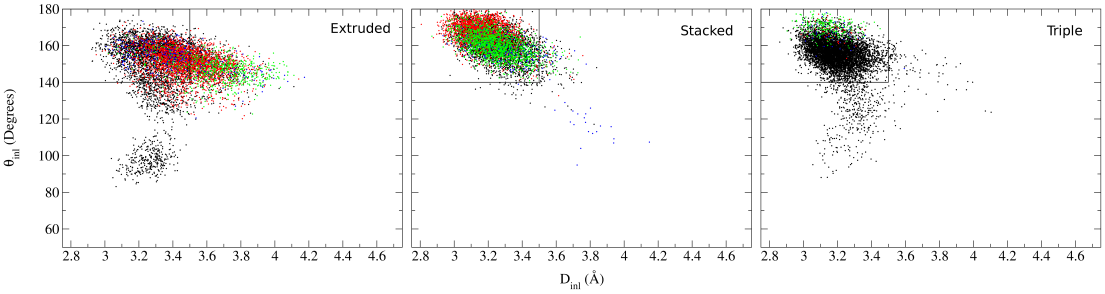



Figure 4.15: Nucleophile attack angle versus nucleophile-phosphate distance with G33 deprotonated (G33N1$^-$) and U-1 in each of the three conformational states (Extruded, Stacked, Triple). Clustering analysis is detailed in Figure 4.5 of the main text. The black outlined region ($\theta_{inl}$ $> 140°$ and $D_{inl} < 3.5$ Å) and indicates active "in-line" conformations that favor catalysis.

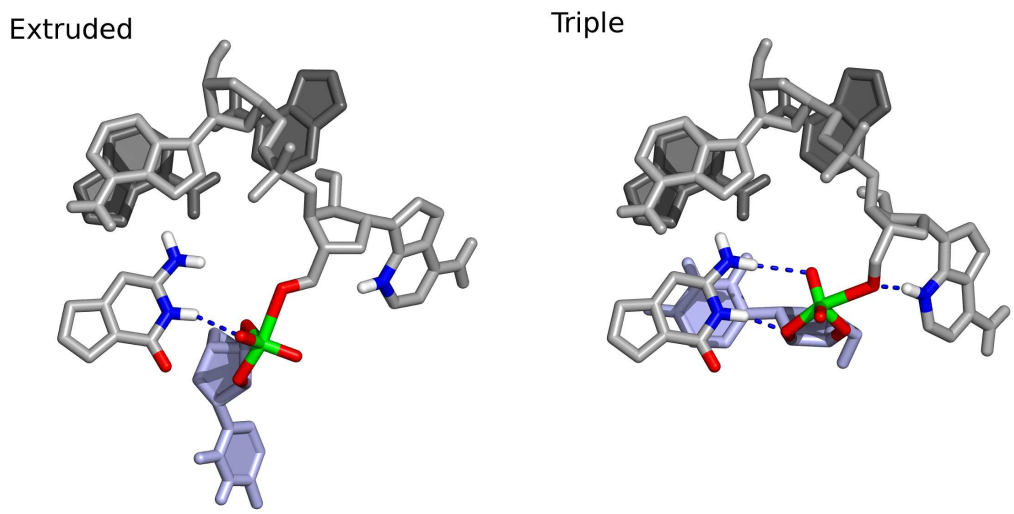



Figure 4.16: Representative structure of the proposed phosphorane transition state models along with the characteristic hydrogen bonding network, for both the U-1 extruded (a) and triple (b) states.

Table 4.1: Summary of free energy calculations for model nucleotide triplexes and each leg of the thermodynamic cycle presented in Figure 4.8. [a]Experimental p$K_a$ values assigned to correspond with the free energy to deprotonate the N3 site for adenine[4] or the N1 site of guanine[5] in their respective triplexes. [b]The adjusted p$K_a$ values after evenly distributing the residual -0.53 kcal/mol across all four microstates.

| **System** | **ΔG (kcal/mol)** | **p$K_a$ Shift** | **p$K_a$** | **Closed Cycle[b] p$K_a$** |
|---|---|---|---|---|
| $CH_3$-pU-**pA**-pA-$CH_3$ | -68.62 ± 0.03 | - - | 1.5 ± 0.3[a] | - - |
| $CH_3$-pG-**pG**-pA-$CH_3$ | -113.75 ± 0.11 | - - | 9.4 ± 0.2[a] | - - |
| ${}_{AH^+}E_{BH} \rightarrow {}_{A}E_{BH}$ | -61.84 ± 0.09 | 4.94 ± 0.07 | 6.44 ± 0.31 | 6.53 ± 0.31 |
| ${}_{AH^+}E_{BH} \rightarrow {}_{AH^+}E_{B^-}$ | -115.54 ± 0.10 | -1.31 ± 0.11 | 8.09 ± 0.23 | 7.99 ± 0.23 |
| ${}_{AH^+}E_{B^-} \rightarrow {}_{A}E_{B^-}$ | -60.27 ± 0.09 | 6.08 ± 0.07 | 7.58 ± 0.31 | 7.48 ± 0.31 |
| ${}_{A}E_{BH} \rightarrow {}_{A}E_{B^-}$ | -114.51 ± 0.10 | -0.56 ± 0.11 | 8.84 ± 0.23 | 8.94 ± 0.23 |

# Chapter 5

# Cleaning Up Mechanistic Debris Generated by Twister Ribozymes Using Computational RNA Enzymology

Reprinted with permission from Colin S. Gaines, Timothy J. Giese, and Darrin M. York. Cleaning Up Mechanistic Debris Generated by Twister Ribozymes Using Computational RNA Enzymology. *ACS Catal.*, 9(7):5803–5815, 2019. 

The catalytic properties of RNA have been a subject of fascination and intense research since their discovery over 30 years ago. Very recently, several classes of nucleolytic ribozymes have emerged and been characterized structurally. Among these, the twister ribozyme has been center-stage, and a topic of debate about its architecture and mechanism owing to conflicting interpretations of different crystal structures and in some cases conflicting interpretations of the same functional data. In the present work, we attempt to clean up the mechanistic "debris" generated by twister ribozymes using a comprehensive computational RNA enzymology approach aimed to provide a unified interpretation of existing structural and functional data. Simulations in the crystalline environment and in solution provide insight into the origins of observed differences in crystal structures and coalesce on a common active site architecture and dynamical ensemble in solution. We use GPU-accelerated free energy methods with enhanced sampling to ascertain microscopic nucleobase p$K_a$ values of the implicated general acid and base, from which predicted activity-pH profiles can be compared directly with experiments. Next, *ab initio* quantum mechanical/molecular mechanical (QM/MM) simulations with full dynamic solvation under periodic boundary conditions are used to determine mechanistic pathways through multidimensional free energy landscapes for the reaction. We then characterize the rate-controlling transition state and make predictions about kinetic isotope effects and linear free energy relations. Computational mutagenesis is performed to explain the origin of rate effects caused by chemical modifications and make experimentally testable predictions. Finally, we provide evidence that helps to resolve conflicting issues related to the role of metal ions in catalysis. Throughout each stage, we highlight how a conserved L-platform structural motif, together with a key L-anchor residue, forms the characteristic active site scaffold enabling each of the

catalytic strategies to come together for not only the twister ribozyme but also the majority of the known small nucleolytic ribozyme classes.

## 5.1 Introduction

The mechanisms whereby molecules of RNA can catalyze chemical reactions in biology has been a topic of tremendous interest, and growing impact since its discovery over three decades ago. A predictive understanding of the mechanisms of RNA catalysis in natural biological contexts can ultimately be transferred to synthetic systems such as xeno nucleic acids[110] or Hachimoji DNA and RNA[111] that have promise for future therapeutic and synthetic biological applications. Much of what is known about these mechanisms has been gleaned from detailed experimental and computational studies of small nucleolytic ribozymes that catalyze site-specific cleavage (and ligation) of RNA.[26,56,57] These ribozymes are widespread in both bacterial and human genomes[28–31] where they likely play complex roles in RNA processing and regulation of gene expression and have impact in biotechnology[55,112], medicine[23,24,113], and theories into the origins of life itself[42,44,46]. In the last 5 years, the number of known naturally occurring nucleolytic ribozyme classes has roughly doubled, sparking a surge of experimental effort aimed toward their structural and functional characterization.[30,31] This wealth of new information promises to reveal new insight into the diverse array of catalytic mechanisms available to RNA[114], including common themes and possible evolutionary connections between ribozyme classes.

Among the newly discovered ribozyme classes, the twister ribozyme stands apart as a system that has attracted a great deal of attention and ignited several debates in the literature. Recently, experts have brought to the forefront a critical barrier to progress in the field: the "mechanistic debris" generated by twister ribozymes, created in the wake of the rush to unveil details of its catalytic mechanism.[3] This debris arises from diverse biophysical data sets that lead to divergent structural models, conflicting interpretations of essentially the same biochemical data, and inconsistent use of different terms to discuss the same catalytic effects.

Twister ribozymes catalyze RNA transphosphorylation that leads to site-specific cleavage of the RNA phosphodiester backbone. This is a fundamentally important reaction in biology that is catalyzed by naturally occurring nucleolytic ribozymes (hammerhead[115–117], hairpin[118,119], HDV[120,121], VS[122,123], *glmS*[124,125], twister[6,30], pistol[31,74], TS[7,31], and hatchet[31] ribozymes), and protein enzymes (e.g., RNase A[126]), as well as artificially engineered DNA enzymes (e.g., 8-17 DNAzyme[127]). In this reaction, the 2′O makes a nucleophilic attack on the phosphorus

atom of the adjacent scissile phosphate to form a pentavalent transition state (or metastable intermediate), followed by departure of the O5′ leaving group to produce 2′,3′-cyclic phosphate and 5′OH cleavage products. To achieve catalytic rate enhancements that range from $10^6$ to $10^8$ for RNA enzymes[1,2], and several orders of magnitude greater for RNase A[126], enzymes employ up to four mutually enhancing catalytic strategies (designated $\alpha$, $\beta$, $\gamma$ and $\delta$)[1]:

- $\alpha$, arrangement of the O2′ nucleophile, P atom and O5′ leaving group in an active in-line attack geometry;
- $\beta$, electrostatic stabilization (neutralization/protonation) of the negative charge accumulation on the nonbridging phosphoryl oxygens (NPOs) at the pentavalent transition state;
- $\gamma$, activation (deprotonation) of the 2′OH to facilitate nucleophilic attack;
- $\delta$, stabilization (neutralization/protonation) of the accumulating charge of the O5′ to facilitate leaving group departure.

In the case of the twister ribozyme, there is strong evidence from functional studies that $\gamma$ and $\delta$ strategies occur through a general base/acid mechanism involving two highly conserved nucleobase residues: the N1 position of G33 and N3 position of A1.[128] The latter general acid strategy has never before been implicated, as the most likely acid site on an adenine nucleobase is at the N1 position rather than the N3[4], due to the unperturbed microscopic p$K_a$ of N1 being significantly closer to neutrality. The proposition of the N3 position of A1 as the general acid was initially met with some controversy[3], primarily due to the interpretations of different sets of crystallographic data and in some cases different interpretations of functional data sets. Further questions remained unanswered, such as: How does the twister ribozyme adopt a catalytically active conformation in solution? What are the origins of the apparent p$K_a$ shifts of the general acid and base, and how are they related to microscopic p$K_a$ values? What elements serve to stabilize the transition state, and how does this affect bonding? What is the role of solvent components, including divalent metal ions, in the reaction? While new insight into these questions has been provided by recent experiments[128], detailed answers must derive from rigorous computational modeling that provides a unified atomic-level interpretation of the current body of experimental data.

Multiscale modeling/simulation provide powerful tools to aid in the mechanistic interpretation of experimental data on enzymes[129,130] and ribozymes.[100] However, it must be remembered that even the most rigorous simulations, in the end, rely on models and are only meaningful if they are able to explain a broad range of experimental data for the system under study.

Here, we apply a comprehensive computational RNA enzymology approach[100] to clean up the mechanistic debris, as defined previously, generated by twister ribozymes and hopefully open the door to a unified interpretation of the current body of structural and functional data such that a consensus view of the mechanism can emerge.

In this approach, we first consider available crystallographic structures to explore the origin of their structural differences and perform crystal simulations to validate our simulation models and provide a baseline for discussion of predicted rearrangements that occur in solution. We then use molecular dynamics (MD) simulations to derive a structural and dynamical model of the catalytically active conformation and protonation state in solution that is consistent with a wide range of functional data. Using GPU-accelerated free energy methods, we characterize the probability of observing the ribozyme in its active state as a function of pH and validate the model by comparison with the experimentally measured activity-pH profiles. Departing from the active state, we determine the intrinsic reaction free energy barrier and catalytic pathway for the chemical steps of the reaction using *ab initio* combined quantum mechanical/molecular mechanical (aiQM/MM) simulations[100,129,130]. We then make experimentally testable predictions of the heavy-atom kinetic isotope effects based on the calculated transition state ensemble. We provide resolution of issues regarding the catalytic role of a divalent metal ion in the active site and make functional predictions that can be further tested with experimental mutagenesis. Throughout, we discuss the twister ribozyme mechanism using the simplified framework of four fundamental catalytic strategies for RNA transphosphorylation discussed above and draw important conclusions about how catalytic RNAs exhibit both similarities and fundamental differences to RNA-cleaving enzymes in the protein world. Finally, we demonstrate how the combination of all four catalytic strategies is brought together and enabled by the L-platform motif, which forms a characteristic scaffold in the active site of not just the twister ribozyme, but also most currently known small nucleolytic ribozyme classes.

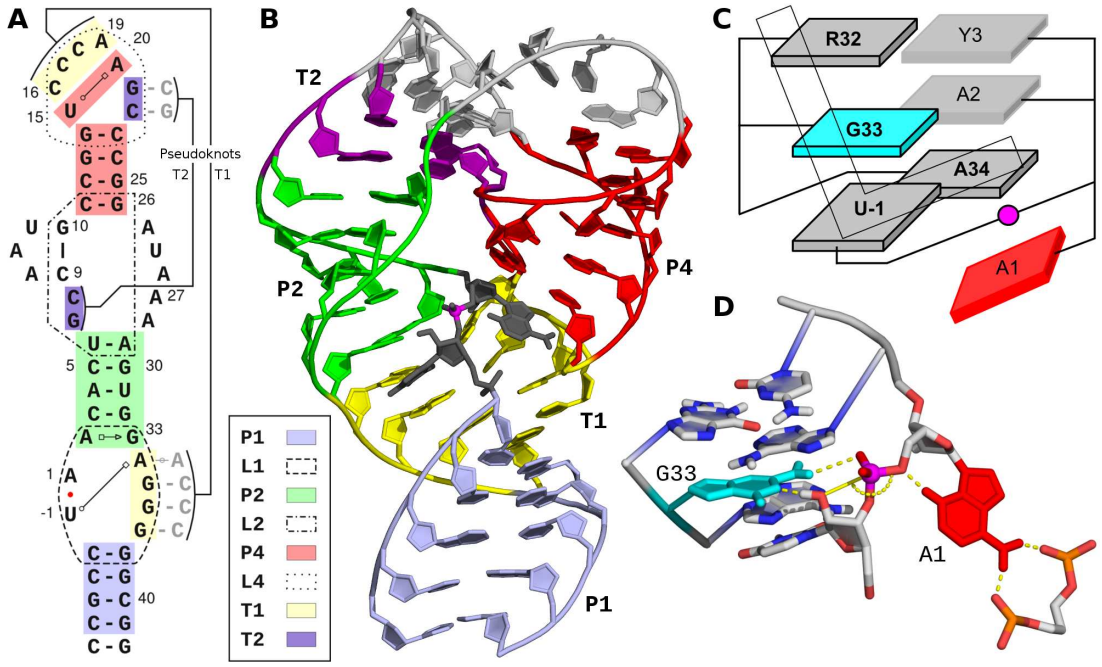



Figure 5.1: The structure of the twister ribozyme and active site L-platform motif. (**a**) The sequence and secondary structure of the twister ribozyme (PDB ID: 4OJI[6]), highlighting secondary structure elements (stems, loops, and pseudoknots) that are directly comparable across crystallographic structures. (**b**) Simulation snapshot showing the global fold of the twister ribozyme in a catalytically active state in solution, with color scheme matching that in **a**. (**c**) Cartoon-block schematic showing active site base pairing that forms the L-platform motif for the twister ribozyme. The general acid and base are highlighted in red and blue respectively, with the scissile phosphate in magenta. The bolded residues (U-1 and residues 32-34) form the 'L' of this motif, while A2 and Y3 (gray) constitute the "L-anchor" that serves to anchor the general base. (**d**) Zoom in of snapshot from **b** highlighting the base pairing and hydrogen bonding around the scissile phosphate characteristic of the L-platform motif in twister active site. Residues depicted are the same as in **c**, with the addition of the phosphates of N16/17 shown anchoring A1 in the *syn* conformation.

## 5.2 Results

### 5.2.1 Discrepancies in crystal structures stem from packing that disrupts weak helices.

The recent discovery of the twister ribozyme from comparative genomics[30] sparked the generation of a wealth of structural data from X-ray crystallography[6,9,10,73] that has been discussed in a recent review[131]. The twister ribozyme secondary structure (Figure 5.1a) consists of three alternating stems (P1, P2, and P4) and loops (L1, L2, and L4) which are organized into a compact fold (Figure 5.1b) by the tertiary contacts formed by two pseudoknots T1 and T2. The scissile phosphate contained within the L1 loop is then positioned in the center of the ribozyme where, in addition to stacking interactions, a series of non-standard (i.e., not canonical WC) base pairs and hydrogen bonds form the active site (Figure 5.1d). The architecture of the functional active site is supported by an L-platform/L-anchor motif (Figure 5.1c) that acts as a central scaffold for positioning key nucleobase residues and enabling all four catalytic strategies to come together.

Currently, there are five different crystal structures (PDB IDs 4QJH[9], 4OJI[6], 4RGE[10], 5DUN[73], and 4QJD[9]) four of which are significantly distinct (4RGE and 5DUN differ only in the deactivating mutation, 2′-deoxy and 2′O-methyl, respectively), available of the twister ribozyme (Figure 5.2), all of which require a conformational rearrangement to arrive at a catalytically active state. The two major areas of difference in the structures involve the folding of the P1 stem and the position of the uracil (U-1) immediately upstream of the scissile phosphate. A comparative analysis of these crystal structures has been presented by Gebetsberger and Micura[131] and below we provide additional insight into the origins of structural differences and their mechanistic interpretation.

Focusing first on the P1 stem, a clear trend emerges: the major discrepancies in the structural models stem from crystal packing that disrupts weak helices. The two sequences that have the longest CG-rich P1 stems, 4QJH (Figure 5.2a) and 4OJI (Figure 5.2b) with 8 bp and 5 bp, respectively, are fully intact and pack by coaxially stacking in the crystal, whereas the two sequences 4RGE (Figure 5.2c) and 4QJH (Figure 5.2d) with 4 bp stems of exclusively weak base pairs (A-U or G-U) are seen to have P1 stems whose 5′ strand is displaced (partially in the 4RGE and fully in 4QJD) as a result of crystal packing. A further distinction of the 4RGE sequence is that in the crystal U-2 and U-5 form base triples involving *trans* WC/H and *cis* WC/H base pairs with A34 and A35, respectively; the 4QJD sequence with guanines at

positions -2 and 35 does not form these additional base pairing interactions with the L1 loop.

Adding to the puzzle, recent FRET studies of both the *Oryza sativa*[132] and *env22*[133] twister ribozymes (corresponding to the 4OJI and 4RGE structures, respectively) provided conflicting evidence for the importance of the phylogenetically conserved P1 stem in the folding the ribozyme. For the 4OJI sequence, folding of the critical T2 pseudoknot was correlated with a fully intact and strongly base paired P1 stem, while for the 4RGE sequence the P1 stem is neither essential for folding nor efficient cleavage activity (as show by studies where the P1 stem was eliminated in its entirety[73]). As will be discussed below, this data is reconciled by considering how a misfolded P1 stem may interfere with local conformational rearrangements that result in the formation of the essential L-platform motif within the active site.

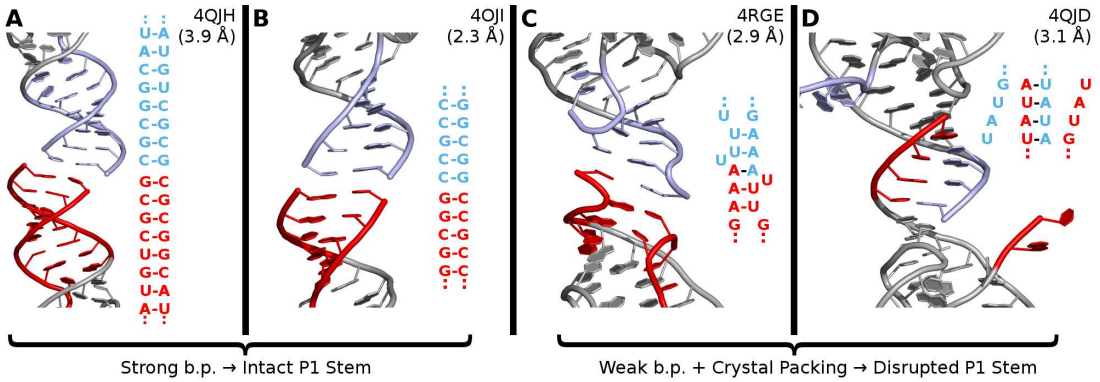



Figure 5.2: Comparisons of twister ribozyme P1 stem crystal packing. Cartoon representations of P1 stems of two symmetry related monomers colored in light blue and red for each of four crystallized sequences. (**a**) PDB ID: 4QJH[9] at 3.9 Å. Eight base pair P1 stem with majority C-G pairs remains intact and coaxially stacks in the crystal. (**b**) PDB ID: 4OJI[6] at 2.3 Å. P1 stem is comprised of all C-G base pairs, is fully intact and coaxially stacks in the crystal, similar to the 4QJH structure shown in **a**. (**c**) PDB ID: 4RGE[10] at 2.9 Å. Middle two U-A base pairs remain intact, while U-2 and U-4 form base triples in the L1 internal loop. In addition to coaxial stacking, the position of A41 and A41′ appear to be an average of two possible orientations for a WC/H base pair between those residues. (**d**) PDB ID: 4QJD[9] at 3.1 Å. The 3′ strand of the P1 stem (5′-UAUA-3′) is complementary with the equivalent strand in a symmetry related monomer (3′-AUAU-5′), leading to base pairing across the monomers.

Moving now to the active site contained within the L1 loop immediately adjacent to the P1 stem, it is observed that U-1 is extruded from any helical stacking in every structure. Further,

none of the crystal structures, except 4RGE, have the U-1 residue positioned such that the O2′ nucleophile would be reasonably in-line with the scissile phosphate. However, computational studies[134] departing from the 4RGE structure suggest that, in solution, in-line fitness is not maintained and an active state is not achieved. Given that in-line attack by U-1:O2′ is essential for catalysis, it is clear that some local conformational rearrangement is required for each crystal structure to form a catalytically active state in solution.

Toward this end, we depart from the 4OJI structure, which is not only the highest resolution structure at 2.3 Åbut also has strong base pairing in the P1 stem that appears less perturbed by crystal packing compared to the other crystal structures. To validate our simulation models, we performed simulations in the crystalline environment to compare the structure and fluctuations directly to crystallographic data (Figure 5.3). The simulated and experimental structures were overall very close (root-mean-square deviation, RMSD 0.80 Å), as were the B-values (root-mean-square error, RMSE 12.98 $Å^2$, R=0.92). These results lent credence to our simulation models and bolstered confidence in our prediction of the conformational rearrangements (described in the Section 5.2.2) that occur in solution resulting from removal of crystal packing restraints.

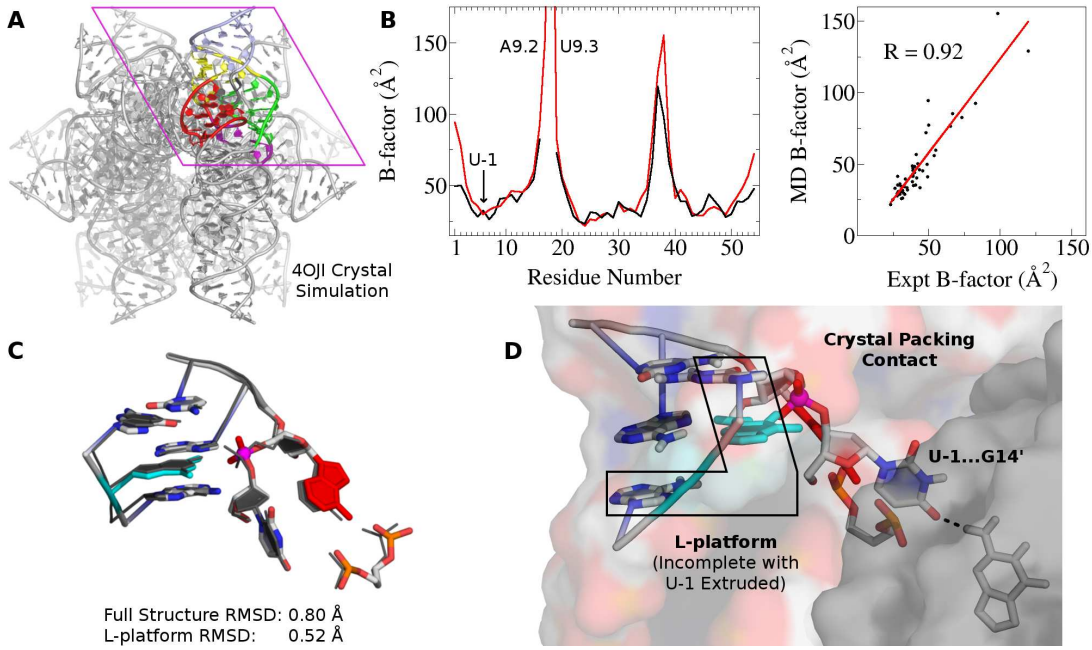



Figure 5.3: Crystal Simulation of PDB ID: 4OJI twister ribozyme. (**a**) Unit cell of 4OJI structure containing 12 asymmetric units used for crystal simulation viewed along six-fold symmetry axis. (**b**) Comparison of simulated and experimental B-factors, from 270 ns of MD crystal simulation. B-factors are calculated for each residue from a single combined trajectory where the full ensemble of structures (12 asymmetric units) is considered after applying the appropriate symmetry operations. (**c**) Overlay of average simulation (colored) and experimental (black) structures for the active site residues considered as part of the L-platform/L-anchor motif. (**d**) Crystal packing contact between U-1 and G14′ that stabilizes the conformation of U-1 where it is displaced from the heel of the L-platform.

### 5.2.2 Local conformational rearrangement is required to form a catalytically active state in solution.

In the 4OJI crystal structure, U-1 is observed to be involved in a crystal packing contact with G14′ (symmetry-related monomer). Crystal simulations (Figure 5.3b) indicate that hydrogen bonding between U-1 and G14′ locks U-1 in an extruded state that prevents in-line fitness (Figure 5.4a). In solution, the extruded state remains populated and becomes more flexible in the absence of the packing restraints (Figure 5.4b). Long-time simulations reveal that the twister active site can undergo a local conformational rearrangement whereby U-1 stacks under

G33 and forms a tWH/tWW (U-1/A34/A19) base triple that enables good in-line fitness to be achieved[135]. There exists an intermediate conformational state between the extruded and triple states, whereby U-1 stacks with G33 but is not within hydrogen bonding distance of A34 that also sufficiently enables in-line alignment of the nucleophile with the scissile phosphate. Over these unrestrained long-time simulations[135] all three states are sampled, demonstrating the conformational flexibility of the U-1 residue in solution relative to in the crystal structures. Connecting back to the discussion of the structural data, the positioning of the U-1 residue at the heel of the L-platform provides a model that can begin to unify the interpretation of these experiments. For the positioning of the inherently flexible U-1 residue in the base triple (U-1/A34/A19 and stacked with G33) to occur, the Hoogsteen edge of A34 must be accessible. The evidence suggests that this can be achieved either with a strongly base paired P1 stem, as seen with the *Oryza sativa* ribozyme, or by elimination of the P1 stem to prevent U-2 from competing for the triple with A34 and A19 observed with the *env22* construct.

With U-1 in the base triple (with G33 and A1 nucleobases in their standard/neutral protonation states), a significant sampling of conformations where the nucleophile is poised for in-line attack (Figure 5.4c) is observed. When the general base and acid residues are in their active protonation states (G33:N1$^-$ and A1:N3H$^+$), there is a considerable enhancement of the in-line fitness (Figure 5.4d) that is supported by stable hydrogen bonding with the O2′ nucleophile and O5′ leaving group. The same trend is seen with U-1 restrained to stack with G33 without forming the base triple[135]. It should be noted that in solution the U-1 residue has been seen to interconvert between these conformational states[135], and therefore, this residue was restrained to focus on exploring in-line fitness as a function of conformational and protonation state. In either the "stacked" or base "triple" conformation, G33:N1 is poised to act as the general base to activate the nucleophile, and A1:N3 is positioned to act as a general acid catalyst to donate a proton to the leaving group. In previous work, both the "stacked" and "triple" states have been considered and discussed in detail[135]. Here, we focus on the more structured U-1/A34/A19 base triple that, together with the active G33:N1$^-$/A1:N3H$^+$ protonation state (Figure 5.4d defines the "active state" of the ribozyme. This state is used later as and departure point for aiQM/MM simulations of the chemical steps of the reaction. The observed rate of cleavage is then directly proportional to the probability of observing the ribozyme in its active state. Characterization of this probability is challenging since both protonation and conformation state elements are coupled to one another and strongly dependent on pH and ionic conditions. In the next two sections (Sections 5.2.3 and 5.2.4) we consider each of these states and explore the nature of their coupling.

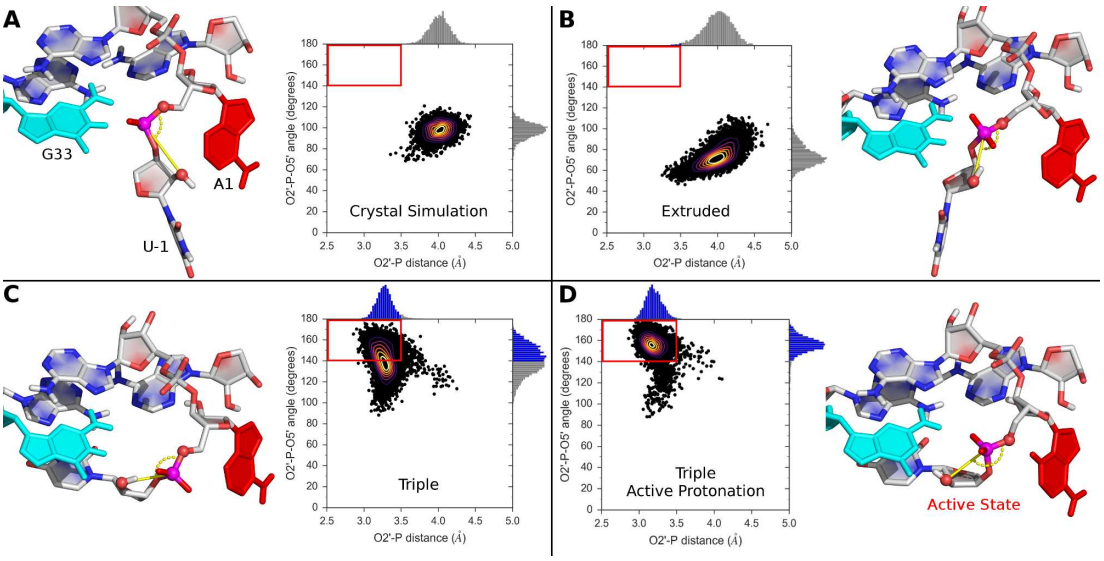



Figure 5.4: In-line fitness of the twister ribozyme. Plots of nucleophilic attack parameters: O2′-P-O5′ angle vs. O2′-P distance, with the red box in the upper left corner indicating high "in-line fitness". Representative structures corresponding to each plot are shown, highlighting the alignment of the active site residues and scissile phosphate. Reactive atoms shown as spheres (oxygens in red and phosphate in magenta). General acid and base colored light blue and red, respectively. (**a**) Results from 270 ns crystal simulation of the 4OJI structure, the nucleophile was modeled in at each frame for analysis alone. (**b-d**) Results for the 4OJI sequence in a solution environment with the nucleophile included in the 75 ns simulation, where U-1 is restrained to either be (**b**) extruded from the active site (similar to the crystal), or (**c,d**) forming a base triple within the L1 loop. In simulations **a-c** the presumptive acid and base are in the neutral protonation state, whereas the results shown in panel **d** are from a simulation of the active state where the G33 is deprotonated at the N1 position and A1 is protonated at the N3 position. Contours are drawn to highlight the density of overlapping data points, while the corresponding distributions are colored blue for angles or distances within the region indicating high "in-line fitness" (O2′-P-O5′ angle $> 140°$ and O2′-P distance $< 3.5$ Å).

### 5.2.3 Interpretation of activity-pH and p$K_a$ data.

Activity-pH datasets have been collected for a "wild-type" bimolecular type P3 twister ribozyme construct and a variety of mutants over a wide range of pH values under ion concentrations of 10 mM $MgCl_2$, 100 mM NaCl and 0.05 mM EDTA.[6,128] Additionally, NMR measurements for the

p$K_a$ of the presumptive general acid (A1) have been collected by measuring the chemical shift of $^{13}$C2-labeled-A1 both as part of a bimolecular twister construct as well as the substrate strand alone.[73] For the twister complex, the measured microscopic p$K_a$ is 5.1 and likely corresponds to the catalytic N3 site (although the experiment is unable to distinguish between protonation at either N1 or N3). Similar attempts were made to measure the microscopic p$K_a$ of the general base (G33) but were hindered by instability of the RNA at pH values above 9.5; however, the current body of evidence suggests the p$K_a$ of G33 is unshifted toward neutrality in the ribozyme.[73]

To aid in the interpretation of the experimental activity-pH data, we consider a series of three successively more complex models[136]: (1) a simple, non-cooperative model, (2) a cooperative model that allows coupling of general acid and base protonation states, and (3) an influencer model that further couples protonation and conformational states. The free energy differences for each leg of the thermodynamic cycle (Figure 5.7) corresponding to the cooperative model were calculated in an effort to estimate both the microscopic p$K_a$ values for the general acid and base in the ribozyme environment and the coupling between them. The predicted p$K_a$ values for A1:N3 and G33:N1 are 5.75 $\pm$0.23 and 9.24 $\pm$0.18, respectively, with a the coupling between them ($\Delta$p$K_{coop}$) of 0.21. While these estimates are reasonably consistent with the spectroscopic values, we opted to use the data from our simulations in the most conservative fashion. Therefore, the experimental p$K_a$ values were used as constraints, in addition to the calculated coupling, such that each of these models has three "free" parameters (Table 5.1) used to fit the data with the same statistical quality ($R^2 = 0.9894$), and are described in detail in the 5.5.

The only model that is able to fit the activity-pH data[6,128] with apparent p$K_a$ values of the general acid and base (6.8 and 9.5, respectively) while accounting for the microscopic p$K_a$ values from NMR measurements (5.1 and 9.5, respectively) and predicting responses to mutational effects, is the conformational influencer model shown in Figure 5.5a. This model predicts that the rate constant associated with the pH-independent rate of cleavage (a function of both the intrinsic rate and any other pH independent behavior), $k_{cl}$, is $\geq$ 200,000 min$^{-1}$, corresponding to a barrier of $\leq$ 12.6 kcal/mol. It is important to note that the activity-pH data considered in this analysis[6,128] was collected for a twister construct with guanine at the -1 position, rather than the strongly preferred uracil (90% conserved identity)[30]. This mutation was necessary for accurate kinetics measurements as the sequences with uracil cleaved too rapidly to be accurately measured (T. J. Wilson and D. M. J. Lilley, personal communication). As such, the estimate for the intrinsic barrier for the wild type twister ribozyme with uracil at the -1 position is expected

to be less than 12.6 kcal/mol, the upper bound predicted for the pH-independent rate of the G-1 mutant.

In addition to explaining a broader range of experimental data, our simulations provide support for the conformational influencer model. As demonstrated above, and in previous work[135], the positioning of U-1 at the heel of the L-platform is critical to forming a catalytically active state in solution. Therefore, we hypothesize that a pH-dependent equilibrium between the extruded and triple states (or at the very least stacking of U-1 with G33) of U-1 may provide the underlying physical basis for the conformational influencer. Given that this proposed model for the active conformation of U-1 is characterized not only by hydrogen bonding with A34, but also by stacking G33, this hypothesis could be tested with 2-aminopurine fluorescence spectroscopy. Twister ribozyme constructs that have a weak base pairing P1 stem or no P1 stem at all have already been shown to cleave with 2-aminopurine as the -1 residue with only a mild decrease in the observed rate[73]. Therefore, it is reasonable to propose extending these fluorescence experiments to explore the pH dependence with the twister ribozyme construct that contains a strongly base paired P1 stem (and for which the kinetics data was collected and subsequent computational modeling was performed), as they could directly assess whether the local rearrangement that completes the L-platform is in fact the conformational influencer. In any event, this conformational influencer model enables the consistent interpretation of the currently available experimental data and, when combined with our simulations, establishes a model for the active state in solution that serves as a departure point to further probe the catalytic chemical steps of the reaction.

### 5.2.4 Quantum free energy simulations predict stepwise nucleophile activation followed by a concerted mechanism of nucleophilic substitution with partial proton transfer in the rate-controlling transition state.

In studying the twister ribozyme chemical mechanism, we consider the general reaction scheme:

$$E + \mathrm{S} \overset{\Delta G_b}{\rightleftharpoons} \mathrm{ES_u} \overset{\Delta G_f}{\rightleftharpoons} ES_f \overset{\Delta G_{active}}{\rightleftharpoons} \mathrm{ES_r^*} \rightarrow [ES]_1^{\ddagger} \longrightarrow \mathrm{ES_{AP}} \rightarrow [ES]_2^{\ddagger} \rightarrow EP \overset{\Delta G_{b,p}}{\rightleftharpoons} \mathrm{E} + \mathrm{P} \tag{5.1}$$

where “E”, “S” and “P” represent the enzyme, substrate, and product, respectively, and the subscripts “u” and “f” represent unfolded and folded states, respectively. In the case of the self-cleaving twister ribozyme, we omit discussion of substrate binding and product release (shown in gray) and depart from the folded ground state ($ES_f$). This folded state is in equilibrium with

the rarely populated, reactant active state ($ES_r^*$) that is catalytically competent to carry out chemistry. The reaction then proceeds through a first transition state ($[ES]_1^\ddagger$) corresponding to activation of the 2′OH nucleophile by the general base to arrive at an "activated precursor" ($ES_{AP}$) intermediate. From the $ES_{AP}$ intermediate, the reaction follows a pathway proceeding through a second transition state ($[ES]_2^\ddagger$) to arrive at the 2′,3′-cyclic phosphate and 5′OH product ($EP$).

In Section 5.2.3, we used molecular dynamics and alchemical free energy simulations together with activity-pH and NMR data to establish bounds for the experimentally estimated free energy for forming the active state in solution at optimal pH ($\Delta G_{active} \geq 6.18$ kcal/mol) and the pH-independent free energy barrier that includes the rate-controlling chemical step of the reaction ($\leq 12.61$ kcal/mol). In the present section, we use multiscale quantum mechanical simulations to explore the free energy surfaces corresponding to the chemical steps of the reaction, enabling prediction of pathways and free energy barriers, and providing an atomic-level interpretation of mechanism[129,130]. Specifically, we use *ab initio* combined quantum mechanical/molecular mechanical simulations[100,129,130] with rigorous long-ranged electrostatics under full periodic boundary conditions[137] to determine 2D free energy profiles[109] along relevant reaction coordinates (see Section 5.4 for details). Similar aiQM/MM methods have been applied very recently to gain insight into mechanisms of phosphoryl transfer in RNA polymerase II[138]. For the twister ribozyme, we consider three general reaction coordinates (Figure 5.5b): a phosphoryl transfer, general base and general acid reaction coordinate. The phosphoryl transfer coordinate ($R_3 - R_4$) is the difference of the P-O5′ leaving group ($R_3$) and O2′-P nucleophile attack ($R_4$) distances. Analogously, the general base ($R_1 - R_2$) and acid ($R_5 - R_6$) coordinates are the corresponding difference distances involving proton transfer from the nucleophile to G33:N1, and from A1:N3 to the leaving group, respectively. We consider separately coupling of the phosphoryl reaction coordinate with the general base (Figure 5.5c) and general acid (Figure 5.5d and Table 5.2) coordinates.

As discussed earlier, MD simulations predict both a "stacked" and base "triple" state that exhibit high in-line fitness[135]. The "stacked" state lacks the tWH base pairing interaction between U-1 and A34 present in the base triple, making the former more conformationally dynamic. As discussed earlier, we have adopted the base triple conformation, along with the active G33:N1$^-$/A1:N3H$^+$ protonation state, as the departure point for aiQM/MM simulations. To map out the free energy profiles for the chemical steps of the reaction, many computationally intensive aiQM/MM simulations need to be carried out, and each of these simulations are conducted over much shorter time scale than the long-time classical MD simulations used to

study the conformational dynamics of the system. As such, the more structured base triple state is better suited as a departure point for the aiQM/MM simulations. Further, tests of the initial (general base) step of the reaction departing from both the stacked and base triple states indicate that the free energy barrier for nucleophile activation is $\sim$ 1 kcal/mol lower for the base triple than the stacked state.

The first free energy profile (Figure 5.5c) shows the coupling of the general base activation (Y-axis) with progression of phosphoryl transfer (X-axis) along the minimal free energy pathway. It is clear that the general base activation of the nucleophile is essentially fully complete prior to the initial nucleophilic attack of the phosphoryl transfer. This is evident from the vertical line in the 2D map that connects ($\mathrm{ES_r^*} \rightleftharpoons [\mathrm{ES}]_1^{\ddagger} \rightleftharpoons \mathrm{ES_{AP}}$) and indicates this step occurs with no contribution from the phosphoryl transfer coordinate. From linear free energy relations[139], this would correspond to a Brønsted coefficient, $\beta \simeq 1$, resulting from activation of the nucleophile in a pre-equilibrium step. Given the stepwise nature of this part of the reaction, general base activation of the nucleophile can be separated from phosphoryl transfer progression, and this step makes an additive free energy contribution of 2.29 kcal/mol to arrive at the activated precursor intermediate ($ES_{AP}$).

The second free energy profile (Figure 5.5d) shows the coupling of the proton donatation to the leaving group from the general acid (Y-axis) with progression of phosphoryl transfer (X-axis) along the minimal free energy pathway. Unlike the general base activation, the general acid and phosphoryl transfer steps are concerted, as indicated by the sigmoidal shape of the minimum free energy path and finite slope at the transition state $[ES]_2^{\ddagger}$. The barrier to arrive at the transition state from the $ES_{AP}$ intermediate is 7.34 kcal/mol. The rate-controlling transition state $[ES]_2^{\ddagger}$ occurs at a phosphoryl transfer coordinate slightly greater than 0, indicating P-O5′ bond cleavage is very slightly more progressed in terms of bond order relative to O2′-P bond formation, and that these processes are nearly synchronous in the transition state. As will be discussed below, this is in contrast to nonenzymatic RNA cleavage under alkaline conditions and to a lesser extent cleavage catalyzed by RNase A as predicted by theoretical calculations[140] and supported by kinetic isotope effect measurements[141].

Combining the free energy estimates for the general base activation and phosphoryl transfer/general acid steps, the simulations predict an overall intrinsic reaction free energy barrier of 9.63 kcal/mol to arrive at the rate-controlling transition state departing from the active state (i.e., $\mathrm{ES_r^*} \rightleftharpoons [\mathrm{ES}]_2^{\ddagger}$). This is less than the upper bound ($\leq$ 12.61 kcal/mol) estimated from modeling of the activity-pH data in the previous section (Section 5.2.3). It is not unexpected that the QM/MM simulations of the 4OJI sequence predict that the intrinsic reaction barrier

is well below the bound derived from the activity-pH data, because the kinetics were measured with a slower G-1 construct, as discussed previously. The important implication of the calculated free energy barrier falling below the estimated bounds is that the predicted pathway from QM/MM simulations corresponds to a feasible mechanism consistent with experimental constraints.

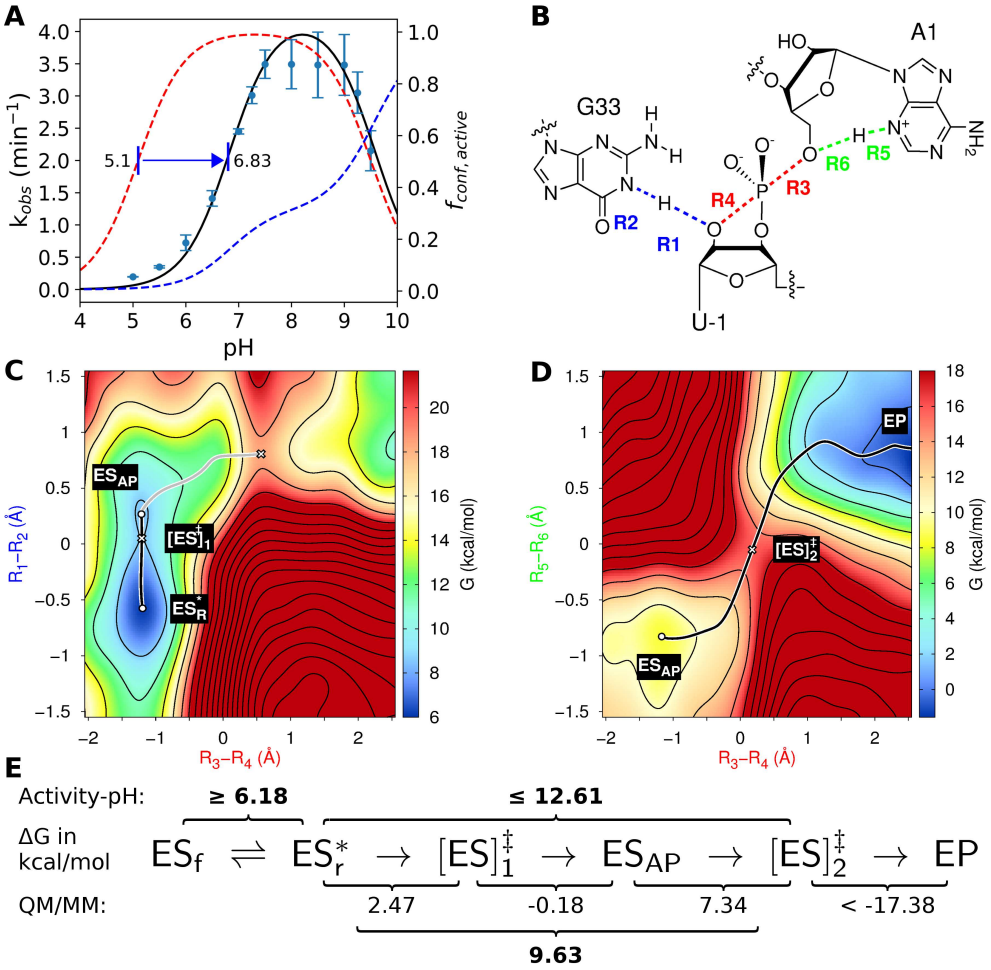



Figure 5.5: Model for catalytic mechanism of the twister ribozyme. (**a**) Experimental activity-pH profile[6]. Black: Conformational influencer model fit to the experimental data for the WT twister construct[6]. Red: Cooperative model (scaled) with p$K_a$ values assumed to be 5.1 and 9.5 with 0.21 units of coupling. Blue: Fraction of the active conformational state for the pH-dependent L-platform formation that produces an apparent p$K_a$ shift of the general acid from 5.1 to 6.83. (**b**) QM/MM reaction coordinates for: General base (R1-R2, blue), Phosphoryl transfer (R3-R4, red) and General acid (R5-R6, green) steps. (**c, d**) 2D *ab initio* QM/MM free energy profile for (**c**) the general base and phosphoryl transfer steps, and (**d**) general acid and phosphoryl transfer steps. The two profiles intersect at the local free energy minimum (ES$_{AP}$), and together indicate a stepwise nucleophile activation followed by a concerted nucleophilic substitution with partial proton transfer in the rate-controlling transition state. (**e**) Estimated free energies (kcal/mol) for the proposed catalytic mechanism from both the conformational influencer model (**a**) and the QM/MM simulations (**c,d**).

It should be pointed out that one of the high-resolution crystal structures (4RGE) of Patel and Micura[10] identifies a $Mg^{2+}$ ion bound at the active site, whereas the other high-resolution structure (4OJI) of Lilley[6] had an active site devoid of $Mg^{2+}$ ions. Stereospecific phosphorothioate substitution experiments with thiophilic metal ion rescue have been generated by both of these labs, in addition to the Breaker lab,[30,73,128] and discussed in a recent perspective by Breaker[3] that concluded the pro-$R_P$ NPO of the scissile phosphate does not require inner-sphere coordination of a divalent ion for catalysis (but likely is stabilized by hydrogen bonding with the exocyclic amine of G33), consistent with our simulations. The pro-$S_P$ position also does not appear to require inner-sphere coordination of a metal ion for catalysis as substitution of sulfur at this position has essentially no effect on maximum rate[3], although this position does appear to exhibit a modest thiophilic rescue[73].

Our simulation results provide insight into this important observation. Simulations predict the electrostatic environment created by the active site strongly attracts $Na^+$ ions from solution that interact with both G33:O6 and the pro-$S_P$ NPO of the scissile phosphate and on average follow the developing charge as it migrates from the nucleophile to the leaving group along the reaction coordinate (Figure 5.8). In this way, our simulations support the notion that the pro-$S_P$ NPO may be electrostatically stabilized by metal ions in solution (including $Mg^{2+}$ as observed in the 4RGE structure[10]) but not in such a way that requires direct inner-sphere coordination as confirmed by the phosphorothioate/metal rescue experiments[3]. It is nevertheless possible that phosphorothioate substitution at the pro-$S_P$ NPO creates a thiophilic divalent ion binding site that modestly enhances the electrostatic stabilization of the transition state ($\beta$ catalysis), leading to a modest "rescue" enhancement, but is not strictly required for the native substrate. In Section 5.2.5, we develop experimental tests that can serve to further validate our mechanism and pathway.

### 5.2.5 Simulation models lead to experimentally testable mutation and kinetic isotope effect predictions.

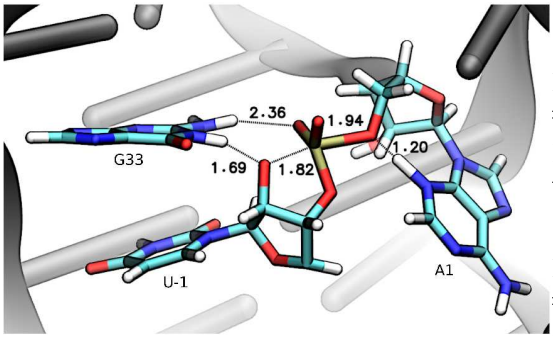
The molecular simulation models for the active state and catalytic pathway presented here provide a foundation from which to make experimentally testable predictions. Toward that end, we consider a set of mutations to the catalytic nucleobases (G33 and A1) that, to our knowledge, have not been measured and make predictions of how those mutations would affect both the pH-dependent probability of forming the active state and the intrinsic reaction barrier (Figure 5.6, top). Here we propose chemically precise nucleobase modifications that shift the

p$K_a$ of G33:N1 while preserving the hydrogen bond interface at the Watson-Crick edge so as not to directly impact hydrogen bonding within the active site. Changes in the pH-dependent active state probability can be directly predicted from the p$K_a$ shift using the influencer model fit to the activity-pH data. Changes in the intrinsic rate can be determined from repeating QM/MM simulations of the relevant steps of the reaction and measuring the free energy differences.

The G33(7cG) mutation (i.e., 7-deazaguanine-33) shifts the p$K_a$ of N1 up by 1.1 p$K$ units, lowering the probability of the active protonation state and thus raising the $\Delta$G for adopting the active state by 1.5 kcal/mol. However, by shifting the p$K_a$ of the general base up, away from neutrality, the difference between its p$K_a$ and that of the nucleophile is reduced. This is reflected in a slightly lower free energy difference between the $ES_r^*$ and $ES_{AP}$ states and a reduction of the intrinsic barrier by 0.24 kcal/mol. The end result is a predicted $k_{obs} \leq 93$ min$^{-1}$ or a 8.5-fold reduction in the rate constant compared to the wild type twister ribozyme. Similar competing effects are seen with the G33(6sG) mutation (i.e., 6-thioguanine-33), with the p$K_a$ being shifted towards neutrality by 1.1 p$K$ units instead of towards the nucleophile. The increased probability of the active protonation state is overshadowed by a significant increase in the free energy associated with the intrinsic rate, leading to a predicted 9-fold decrease in the observed rate. These modest perturbations to the p$K_a$ of the general base, correspond to modest decreases in the predicted rate constant. As an additional validation of our model, a more extreme shift in the p$K_a$ (-5.9 units) of the general base was tested, via mutation to 2-aminopurine. Our prediction of a 63-fold reduction in k$_obs$ matches closely with the experimentally measured value, at pH 7, corresponding to a 72-fold decrease in the rate.[6]

The interpretation of the predictions of the intrinsic barrier for the mutations of the general acid are more straightforward compared to the general base. The A1(3cA), general acid knock-out mutation results in a predicted intrinsic barrier of roughly 30 kcal/mol (Figure 5.6, top). An intrinsic barrier of this magnitude is on par with the background rate of cleavage.[142] When this mutation measured experimentally[128] the rate is four orders of magnitude slower at neutral pH and undetectable at low pH, as predicted from the simulations. Further, A1(7cA) mutation led to enhanced activity and a shift in the activity-pH profile aligned with the expected microscopic p$K_a$ shift of the N1 and N3 endocyclic nitrogens. We performed QM/MM simulations with the A1(3cA) knockout mutation and recapitulated the expected activity loss, and then repeated simulations with a A1(3cA) knockout in addition to a 5′ thio enhanced leaving group chemical modification, which was predicted to rescue activity (Figure 5.6).

| | WT (9.5) | G33(7cG) (10.6) | G33(6sG) (8.7) | G33(2AP) (3.6) | A1(3cA) | A1(3cA) + O5'S |
|---|---|---|---|---|---|---|
| $\Delta G_{active}$ | $\geq$ 6.18 | $\geq$ 7.68 | $\geq$ 5.09 | $\geq$ 2.31 | $\geq$ 3.87 | $\geq$ 3.87 |
| $\Delta G_{int}$ | 9.63 | 9.39 | 12.04 | 15.94 | 29.99 | 5.19 |
| $\Delta G_{obs}$ | $\geq$ 15.81 | $\geq$ 17.07 | $\geq$ 17.13 | $\geq$ 18.25 | $\geq$ 33.86 | $\geq$ 9.06 |
| $k_{obs}$ | $\leq$ 800 | $\leq$ 93 | $\leq$ 85 | $\leq$ 13 | nd ($\leq 10^{-7}$) | $\leq$ 7.6 x $10^{7}$ |
| k/k′ | ··· | 8.5 | 9 | 63 | ··· | 1.0 x $10^{-4}$ |




| $^{18}O$ Kinetic Isotope Effects | | | |
|---|---|---|---|
| **Non-Enzymatic** | $^{18}k_{NUC}$ | $^{18}k_{LG}$ | $^{18}k_{NPO}$ |
| Calc. | 0.977 | 1.042 | 1.004 |
| Expt. | 0.984 | 1.034 | 0.999 |
| **RNase A** | | | |
| Calc. | 0.993 | 1.028 | 1.002 |
| Expt. | 0.994 | 1.014 | 1.001 |
| **Twister** | | | |
| Calc. | 0.997 | 1.002 | 0.999 |

Figure 5.6: Experimental predictions of mutational and kinetic isotope effects. (**Top**) Predictions of mutational effects on the balance of free energy for adoption of the catalytically active state in solution and the intrinsic rate of reaction. $\Delta G_{active}$ is derived at pH 7 from the conformational influencer model and thus only provides a lower bound, while $\Delta G_{int}$ is derived from the QM/MM simulations. For the WT and general base mutants the predicted p$K_a$ for the N1 of residue 33 is listed in parentheses. Free energies are in kcal/mol, while rate constants have units of min$^{-1}$ (**Bottom Left**) The PBE0/6-31G* optimized transition state structure of twister ribozyme. (**Bottom Right**) Predictions of kinetic isotope effect values calculated at the PBE0/6-31G* level of theory along with experimental values for the non-enzymatic model reaction and RNase A.[11]

As a final prediction, kinetic isotope effects (KIEs) for the WT mechanism of the twister

ribozyme were calculated (Figure 5.6, bottom). KIEs report on the relative reaction rate constants $k/k'$ between isotopologues where $k$ and $k'$ are the pseudo first-order rate constants for the light and heavy isotopologues, respectively[71]. Measurement of KIEs offer the most sensitive mechanistic probe of changes in bonding that occurs in proceeding from the reactant state minimum through the rate controlling transition state. KIEs arise from subtle nuclear quantum effects that are responsive to changes in electronic potential energy surfaces and especially bond order and typically require computational approaches to provide a meaningful atomic-level interpretation[141,143–145].

Phosphoryl transfer reaction mechanisms have been studied extensively with KIEs[71,141,143–145], particularly with $^{18}$O isotope substitution at the nucleophile and leaving group positions (1° isotope effects), and non-bridge phosphoryl oxygen (2° isotope effect positions). The most straight-forward interpretation of these KIEs (Figure 5.10) is that if the bonding environment of an isotopologue becomes more "loose" in proceeding from reactant to transition state (e.g., if the average bond order associated with an isotopic position decreases), differences in the zero-point energy will cause the reaction to be slower for the heavier isotopologue, leading to a "normal" KIE value ($k/k'$>1). Conversely, if progression to the transition state leads to a more "tight" bonding environment, this will lead to an "inverse" KIE value ($k/k' < 1$).

The RNA cleavage (2′-O-transphosphorylation) reaction catalyzed by the twister ribozyme is also catalyzed by the protein enzyme RNase A[126]. Recently, KIEs have been measured/calculated for the uncatalyzed reaction[143,146] as well as catalyzed by RNase A[141,144,147], and in the presence of $Zn^{2+}$ ions[145,148]. These reactions proceed via a largely associative mechanism, as is typical for transesterification and hydrolysis of phosphate diesters[71,141,149], involving nucleophilic attack of the O2′ that proceeds through a dianionic pentavalent transition state. The large inverse O2′ and large normal O5′ 1° KIEs for the uncatalyzed reaction under alkaline conditions have been interpreted to suggest a considerably late transition state characterized by an almost fully formed O2′-P bond and an almost fully broken P-O5′ bond. Catalysis by RNase A leads to a less pronounced inverse O2′ and normal O5′ KIEs, and a transition state characterized by less P-O5′ bond cleavage and partial proton transfer from the general acid His119 to the O5′ leaving group. Catalysis by $Zn^{2+}$ ions has a similar KIE signature[145].

In the case of twister ribozyme, we predict even less pronounced inverse O2′ and normal O5′ KIEs than in RNase A, corresponding to a transition state that has a slightly less fully formed O2′-P bond, and less fully broken P-O5′ bond with significant degree of proton transfer to the leaving group. The explanation for this prediction is fairly simple when put into the context of general acid catalysis. In RNase A, the general acid is a histidine residue with unshifted

p$K_a$ of around 6[8]. In twister, the general acid is the N3 position of an adenine residue which is much more acidic (unshifted p$K_a$ less than 2). The greater acidity of the general acid in twister ribozyme makes it more reactive, causing the proton transfer to the leaving group to occur more readily and thus earlier along the reaction coordinate. Consequently, the twister ribozyme catalyzed reaction has less O2′-P bond formation and less P-O5′ bond cleavage and more advanced proton transfer in the transition state relative to that of the reaction catalyzed by RNase A.

## 5.3 Discussion

Overall, the computational results presented here provide a detailed dynamical model of twister ribozyme catalysis that unifies the interpretation of the current body of structural and functional data, and makes several experimentally testable predictions. Like other nucleolytic ribozymes, the twister ribozyme catalyzes RNA cleavage with an impressive rate enhancement relative to the uncatalyzed background rate, nearly on par with its protein enzyme counterparts such as RNase A that have evolved to promote multiple turn-over reactions. A striking feature is that this rate enhancement arises from a fast (RNase A-like) intrinsic cleavage rate counterbalanced by slow (low probability) formation of the catalytically active state itself.

The observed rate constant for UpA bound RNase A is $8.4\times10^4$ min$^{-1}$ (approximate activation free energy, $\Delta G^{\ddagger} \approx 13.2$ kcal/mol), which is at least 2 orders of magnitude faster than the estimated maximum rate constant for twister.[150] To estimate the probability of the active state, molecular simulations of RNase A at constant pH have been used to interpret the activity-pH profiles and have led to the conclusion that there is minimal cooperativity between protonation states of the general base and acid in this system.[8] Using a simple, non-cooperative model for RNase A with apparent p$K_a$ values of 4.88 and 6.95, adoption of the catalytically active state ($ES_f \rightleftharpoons ES_r^*$) represents 2.83 kcal/mol.[151] The intrinsic rate according to this model would correspond to a $\Delta G$ value of 10.3 kcal/mol, which is on par with the estimates for the twister ribozyme. A similar analysis for the VS ribozyme[26] shows that this same balance is critical for the mechanism of that ribozyme. Even with imperfectly tuned microscopic p$K_a$ values for the acid and base (at least two units removed from neutrality in the case of twister), the predicted intrinsic rate of these RNA enzymes is comparable to that of the protein enzyme analog RNase A.

Additionally, as a valuable compliment, the work of Świderek et al. presents estimates for the "intrinsic rate" of the twister ribozyme for a mechanism departing directly from the

crystallographically observed (most probable, but catalytically inactive) state.[152] Their explored mechanism relies on using one of the NPOs as proton shuttle, since in the most probable conformational and protonation state, neither G33 nor A1 are poised to act in a catalytic role. The calculated intrinsic barrier to reaction for this mechanism is ∼30 kcal/mol. This is equivalent to the uncatalyzed/background rate of cleavage[142], further highlighting the need for the ribozyme to adopt an improbable but catalytically active state in solution.

The realization that the catalytically active state of ribozymes often may be highly improbable demands caution in the interpretation of structural data. X-ray crystal structures of nucleolytic ribozymes have been critically important to the field in advancing our understanding of mechanism. However, since X-ray data depict static structures of deactivated ribozymes in crystalline environments, the degree to which they represent a dynamic active state in solution is, at best, speculative. In the case of the twister ribozyme, all of the available crystal structures require at least a local re-arrangement in order to adopt a catalytically active conformation. As discussed previously, molecular simulations[134] and quantum mechanical calculations[152] departing from the crystal structure and not realizing a catalytically active state led to negative or conflicting results. Indeed, the recent review by Breaker outlining "mechanistic debris generated by twister ribozymes"[3] warns that theoretical investigations departing from disparate structural models yield different predictions about mechanism and thwart efforts to unify conclusions. The present study thus pays special attention, and indeed sheds light on, the origin of the differences in the currently available X-ray crystal structures (e.g., weak base pairing that can lead to disruption of the P1 stem in the crystal) and goes on to identify a local conformational re-arrangement that leads to a catalytically active state that is consistent with known general catalytic strategies.[1]

## 5.4 Methods

The present work takes a comprehensive computational RNA enzymology approach[100] to study the catalytic mechanism of the twister ribozyme that combines (1) long-time molecular dynamics simulations in both a crystalline environment[153] and in solution at several stages along the catalytic reaction pathway, (2) GPU-accelerated alchemical free energy simulations, (3) multi-dimensional *ab initio* combined quantum mechanical/molecular mechanical (QM/MM) simulations, and (4) computational mutagenesis and kinetic isotope effect calculations. All molecular dynamics simulations were performed using the AMBER 18 package[154], in particular the GPU-accelerated simulation engine (PMEMD)[155], using the AMBER ff99OL3 RNA

force field which includes $\alpha/\gamma$[156] and $\chi$[157] dihedral modifications to the standard AMBER ff99 force field[158,159]. The solvent environment was modeled using TIP4P-Ew waters[84] with ion parameters for both monovalent[86] and divalent ions[87] designed for use with this water model. Alchemical free energy simulations were performed using the GPU-accelerated thermodynamic integration(TI)[160,161] method recently implemented into AMBER 18 by our group[155]. The *ab initio* QM/MM simulations were performed using code implemented in-house within a development version SANDER MD program[154] and were conducted in explicit solvent under full periodic boundary conditions using the recently introduced ambient potential composite Ewald method[137] for rigorous long-ranged electrostatics. Two-dimensional profiles were analyzed using the 2D variational free energy profile (vFEP) method[109]. QM/MM simulations and kinetic isotope effect calculations were performed using *ab initio* PBE0/6-31G* density functional quantum model[162,163]. Simulations were performed using a wide array of national production cyberinfrastructure provided by the NSF, NIH and Rutgers University. Full details for all computations in this work are provided in Section 5.5. A movie summarizing this work on the twister ribozyme is included in Appendix A.3.

## 5.5 Supporting Information for: Cleaning Up Mechanistic Debris Generated by Twister Ribozymes Using Computational RNA Enzymology

### 5.5.1 Supporting Methods

#### 5.5.1.1 General MD Simulation Procedures

All molecular dynamics simulations were performed using the AMBER 18 package[154], in particular the GPU-accelerated simulation engine (PMEMD)[155], using the AMBER ff99OL3 RNA force field which includes $\alpha/\gamma$[156] and $\chi$[157] dihedral modifications to the standard AMBER ff99 force field[158,159]. The solvent environment was modeled using TIP4P-Ew waters[84] with ion parameters for both monovalent[86] and divalent ions[87] designed for use with this water model. All solution phase simulations solvated the RNA (PDB ID: 4OJI)[6] in a truncated octahedral cell with at least 12Å between the solute and the edge of the cell. In addition to the crystallographically observed $Mg^{2+}$ and $Na^{+}$ ions to neutralize total charge, the bulk NaCl concentration was modeled at 140 mM to mimic near-physiological conditions. For the 54 residue twister ribozyme, the solvent box included 10,478 waters bringing the total number of atoms to approximately 44,000. Unless otherwise stated, these simulations were run in the NPT ensemble at 298K

and 1 bar. The temperature and pressure were maintained using a langevin thermostat with a collision frequency of 5.0 $ps^{-1}$ and the Berendsen barostat with a pressure relaxation time of 1 ps. Periodic boundary conditions were used with a 12Ånon-bonded cutoff and long-range electrostatics were accounted for by particle mesh Ewald[164]. All trajectories were generated with a 1 fs time step and bonds involving hydrogens were constrained using SHAKE[165].

#### 5.5.1.2 Solution MD Equilibration - Pre-annealing stage

In keeping with an equilibration protocol previously described in the ribozyme literature[88], the solvent environment (excluding crystallographically observed divalent metal ions) were energy minimized for 2000 steps with all other heavy atoms (solute) restrained to their initial (crystallographic) coordinates with a 50 kcal/mol-$Å^2$ harmonic potential. With the solute restraints maintained throughout the solvent equilibration procedure, the system was heated to 298 K at a rate of 1 K/ps and held at 298 K for 500 ps with constant volume.

#### 5.5.1.3 Solution MD Equilibration - Annealing stage

The solvent was annealed in two cycles of heating and cooling with constant volume and then equilibration of the density at constant pressure as follows: First, the system was heated to 600 K at 1 K/ps and held at high temperature for 500 ps. Next, the temperature was returned to 298 K, again at a rate of 1 K/ps and held for 1.5 ns. Finally, the density was equilibrated by maintaining constant temperature (298 K) and pressure (1 bar) for 3 ns.

#### 5.5.1.4 Solution MD Equilibration - Solute relaxation stage

With the solvent now equilibrated, the full system was energy minimized for 2000 steps. The force constants used for the solute restraints (now using the energy minimized coordinates as the reference) were stepped down in five stages (50, 25, 10, 5, and finally 2 kcal/mol-$Å^2$) allowing the system to relax for 100 ps at each stage.

#### 5.5.1.5 Crystal Simulation - System Setup

For the simulations of the PDB ID: 4OJI crystal, a unit cell consisting of 12 asymmetric units arranged in the $P6_522$ space group and all crystallographically observed ions and small molecules (acetate and ammonium) was constructed using the XtalUtilies packaged in AMBER 16. Two residues (A17 and U18) in the 4OJI structure were not resolved and were manually modeled in such that clashes when crystal packing were avoided. Ammonium acetate and magnesium

chloride at the experimental bulk concentrations of 0.1 and 0.02 M, were also included. The solvent box was initially set to the dimensions of the unit cell (a = b= 55.38 Å, c = 175.47 Å; $\alpha = \beta = 90°$, $\gamma = 120°$), then waters were added and equilibrated in an iterative fashion until the average unit cell volume closely matched the crystallographic volume under constant pressure conditions (1 bar). During this solvent equilibration procedue, the coordinates of the RNA were restrained with a 50 kcal/mol-Å$^2$ harmonic potential. First the system was energy minimized for 2000 steps, and then heated to the crystallization temperature (280 K) at a rate of 1 K/ps. The unit cell volume was evaluated over the course of 20 ns at constant pressure for consistency with the experimental unit cell volume. With 9431 waters (61505 atoms in total), the simulated volume was on average within 0.067% of the experimental value.

#### 5.5.1.6 Crystal Simulation - Simulation and Analysis

Following the solvent equilibration, the full system was minimized and restraints on the RNA were relaxed according to the solute relaxation procedure described above. A 270 ns trajectory was generated matching the general MD simulation parameters, except for setting the temperature to 280 K. The average structure and B-factors were determined by applying the appropriate reverse symmetry operations to each of the twelve monomers in the simulated unit cell and combining their individual trajectories into one without further alignment. In this way the fluctuations of the crystal lattice, not just the asymmetric units alone, are preserved. To generate the distribution for in-line fitness presented in Figure 5.4a, the O2′ nucleophile was modeled in to every frame of the combined crystal simulation trajectory in order to calculate the O2′-P distance and O2′-P-O5′ angle.

#### 5.5.1.7 Solution Simulations

Unless otherwise stated, all solution simulations were performed using the general MD simulation procedures described above.

#### 5.5.1.8 In-line fitness

The remaining three distributions exploring in-line fitness as a function of conformation and protonation state, were generated from 75 ns simulations where the position of U-1 is restrained. For the extruded conformation (Figure 5.4b), the distance between U-1:N3 and A34:N7 was restrained to be greater than 7Å, while for the two triple conformations (Figure 5.4c,d) this distance was restrained to be less than 3.75Å. A half-harmonic potential with a force constant of

10 kcal/(mol Å$^2$) was used to enforce these restraints. The average structure from the simulation with the ribozyme in both the active conformational state (triple) and the active protonation state (G33:N1$^-$ / A1:N3H$^+$) serves as the starting point for the simulations detailed in the following sections as this structure represents the catalytically active state in solution for the twister ribozyme.

#### 5.5.1.9 p$K_a$ values calculated from free energy simulations

For the simulated cooperative model, the free energy difference between each of the four protonation states was computed using Thermodynamic Integration (TI)[155,160,161]. The alchemical states differ only in their choice of MM charges (the acidic protons do not contribute to the Lennard-Jones energy). To determine the set MM charges used in the TI calculations, simulations of the four protonation states were performed, and 160 frames from each trajectory were used in a restrained electrostatic potential (RESP) charge-fitting procedure. For each of the active site residues, a QM/MM geometry optimization was performed using the PBE0/6-31G* Hamiltonian, followed by evaluation of the electrostatic potential on a molecular surface surrounding the QM residue, as it is polarized by the MM ribozyme environment. The fitting procedure produces four sets of G33 and A1 nucleobase charges (corresponding to the four protonation states), and a single set of RESP-fit charges for U1, A2, A33 for use in all protonation states. Fitting constraints were applied to preserve the nucleobase net charges and to retain common nonbridge oxygen charges.

The closed thermodynamic cycle connecting the four protonation states is divided into 44 $\lambda$-window intermediate states, each of which is simulated for 60 ns (1 fs timestep) in the canonical ensemble at 298 K using a Langevin thermostat (5 ps collision frequency). The 44 simulations are run in tandem, and equilibrated among each other through the use of Hamiltonian replica exchange, using an exchange-attempt-rate of 20 ps/attempt. The first 20 ns is discarded as equilibration, and the final 40 ns are used to evaluate four free energy differences (corresponding to the four edges of the thermodynamic cycle). The thermodynamic cycle was evaluated three independent times by varying the initial velocities. The calculated free energies are corrected by reperforming the above procedure on isolated adenine monophosphate (AMP) and guanine monophosphate (GMP) in an aqueous environment. The free energy of these reference simulations are removed and replaced by experiment using measured p$K_a$ values. For example, the p$K_a$ of A1 when G33 is neutral is:

$$\mathrm{p}K_{a,G}(A) = \frac{\Delta G_{RNA,G}(AH^+ \rightarrow A) - \Delta G_{AMP,G}(AH^+ \rightarrow A)}{1.363} + \mathrm{p}K_{a,\mathrm{expt.}}(A) \tag{5.2}$$

The error in the p$K_{a,G}(A)$ value is propagated from the standard error of the three independent trials and the residual free energy needed for the thermodynamic cycle to close (-0.39 kcal/mol) was evenly distributed across all four protonation states. Additionally, the coupling between the acid and base p$K_a$s are calculated from these free energies as described in the thermodynamic cycle shown as Figure 5.7.

#### 5.5.1.10 QM/MM - 2D free energy profiles

The *ab initio* QM/MM simulations were performed using code implemented in-house within a development version SANDER MD program.[154] Departing from the catalytically active state in solution determined from the aforementioned MD simulation, structures corresponding to 17 windows along the general base coordinate (R1-R2) from -1.60 to 1.60, with 0.2Å spacing were generated in series. These windows were then equilibrated for 2 ps using the AM1/d-PhoT hamiltonian[166] with additional sugar pucker corrections[167]. From each of these windows an equivalent process was carried out to generate windows spaced 0.4Å apart ranging from -2.4 to 2.4 along the phosphoryl transfer coordinate (R3-R4), thus generating a 2D grid of 221 windows. Production simulations for this 2D grid were then carried out using the PBE0/6-31G* level of theory for the QM region and the AMBER FF14SB force field to describe the MD region. The QM region used these free energy calculations included the two residues flanking the scissile phosphate (U-1 and A1) in addition to the nucleobase G33, totalling 70 atoms in addition to the ∼44,000 MD atoms. All QM/MM simulations were run in the NVT ensemble at 300K using a Langevin thermostat (5 ps collision frequency). Long-range electrostatics were treated with the ambient potential composite Ewald method[137]. A 12 Å real-space cutoff was used, and the reciprocal-space energy was solved using a 1 point/Å uniform fast Fourier transform grid. The wavefunction was converged to within a commutator error of $10^{-7}$ au at each step of dynamics. Harmonic restraints with force constants of 100 kcal/(mol Å$^2$) were applied to the reaction coordinates with an additional half-harmonic restraint preventing proton transfer from the general acid. The statistics gathered from these initial 2 ps production simulations were then analyzed using the vFEP software[109] and a minimum free energy pathway along the resulting surface was determined. This pathway was then refined by simulating 80 windows on the grid nearest to the pathway for an additional 8 ps. The refined 2D free energy surface for the general base and phosphoryl transfer coordinates was again generated using the vFEP software on the now 1.1 ns of data. An equivalent procedure was carried out to explore the coupling of the phosphoryl transfer (R3-R4) and general acid (R5-R6) coordinates. For this profile, the 2D grid had windows from -2.4 to 2.4 spaced 0.4Å apart for both the phosphoryl

transfer and general acid coordinates.

#### 5.5.1.11 QM/MM - computational mutagenesis

With the activation of the nucleophile by the general base being step-wise with the phosphoryl transfer, one dimensional free energy simulations of the general base step were sufficient to examine the effects of G33 mutations on the intrinsic rate. Departing from the 17 windows initially equilibrated with AM1/d-PhoT for the general base coordinate of the 2D grid, the 7-deaza, 6-thio, and 2-aminopurine mutations of G33 were made and re-equilibrated for 2 ps. 10 ps of production sampling was then completed at the PBE0/6-31G* level of theory and 1D free energy profiles were generated using vFEP. An additional half-harmonic restraint was applied to the general acid proton to match the procedure carried out for the 2D profiles. For the general acid knockout and rescue simulations, the 29 windows along the phosphoryl transfer coordinate ranging from -2.8 to 2.8 with 0.2Å spacing with a half-harmonic restraining proton transfer from the general base were generated in the same fashion as the 2D grid. The umbrella windows for each of these systems were then equilibrated with AM1/d-PhoT for 2 ps and 10 ps of production statistics using PBE0/6-31G* to describe the QM region were collected. The barriers from each of these simulations were determined from the 1D free energy profile generated by vFEP.

#### 5.5.1.12 Kinetic Isotope Effect calculations

The Twister and RNase A Kinetic Isotope Effect (KIE) values shown in Figure 5.6 are obtained by performing QM/MM geometry optimizations to yield minimum and transition state structures on the adiabatic potential energy surface using the PBE0/6-31G* QM Hamiltonian and computing the Hessians from finite differentiation of the atomic gradients (0.005 Bohr step length). The DL-Find optimization library was used to perform the geometry optimizations. For Twister, 959 atoms were included in the optimizations and Hessian calculation. 70 of the atoms are the QM region and the remainder are the MM atoms nearest the QM region. The remaining MM atoms in the system are fixed in space to their initial positions, obtained from the free energy simulation trajectory. Similarly, the RNase A optimizations included 1147 atoms (105 of which are QM). The geometry optimizations and Hessian evaluations of the non-enzymatic reaction shown in Figure 5.6 are performed with PBE0/6-31G* in PCM implicit solvent using the Gaussian 09 software.

Normal mode frequencies are obtained by diagonalizing the mass-weighted Hessian upon removal of the translational and rotational degrees of motion. The normal mode frequencies are then used in the Bigeleisen equation to evaluate the KIE value.[168]

### 5.5.2 Models for interpretation of activity-pH and p$K_a$ data

When applying the following models to the experimental activity-pH and p$K_a$ data there criteria were used to evaluate the quality of for each model that extends beyond statistical measures to give more insight into the underlying physical picture. These criteria are as follows:

1. Quality of fit - Using a statistical measure (e.g. $R^2$), how well does the model fit the experimental activity-pH profile
2. Microscopic p$K_a$s - Can the model incorporate spectroscopic measurements of microscopic p$K_a$s while fitting to the activity-pH data?
3. Mutations - Does the model respond to mutations of the ribozyme sequence, and thus different microscopic p$K_a$s of relevant residues in the same manner as the experimentally observed activity-pH profiles?

#### 5.5.2.1 Model 1: Simple, non-cooperative

The simplest and most commonly used model in interpreting activity-pH profiles is a non-cooperative model (Case 1, Frankel and Bevilacqua[136]). The non-cooperative model includes the general acid and general base as the only contributors to the pH dependence of ribozyme activity. Furthermore, this model considers the p$K_a$s of the acid and base to be independent of one another. The equation used to describe this model and fit to the experimental activity-pH profile is as follows:

$$k_{\mathrm{obs}} = k_{\mathrm{cl}} \ / \ 1 + 10^{(pK_{\mathrm{a,B}} - pH)} + 10^{(pK_{\mathrm{a,B}} - pK_{\mathrm{a,A}})} + 10^{(pH - pK_{\mathrm{a,A}})} \tag{5.3}$$

Where, $k_{obs}$ is the experimentally observed rate at each pH, $k_{cl}$ is the pH independent rate of cleavage, and the "apparent" p$K_a$s of the acid and base are p$K_a$,A and p$K_a$,B, respectively.

Fitting of this model to each of the two, currently available, experimental activity-pH profiles for 'wild-type' twister ribozymes[6,30] gives the same qualitative picture: an apparent p$K_a$ around 6.8-6.9 (assigned to the general acid) and a general base p$K_a$ at or above 9.5. While the statistical quality of fit is quite high ($R^2 = 0.9894$), this model can only characterize the apparent p$K_a$s of the acid and base and, as such, cannot incorporate the spectroscopically measured p$K_a$s into the physical picture for the catalytic mechanism of the twister ribozyme. It turns out that for the general base, the apparent p$K_a$ of 9.5 is not at odds with the NMR measurements and it is not unreasonable to assume that this apparent p$K_a$ is actually the microscopic p$K_a$ of G33. This is not the case for the general acid A1. A similiar assumption that the apparent

p$K_a$ of 6.9 is the microscopic p$K_a$ of A1:N3 would require the active site to shift the p$K_a$ of this site (from 1.5)[4] up by 5.4 units. While the active site is highly electronegative and would be expected to shift the p$K_a$ of the adenine towards neutrality, the magnitude of this shift is unreasonable especially in light of the NMR measurements. As for the final criteria, mutational studies, the non-cooperative model would respond in kind to shifts in the microscopic p$K_a$s of mutated residues if and only if it is assumed that the apparent p$K_a$s of the acid and base are the microscopic p$K_a$s for those residues - a faulty assumption.

Despite the limitations of this model, it does provide an initial estimate (Table 5.1) for both the probability of the active state in solution, as well as, the intrinsic rate of reaction. It is important to note that the free energy associated with adoption of the catalytically active state is underestimated by this model (and therefore overestimates the intrinsic reaction barrier) since it assumes the microscopic p$K_a$ of the acid is unreasonably high.

#### 5.5.2.2 Model 2: Cooperative

The cooperative model (Case 2, Frankel and Bevilacqua[136]) builds on the simple, non-cooperative model by allowing the p$K_a$s of the acid and base to couple to one another, increasing the probability of the active protonation state of each residue when the other is in its active (reverse) protonation state. The equation used to describe this model and fit to the experimental activity-pH profile is as follows:

$$k_{\mathrm{obs}} = k_{\mathrm{cl}} \ / \ 1 + 10^{(pK_{\mathrm{a,B}} - \Delta pK_{coop} - pH)} + 10^{(pK_{\mathrm{a,B}} - \Delta pK_{coop} - pK_{\mathrm{a,A}})} + 10^{(pH - pK_{\mathrm{a,A}} + \Delta pK_{coop})} \quad (5.4)$$

Where, p$K_a$s A and B are the apparent p$K_a$s of the acid and base in the background of the other residue in its standard-neutral protonation state, while $\Delta$p$K_{coop}$ is an additional free parameter that represents the cooperativity between the acid and base.

With this additional free parameter, $\Delta$p$K_{coop}$, there is a now an infinite set of solutions for this model when fitting to the experimental data. These solutions have apparent p$K_a$s that are indentical to the non-cooperative solution (6.83 and 9.58), but the pH independent rate ($k_{cl}$) can fully compensate for any value for $\Delta$p$K_{coop}$. As for how this model performs based on the criteria described above, the picture is the same as with the non-cooperative model: identical statistical quality of fit ($R^2 = 0.9894$) and neither can accomodate the NMR measurements for the p$K_a$ of A1. This model does, however, improve the physical description of the catalytic mechanism for the twister ribozyme by allowing the protonation states of the acid and base to couple with one other. The question is then, is there an estimate for the magnitude of the coupling between them?

As was discussed by Frankel and Bevilacqua[136], this attractive cooperativity between the acid and base is not detectable using the typical spectroscopic techniques for measuring p$K_a$s (e.g. the NMR measurements discussed, previously). However, the value of $\Delta$pK$_{coop}$ along with estimates for the microscopic p$K_a$s of the catalytic residues can be determined by simulating the thermodynamic cycle (Front face of Figure 5.7) that underlies this model. We used newly developed GPU-accelerated free energy methods[155,160,161] to calculate the probability of each of the four protonation states, with the ribozyme in the active conformational state (U-1 forming a triple to complete the L-platform). From this set of calculations, the predicted p$K_a$s for the general acid (A1:N3) and base (G33:N1) are 5.75 $\pm$0.23 and 9.24 $\pm$0.18, respectively and compare well to the spectroscopically measured p$K_a$s of 5.1 and $\geq$ 9.5. The coupling between these residues, $\Delta$pK$_{coop}$ was estimated to be 0.21 or 0.3 kcal/mol. This estimate for $\Delta$pK$_{coop}$ is then used as a constraint for this model and the influencer model that follows.

#### 5.5.2.3 Model 3: Conformational Influencer

One possibility for building upon the cooperative model is to include a pH dependent conformational event. For this model, the conformational event can be characterized as having just two states, one of which is 'active', while the other is not catalytically competent. In the case of the twister ribozyme, these two conformational states would correspond to the position of U-1; 'active' while in a base triple with A34/A19 and 'inactive' when extruded from its position at the heel of the L-platform. The equation used to describe this model and fit to the experimental activity-pH profile is as follows:

$$\begin{aligned} k_{\text{obs}} &= k_{\text{cl}} \ / \ Q \\ Q &= (1 + W_{AH^+/B^-,\ inactive}) \ \times \ 1 \\ &+ \ (1 + W_{AH^+/BH,\ inactive}) \ \times \ 10^{(pK_{\text{a,B}} - \Delta pK_{coop} - pH)} \\ &+ \ (1 + W_{A/BH,\ inactive}) \ \times \ 10^{(pK_{\text{a,B}} - \Delta pK_{coop} - pK_{\text{a,A}})} \\ &+ \ (1 + W_{A/B^-,\ inactive}) \ \times \ 10^{(pH - pK_{\text{a,A}} + \Delta pK_{coop})} \end{aligned} \tag{5.5}$$

Where, $W_{X/Y,\ inactive}$ is derived from the relative probability of the inactive conformational state for each protonation state of the acid (X = A or AH$^+$) and base (Y = BH or B$^-$). The additional parameters in this model necessitate additional constraints, in order to arrive at a single meaningful solution. The initial constraints and assumptions made for this model are as follows:

1. The microscopic p$K_a$ for the acid is constrained to the NMR value: p$K_a$ = 5.1

2. The microscopic p$K_a$ for the base is constrained to the apparent p$K_a$ from the WT twister activity-pH profile / an unshifted guanine: p$K_a$ = 9.5

3. The cooperativity between the acid and base is constrained to the calculated value from the simulated thermodynamic cycle: $\Delta$pK$_{coop}$ = 0.21.

4. The reaction proceeds to products only from the state where the ribozyme is both in the active protonation state and active conformational state ($_{AH^+}E_{B^-,\ active}$).

Under these same assumptions the cooperative model would produce a bell-shaped activity-pH profile as shown as the red curve in Figure 5.5a. However, by coupling a conformational event to each of the four protonation states, this expanded model can make the microscopic p$K_a$s for the acid and base appear shifted from their actual values to fit the experimental data. In general, the apparent shift towards higher p$K_a$s can be achieved when the probability of the active conformational state increases with increasing pH (Figure 5.5a, blue). However, to arrive at a unique solution for this pH-dependent conformational model a few additional constraints were needed. First, the model is insensitive to the value of $W_{AH^+/B^-,\ inactive}$ due to low probability of active protonation state at any pH (relative to the other protonation states), and thus a minimal contribution of this term to the partition function. Second, the shift in the apparent p$K_a$s depends only on the probabilities of the inactive conformation in each of the two charged states relative to that in the standard-neutral state. Therefore, two additional constraints fixing $W_{AH^+/B^-,\ inactive}$ and $W_{A/B^-,\ inactive}$ to both be zero were included. It is interesting to note that when unconstrained $W_{A/B^-,\ inactive}$ optimizes towards zero. However, the true advantage of these constraints is that by excluding inactive conformations in these two protonation states, the model provides a single solution that corresponds to maximizing the probability of the catalytically active state and thus a lower bound on the predicted intrinsic rate of reaction. Under this full set of constraints, only three free parameters remain and as such the fit for this conformational influencer model (Figure 5.5a, black) is now indistinguishable ($R^2$ = 0.9894) from that of the non-cooperative and cooperative models (each model having exactly three free parameters).

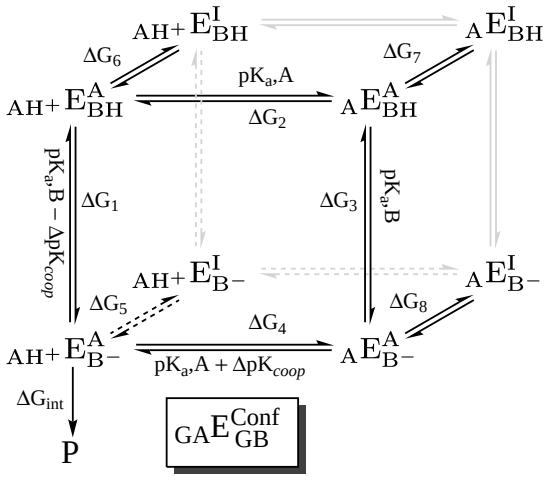



Figure 5.7: Thermodynamic cycle for the conformational influencer model, where the front face of the cube corresponds to the cooperative model. At each vertex of the cube the protonation state of the general acid and base are shown as subscripts, while the superscript indicates whether the ribozyme is in either the active (A) or inactive (I) conformational state.

Table 5.1: Parameters for and free energies derived from fitting each of the three models to the experimental activity-pH data.[6]

| | Non-coopertative | | Cooperative | | Conformational Influencer | |
|---|---|---|---|---|---|---|
| Fixed Parameters | ... | | $\Delta pK_{coop}$ | 0.21 | $pK_a$,a | 5.1 |
| | | | | | $pK_a$,b | 9.5 |
| | | | | | $\Delta pK_{coop}$ | 0.21 |
| | | | | | $\Delta G_5$ | $\infty$ |
| | | | | | $\Delta G_8$ | $\infty$ |
| Free Parameters | $pK_a$,a | 6.83 | $pK_a$,a | 6.83 | $\Delta G_6$ | -3.05 kcal/mol |
| | $pK_a$,b | 9.58 | $pK_a$,b | 9.58 | $\Delta G_7$ | -0.46 kcal/mol |
| | $k_{cl}$ | 2400 $min^{-1}$ | $k_{cl}$ | 1500 $min^{-1}$ | $k_{cl}$ | 210,000 $min^{-1}$ |
| $\Delta G_{active}$ (kcal/mol) | | 3.75 | | 3.46 | | $\geq$ 6.18 |
| $\Delta G_{int}$ (kcal/mol) | | 15.26 | | 15.55 | | $\leq$ 12.61 |

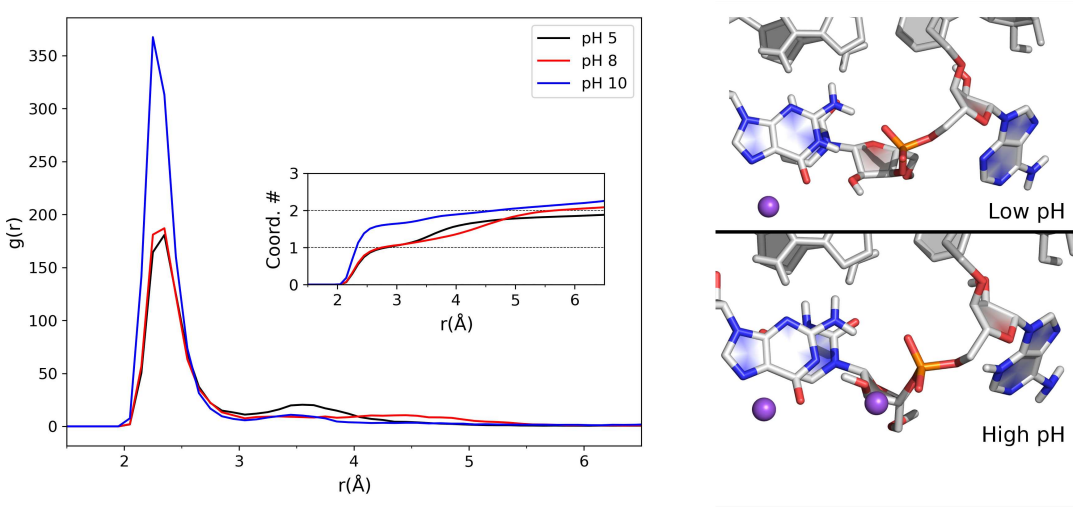



Figure 5.8: Radial distribution function and representative snapshots of the active site ion environment as a function of pH, from simulated thermodynamic cycle corresponding to the cooperative activity-pH model. The active state in solution used as the departure point for the QM/MM free energy calculations contains ions at both the Hoogsteen edge of G33 and bridging the O2′ nucleophile and pro-$S_{\mathrm{P}}$ NPO, as seen in the high pH snapshot.

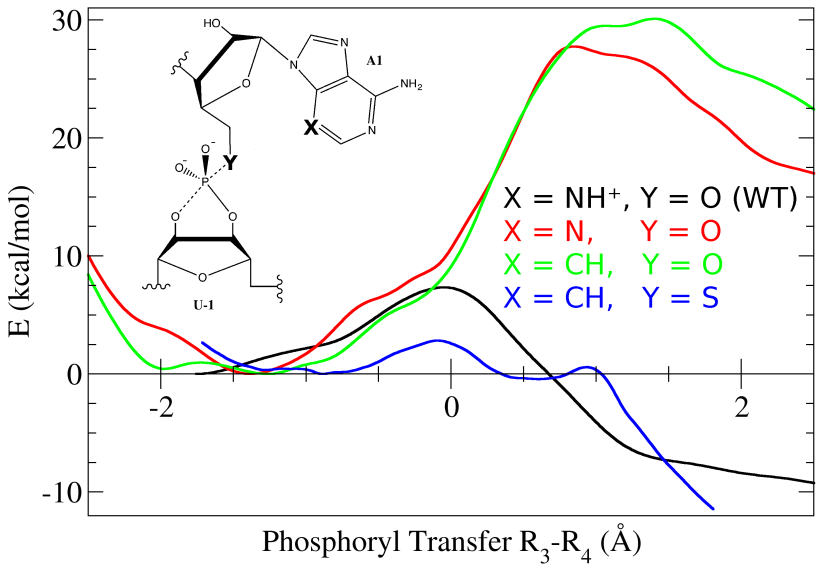



Figure 5.9: 1D QM/MM free energy calculations of the phosphoryl transfer step testing a variety of mechanistic pathways probing the role of A1:N3 as the general acid. Knockout of this site either by deprotonation (X = N, red) or by deaza substitution (X=CH, green) results in a predicted total loss of activity. However, these calculations predict that an enhanced, 5′ thio, leaving group (X = CH / Y = S, blue) would have a significant rescue effect and other processes along the reaction path would likely become rate limiting.

Table 5.2: Results from the conformational influencer model and QM/MM free energy calculations. The general reaction scheme is considered here, where "E", "S" and "P" represent the enzyme, substrate, and product, respectively, the subscripts "u" and "f" represent unfolded and folded states, respectively. The bounds on the free energies for adopting the catalytically active state ($\Delta G_{active}$) and the intrinsic rate of reaction ($\Delta G_{int}$) are derived from fitting of the conformational influencer model to the experimental activity-pH data. This lower bound on the free energy for adopting the $ES_r^*$ state is then taken as the starting point for the QM/MM calculations. The QM/MM estimate for $\Delta G_{int}$ assumes the general base proton transfer ($ES_r^* \rightleftharpoons ES_{AP}$) is a pre-equilibrium step.

$$E + \mathrm{S} \xrightleftharpoons{\Delta G_b} \mathrm{ES_u} \xrightleftharpoons{\Delta G_f} ES_f \xrightleftharpoons{\Delta G_{active}} \mathrm{ES_r^*} \rightarrow [ES]_1^\ddagger \longrightarrow \mathrm{ES_{AP}} \rightarrow [ES]_2^\ddagger \rightarrow EP \xrightleftharpoons{\Delta G_{b,p}} \mathrm{E + P}$$

| Activity-pH | | $\Delta G$ (kcal/mol) |
|---|---|---|
| $ES_f \rightleftharpoons ES_r^*$ | $(\Delta G_{active})$ | $\geq$ **6.18** |
| $ES_r^* \rightarrow [ES]_2^\ddagger$ | $(\Delta G_{int})$ | $\leq$ **12.61** |

| QM/MM | $R_1$ - $R_2$ Å | $R_3$ - $R_4$ Å | $R_5$ - $R_6$ Å | G kcal/mol | $\Delta G_{\mathrm{fwd}}$ kcal/mol | $\Delta G_{\mathrm{rev}}$ kcal/mol |
|---|---|---|---|---|---|---|
| $ES_r^*$ | -0.58 | -1.19 | -0.83 | **6.18** | $\cdots$ | $\cdots$ |
| $[ES]_1^\ddagger$ | 0.05 | -1.21 | -0.83 | 8.65 | 2.47 | 0.18 |
| $ES_{AP}$ | 0.27 | -1.21 | -0.83 | 8.47 | $\cdots$ | $\cdots$ |
| $[ES]_2^\ddagger$ | 0.27 | 0.18 | -0.05 | 15.81 | 7.34 | $>$ 17.38 |
| $EP$ | 0.27 | 2.57 | 0.86 | $<$ -1.57 | $\cdots$ | $\cdots$ |
| $ES_r^* \rightarrow [ES]_2^\ddagger$ | $(\Delta G_{int})$ | | | | **9.63** | |

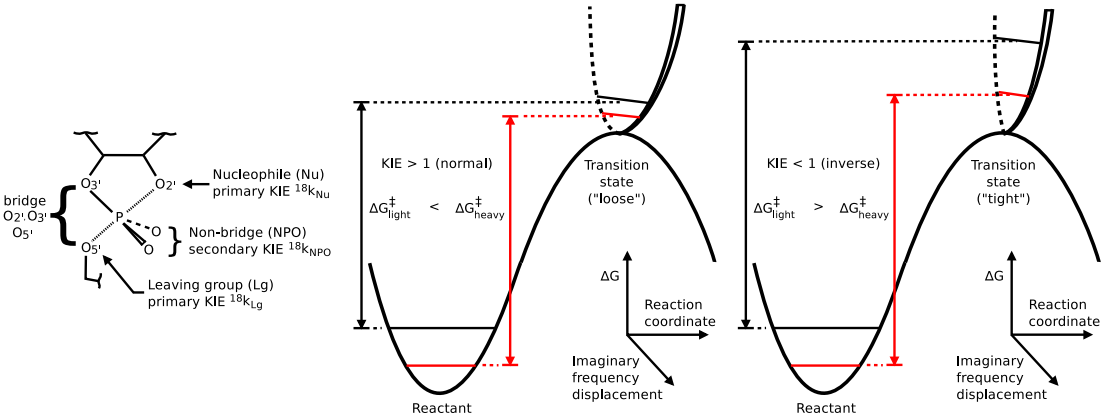



Figure 5.10: Schematic of the reaction site and relationship between kinetic isotope values with the character of the transition state.

# Chapter 6

# The L-platform/L-scaffold Framework: A Blueprint for RNA-cleaving Nucleic Acid Enzyme Design

Reprinted with permission from Colin S. Gaines, Joe A. Picirilli, and Darrin M. York. The L-platform/L-scaffold framework: a blueprint for RNA-cleaving nucleic acid enzyme design. *RNA*, 26:111–125, 2020. 

We develop an L-platform/L-scaffold framework we hypothesize may serve as a blueprint to facilitate site-specific RNA-cleaving nucleic acid enzyme design. Building on the L-platform motif originally described by Suslov and coworkers, we identify new critical scaffolding elements required to anchor a conserved general base guanine ("L-anchor") and bind functionally important metal ions at the active site ("L-pocket"). Molecular simulations, together with a broad range of experimental structural and functional data, connect the L-platform/L-scaffold elements to necessary and sufficient conditions for catalytic activity. We demonstrate that the L-platform/L-scaffold framework is common to five of the nine currently known naturally occurring ribozyme classes (Twr, HPr, VSr, HHr, Psr), and intriguingly from a design perspective, the framework also appears in an artificially engineered DNAzyme (8-17dz). The flexibility of the L-platform/L-scaffold framework is illustrated on these systems, highlighting modularity and trends in the variety of known general acid moieties that are supported. These trends give rise to two distinct catalytic paradigms, building on the classifications proposed by Wilson and coworkers and named for the implicated general base and acid. The "G+A" paradigm (Twr, HPr, VSr) exclusively utilizes nucleobase residues for chemistry, and the "G+$M^+$" paradigm (HHr, 8-17dz, Psr) involves structuring of the "L-pocket" metal ion binding site for recruitment of a divalent metal ion that plays an active role in the chemical steps of the reaction. Finally, the modularity of the L-platform/L-scaffold framework is illustrated in the VS ribozyme where the "L-pocket" assumes the functional role of the "L-anchor" element, highlighting a distinct mechanism, but one that is functionally linked with the hammerhead ribozyme.

## 6.1 Introduction

Nucleolytic ribozymes are small catalytic RNAs that site-specifically cleave the sugar-phosphate backbone of their RNA substrates[56,57,81,99,169,170]. This reaction is ubiquitous in biology[171] and has importance for biochemical tools and biomedical technology[51,55,60,63,172,173]. Currently, there are nine known naturally occurring nucleolytic ribozyme classes: hammerhead (HHr)[115–117], hairpin (HPr)[118,119], hepatitis delta virus (HDVr)[120,121], Varkud satellite (VSr)[122,123], *glmS*[124,125], twister (Twr)[6,30], twister sister or simply "TS" (TSr)[7,31], pistol[31,74] (Psr), and hatchet (Htr)[31] ribozymes. However, molecules of DNA can also be artificially engineered to catalyze RNA cleavage (e.g. 8-17 DNAzyme or 8-17dz)[127,174], and these DNAzymes have some potential advantages owing to their greater stability and ease of synthesis[175]. A broad understanding of the detailed mechanisms of these nucleic acid enzymes provides a foundation from which general design principles may emerge that are transferable to nonbiological contexts[25,47,56,57,170]. A novel comparative analysis of nucleolytic ribozyme active sites has lead to new insight into the diverse array of catalytic strategies employed by these systems[114] and has motivated closer examination of the platforms that form their catalytic cores.

Recently, Suslov et al. identified a common active site motif, designated the "L-platform", in the VSr, HPr, and HHr ribozymes[123] based on structural data available at the time. This four-residue motif involved base stacking interactions that sandwich a strictly conserved guanine and position it adjacent to the scissile phosphate where it is poised to act as a general base in catalyzing the 2′-O-transphosphorylation reaction. As originally described in the context of the VSr, this stem within the L-platform is stabilized by a purine-purine base pair at the base of the "L".

However, the "L" alone is insufficient to establish a catalytic core for these ribozymes. Further details about how this motif could be expanded to describe the full active site and potentially serve as a generalizable platform to facilitate design are difficult to derive solely from crystallographic data, as this data is usually not sufficient to unambiguously identify the catalytically active states of the enzymes[14,16,19,20,176]. Rather, rigorous computational simulations together with structural and functional data are required to define a meaningful dynamical ensemble for the active state in solution, as well as other states along the reaction coordinate. It is from these experimentally motivated computational studies combined with a comparative view across known ribozyme classes that predictive insight can be gleaned.

Recently, such studies have been carried out by our group for a number of ribozyme classes, including HHr[17,18], HPr[15], HDVr[177], VSr[16], Twr[14], Psr[20], and TSr[176], as well as 8-17dz[19].

These studies have provided detailed atomic-level models of the functionally active state, not apparent from the crystal structures, that explain a wide range of experimental data and have enabled identification of common features and new design principles to emerge. Herein we identify new functionally critical scaffolding elements of the active site architecture that build on the L-platform motif introduced by Suslov et al.[123]. We refer collectively to these elements as the L-scaffold, and provide an in-depth characterization of this new composite L-platform/L-scaffold motif (Figure 6.1). Additionally, we demonstrate that this motif is common to over half of the currently known naturally occurring ribozyme classes (Twr, HPr, VSr, HHr, Psr), as well as a recently structurally characterized artificially engineered DNAzyme (8-17dz). We also provide details about the structural features of the L-platform/L-scaffold motif and connect them to the necessary and sufficient conditions for catalytic activity.

The two main structural elements of the L-scaffold are referred to as the "L-anchor" and "L-pocket" (Figure 6.1). The L-anchor nucleobase functions to position a conserved G (implicated as the general base) to activate the 2′-OH nucleophile through the interaction of the Hoogsteen edge of the L-anchor with the sugar edge of the G. The L-pocket, on the other hand, enables formation of a divalent metal ion binding site. We describe distinguishing features of the L-platform/L-scaffold framework that give rise to two distinct catalytic paradigms (expanding the classification proposed by Wilson et al.[26]) named for the identities of the general base and acid implicated in catalysis. The "G+A" paradigm[26] (Twr, HPr, VSr) exclusively utilizes nucleobase functional groups for chemistry, and by analogy what we will refer to as the "G+$M^+$" paradigm (HHr, 8-17dz, Psr) involves structuring of the L-pocket for recruitment of a divalent metal ion that plays an active role in the chemical steps of the reaction. We also illustrate how the L-platform/L-scaffold framework enables cross-over of themes associated with each paradigm, such as in the case of VSr that uses the L-pocket within the G+A paradigm to recruit a divalent metal ion that substitutes for the missing nucleobase L-anchor and assists in organizing the active site (without directly being involved in chemistry as in the G+$M^+$ paradigm)[16].

We begin by establishing notation to define the prototypical L-platform/L-scaffold framework, and the specific requirements for nucleobase stacking and hydrogen bonding interactions. We then describe distinguishing features of the G+A and G+$M^+$ paradigms, including nucleobase requirements for the L-pocket, and present specific examples for each paradigm. Finally, we provide a comprehensive summary of the mechanisms whereby the L-platform/L-scaffold enables different catalytic strategies, following a recently introduced ontology for discussion of RNA cleavage reactions[68].

## 6.2 Notation and Conventions

In describing the L-platform/L-scaffold composite motif, we will utilize standard IUPAC nucleic acid nomenclature and ambiguity symbols, as well as abbreviations for edge-to-edge base pairing families taken from the works of Leontis et al.[12,13]. In the prototypical L-platform/L-scaffold (Figure 6.1) the motif is shown with two strands, the substrate strand and the general base strand. The nucleotide residue positions are then numbered from -1 to 3 for the substrate strand (with the scissile phosphate as position zero) and continue with 4 through 6 for the general base strand moving in the 5′ to 3′ direction. This general motif numbering scheme is intended to remove the L-platform/L-scaffold motif from the context of any individual ribozyme class and then number it in a way that allows one to refer to specific nucleotides by their relative positions in the motif and corresponding functional roles. In actuality, the consecutiveness of the bases and number of strands recruited to form this active site motif is variable, although those variations often follow discernible trends. Throughout the text, nucleotide residues will generally be referred to by their motif positions with additional ribozyme specific numbering in parentheses when necessary, and atomic numbering will be indicated by subscripts for distinction. For example, "$N_1$" will indicate a nitrogen atom at atomic position 1 of a nucleotide, whereas "N1" refers to a nucleotide residue in position 1 of the general motif and "N1 (C1.1 in HHr)" indicates N1 in the general motif that corresponds specifically to C1.1 in HHr.

In describing the connection between the L-platform/L-scaffold motif and the specific catalytic strategies enabled by it, we will use a simplified framework originally proposed by Emilsson et al.[1] and recently expanded to an ontology for facilitating discussion of mechanisms of RNA-cleaving enzymes in precise atomic-level detail[68]. For the underlying chemical reaction catalyzed by these ribozymes, a 2′-OH nucleophile becomes activated and makes an in-line attack on the adjacent scissile phosphate, which then proceeds through a pentavalent phosphorane transition state to form 2′,3′-cyclic phosphate and 5′-OH cleavage products. These general catalytic strategies to facilitate RNA cleavage are designated alpha ($\alpha$) for the in-line fitness of the 2′-OH nucleophile, beta ($\beta$) for the electrostatic stabilization/protonation of the non-bridging phosphoryl oxygens (NPOs) of the dianionic phosphorane transition state, gamma ($\gamma$) for the deprotonation (activation) of the 2′-OH nucleophile (e.g., by a general base), and delta ($\delta$) for the electrostatic stabilization (neutralization/protonation) of the 5′-O leaving group (e.g., by a general acid). The $\beta$, $\gamma$ and $\delta$ strategies can be further decomposed into primary (1°), secondary (2°), and tertiary (3°) contributions[68].

## 6.3 General Definition of the L-platform/L-scaffold Motif

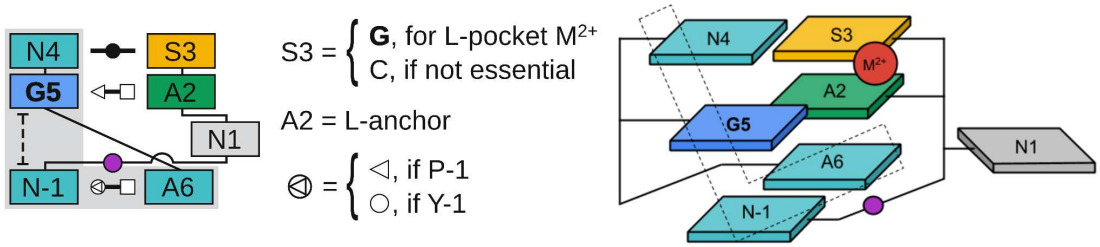



Figure 6.1: Generalized L-platform/L-scaffold composite motif. Symbolic secondary structure (left) and cartoon-block schematic (right) representations of the generalized L-platform/L-scaffold. Residue numbering and base pair symbols[12,13] are described in the Notation section. Strict conservation of nucleobase identity (e.g. for the general base guanine G5) is shown with bold font. Other nucleotides are shown with ambiguity characters representing apparent trends in preferred nucleobase identity across the ribozyme classes (e.g. N1, A2, and S3). The constituent elements of the L-platform (Left: gray background and Right: dotted 'L') motif are colored as follows: General base (G5) in blue, stacking/pairing 'L' nucleotides (N-1, N4 and A6) in teal. The L-scaffold consists of the L-anchor (A2) in green, L-pocket nucleotide (S3) in yellow, and the scissile phosphate in magenta along with the interactions between those residues and the L-platform. The bulged N1 nucleotide is not colored as it does not generally play a specific functional role (with the exception of twister ribozymes).

Here, we define the generalized prototype L-platform/L-scaffold motif as a series of three base pairs, with one bulged residue, organizing two or more strands into a compact catalytic core shown in Figure 6.1. A conserved guanine, G5, acts as the general base in catalysis and is sandwiched in the middle of a three base stack by the N-1 and N4 nucleotides. The N-1 nucleobase (5′ of the scissile phosphate) then sits at the heel of the eponymous "L" and pairs with the Hoogsteen edge of, typically, an adenine denoted as A6. The N1 nucleotide, immediately downstream of the scissile phosphate, is bulged from the motif in the sense that it is not explicitly involved in base pairing within this three base pair stack, independent of consecutiveness. The general base, G5, is then hydrogen bonding along its sugar edge, most commonly in a tSH base pair with A2. Since this interaction is critical for anchoring the central residue of the "L", residue 2 in the prototypical motif will be referred to as the L-anchor.

Finally, the motif is capped by a cWW base pair (or wobble pair) between S3 and N4. For

the ribozymes where a divalent metal ion is critical for function, S3 is conserved as a guanine. The Hoogsteen edge of this guanine then completes a cation binding pocket in conjunction with one of the NPOs of the scissile phosphate as well as one 5′ of the L-anchor (residue 2). This set of ligands defines the L-pocket and consistently promotes binding of a divalent metal ion within the active site. In contrast, the ribozymes (Twr and HPr) for which divalent metals are not required for catalysis, cytosine at position 3 is the preferred nucleobase. It is interesting to note that this is the only nucleobase that cannot bind a metal along its Hoogsteen edge. These trends lead us to characterize nucleotide 3 as S in the general motif, as a means to highlight how the base identity is correlated with the requirement for divalent metal ions in each of the ribozyme classes. In the next sections (Sections 6.3.1 and 6.3.2) we discuss how the L-platform/L-scaffold motif can enable RNA cleavage within two catalytic paradigms designated G+A and G+M$^{+}$ [26], named for the identities of the general base (guanine) and acid (adenine or a metal ion and/or its ligands).

### 6.3.1 G+A Catalytic Paradigm

The Twr, HPr and VSr fit the G+A paradigm and have active sites that exhibit remarkable structural similarities (Figure 6.2). Ribozymes in the G+A paradigm utilize the L-platform/L-scaffold framework as an active site architecture to enable the chemical steps of catalysis to be carried out exclusively by nucleobase functional groups. In addition to the strictly conserved G5 that acts as a general base ($\gamma$ catalysis), a conserved adenine residue acts as a general acid ($\delta$ catalysis), either through the $N_3$ (Twr) or the $N_1$ (HPr and VSr) positions. In order for the adenine to act as an effective general acid through either the $N_1$ or $N_3$ heteroatoms (1° $\delta$ catalysis), it must have a p$K_a$ that is appropriately tuned (2° $\delta$ catalysis), and it must be properly positioned for proton donation to the O5′ leaving group (3° $\delta$ catalysis).

### 6.3.2 G+M$^{+}$ Catalytic Paradigm

The HHr, 8-17dz and Psr enable catalytic strategies within the G+M$^{+}$ paradigm and have similar active site architectures and divalent metal ion binding modes in their active states (Figure 6.3). Ribozymes grouped in the G+M$^{+}$ paradigm all require an L-pocket metal ion binding site that recruits a divalent metal ion to assist in catalysis. The divalent metal ion can aid in stabilizing the O5′ leaving group by providing direct electrostatic stabilization as a Lewis acid (1° $\delta$ catalysis), or as a general acid acting through p$K_a$ tuning (activation) of a coordinated water molecule or functional group such as an 2′-OH (2° $\delta$ catalysis). Finally, the

L-pocket metal ion can play a structural role in bringing the required elements for $\delta$ catalysis together, orienting the scissile phosphate and positioning the O5′ leaving group (3° $\delta$ catalysis).

### 6.3.3 Canonical Sequences

The canonical sequences for each ribozyme class, as depicted in the symbolic secondary structure diagrams (Figures. 6.2 and 6.3, left) were derived from bioinformatics, when available, or through evaluating mutation and in vitro selection experiment results. The canonical sequences are taken from sequence alignments for Twr[30], HHr[178], and Psr[31], whereas the 8-17 DNAzyme sequence, presented in the original crystallographic work[174], was adapted from commonly observed variations and mutagenesis experiments. In a similar fashion, an analysis of mutational and in vitro selection experiments was conducted for HPr and VSr in order to arrive at a canonical sequence for the active site residues.

For VSr, mutations that resulted in a decrease in activity greater than 100-fold[179,180], were rejected from the canonical sequence at that position. From the in vitro selection experiments, all nucleobases for which there was more than one sequence that exhibited high activity, that is, $> 50\%$ activity in the cleavage assay[181], were included in the canonical sequence. Similar criteria were applied in developing the canonical sequence for HPr. However, there are two nucleotides in the HPr active site for which there is not a strict requirement for a specific nucleobase, the L-anchor (N2) and L-pocket (N3) nucleotides. The mutational data nonetheless suggest that there are clear, but subtle trends for the preferred nucleobases: U2 and C3 depicted in Figure 6.2. For the L-anchor nucleotide, uracil (the "wild-type" residue) is fastest, followed by adenine (24-fold lower in activity); guanine and cytosine were both observed as 73-fold down in activity[182]. However, Perez-Ruiz et al. suggest that sequences with cytosine at the N2 position were non-cleavable in their assay[183], which would be consistent with the proposed tHS base pairing between the L-anchor and the general base guanine[13]. For the L-pocket nucleobase in HPr, all nucleobases are tolerated (decrease in activity $<$ 10-fold in all cases). However, the observed rate was negatively correlated with the ability for this nucleobase to bind a $Mg^{2+}$ ion at its Hoogsteen edge (G:$N_7$ $>$ A:$N_7$ $>$ U:$O_4$ $>$ C:"C-H" edge)[184], motivating our decision to suggest cytosine in the canonical sequence.

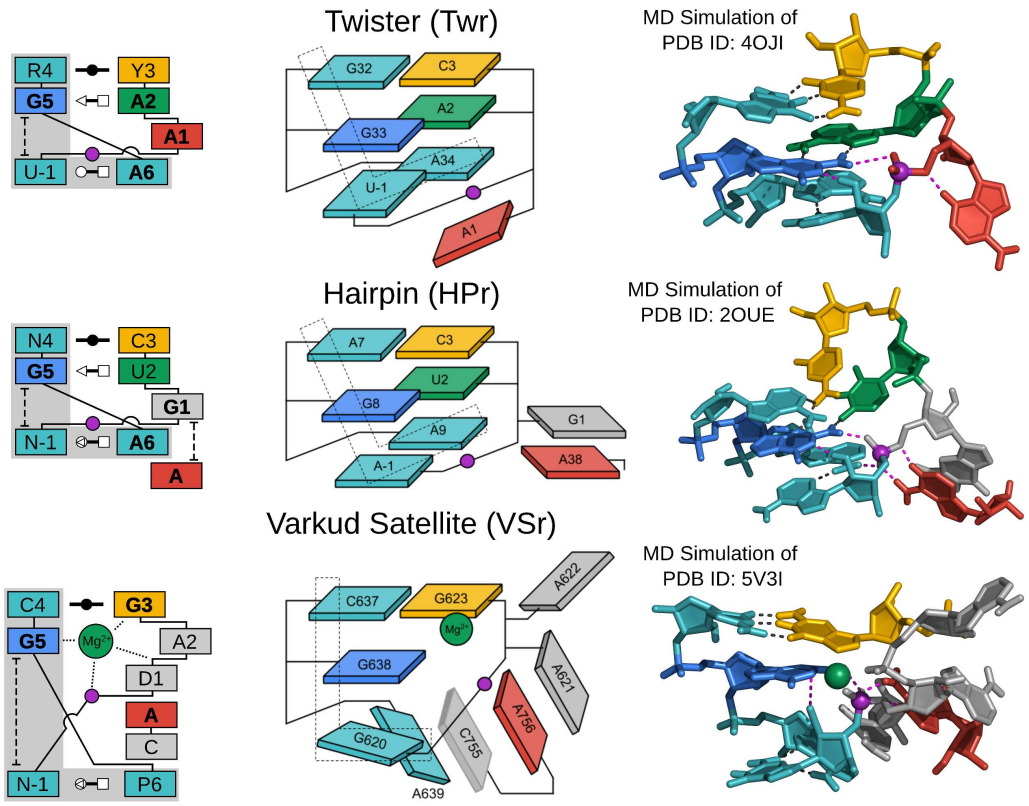



Figure 6.2: G+A Paradigm. Symbolic secondary structure (left), cartoon-block schematic (middle) and 3D atomic (right) representations of the specific L-platform/L-scaffold composite motif for ribozymes of the "G+A" paradigm (Twr, HPr and VSr), categorized as such with a guanine (G in blue) and adenine (A in red) implicated as the general base and acid, respectively. Generalized nucleotide residue numbering is used for the secondary structure (left) as in Fig. 6.1, whereas the cartoon-block schematic (middle) uses the canonical residue numbering of the specific ribozyme being illustrated. Base pair symbols[12,13] are described in the Notation section, and the color scheme is the same as in Fig. 6.1: General base (G5) in blue, stacking/pairing 'L' residues in teal, L-anchor in green, L-pocket nucleobase in yellow, and the scissile phosphate in magenta. Structural h-bonds are in black and h-bonds implicated in the catalytic mechanism are shown in magenta. The active site metal ion in VSr that plays an organizational role as the L-anchor is shown in green. Bold font is used for residues conserved with a frequency greater than 97% for ribozymes where consensus sequence information is available, as well as residues that are otherwise critical for activity. 3D atomic representations derived from MD simulations of each ribozyme: Twr[14], HPr[15], and VSr[16].

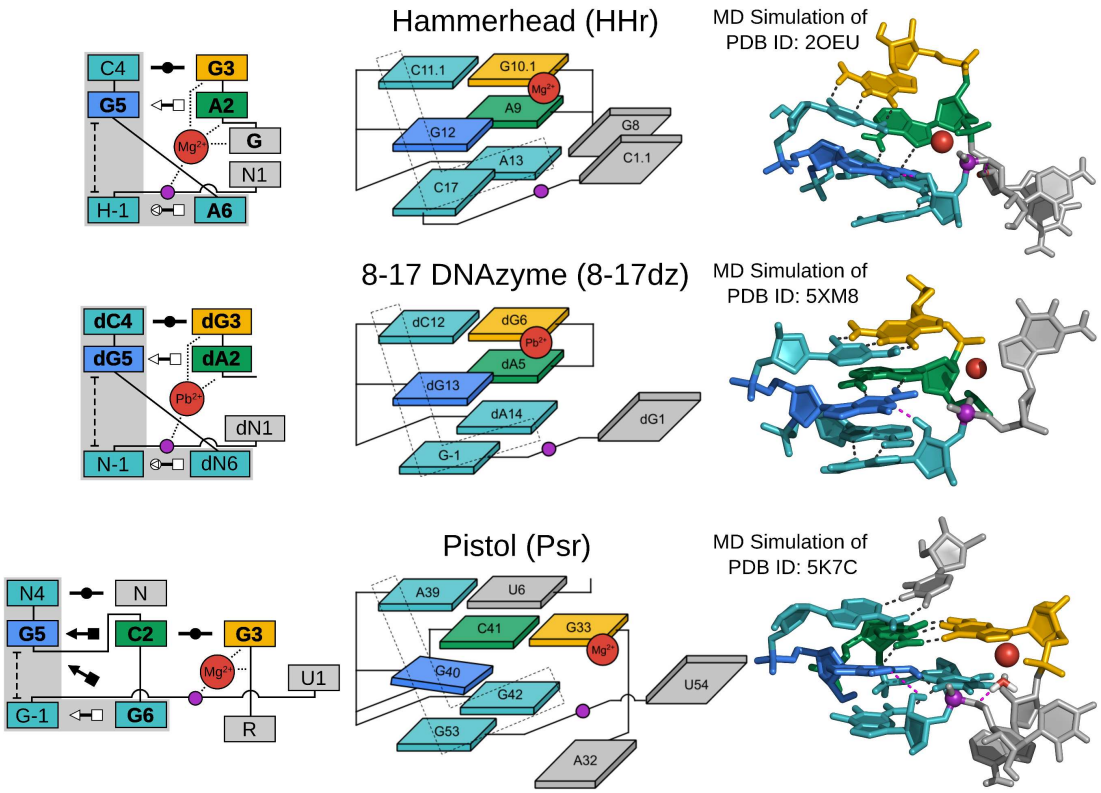



Figure 6.3: G+M$^+$ Paradigm. Symbolic secondary structure (left), cartoon-block schematic (middle) and 3D atomic (right) representations of the specific L-platform/L-scaffold composite motif for ribozymes and DNAzyme of the "G+M$^+$" paradigm (HHr, 8-17dz and Psr), categorized as such with a guanine (G in blue) and a metal ion (M$^+$ in red) implicated as the general base and acid, respectively. Generalized nucleotide residue numbering is used for the secondary structure (left) as in Fig. 6.1, whereas the cartoon-block schematic (middle) uses the canonical residue numbering of the specific ribozyme/DNAzyme being illustrated. Base pair symbols[12,13] are described in the Notation section, and the color scheme is the same as in Fig. 6.1: General base (G5/dG5) in blue, stacking/pairing 'L' residues in teal, L-anchor in green, L-pocket nucleobase in yellow, and the scissile phosphate in magenta. Structural h-bonds are in black and h-bonds implicated in the catalytic mechanism are shown in magenta. The active site metal ion that plays a role in the chemical steps of the reaction is shown in red. Bold font is used for residues conserved with a frequency greater than 97% for ribozymes where consensus sequence information is available, as well as residues that are otherwise critical for activity. 3D atomic representations derived from MD simulations of each ribozyme: HHr[17,18], 8-17dz[19], and Psr[20].

## 6.4 Survey of Catalytic Strategies Enabled by the L-platform/L-scaffold in RNA and DNA Enzymes

In the previous sections, we introduced the prototypical L-platform/L-scaffold composite motif and the two catalytic paradigms that it supports. Now we will illustrate the tremendous versatility of this motif as a framework for design by providing a comprehensive survey of how these catalytic strategies are exploited to facilitate site-specific RNA cleavage via 2′-O-transphosphorylation in five known ribozyme classes as well as an engineered DNAzyme. Tables 6.1 and 6.2 list the specific positions and residue interactions that are implicated in each of the catalytic strategies. These are now summarized and discussed within the context of design principles for each of the ribozymes and DNAzyme in Figures 6.2 and 6.3.

### 6.4.1 $\alpha$ Catalysis

In-line alignment of the nucleophile, scissile phosphate, and leaving group is highly correlated with the splaying apart of the two nucleobases flanking the scissile phosphate, N-1 and N1. In the L-platform motif, the N-1 nucleotide is positioned at the heel of the "L" by stacking with the general base, G5, as well as pairing with the nucleobase at position 6 at the toe of the "L". In the G+A paradigm where exactly two strands form the active site, N1 of the L-scaffold is bulged from the motif by interactions between the L-anchor (residue 2 or the L-pocket $Mg^{2+}$ in VSr) and the general base. The N1 nucleotide is similarly bulged from the motif in the G+$M^+$ paradigm where additional strands are recruited to position residues 2 and 3. In both paradigms, residue 1 is then involved in ribozyme specific tertiary interactions in order to achieve the requisite splay in conjunction with N-1.

While it has been suggested that ideal in-line alignment alone is unlikely to provide more than a 100-fold increase in the catalytic rate[1] relative to typical non-catalytic RNAs, constraining the positions of the flanking bases serves not only to improve in-line fitness, but also to localize the reactive atoms within the active site. This is further facilitated by interaction of the general base exocyclic amine, G5:$N_2$, with an NPO of the scissile phosphate in the G+A paradigm and Psr. Whereas, for HHr and 8-17dz, this same exocyclic amine may instead interact with the nucleophile prolonging the lifetime of its deprotonated form, while helping to align it with the scissile phosphate. As will prove to be a common trend with many of the components of the L-platform/L-scaffold motif, this functional group (G5:$N_2$) likely contributes to multiple catalytic strategies and will be discussed in greater detail throughout the following sections. Finally, it is interesting to note that while the identities of neither N-1 nor N1 are conserved

| Catalytic Strategy | Function | G+A | | | G+M$^+$ | | |
|---|---|---|---|---|---|---|---|
| | | Twr | HPr | VSr | HHr | 8-17 dz | Psr |
| $\alpha$ | Splay | U-1 tWH A34<br>bulged A1 | A-1 tSH A9<br>bulged G1 | G620 tSH A639<br>bulged A621 | C17 tWH A13<br>bulged C1.1 | G-1 tSH dA14<br>bulged dG1 | G53 tSH G42<br>bulged U54 |
| | GB...NPO | G33:N2...A1:pro-$R_p$ | G8:N2...G1:pro-$S_p$ | G638:N2...A621:pro-$R_p$ | | | G40:N2...U54:pro-$R_p$ |
| | GB...Nuc | | | | G12:N2...C17:O2′ | dG13:N2...G-1:O2′ | |
| 1°$\gamma$ | General base | G33:N1$^-$ | G8:N1$^-$ | G638:N1$^-$ | G12:N1$^-$ | dG13:N1$^-$ | G40:N1$^-$ |
| 2°$\gamma$ | p$K_a$ base | M$^+$ ∗ G33:H-edge | M$^+$ ∗ G8:H-edge | Mg$^{2+}$ ∘ G638:O6 | M$^+$ ∗ G12:H-edge | M$^+$ ∗ dG13:H-edge | M$^+$ ∗ G40:H-edge |
| | p$K_a$ O2′ | M$^+$ ∗ U-1:O2′ | M$^+$ ∗ A-1:O2′ | Mg$^{2+}$ ∘ G620:O2′ | M$^+$ ∗ C17:O2′<br>G12:N2...C17:O2′ | M$^+$ ∗ G-1:O2′<br>dG13:N2...G-1:O2′ | M$^+$ ∗ G53:O2′ |
| 3°$\gamma$ | L-anchor | G33 tSH A2 | G8 tSH U2 | Mg$^{2+}$ ∘ G638:O6 | G12 tSH A9 | dG13 tSH dA5 | G40 cSH C41 and/or<br>G40 cSH G42 |
| | L-pocket<br>(binds anchor) | | | Mg$^{2+}$ • A621:pro-$S_p$<br>Mg$^{2+}$ • A622:pro-$R_p$<br>Mg$^{2+}$ • G623:H-edge | | | |
| | GB...NPO | G33:N2...A1:pro-$R_p$ | G8:N2...G1:pro-$S_p$ | G638:N2...A621:pro-$R_p$ | | | G40:N2...U54:pro-$R_p$ |
| | | | | | G12:N2...C17:O2′ | dG13:N2...G-1:O2′ | |
| | O2′ position | U-1 tWH A34<br>U-1 stack with G33 | A-1 tSH A9<br>A-1 stack with G8 | G620 tSH A639<br>G620 stack with G638 | C17 tWH A13<br>C17 stack with G12 | G-1 tSH dA14<br>G-1 stack with dG13 | G53 tSH G42<br>G53 stack with G40 |

Table 6.1: The catalytic strategies and functional role of each interaction presented in this table are described in greater detail in the corresponding sections (e.g $\alpha$ Catalysis) of the main text. Nucleobase numbering follows the ribozyme class specific number schemes (Figs. 6.2 and 6.3, middle). For metal ion binding, ∗ indicates territorial/transient binding, ∘ indicates outer-sphere coordination, and • indicates direct/inner sphere coordination. Base pair abbreviations are derived from the edge-to-edge pairing families. Hydrogen bonding between donor 'D' of residue X and acceptor 'A' of residue Y is indicated by X:D...Y:A. Interactions described in this table for the $\alpha$ and $\gamma$ catalytic strategies are derived from molecular simulation models of the active state in solution for each ribozyme.

| Catalytic Strategy | Function | G+A | | | G+$M^+$ | | |
|---|---|---|---|---|---|---|---|
| | | Twr | HPr | VSr | HHr | 8-17 dz | Psr |
| 1°$\beta$ | Ion | | | $Mg^{2+}$ • A621:pro-$S_p$ | $Mg^{2+}$ • C1.1:pro-$R_p$ | $Pb^{2+}$ • dG1:pro-$R_p$ | |
| 2°$\beta$ | Ion | $M^+$ ∗ A1:pro-$S_p$ | | | | | $Mg^{2+}$ ∘ U54:pro-$R_p$ |
| | H-bonding | G33:N2...A1:pro-$R_p$ | G8*:N2...G1:pro-$S_p$<br>A38:N6...G1:pro-$R_p$<br>A9:N6...G1:pro-$R_p$ | G638:N2...A621:pro-$R_p$<br>A756:N6...A621:pro-$R_p$ | | | G40:N2...U54:pro-$R_p$ |
| 1°$\delta$ | General acid | A1:N3$H^+$ | A38:N1$H^+$ | A756:N1$H^+$ | $Mg^{2+}$/$Mg^{2+}$ • G8:O2′ | $Pb^{2+}$/$Pb^{2+}$ • $OH_2$ | $Mg^{2+}$ • $OH_2$ |
| 2°$\delta$ | p$K_a$ acid | A1:N6...C16/C17:NPOs | A38:N6...G1:pro-$R_p$ | A756:N6...A621:pro-$R_p$ | | | |
| 3°$\delta$ | Acid anchor | A1:N6...C16/C17:NPOs<br>A1 $\chi$ syn | A38:N6...G1:pro-$R_p$<br>A38 stack with G1 | A756:N6...A621:pro-$R_p$<br>A756 stack with A621 | | | |
| | L-pocket | | | | $Mg^{2+}$ • C1.1:pro-$R_p$<br>$Mg^{2+}$ • A9:pro-$R_p$<br>$Mg^{2+}$ ∘ G10.1:H-edge | $Pb^{2+}$ • dG1:pro-$R_p$<br>$Pb^{2+}$ ∘ dA5:pro-$R_p$<br>$Pb^{2+}$ ∘ dG6:H-edge | $Mg^{2+}$ ∘ U54:pro-$R_p$<br>$Mg^{2+}$ ∘ G33:pro-$R_p$<br>$Mg^{2+}$ • G33:H-edge |

Table 6.2: The catalytic strategies and functional role of each interaction presented in this table are described in greater detail in the corresponding sections (e.g $\alpha$ Catalysis) of the main text. Nucleobase numbering follows the ribozyme class specific number schemes (Figs. 6.2 and 6.3, middle). For metal ion binding, ∗ indicates territorial/transient binding, ∘ indicates outer-sphere coordination, and • indicates direct/inner sphere coordination. Base pair abbreviations are derived from the edge-to-edge pairing families. Hydrogen bonding between donor 'D' of residue X and acceptor 'A' of residue Y is indicated by X:D...Y:A. Interactions described in this table for the $\beta$ and $\delta$ effects come from molecular simulation models of the transition state.

across ribozyme classes, the base pairing arrangement involving N-1 and the Hoogsteen edge of residue 6 at the base of the "L" is observed in all cases.

### 6.4.2 $\beta$ Catalysis

In the context of ribozyme catalysis at physiological pH, stabilization of the negative charge accumulation on scissile phosphate NPOs ($\beta$ catalysis) can be achieved either through direct coordination of a divalent metal ion (1° $\beta$) or through hydrogen bond donation by a nucleobase or indirect (outer-sphere) interaction with metal ion (2° $\beta$) - and commonly both. While mechanistic pathways that involve direct protonation of an NPO (a form of 1° $\beta$ catalysis) can not be definitively ruled out, for the ribozymes considered here, the current body of evidence does not support this as a likely mechanism.

Phosphorothioate substitutions have proven particularly useful in elucidating the interactions with the NPOs, although measurements can be difficult to interpret without complementary site-specific chemical modifications and computational simulations, as was notably the case with the twister ribozyme[3]. For Twr, the functional data suggest that direct ion coordination to the scissile phosphate NPO is not required for catalysis[128]. That being said, inner-sphere coordination at the pro-$S_P$ position has been observed in some crystal structures[10,73] and simulations suggest Twr has an electronegative active site that can attract metal ions nonspecifically to enhance electrostatic stabilization[14]. Furthermore, removing the exocyclic amine of G5 (G33 in Twr) by substitution with inosine eliminated the stereospecific thio effect where the wild-type construct was two orders of magnitude slower with sulfur substitution at the pro-$R_P$ NPO than at the pro-$S_P$[128].

However, it was also noted that the addition of thiophilic metals (e.g. $Mn^{2+}$ or $Cd^{2+}$) had a negligible impact on the activity of both phosphorothioate substrates, indicating that divalent metal ion interactions with the NPOs are likely to involve indirect coordination, as suggested by the computational modeling[14]. In summary, the current model for the twister ribozyme (Table 6.2) proposes that both the exocyclic amine of G5 (G33 in Twr) donating a hydrogen bond to the pro-$R_P$ NPO and non-specific interactions between monovalent ions and the pro-$S_P$ NPO contribute to $\beta$ catalysis.

In the case of VSr, a functionally important divalent metal ion binds to the pro-$S_P$ oxygen, but experiments and simulations suggest the role of this ion is primarily to organize the active site[16]. It is only in the case of HPr that both NPOs of the scissile phosphate are saturated with hydrogen bonding from nucleobases. Thus for the ribozymes observed to fit the G+A paradigm,

it appears that the pro-$R_P$ NPO is involved exclusively in hydrogen bonding, whereas the pro-$S_P$ NPO is either available to interact with a metal ion (Twr and VSr) or other hydrogen bonding interactions (HPr). The most notable hydrogen bond donor, in this context, is the exocyclic amine of the general base guanine (G5:$N_2$). As seen in Tables 6.1 6.2 and this interaction plays multiple roles, being involved not just in $\beta$ catalysis by stabilizing charge through hydrogen bonding at the transition state, but also in $\alpha$ catalysis (GB-NPO contact) by localizing the reactive atoms in the active site and $\gamma$ catalysis by both positioning G5 (anchor) and localizing the reactive atoms (GB-NPO contact).

In the G+$M^+$ paradigm, the role of G5:$N_2$ cannot be as cleanly delineated. Crystallographic evidence for Psr suggests that G5:$N_2$ donates a hydrogen bond to the pro-$R_P$ NPO, whereas for 8-17dz it is the neutral/protonated G5:$N_1$ position, and for HHr neither hydrogen bond donor along the Watson-Crick edge of G5 is close enough to the pro-$R_P$ NPO (though, a water bridged contact could be present). For all three ribozymes, divalent metal ions have not been observed crystallographically directly bound to the pro-$R_P$ NPO. However, these crystal structures do not provide a complete, representative picture of the active state in solution. In particular for 8-17dz and HHr, both functional data[185–187] and simulation[18,19] results suggest that a divalent metal ion directly coordinates the pro-$R_P$ NPO in the catalytically active state. Furthermore, there is evidence that the pro-$R_P$ NPO may bind this metal in the ground state for HHr[188]. Hence, for 8-17dz and HHr, the most likely model of the active state involves a divalent metal ion that directly coordinates the pro-$R_P$ NPO. Interestingly, for Psr, there is a large normal thio effect at the pro-$R_P$ position of the scissile phosphate, that is neither fully rescuable by thiophilic metals, nor by the G5 inosine mutation (G40 in Psr)[189]. These data supports a model for Psr whereby the L-pocket metal is indirectly coordinating the pro-$R_P$ NPO, while G5:$N_2$ donates an additional hydrogen bond to that oxygen. In contrast to the ribozymes in the G+A paradigm, the coordination of the divalent metal ion (either directly or indirectly) to the pro-$R_P$ NPO of the scissile phosphate observed in the G+$M^+$ paradigm relieves the functional requirement of G5:$N_2$ to hydrogen bond with the same NPO. It is important to note that, in general, the G5:$N_2$ exocyclic amine may contribute not only to $\beta$ catalysis, but also to $\alpha$ and $\gamma$ catalysis (as discussed in their respective sections) making interpretation of G5 inosine mutations challenging without additional functional data and rigorous atomic-level modeling of the active state.

### 6.4.3 $\gamma$ Catalysis

The L-platform/L-scaffold motif enables activation of the nucleophile ($\gamma$ catalysis) using primary, secondary and tertiary contributions that satisfy the following requirements:

- 1°$\gamma$: activation of the nucleophile via abstraction of the proton from the 2′ hydroxyl.
- 2°$\gamma$: dynamic tuning of the p$K_a$ values of the general base and nucleophile.
- 3°$\gamma$: spatial localization of the base, nucleophile, and scissile phosphate, and orientation of hydrogen bonds such that nucleophile activation is productive.

Each of these are described in more detail below.

*1°$\gamma$ catalysis.* Across all ribozyme classes considered here that employ the L-platform/L-scaffold motif, the prevailing model is one where activation of the nucleophile occurs via abstraction of the proton from the 2′ hydroxyl by the $N_1$ heteroatom of the strictly conserved general base guanine, G5[57,81,174,189,190]. Within the admittedly limited data set of all known nucleolytic ribozymes, a guanine proposed to act in this functional role is the most common mechanism for 1°$\gamma$ catalysis; further including the *glmS* ribozyme[191] that does not utilize the L-platform/L-scaffold framework.

*2°$\gamma$ catalysis.* In order for the G5 general base to activate the nucleophile, it must be deprotonated at the $N_1$ atomic position. In solution, the unperturbed p$K_a$ of guanine $N_1$ is 9.2[192]. However in the ribozyme environment, the guanine is held in an electronegative active site with its WC edge near to the scissile phosphate. This, in the absence of other factors, would likely lead to a considerable p$K_a$ up-shift and reduced activity at near-neutral conditions. In addition to tuning of the G5:$N_1$ p$K_a$, an environment that also increases the acidity of the nucleophile (p$K_a$ down-shift) would facilitate proton transfer. The L-platform/L-scaffold enables this to be accomplished by recruitment of metal ions in both the G+A and G+$M^+$ paradigms.

Metal ion interactions with the G5:$O_6$ position stabilize delocalized charge in the ionized $G5^-$ nucleotide causing a down-shift in the p$K_a$. In all systems with the exception of VSr, the electronegative Hoogsteen edge of G5 is left exposed to solvent and attracts metal ions (monovalent and/or divalent) from solution[14,17,19,20,76,193]. The HHr is a particularly illuminating example. Recent crystallographic work by Golden and co-workers has identified the appearance of a divalent metal ion binding site ("G-site") at the Hoogsteen edge of the general base guanine in HHr (G12:H-edge in the ribozyme specific numbering scheme and Table 6.1) at pH 8.5, which was not evident at lower pH[76,193], suggesting metal ion binding is correlated with deprotonation. The apparent p$K_a$ assigned to the general base in HHr is 8.0, which is among

the lowest for the known ribozyme classes. Quantum mechanical calculations and free energy simulations in the absence of G-site binding predicted an up-shift of the microscopic p$K_a$ by 3.7 units, whereas with a $Mg^{2+}$ ion (weakly) bound, the microscopic p$K_a$ shifted down by 1.2 units so as to closely align with the apparent p$K_a$ value of 8.0 derived from activity-pH profiles[17]. Similar G-site metal ion binding modes have been observed crystallographically for the Psr[74,75]. In VSr, on the other hand, the Hoogsteen edge of G5 is not solvent exposed, but rather makes strong outer-sphere contact with a functionally critical divalent metal ion bound to the pro-$S_P$ NPO of the scissile phosphate in addition to the pro-$R_P$ NPO of A2 and the Hoogsteen edge of G3 (L-pocket, binds anchor in Table 6.1)[16,194,195]. This contact would be expected to tune the G5 p$K_a$, in addition to serving a critical role as the L-anchor to organize the active site, as will be discussed in the 3°$\gamma$ section below.

Metal ion interactions can also serve to increase the acidity of the O2′ nucleophile facilitating activation (deprotonation). Here again, the electrostatically strained active sites attract metal ions from solution to assist in catalysis. In all of the ribozymes except VSr, monovalent ions from solution are predicted from MD simulations to be territorially bound[196] to the nucleophile and scissile phosphate, and in some cases form bridging interactions that additionally help to align the nucleophile[19,88]. In this position, these ions tune (down-shift) the p$K_a$ of the nucleophile to increase its acidity and facilitate proton transfer.

It should be emphasized that there are compensating effects related to p$K_a$ tuning as a consequence of metal ion interactions in the context of 2°$\gamma$ catalysis that need to be discussed in terms of both thermodynamics and kinetics. Specifically, in order to enhance catalysis, the thermodynamic gain of p$K_a$ down-shifting of the general base upon metal ion binding at its Hoogsteen edge (i.e., increasing the probability of being deprotonated at physiologically relevant pH) cannot be overcompensated by the kinetic penalty of decreased basicity at G5:$N_1$. Similar considerations would apply to balancing the thermodynamic gain of p$K_a$ down-shifting of the nucleophile and the kinetic penalty of reduced nucleophilicity. Alternately stated, ribozymes can employ more reactive functional groups that have a p$K_a$ values shifted from their ideal catalytic pH such that the disadvantage of low abundance (probability) of the active state is partially compensated by higher reactivity. These effects have been considered in recent QM/MM and free energy simulations and found to lead to overall rate enhancement for HHr[17], Twr[14] and VSr[16]. The Hoogsteen edge of guanine is an inherently weak metal ion binding site[184,197], and it is expected that $Mg^{2+}$ ions are fractionally occupied and have fast exchange rates. Hence, it is possible that metal ion binding kinetics play role in dynamically tuning the p$K_a$ of the general base and 2′-OH to facilitate nucleophile activation.

*3°γ catalysis.* In addition to base stacking within the "L" of the L-platform motif, the L-anchor and L-pocket components of the L-scaffold have a profound effect on $\gamma$ catalysis in each of the ribozyme systems. The most common method of anchoring the general base guanine (G5) is through a *trans* sugar edge/Hoogsteen (tSH) base pair with residue 2 (Figure 6.1). Twr, HHr and 8-17dz each strongly conserve an adenine at position 2, likely because it forms the most stable tSH base pair (three hydrogen bonds) with guanine in this configuration[13]. The HPr also anchors the general base via tSH base pairing and it has been noted that all nucleobases except cytosine are tolerated at position 2[183], as expected for this base pair family.

The pistol ribozyme deviates most significantly from the general L-platform/L-scaffold motif, where a cytosine residue (C41 in Psr) is inserted 3′ of the general base and acts as the L-anchor at position 2 in the motif. This shifts G3 (G33) from the prototypical position forming a cWW with N4 (A39) to instead form a cWW pair with C2 (C41). The functional roles of the nucleobases at each of these positions remain the same, despite the unique connectivity and tertiary fold of Psr. With the inserted C41, Psr anchors the general base guanine G5 (G40) along its sugar edge, but through cSH hydrogen bonds with C2:$N_4$ (C41:$N_4$) and/or G6:$O_6$ (G42:$O_6$).

In the case of VSr, residue 2 is a highly conserved adenine (A622) which would be expected to form the previously discussed tHS base pair with the general base G5 (G638). However, recent crystal structures[123,198] and computational modeling[16] suggest that this nucleotide (A622) is bulged from the active site along with residue 1 (A621) and does not play the role of the L-anchor. Instead, the general base guanine (G638) in VSr is anchored along its Hoogsteen edge through outer-sphere coordination of a $Mg^{2+}$ ion bound in the L-pocket, while the bulged N1 and A2 (A621 and A622, respectively) residues support docking of the VSr dimer. The L-pocket binding site is formed by a guanine residue at the 3 position (G623:H-edge) together with the pro-$S_P$ and pro-$R_P$ NPOs of the scissile and A2 phosphates, respectively (similar to HHr, 8-17dz and Psr in the G+$M^+$ paradigm).

While the general base is anchored with its WC edge available, the nucleophile must be positioned such that the guanine can deprotonate it. Within the L-platform/L-scaffold framework there is a single strategy for positioning the nucleophile: a trans base pair involving N-1 and the Hoogsteen edge of nucleobase 6 to form the foot of the "L" of the L-platform (designated "O2′ position" in Table 6.1). This base pair provides stability for the N-1 residue to stack with the catalytic guanine, G5. It is then the stacking of these two nucleobases, in opposing orientations, that positions the N-1 sugar such that it is well situated for the general base G5 to accept a proton from the O2′ nucleophile. As for the identities of the residues in this base pair, like

residue 2, there seems to be a preference for adenine at the 6 position. This is likely due to the flexibility of having both a hydrogen bond donor and acceptor along the Hoogsteen edge of adenine. However, there are numerous combinations of nucleobases that can form either tWH or tSH base pairs. While this full set of base pairs are not strictly isosteric[13], the alignment of the N-1 sugar relative to G5 is similar and thus there is a wide range of variation in the identity of these nucleobases across the ribozyme classes or even within individual classes.

The exocyclic amine of G5 plays an important role in the hydrogen bond network of the L-platform/L-scaffold. Not only does this amine donate an important hydrogen bond to the L-anchor nucleotide, it can also hydrogen bond to one of the NPOs of the scissile phosphate (Twr, HPr, VSr and Psr) or the nucleophile (HHr and 8-17dz) as discussed previously. In either case, these interactions help to position the nucleophile at a nexus between the general base and the scissile phosphate. Furthermore, this hydrogen bonding network can serve to increase the acidity of the 2′OH nucleophile (similar to the proposed role of Lys41 in RNase A[126]) and/or enhance productive hydrogen bonding by elimination of non-productive, competing hydrogen bond interactions that would hinder its activation by G5:$N_1^-$[114]. Additionally, for the ribozymes that require a divalent metal ion for catalysis (VSr, HHr, Psr, and 8-17dz), the metal ion's interaction with the scissile phosphate may also impact the hydrogen bond network involving the nucleophile and thus contribute to $\gamma$ catalysis in a similar fashion[199]. While it is difficult to create experiments that are able to fully decouple these contributions to the various catalytic strategies, theoretical methods, in many instances, are able to integrate constraints that enable their quantitative deconstruction.

### 6.4.4 $\delta$ Catalysis

A remarkable feature of the L-platform/L-scaffold motif is its flexibility in supporting different acids, particularly in contrast to the stringent requirement for an invariant general base. Identification of common trends in how the different acids are positioned and utilized provides a foundation for future design focused on tailoring the identity of the general acid. As discussed above, the L-platform/L-scaffold supports both G+A and G+$M^+$ paradigms, differentiated primarily by their distinct mechanisms for $\delta$ catalysis - utilizing either a protonated adenine or a divalent metal ion in some way. These distinctions are discussed in terms of 1°, 2° and 3° contributions to $\delta$ catalysis below.

*1°$\delta$ catalysis.* The G+A paradigm, originally coined by the Wilson et al.[26], groups VSr and

HPr which both use the $N_1$ adenine heteroatom[200–203] with Twr that uses the $N_3$ adenine heteroatom[128] for general acid catalysis (Figure 6.2). The G+M$^+$ paradigm (Figure 6.3) includes HHr, 8-17dz and Psr, each of which has a metal ion implicated as playing a critical role in general acid catalysis. This divalent ion is recruited to the active site by electrostatic engineering of the previously defined L-pocket (3°$\delta$ in Table 6.2) formed by the Hoogsteen edge of a guanine in position 3 of the L-platform/L-scaffold, along with one of the NPOs of both residue 2 and the scissile phosphate. In HHr, a $Mg^{2+}$ ion binds in the L-pocket[88,186,188,204] and increases the acidity of the 2′-OH of G8 that then can act as the general acid[205,206], although alternative mechanisms have been suggested where a metal-bound water molecule acts as the acid[76,193] and it is possible under different conditions that both pathways are available[207]. In 8-17dz and Psr, a water molecule coordinating the divalent metal ion bound in the L-pocket ($Pb^{2+}$ and $Mg^{2+}$, respectively) likely acts as the general acid[174,189,208], although simulations suggest that the $Pb^{2+}$ ion in 8-17dz could also function, at least in part, as a Lewis acid[19].

*2°$\delta$ catalysis.* Similar to the need to tune the p$K_a$ of the general base guanine, in the G+A paradigm where the general acid is an adenine, tuning of the p$K_a$ can facilitate $\delta$ catalysis. However, unlike the guanine general base that required down-shifting of the p$K_a$, the general acid in the case of adenine in the G+A paradigm requires an up-shifted p$K_a$. This is somewhat less challenging in the sense that the negative electrostatic environment of the active site, due to phosphate moieties, facilitates up-shifting the p$K_a$ of the general acid by hydrogen bonding with the $N_6$ exocyclic amine of adenine. In Twr the $N_6$ amine of A1 donates dual hydrogen bonds to the NPOs of nucleotide residues 16 and 17 (in the Twr specific numbering scheme), whereas in HPr and VSr, the $N_6$ amine donates a hydrogen bond to the pro-$R_P$ oxygen of the scissile phosphate. In the G+M$^+$ paradigm, there is no obvious special contribution of the L-platform/L-scaffold residues to 2°$\delta$ catalysis (although as discussed above, the metal ion itself can promote 2°$\delta$ catalysis.

*3°$\delta$ catalysis.* The general acid must be held in a position where it is poised to donate a proton to the O5′ leaving group. In the G+A paradigm, the hydrogen bond interaction between the general acid adenine $N_6$ exocyclic amine and the NPOs discussed above for 2°$\delta$ catalysis are also important for holding the general acid in position. In Twr, anchoring of the A1 $N_6$ exocyclic amine is achieved through interactions with the NPOs of two nucleotides involved in a pseudoknot near the active site, together with an uncommon *syn* orientation about the glycosidic bond of A1. This enables positioning of the general acid adenine, where A1 protonated at $N_3$ is oriented such that it can hydrogen bond with the O5′ leaving group and is thus poised

to donate that proton to complete the transphosphorylation reaction. For HPr and VSr, the positioning of the general acid is facilitated by base stacking with the bulged N1 nucleotide of the L-platform/L-scaffold, in addition to hydrogen bonding between the general acid $N_6$ and the pro-$R_P$ oxygen of the scissile phosphate.

*Role of the L-pocket in the G+M$^+$ paradigm.* In the G+M$^+$ paradigm, the positioning of the general acid is enabled by binding of the divalent metal ion in the L-pocket. However there is some variability in both the positioning of the L-pocket ligands (pro-$R_P$ NPO of the scissile phosphate, the Hoogsteen edge of G3, and an additional pro-$R_P$ NPO) within the L-platform/L-scaffold as well as the binding modes to those ligands. In HHr, a $Mn^{2+}$ ion (PDB ID: 2OEU[107]), and a $Mg^{2+}$ ion (PDB ID: 5EAO[76]), each have been observed crystallographically directly coordinated to the $N_7$ of the L-pocket G3 (G10.1 in HHr) as well as the pro-$R_P$ NPO of A2 (A9 in HHr). Phosphorothioate-thiophilic metal ion rescue experiments[186,187], supported by molecular simulations[209,210], suggest that in the active state the scissile phosphate acquires a functionally important inner-sphere coordination with a divalent metal ion that has yet to be observed crystallographically.

Original computational studies carried out by Lee et al.[209,210] developed the first rigorous atomic-level model for the active state of HHr whereby the catalytic metal ion occupies a bridging position between the scissile and A9 phosphates[88]. Crystallographic evidence at the time suggested that a divalent metal ion coordinates G10.1:$N_7$ (L-pocket G3 nucleotide) and the pro-$R_P$ of A9 in the ground state, and simulations predicted that this ion can migrate into the bridging position prior to forming the transition state[18]. In this position, it was discovered the $Mg^{2+}$ ion forms interactions with the 2′OH of G8, increasing its acidity and enabling it to act as a general acid. Further computational mutagenesis[211] and quantum mechanical simulations[212] lent further support for a metal-activated G8:$O_{2'}$ to act as the general acid, and experimental studies by Thomas and Perrin[206] provided convincing evidence the theoretical predictions were correct.

$Cd^{2+}$ rescue experiments on an extended HHr construct from *Schistosoma mansoni* indicate that a catalytic metal ion may occupy this binding mode even in the ground state[188]. Due to structural constraints in the active site, this would preclude the metal ion maintaining direct coordination with $N_7/O_6$ at the Hoogsteen edge of the L-pocket G3. Rather, outer-sphere coordination with the Hoogsteen edge of G3 is more plausible given that direct coordination of a $Mg^{2+}$ ion to the $N_7$ positions of nucleobases is quite rare[197], and substitution of 7-deazaguanine at the G3 position in HHr results in only a modest ∼30-fold decrease in the observed rate[213], leading the authors to conclude that while this site is important[186,214], it is not catalytically

indispensable.

In 8-17dz, a $Pb^{2+}$ ion (PDB ID: 5XM8) was observed with partial occupancy bound at the $O_6$ position of the L-pocket G3 (G6 in 8-17dz)[174]. While less biochemical data are available for this system, it has been reported that stereospecific thio substitution at the pro-$R_P$ NPO of the scissile phosphate can be used to selectively remove the $S_P$ isomer, with the remaining $R_P$ isomer being active in the presence of $Cd^{2+}$[215]. Further, similar to HHr, a 7-deazaguanine substitution at the L-pocket G3 position leads to only a modest ∼25-fold decrease in the observed rate[216]. Together, this suggests the active state of 8-17dz has a catalytic metal ion binding mode similar to that of HHr, involving direct coordination to the pro-$R_P$ NPO of the scissile phosphate, and indirect coordination with the Hoogsteen edge of G3.

In Psr, a $Mg^{2+}$ ion (PDB ID: 5K7C[74], 5KTJ[75], and 6R47[189]) was modeled as directly coordinated to the $N_7$ of the L-pocket G3 (G33 in Psr), and indirectly coordinated to the pro-$R_P$ NPO of the scissile phosphate. Thio substitution experiments indicate that there is a significant, normal effect at the scissile phosphate NPO[217], subsequently identified as the pro-$R_P$ position[189], that unlike with HHr, is not rescuable by $Mn^{2+}$ ions. Furthermore, in contrast to both HHr and 8-17dz, loss of inner-sphere coordination with the $N_7$ that occurs upon 7-deazaguanine substitution leads to a minimum ∼300-fold decrease in the observed rate[208], and may be even substantially more detrimental ($> 10^4$-fold decrease) to activity[189]. Taken together, this suggests that the active state of Psr has the catalytic metal ion maintaining direct coordination to the $N_7$ of the L-pocket G3, while making an outer-sphere contact to the pro-$R_P$ NPO of the scissile phosphate. This pattern of direct/indirect metal ion coordination for Psr with the L-pocket binding ligands is opposite to that for HHr and 8-17dz.

The functional requirement for direct coordination of a divalent metal ion with the Hoogsteen edge of the L-pocket G3 is, however, not unique to Psr or the G+$M^+$ paradigm. The VSr requires a functionally critical divalent metal to bind in the L-pocket (not to serve as the general acid, but rather the role of the missing L-anchor) and is predicted to directly coordinate the Hoogsteen edge of G3 (G623 in VSr). Similar to Psr, 7-deazaguanine mutation at this position in VSr effectively abolishes activity[16]. Hence, in the case of Psr and VSr, there is strong evidence that the active state requires direct $Mg^{2+}$ ion coordination to the $N_7$ position of the L-pocket G3, despite this binding mode being rarely observed crystallographically[197]. As for the remaining ligand defining the L-pocket (prototypically an NPO 5′ of the L-anchor), in the G+$M^+$ paradigm it is consistently the pro-$R_P$ non-bridging oxygen. However, with the L-anchor nucleotide being part of the general base strand in Psr, it is the phosphate 5′ of G3 (rather than A2 in HHr and 8-17dz) that serves as that additional contact for the catalytic

metal ion.

## 6.5 Conclusion

We present a generalized L-platform/L-scaffold active site architecture that serves as a blueprint to facilitate rational design of nucleic acid enzymes that catalyze site-specific RNA cleavage through 2′-O-transphosphorylation. We illustrate how the generalized L-platform/L-scaffold is common to five of the nine currently known naturally occurring ribozymes classes (Twr, HPr, VSr, HHr, Psr) as well as a recently structurally characterized artificially engineered DNAzyme (8-17dz). We identify key base pairing and stacking requirements that enable conserved features to emerge, as well as elements that can tolerate variation both across and within the ribozyme classes to be explored. The L-platform/L-scaffold motif poises an invariant guanine to act as the general base, while leaving the Hoogsteen edge exposed to solvent, enabling recruitment of cations (except for VSr where this is achieved by outer-sphere interactions with a divalent ion bound in the L-pocket). Within this motif, the preference for guanine (over the other naturally occurring nucleobases) to act as the general base is clear. Having both a hydrogen bond donor and an acceptor along the sugar edge provides both stability and flexibility in base pairing with the L-anchor residue. Along the Watson-Crick edge, the exocyclic amine can also hydrogen bond with the non-bridging oxygens of the scissile phosphate helping to both localize the reactive atoms and stabilize negative charge in the transition state. Finally, the p$K_a$ of the $N_1$ site can be dynamically tuned through metal ion interactions at the Hoogsteen edge, facilitating activation of the nucleophile via proton transfer.

In contrast, the identity of the general acid displays almost as much variety as the global folds across the ribozyme classes. Despite the differences, there are still clear trends in how the L-platform/L-scaffold "docks" the various general acids, enabling classification of the ribozymes into one of two paradigms: G+A and G+$M^+$. Within the G+$M^+$ paradigm a divalent metal ion implicated in general acid catalysis is bound by a well-defined set of ligands that form the L-pocket. On the other hand, the structural requirements and even the apparent preference for utilizing adenine as the acid remains less clear. Ultimately, the full extent to which the L-platform/L-scaffold can accommodate different general acids in naturally occurring or synthetic contexts remains to be seen.

Furthermore, the L-platform/L-scaffold motif is common to a large fraction of the currently identified small nucleolytic ribozyme classes, but it is by no means a universal platform. The HDV and twister sister ribozymes have $Mg^{2+}$ implicated as the general base and cytosine has

been proposed as their general acid. While the *glmS* ribozyme utilizes a guanine in the general base role, its active site does not conform to the L-platform/L-scaffold architecture, likely due to additional structural requirements for binding the GlcN6P cofactor. A co-factor independent *glmS* variant has been *in vitro* evolved[218], but retains the wild-type fold and thus does not adopt an L-platform/L-scaffold active site. There are numerous ribozymes both naturally occurring (e.g. hatchet ribozyme) or artificially engineered (e.g. GR5, 10-23, and NaA43 DNAzymes) yet to be structurally characterized and many more yet to be discovered that are certain to provide insight into the general principles of RNA catalysis that could further facilitate rational design. It is also particularly intriguing that through directed evolution 8-17dz converged on the same L-platform/L-scaffold motif as the naturally occurring ribozymes examined here. This begs the question as to the extent to which design principles such as those defined by the L-platform/L-scaffold motif presented here might also be translated into non-biological contexts such as with the recently reported Hachimoji RNA/DNA[111].

A PyMOL script and collection of PDB files comparing the active sites for the six ribozymes that conform to the L-platform/L-scaffold motif is included in Appendix A.4.

# Bibliography


[1] Gail Mitchell Emilsson, Shingo Nakamura, Adam Roth, and Ronald R. Breaker. Ribozyme speed limits. *RNA*, 9:907–918, 2003.

[2] Ronald R. Breaker, Gail Mitchell Emilsson, Denis Lazarev, Shingo Nakamura, Izabela J. Puskarz, Adam Roth, and Narasimhan Sudarsan. A common speed limit for RNA-cleaving ribozymes and deoxyribozymes. *RNA*, 9:949–957, 2003.

[3] Ronald R. Breaker. Mechanistic Debris Generated by Twister Ribozymes. *ACS Chem. Biol.*, 12(4):886–891, 2017. doi: 10.1021/acschembio.7b00010. URL `10.1021/acschembio.7b00010`.

[4] Larisa E. Kapinos, Bert P. Operschall, Erik Larsen, and Helmut Sigel. Understanding the acid-base properties of adenosine: the intrinsic basicities of N1, N3 and N7. *Chem. Eur. J.*, 17:8156–8164, 2011.

[5] Vincenzo Verdolino, Roberto Cammi, Barbara H. Munk, and H. Bernhard Schlegel. Calculation of $pK_a$ Values of Nucleobases and the Guanine Oxidation Products Guanidinohydantoin and Spiroiminodihydantoin using Density Functional Theory and a Polarizable Continuum Model. *J. Phys. Chem. B*, 112:16860–16873, 2008.

[6] Yijin Liu, Timothy J. Wilson, Scott A. McPhee, and David M J. Lilley. Crystal structure and mechanistic investigation of the twister ribozyme. *Nat. Chem. Biol.*, 10:739–744, 2014.

[7] Yijin Liu, Timothy J. Wilson, and David M. J. Lilley. The structure of a nucleolytic ribozyme that employs a catalytic metal ion. *Nat. Chem. Biol.*, 13:508–513, 2017. doi: doi:10.1038/nchembio.2333. URL `https://www.nature.com/nchembio/journal/v13/n5/full/nchembio.2333.html`.

[8] Thakshila Dissanayake, Jason M. Swails, Michael E. Harris, Adrian E. Roitberg, and Darrin M. York. Interpretation of pH-Activity Profiles for Acid-Base Catalysis from Molecular Simulations. *Biochemistry*, 54:1307–1313, 2015.

[9] Daniel Eiler, Jimin Wang, and Thomas A. Steitz. Structural basis for the fast self-cleavage reaction catalyzed by the twister ribozyme. *Proc. Natl. Acad. Sci. USA*, 111(36):13028–13033, 2014.

[10] Aiming Ren, Marija Košutić, Kanagalaghatta R. Rajashankar, Marina Frener, Tobias Santner, Eric Westhof, Ronald Micura, and Dinshaw J. Patel. In-line alignment and Mg2+ coordination at the cleavage site of the env22 twister ribozyme. *Nat. Commun.*, 5: 5534–5544, 2014.

[11] Michael E Harris, Joseph A Piccirilli, and Darrin M York. Integration of kinetic isotope effect analyses to elucidate ribonuclease mechanism. *Biochim. Biophys. Acta*, 1854:1801–1808, November 2015. ISSN 0006-3002. doi: 10.1016/j.bbapap.2015.04.022.

[12] N. B. Leontis and E. Westhof. Geometric nomenclature and classification of RNA base pairs. *RNA*, 7:499–512, 2001.

[13] Neocles B. Leontis, Jesse Stombaugh, and Eric Westhof. The non-watson-crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, 30:3497–3531, 2002.

[14] Colin S. Gaines, Timothy J. Giese, and Darrin M. York. Cleaning Up Mechanistic Debris Generated by Twister Ribozymes Using Computational RNA Enzymology. *ACS Catal.*, 9(7):5803–5815, 2019.

[15] Hugh Heldenbrand, Pawel A. Janowski, George Giambaşu, Timothy J. Giese, Joseph E. Wedekind, and Darrin M. York. Evidence for the role of active site residues in the hairpin ribozyme from molecular simulations along the reaction path. *J. Am. Chem. Soc.*, 136: 7789–7792, 2014.

[16] Abir Ganguly, Benjamin P. Weissman, Timothy J. Giese, Nan-Sheng Li, Shuichi Hoshika, Rao Saieesh, Steven A. Benner, Joseph A. Picirilli, and Darrin M. York. Confluence of theory and experiment reveals the catalytic mechanism of the Varkud satellite ribozyme. *Nature chemistry*, 12:193–201, 2020.

[17] Haoyuan Chen, Timothy J Giese, Barbara L Golden, and Darrin M York. Divalent Metal Ion Activation of a Guanine General Base in the Hammerhead Ribozyme: Insights from Molecular Simulations. *Biochemistry*, 56:2985–2994, June 2017.

[18] Tai-Sung Lee, Kin-Yiu Wong, George M. Giambasu, and Darrin M. York. *Bridging the gap between theory and experiment to derive a detailed understanding of hammerhead ribozyme catalysis*, volume 120, chapter 2, pages 25–91. 2013.

[19] Şölen Ekesan and Darrin M. York. Dynamical ensemble of the active state and transition state mimic for the RNA-cleaving 8-17 DNAzyme in solution. *Nucleic Acids Res.*, 47(19): 10282–10295, 09 2019.

[20] Ken Kostenbader and Darrin M. York. Molecular simulations of the pistol ribozyme: unifying the interpretation of experimental data and establishing functional links with the hammerhead ribozyme. *RNA*, 25(11):1439–1456, 2019.

[21] K. Kruger, P.J. Grabowsky, A.J. Zaug, J. Sands, D.E. Gottschling, and T.R. Cech. Self-splicing RNA: autoexcision and autocyclisation of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell*, 31:147–157, 1982.

[22] Thomas R. Cech and Joan A. Steitz. The Noncoding RNA Revolution – Trashing Old Rules to Forge New Ones. *Cell*, 157:77–94, 2014.

[23] Bruce A Sullenger and Smita Nair. From the RNA world to the clinic. *Science*, 352(6292): 1417–1420, 2016.

[24] Cameron M. Kim and Christina D. Smolke. Biomedical applications of RNA-based devices. *Curr. Opin. Biomed. Eng.*, 4:106–115, 2017.

[25] Randi M Jimenez, Julio A Polanco, and Andrej Lupták. Chemistry and biology of self-cleaving ribozymes. *Trends in biochemical sciences*, 40(11):648–661, 2015.

[26] Timothy J. Wilson, Yijin Liu, and David M. J. Lilley. Ribozymes and the mechanisms that underlie RNA catalysis. *Front. Chem. Sci. Eng.*, 10:178–185, 2016.

[27] Christina E. Weinberg, Zasha Weinberg, and Christian Hammann. Novel ribozymes: discovery, catalytic mechanisms, and the quest to understand biological function. *Nucleic Acids Res.*, 47:9480–9494, 2019.

[28] Christian Hammann, Andrej Luptak, Jonathan Perreault, and Marcos De La Peña. The ubiquitous hammerhead ribozyme. *RNA*, 18:871–885, 2012.

[29] Chiu-Ho T. Webb, Nathan J. Riccitelli, Dana J. Ruminski, and Andrej Lupták. Widespread occurrence of self-cleaving ribozymes. *Science*, 326(5955):953–, November 13, 2009 2009. URL `http://www.sciencemag.org/cgi/content/abstract/326/5955/953`.

[30] Adam Roth, Zasha Weinberg, Andy G Y Chen, Peter B Kim, Tyler D Ames, and Ronald R Breaker. A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat. Chem. Biol.*, 10:56–62, 2014.

[31] Zasha Weinberg, Peter B. Kim, Tony H. Chen, Sanshu Li, Kimberly A. Harris, Christina E. Lünse, and Ronald R. Breaker. New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nat. Chem. Biol.*, 11:606–610, 2015. doi: 10.1038/nchembio.1846. URL `http://www.nature.com/nchembio/journal/v11/n8/full/nchembio.1846.html`.

[32] Ronald R. Breaker and Gerald F. Joyce. The Expanding View of RNA and DNA Function. *Chem. Biol.*, 21:1059–1065, 2014. doi: 10.1016/j.chembiol.2014.07.008. URL `http://dx.doi.org/10.1016/j.chembiol.2014.07.008`.

[33] Thomas A. Steitz and Peter B. Moore. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem. Sci.*, 28(8):411–418, August 2003.

[34] Harry F. Noller. RNA structure: reading the Ribosome. *Science*, 309:1508–1514, 2005.

[35] Andrei Korostelev, Dmitri N. Ermolenko, and Harry F. Noller. Structural dynamics of the ribosome. *Curr. Opin. Chem. Biol.*, 12(6):674–683, December 2008. doi: 10.1016/j.cbpa.2008.08.037.

[36] T Martin Schmeing and V Ramakrishnan. What recent ribosome structures have revealed about the mechanism of translation. *Nature*, 461(7268):1234–1242, 2009. URL `http://www.nature.com/nature/journal/v461/n7268/abs/nature08403.html`.

[37] Marina V. Rodnina. *Ribozymes and RNA Catalysis*, chapter Peptidyl transferase mechanism: the ribosome as a ribozyme, pages 270–294. RSC Biomolecular Series. RSC Publishing, Cambridge, 2008.

[38] Saba Valadkhan. Role of the snRNAs in spliceosomal active site. *RNA Biol.*, 7(3):345–353, 2010.

[39] Aaron A. Hoskins and Melissa J. Moore. The spliceosome: a flexible, reversible macromolecular machine. *Trends Biochem. Sci.*, 37(5):179–188, May 2012. doi: 10.1016/j.tibs.2012.02.009.

[40] Panagiotis Papasaikas and Juan Valcárcel. The spliceosome: the ultimate RNA chaperone and sculptor. *Trends in biochemical sciences*, 41(1):33–45, 2016.

[41] Chuangye Yan, Jing Hang, Ruixue Wan, Min Huang, Catherine CL Wong, and Yigong Shi. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science*, 349(6253):1182–1191, 2015.

[42] Walter Gilbert. The RNA World. *Nature*, 319:618, 1986.

[43] Raymond F. Gesteland, Thomas R. Cech, and John F. Atkins. *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA*. Cold Spring Harbor Laboratory Press, New York, 2nd edition, 1999.

[44] Timothy J. Wilson and David M. J. Lilley. The evolution of ribozyme chemistry. *Science*, 323(5920):1436–1438, 2009.

[45] Xi Chen, Na Li, and Andrew D. Ellington. Ribozyme catalysis of metabolism in the RNA world. *Chem. Biodivers.*, 4:633–655, 2007.

[46] Ádám Kun, András Szilágyi, Balázs Könnyű, Gergely Boza, István Zachar, and Eörs Szathmáry. The dynamics of the RNA world: insights and challenges. *Annals of the New York Academy of Sciences*, 1341:75–95, 2015.

[47] Thomas R. Cech. Ribozyme engineering. *Curr. Opin. Struct. Biol.*, 2:605–609, 1992.

[48] Nassim Usman, Leonid Beigelman, and James A. McSwiggen. Hammerhead ribozyme engineering. *Curr. Opin. Struct. Biol.*, 1:527–533, 1996.

[49] Michael Famulok. Oligonucleotide aptamers that recognize small molecules. *Curr. Opin. Struct. Biol.*, 9:324–329, 1999.

[50] Garrett A. Soukup and Ronald R. Breaker. Engineering precision RNA molecular switches. *Proc. Natl. Acad. Sci.*, 96:3584–3589, 1999.

[51] Shungo Kobori and Yohei Yokobayashi. Analyzing and Tuning Ribozyme Activity by Deep Sequencing To Modulate Gene Expression Level in Mammalian Cells. *ACS Synth. Biol.*, 7:371–376, 2018.

[52] Darko Balke, Robert Hieronymus, and Sabine Müller. *Challenges and Perspectives in Nucleic Acid Enzyme Engineering*, pages 1–16. Adv. Biochem. Engin./Biotechnol. Springer AG, 2017.

[53] Sungho Jang, Sungyeon Jang, Jina Yang, Sang Woo Seo, and Gyoo Yeol Jung. RNA-based dynamic genetic controllers: development strategies and applications. *Current Opinion in Biotechnology*, 53:1–11, 2018.

[54] Motiar Rahman, Shigeyoshi Matsumura, and Yoshiya Ikawa. Artificial RNA Motifs Expand the Programmable Assembly between RNA Modules of a Bimolecular Ribozyme Leading to Application to RNA Nanostructure Design. *Biology*, 6:37, 2017.

[55] Michele Felletti, Julia Stifel, Lena A. Wurmthaler, Sophie Geiger, and Jorg S. Hartig. Twister ribozymes as highly versatile expression platforms for artificial riboswitches. *Nat. Commun.*, 7:1–8, 2016.

[56] W Luke Ward, Kory Plakos, and Victoria J. DeRose. Nucleic acid catalysis: metals, nucleobases, and other cofactors. *Chem. Rev.*, 114(8):4318–4342, 2014.

[57] David M. J. Lilley. How RNA acts as a nuclease: some mechanistic comparisons in the nucleolytic ribozymes. *Biochem. Soc. Trans.*, 45:683–691, 2017.

[58] Garrett A. Soukup and Ronald R. Breaker. Allosteric nucleic acid catalysts. *Curr. Opin. Struct. Biol.*, 10(3):318–325, 2000.

[59] Sukeerthi Seetharaman, Maris Zivarts, Narasimhan Sudarsan, and Ronald R. Breaker. Immobilized RNA switches for the analysis of complex chemical and biological mixtures. *Nat. Biotech.*, 19:336–341, 2001.

[60] Narendra K. Vaish, Fand Dong, Lori Andrews, Rebecca E. Schweppe, Natalie G. Ahn, Lawrence Blatt, and Scott D. Seiwert. Monitoring post-tranlation modification of proteins with allosteric ribozymes. *Nat. Biotech.*, 20:810–815, 2002.

[61] Lucien Junior Bergeron and Jean-Pierre Perreault. Target-dependent on/off switch increases ribozyme fidelity. *Nucleic Acids Res.*, 33:1240–1248, 2005.

[62] Michele Felletti and Jörg S Hartig. Ligand-dependent ribozymes. *Wiley Interdisciplinary Reviews: RNA*, 8(2):1395, 2017.

[63] Ronald R. Breaker. Engineered allosteric ribozymes as biosensor components. *Curr. Opin. Biotechnol.*, 13(1):31–39, 2002.

[64] Joanne M. Yeakley, Jian-Bing Fan, Dennis Doucet, Lin Luo, Eliza Wickham, Zhen Ye, Mark S. Chee, and Xiang-Dong Fu. Profiling alternative splicing on fiber-optic arrays. *Nat. Biotechnol.*, 20:353–358, April 2002.

[65] Robert Penchovsky. Computational design of allosteric ribozymes as molecular biosensors. *BiotechAdv*, 32:1015–1027, 2014.

[66] Wenhu Zhou, Po-Jung Jimmy Huang, Jinsong Ding, and Juewen Liu. Aptamer-based biosensors for biomedical diagnostics. *Analyst*, 139(11):2627–2640, 2014.

[67] Yuan-Yuan ZHANG, Hui CHENG, Sun Yan, Jin-E WANG, Wu Zheng-Yan, and Pei Ren-Jun. Engineering of Thiamine Pyrophosphate Fluorescent Biosensors Based on Ribozyme Switches in Mammalian Cells. *Chinese Journal of Analytical Chemistry*, 45(2):157–162, 2017.

[68] Philip C. Bevilacqua, Michael E. Harris, Joseph A. Piccirilli, Colin Gaines, Abir Ganguly, Ken Kostenbader, Şölen Ekesan, and Darrin M. York. An Ontology for Facilitating Discussion of Catalytic Strategies of RNA-Cleaving Enzymes. *ACS Chem. Biol.*, 14:1068–1076, 2019.

[69] Joachim Schnabl and Roland K. O. Sigel. Controlling ribozyme activity by metal ions. *Curr. Opin. Chem. Biol.*, 14(2):269–275, Apr 2010. doi: 10.1016/j.cbpa.2009.11.024. URL `http://dx.doi.org/10.1016/j.cbpa.2009.11.024`.

[70] Tuomas Lönnberg and Harri Lönnberg. Chemical models for ribozyme action. *Current Opinion in Chemical Biology*, 9:665–673, 2005.

[71] Alvan C. Hengge. Isotope effects in the study of phosphoryl and sulfuryl transfer reactions. *Acc. Chem. Res.*, 35:105–112, 2002.

[72] Jamie L Bingaman, Sixue Zhang, David R Stevens, Neela H Yennawar, Sharon Hammes-Schiffer, and Philip C Bevilacqua. The GlcN6P Cofactor Plays Multiple Catalytic Roles in the glmS Ribozyme. *Nat. Chem. Biol.*, 13(4):439, 2017.

[73] Marija Košutić, Sandro Neuner, Aiming Ren, Sara Flür, Christoph Wunderlich, Elisabeth Mairhofer, Nikola Vušurović, Jan Seikowski, Kathrin Breuker, Claudia Höbartner, Dinshaw J. Patel, Christoph Kreutz, and Ronald Micura. A Mini-Twister Variant and Impact of Residues/Cations on the Phosphodiester Cleavage of this Ribozyme Class. *Angew. Chem. Int. Ed.*, 54:15128–15133, 2015. doi: 10.1002/anie.201506601.

[74] Aiming Ren, Nikola Vusurovic, Jennifer Gebetsberger, Pu Gao, Michael Juen, Christoph Kreutz, Ronald Micura, and Dinshaw Patel. Pistol Ribozyme Adopts a Pseudoknot Fold Facilitating Site-specific In-line Cleavage. *Nat. Chem. Biol.*, 12:702–708, 2016.

[75] Laura A. Nguyen, Jimin Wang, and Thomas A. Steitz. Crystal structure of Pistol, a class of self-cleaving ribozyme. *Proc. Natl. Acad. Sci. USA*, 114:1021–1026, 2017.

[76] Aamir Mir and Barbara L. Golden. Two Active Site Divalent Ions in the Crystal Structure of the Hammerhead Ribozyme Bound to a Transition State Analogue. *Biochemistry*, 55:

633–636, 2016. doi: 10.1021/acs.biochem.5b01139. URL http://pubs.acs.org/doi/abs/10.1021/acs.biochem.5b01139.

[77] Shu-ichi Nakano, David J. Proctor, and Philip C. Bevilacqua. Mechanistic characterization of the HDV genomic ribozyme: Assessing the catalytic and structural contributions of divalent metal ions within a multichannel reaction mechanism. *Biochemistry*, 40:12022–12038, 2001.

[78] Subha Das and Joseph Piccirilli. General acid catalysis by the hepatitis delta virus ribozyme. *Nat. Chem. Biol.*, 1(1):45–52, 2005.

[79] Jui-Hui Chen, Rieko Yajima, Durga M. Chadalavada, Elaine Chase, Philip C. Bevilacqua, and Barbara L. Golden. A 1.9 Å crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and general acid mechanisms contribute to phosphodiester cleavage. *Biochemistry*, 49(31):6508–6518, 2010.

[80] Tyler Luchko, Sergey Gusarov, Daniel R. Roe, Carlos Simmerling, David A. Case, Jack Tuszynski, and Andriy Kovalenko. Three-dimensional molecular theory of solvation coupled with molecular dynamics in AMBER. *J. Chem. Theory Comput.*, 6:607–624, 2010.

[81] Jesse C. Cochrane and Scott A. Strobel. Catalytic strategies of self-cleaving ribozymes. *Acc. Chem. Res.*, 41:1027–1035, 2008.

[82] D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D.S Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Götz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, Tai-Sung Lee, S. Le Grand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, and P. A. Kollman. *AMBER 14*. University of California, San Francisco, San Francisco, CA, 2014.

[83] Romelia Salomon-Ferrer, Andreas W. Götz, Duncan Poole, Scott Le Grand, and Ross C. Walker. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent Particle Mesh Ewald. *J. Chem. Theory Comput.*, 9:3878–3888, 2013.

[84] Hans W. Horn, William C. Swope, Jed W. Pitera, Jeffry D. Madura, Thomas J. Dick, Greg L. Hura, and Teresa Head-Gordon. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.*, 120(20):9665–9678, 2004.

[85] James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.*, 11:3696–3713, 2015.

[86] In Suk Joung and Thomas E. Cheatham III. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, 112:9020–9041, 2008.

[87] Pengfei Li, Benjamin P. Roberts, Dhruva K. Chakravorty, and Kenneth M. Merz, Jr. Rational design of Particle Mesh Ewald compatible Lennard-Jones parameters for +2 metal cations in explicit solvent. *J. Chem. Theory Comput.*, 9:2733–2748, 2013.

[88] Tai-Sung Lee, George M. Giambaşu, Carlos P. Sosa, Monika Martick, William G. Scott, and Darrin M. York. Threshold Occupancy and Specific Cation Binding Modes in the Hammerhead Ribozyme Active Site are Required for Active Conformation. *J. Mol. Biol.*, 388:195–206, 2009.

[89] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.*, 91:6269–6271, 1987.

[90] David M. J. Lilley. Structure, folding and mechanisms of ribozymes. *Curr. Opin. Struct. Biol.*, 15:313–323, 2005.

[91] Philip C Bevilacqua and Rieko Yajima. Nucleobase catalysis in ribozyme mechanism. *Curr. Opin. Chem. Biol*, 10:455–464, 2006.

[92] Scott A. Strobel and Jesse C. Cochrane. RNA catalysis: ribozymes, ribosomes, and riboswitches. *Curr. Opin. Chem. Biol.*, 11(6):636–643, Dec 2007. doi: 10.1016/j.cbpa.2007.09.010. URL `http://dx.doi.org/10.1016/j.cbpa.2007.09.010`.

[93] Martha J. Fedor. Comparative enzymology and structural biology of RNA self-cleavage. *Annu. Rev. Biophys.*, 38:271–299, 2009.

[94] K. H. Link and R. R. Breaker. Engineering ligand-responsive gene-control elements: Lessons learned from natural riboswitches. *Gene Therapy*, 16:1189–1201, 2009.

[95] Xi Chen and Andrew D. Ellington. Design principles for ligand-sensing, conformation-switching ribozymes. *PLoS Comput Biol*, 5:1000620, 2009.

[96] Jacques Fastrez. Engineering Allosteric Regulation into Biological Catalysts. *ChemBioChem*, 10:2824–2835, 2009.

[97] Beatrix Suess and Julia E. Weigand. Engineered riboswitches: Overview, problems and trends. *RNA Biol*, 5:1–6, 2008.

[98] A. R. Ferré-D'Amaré. Use of a coenzyme by the glmS ribozyme-riboswitch suggests primordial expansion of RNA chemistry by small molecules. *Phil. Trans. R. Soc. B*, 366: 2942–2948, 2011.

[99] Elizabeth A. Doherty and Jennifer A. Doudna. Ribozyme Structures and Mechanisms. *Annu. Rev. Biophys. Biomol. Struct.*, 30:457–475, 2001.

[100] Maria T. Panteva, Thakshila Dissanayake, Haoyuan Chen, Brian K. Radak, Erich R. Kuechler, George M. Giambaşu, Tai-Sung Lee, and Darrin M. York. Multiscale Methods for Computational RNA Enzymology. In Shi-Jie Chen and Donald H. Burke-Aguero, editors, *Methods in Enzymology*, volume 553, chapter 14. Elsevier, 2015.

[101] Garrett A. Soukup and Ronald R. Breaker. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA*, 5:1308–1325, 1999.

[102] Jennifer L. Wilcox and Philip C. Bevilacqua. A Simple Florescence Method for $pK_a$ Determination in RNA and DNA Reveals Highly Shifted $pK_a$'s. *J. Am. Chem. Soc.*, 135 (20):7390–7393, 2013.

[103] Júlia Viladoms, Lincoln G. Scott, and Martha J. Fedor. An active-site guanine participates in glmS ribozyme catalysis in its protonated state. *J. Am. Chem. Soc.*, 133(45):18388–18396, 2011.

[104] Garrett B. Goh, Jennifer L. Knight, and Charles L. Brooks III. pH-dependent dynamics of complex RNA macromolecules. *J. Chem. Theory Comput.*, 9:935–943, 2013.

[105] Philip C. Bevilacqua. Mechanistic considerations for general acid-base catalysis by RNA: Revisiting the mechanism of the hairpin ribozyme. *Biochemistry*, 42:2259–2265, 2003.

[106] Darrin M. York, Weitao Yang, Hsing Lee, Tom Darden, and Lee G. Pedersen. Toward the accurate modeling of DNA: the importance of long-range electrostatics. *J. Am. Chem. Soc.*, 117:5001–5002, 1995.

[107] Monika Martick, Tai-Sung Lee, Darrin M. York, and William G. Scott. Solvent structure and hammerhead ribozyme catalysis. *Chem. Biol.*, 15:332–342, 2008.

[108] Pawel A. Janowski, David S. Cerutti, James Holton, and David A. Case. Peptide Crystal Simulations Reveal Hidden Dynamics. *J. Am. Chem. Soc.*, 135(21):7938–7948, 2013.

[109] Tai-Sung Lee, Brian K. Radak, Anna Pabis, and Darrin M. York. A new maximum likelihood approach for free energy profile construction from molecular simulations. *J. Chem. Theory Comput.*, 9:153–164, 2013.

[110] Irina Anosova, Ewa A. Kowal, Matthew R. Dunn, John C. Chaput, Wade D. Van Horn, and Martin Egli. The structural diversity of artificial genetic polymers. *Nucleic Acids Res.*, 44(3):1007–1021, 2016. doi: 10.1093/nar/gkv1472. URL `https://dx.doi.org/10.1093/nar/gkv1472`.

[111] Shuichi Hoshika, Nicole A. Leal, Myong-Jung Kim, Myong-Sang Kim, Nilesh B. Karalkar, Hyo-Joong Kim, Alison M. Bates, Norman E. Watkins, Holly A. SantaLucia, Adam J. Meyer, Saurja DasGupta, Joseph A. Piccirilli, Andrew D. Ellington, John SantaLucia, Millie M. Georgiadis, and Steven A. Benner. Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science*, 363(6429):884–887, 2019. ISSN 0036-8075. doi: 10.1126/science.aat0971. URL `http://science.sciencemag.org/content/363/6429/884`.

[112] Juewen Liu, Zehui Cao, and Yi Lu. Functional Nucleic Acid Sensors. *Chem. Rev.*, 109:1948–1998, 2009.

[113] Y Zhang, J Wang, H Cheng, Y Sun, M Liu, Z Wu, and R Pei. Conditional control of suicide gene expression in tumor cells with theophylline-responsive ribozyme. *Gene therapy*, 24(2):84, 2017.

[114] Daniel D. Seith, Jamie L. Bingaman, Andrew J. Veenis, Aileen C. Button, and Philip C. Bevilacqua. Elucidation of Catalytic Strategies of Small Nucleolytic Ribozymes from Comparative Analysis of Active Sites. *ACS Catal.*, 8:314–327, 2018.

[115] Gerry A. Prody, John T. Bakos, Jamal M. Buzayan, Irving R. Schneider, and George Bruening. Autolytic processing of dimeric plant virus satellite RNA. *Science*, 231:1577–1580, 1986.

[116] H. W. Pley, K. M. Flaherty, and D. B. McKay. Three-dimensional structure of a hammerhead ribozyme. *Nature*, 372:68–74, 1994.

[117] William G. Scott, James B. Murray, John R. P. Arnold, Barry L. Stoddard, and Aaron Klug. Capturing the structure of a catalytic RNA intermediate: The Hammerhead Ribozyme. *Science*, 274:2065–2069, 1996.

[118] J. M. Buzayan, W. L. Gerlach, and G. Bruening. Nonenzymatic cleavage and ligation of RNAs complementary to a plant virus sattelite RNA. *Nature*, 323:349–353, 1986.

[119] Peter B. Rupert, Archna P. Massey, Snorri Th. Sigurdsson, and Adrian R. Ferré-D'Amaré. Transition State Stabilization by a Catalytic RNA. *Science*, 298:1421–1424, 2002.

[120] L. Sharmeen, M. Y. Kuo, G. Dinter-Gottlieb, and J. Taylor. Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J. Virol.*, 62:2674–2679, 1988.

[121] Adrian R. Ferré-D'Amaré, Kaihong Zhou, and Jennifer A. Doudna. Crystal structure of a hepatitis delta virus ribozyme. *Nature*, 395:567–574, 1998.

[122] Barry J. Saville and Richard A. Collins. A site-specific self-cleavage reaction performed by a novel RNA in neurospora mitochondria. *Cell*, 61:685–696, 1990.

[123] Nikolai B. Suslov, Saurja DasGupta, Hao Huang, James R. Fuller, David M J. Lilley, Phoebe A. Rice, and Joseph A. Piccirilli. Crystal Structure of the Varkud Satellite Ribozyme. *Nat. Chem. Biol.*, 11:840–846, 2015.

[124] Wade C. Winkler, Ali Nahvi, Adam Roth, Jennifer A. Collins, and Ronald R. Breaker. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, 428: 281–286, 2004.

[125] Daniel J. Klein and Adrian R. Ferré-D'Amaré. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science*, 313(5794):1752–1756, Sep 2006. doi: 10.1126/science.1129666. URL `http://dx.doi.org/10.1126/science.1129666`.

[126] Ronald T. Raines. Ribonuclease A. *Chem. Rev.*, 98:1045–1066, 1998.

[127] Stephen W. Santoro and Gerald F. Joyce. A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci. USA*, 94:4262–4266, 1997.

[128] Timothy J. Wilson, Yijin Liu, Christof Domnick, Stephanie Kath-Schorr, and David M. J. Lilley. The Novel Chemical Mechanism of the Twister Ribozyme. *J. Am. Chem. Soc.*, 138(19):6151–6162, 2016. doi: 10.1021/jacs.5b11791. URL `http://pubs.acs.org/doi/abs/10.1021/jacs.5b11791`.

[129] Mireia Garcia-Viloca, Jiali Gao, Martin Karplus, and Donald G. Truhlar. How enzymes work: Analysis by modern rate theory and computer simulations. *Science*, 303:186–195, 2004.

[130] Jiali Gao, Shuhua Ma, Dan T. Major, Kwangho Nam, Jingzhi Pu, and Donald G. Truhlar. Mechanisms and Free Energies of Enzymatic Reactions. *Chem. Rev.*, 106(8):3188–3209, 2006.

[131] Jennifer Gebetsberger and Ronald Micura. Unwinding the twister ribozyme: from structure to mechanism. *WIREs RNA*, 8:1402, 2016. doi: 10.1002/wrna.1402.

[132] Subrata Panja, Boyang Hua, Diego Zegarra, Taekjip Ha, and Sarah A. Woodson. Metals induce transient folding and activation of the twister ribozyme. *Nat. Chem. Biol.*, 13: 1109–1114, 2017.

[133] Nikola Vusurovic, Roger B. Altman, Daniel S. Terry, Ronald Micura, and Scott C. Blanchard. Pseudoknot Formation Seeds the Twister Ribozyme Cleavage Reaction Coordinate. *J. Am. Chem. Soc.*, 139(24):8186–8193, 2017.

[134] Melek N. Ucisik, Philip C. Bevilacqua, and Sharon Hammes-Schiffer. Molecular Dynamics Study of Twister Ribozyme: Role of Mg2+ Ions and the Hydrogen-Bonding Network in the Active Site. *Biochemistry*, 55(27):3834–3846, 2016. doi: 10.1021/acs.biochem.6b00203. URL `http://pubs.acs.org/doi/abs/10.1021/acs.biochem.6b00203`.

[135] Colin S. Gaines and Darrin M. York. Ribozyme Catalysis with a Twist: Active State of the Twister Ribozyme in Solution Predicted from Molecular Simulation. *J. Am. Chem. Soc.*, 138(9):3058–3065, 2016. doi: 10.1021/jacs.5b12061. URL `http://dx.doi.org/10.1021/jacs.5b12061`.

[136] Erica A. Frankel and Philip C. Bevilacqua. Complexity in pH-Dependent Ribozyme Kinetics: Dark p$K_{\mathrm{a}}$ Shifts and Wavy Rate-pH Profiles. *Biochemistry*, 57:483–488, 2018.

[137] Timothy J. Giese and Darrin M. York. Ambient-Potential Composite Ewald Method for ab Initio Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulation. *J. Chem. Theory Comput.*, 12:2611–2632, 2016.

[138] Carmen Ka Man Tse, Jun Xu, Liang Xu, Fu Kit Sheong, Shenglong Wang, Hoi Yee Chow, Xin Gao, Xuechen Li, Peter Pak-Hang Cheung, Dong Wang, Yingkai Zhang, and Xuhui Huang. Intrinsic cleavage of RNA polymerase II adopts a nucleobase-independent mechanism assisted by transcript phosphate. *Nat. Catal.*, 2(3):228–235, 2019.

[139] William P. Jencks. A Primer for the Bema Hapothle. An Empirical Approach to the Characterization of Changing Transition-State Structures. *Chem. Rev.*, 85(6):511–527, 1985.

[140] Ming Huang, Thakshila Dissanayake, Erich Kuechler, Brian K. Radak, Tai-Sung Lee, Timothy J. Giese, and Darrin M. York. A Multidimensional B-Spline Correction for

Accurate Modeling Sugar Puckering in QM/MM Simulations. *J. Chem. Theory Comput.*, 13(9):3975–3984, 2017.

[141] Mike E. Harris, Joseph A. Piccirilli, and Darrin M. York. Enzyme transition states from theory and experiment. *Biochim Biophys Acta*, 1854:1727–1728, 2015. PMID: 26302659.

[142] Yingfu Li and Ronald R. Breaker. Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2'-Hydroxyl Group. *J. Am. Chem. Soc.*, 121:5364–5372, 1999.

[143] Kin-Yiu Wong, Hong Gu, Shuming Zhang, Joseph A. Piccirilli, Michael E. Harris, and Darrin M. York. Characterization of the reaction path and transition states for RNA transphosphorylation models from theory and experiment. *Angew. Chem. Int. Ed.*, 51: 647–651, 2012.

[144] Hong Gu, Shuming Zhang, Kin-Yiu Wong, Brian K. Radak, Thakshila Dissanayake, Daniel L. Kellerman, Qing Dai, Masaru Miyagi, Vernon E. Anderson, Darrin M. York, Joseph A. Piccirilli, and Michael E. Harris. Experimental and computational analysis of the transition state for ribonuclease A-catalyzed RNA 2'-*O*-transphosphorylation. *Proc. Natl. Acad. Sci. USA*, 110:13002–13007, 2013.

[145] Haoyuan Chen, Joseph A. Piccirilli, Michael E. Harris, and Darrin M. York. Effect of $Zn^{2+}$ binding and enzyme active site on the transition state for RNA 2'-O-transphosphorylation interpreted through kinetic isotope effects. *Biochim. Biophys. Acta, Proteins Proteomics*, 1854:1795–1800, 2015.

[146] Benjamin P. Weissman, Nan-Sheng Li, Darrin M. York, Michael Harris, and Joseph A. Piccirilli. Heavy atom labeled nucleotides for measurement of kinetic isotope effects. *Biochim. Biophys. Acta*, 1854:1737–1745, 2015.

[147] Daniel L. Kellerman, Darrin M. York, Joseph A. Piccirilli, and Michael E. Harris. Altered (transition) states: mechanisms of solution and enzyme catalyzed RNA 2*t*-O-transphosphorylation. *Curr. Opin. Chem. Biol.*, 21:96–102, aug 2014. doi: 10.1016/j.cbpa.2014.06.010. URL `http://dx.doi.org/10.1016/j.cbpa.2014.06.010`.

[148] Shuming Zhang, Hong Gu, Haoyuan Chen, Emily Strong, Edward W. Ollie, Daniel Kellerman, Danni Liang, Masaru Miyagi, Vernon E. Anderson, Joseph A. Piccirilli, Darrin M.

York, and Michael E. Harris. Isotope effect analyses provide evidence for an altered transition state for RNA 2'-O-transphosphorylation catalyzed by $Zn^{2+}$. *Chem. Commun.*, 52: 4462–4465, 2016.

[149] Denise M. Perreault and Eric V. Anslyn. Unifying the Current Data on the Mechanism of Cleavage-Transesterification of RNA. *Angew. Chem. Int. Ed.*, 36:432–450, 1997.

[150] Stephen B. delCardayré and Ronald T. Raines. Structural Determinants of Enzymic Processivity. *Biochemistry*, 33(20):6031–6037, 1994. doi: 10.1021/bi00186a001. URL `http://dx.doi.org/10.1021/bi00186a001`.

[151] Chiwook Park and Ronald T. Raines. Catalysis by ribonuclease A is limited by the rate of substrate association. *Biochemistry*, 42:3509–3518, 2003.

[152] Katarzyna Swiderek, Sergio Marti, and Inaki Tunon. Molecular mechanism of the site-specific self-cleavage of the RNA phosphodiester backbone by a twister ribozyme. *Theor. Chem. Acc.*, 136(21):0–0, 2017. doi: 10.1007/s00214-017-2060-8. URL `https://link.springer.com/article/10.1007/s00214-017-2060-8`.

[153] Şölen Ekesan and Darrin M. York. Framework for conducting and analyzing crystal simulations of nucleic acids to aid in modern force field evaluation. *J. Phys. Chem. B*, 123(22):4611–4624, 2019.

[154] D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, Tai-Sung Lee, S. Le-Grand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York, and P. A. Kollman. *AMBER 18*. University of California, San Francisco, San Francisco, CA, 2018.

[155] Tai-Sung Lee, David S. Cerutti, Dan Mermelstein, Charles Lin, Scott LeGrand, Timothy J. Giese, Adrian Roitberg, David A. Case, Ross C. Walker, and Darrin M. York. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.*, 58(10):2043–2050, 2018.

[156] Alberto Pérez, Iván Marchán, David Svozil, Jiri Sponer, Thomas E. Cheatham III, Charles A. Laughton, and Modesto Orozco. Refinement of the AMBER force field for nucleic acids: Improving the description of $\alpha/\gamma$ conformers. *Biophys. J.*, 92:3817–3829, 2007.

[157] Marie Zgarbová, Michal Otyepka, Jiří Šponer, Arnošt Mládek, Pavel Banáš, Thomas E. Cheatham III, and Petr Jurečka. Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.*, 7:2886–2902, 2011.

[158] Thomas E. Cheatham, III, Piotr Cieplak, and Peter A. Kollman. A Modified Version of the Cornell et al. Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J. Biomol. Struct. Dyn.*, 16(4):845–862, 1999.

[159] Junmei Wang, Piotr Cieplak, and Peter A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic biological molecules. *J. Comput. Chem.*, 21(12):1049–1074, 2000.

[160] Tai-Sung Lee, Yuan Hu, Brad Sherborne, Zhuyan Guo, and Darrin M York. Toward Fast and Accurate Binding Affinity Prediction with pmemdGTI: An Efficient Implementation of GPU-Accelerated Thermodynamic Integration. *J. Chem. Theory Comput.*, 13:3077–3084, June 2017. ISSN 1549-9626. doi: 10.1021/acs.jctc.7b00102.

[161] Timothy J. Giese and Darrin M. York. A GPU-Accelerated Parameter Interpolation Thermodynamic Integration Free Energy Method. *J. Chem. Theory Comput.*, 14:1564–1582, 2018.

[162] John P. Perdew, Matthias Ernzerhof, and Kieron Burke. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.*, 105:9982–9985, 1996.

[163] Carlo Adamo and Gustavo E. Scuseria. Accurate excitation energies from time-dependent density functional theory: Assessing the PBE0 model. *J. Chem. Phys.*, 111:2889–2899, 1999.

[164] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Lee Hsing, and Lee G. Pedersen. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103(19):8577–8593, 1995.

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

[166] Kwangho Nam, Qiang Cui, Jiali Gao, and Darrin M. York. Specific reaction parametrization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O, and P atoms. *J. Chem. Theory Comput.*, 3:486–504, 2007.

[167] Ming Huang, Timothy J. Giese, Tai-Sung Lee, and Darrin M. York. Improvement of DNA and RNA Sugar Pucker Profiles from Semiempirical Quantum Methods. *J. Chem. Theory Comput.*, 10(4):1538–1545, 2014. doi: 10.1021/ct401013s.

[168] Jacob Bigeleisen and Max Wolfsberg. Theoretical and experimental Aspects of Isotope Effects in Chemical Kinetics. *Adv. Chem. Phys.*, 1:15–76, 1958.

[169] Adrian R. Ferré-D'Amaré and William G. Scott. Small Self-cleaving Ribozymes. *Cold Spring Harb Perspect Biol*, 2(10):a003574, Oct 2010. doi: 10.1101/cshperspect.a003574. URL `http://dx.doi.org/10.1101/cshperspect.a003574`.

[170] Aiming Ren, Ronald Micura, and Dinshaw J. Patel. Structure-based mechanistic insights into catalysis by small self-cleaving ribozymes. *Curr. Opin. Chem. Biol.*, 41:71–83, 2017.

[171] Jonathan K. Lassila, Jesse G. Zalatan, and Daniel Herschlag. Biological phosphoryl-transfer reactions: Understanding mechanism and catalysis. *Annu. Rev. Biochem.*, 80: 669–702, 2011.

[172] Simon Auslander, Patrick Ketzer, and Jorg S. Hartig. A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression. *Mol. Biosyst.*, 6:807–814, 2010. doi: 10.1039/B923076A. URL `http://dx.doi.org/10.1039/B923076A`.

[173] Simon Ausländer and Martin Fussenegger. Synthetic RNA-based switches for mammalian gene expression control. *Curr. Opin. Biotechnol.*, 48:54–60, 2017.

[174] Hehua Liu, Xiang Yu, Yiqing Chen, Jing Zhang, Baixing Wu, Lina Zheng, Phensinee Haruehanroengra, Rui Wang, Suhua Li, Jinzhong Lin, Jixi Li, Jia Sheng, Zhen Huang, Jinbiao Ma, and Jianhua Gan. Crystal Structure of an RNA-Cleaving DNAzyme. *Nat. Commun.*, 8(1):2006–2015, 2017.

[175] Ronald R. Breaker and Gerald F Joyce. A DNA enzyme that cleaves RNA. *Chem. Biol.*, 1(4):223–229, 1994.

[176] Colin S. Gaines and Darrin M. York. Model for the Functional Active State of the TS Ribozyme from Molecular Simulation. *Angew. Chem. Int. Ed.*, 129:13577–13580, 2017. doi: https://doi.org/10.1002/anie.201705608.

[177] Tai-Sung Lee, Brian K. Radak, Michael E. Harris, and Darrin M. York. A Two-Metal-Ion-Mediated Conformational Switching Pathway for HDV Ribozyme Activation. *ACS Catal.*, 6(3):1853–1869, 2016. doi: 10.1021/acscatal.5b02158.

[178] Jonathan Perreault, Zasha Weinberg, Adam Roth, Olivia Popescu, Pascal Chartrand, Gerardo Ferbeyre, and Ronald R Breaker. Identification of hammerhead ribozymes in all domains of life reveals novel structural variations. *PLoS Comput. Biol.*, 7(5):1002031, May 2011. doi: 10.1371/journal.pcbi.1002031. URL `http://dx.doi.org/10.1371/journal.pcbi.1002031`.

[179] H. C. Guo, D. M. De Abreu, E. R. Tiller, B. J. Saville, J. E. Olive, and R. A. Collins. Nucleotide sequence requirements for self-cleavage of Neurospora VS RNA. *J. Mol. Biol.*, 232:351–361, 1993.

[180] Timothy J. Wilson, Aileen C. McLeod, and David M. J. Lilley. A guanine nucleobase important for catalysis by the VS ribozyme. *EMBO J.*, 26:2489–2500, 2007.

[181] Angela A. Andersen and Richard A. Collins. Rearrangement of a Stable RNA Secondary Structure during VS Ribozyme Catalysis. *Mol. Cell*, 5(1):469–478, 2000.

[182] Richard Shippy, Andrew Siwkowski, and Arnold Hampel. Mutational Analysis of Loops 1 and 5 of the Hairpin Ribozyme. *Biochemistry*, 37:564–570, 1998.

[183] Mercedes Pérez-Ruiz, Alicia Barroso-delJesus, and Berzal-Herranz, Alfredo. Specificity of the Hairpin Ribozyme: Sequence Requirements Surrounding the Cleavage Site. *J. Biol. Chem.*, 274:29376–29380, 1999.

[184] Roland K O Sigel and Helmut Sigel. A stability concept for metal ion coordination to single-stranded nucleic acids and affinities of individual sites. *Acc. Chem. Res.*, 43(7):974–984, Jul 2010. doi: 10.1021/ar900197y. URL `http://dx.doi.org/10.1021/ar900197y`.

[185] George Slim and Michael J. Gait. Configurationally defined phosphorothioate-containing oligoribonucleotides in the study of the mechanism of cleavage of hammerhead ribozymes. *Nucleic Acids Res.*, 19:1183–1188, 1991.

[186] Shenglong Wang, Katrin Karbstein, Alessio Peracchi, Leonid Beigelman, and Daniel Herschlag. Identification of the Hammerhead Ribozyme Metal Ion Binding Site Responsible for Rescue of the Deleterious Effect of a Cleavage Site Phosphorothioate. *Biochemistry*, 38(43):14363–14378, 1999.

[187] E. M. Osborne, J. E. Schaak, and V. J. Derose. Characterization of a native hammerhead ribozyme derived from schistosomes. *RNA*, 11:187–196, 2005.

[188] W Luke Ward and Victoria J Derose. Ground-state coordination of a catalytic metal to the scissile phosphate of a tertiary-stabilized Hammerhead ribozyme. *RNA*, 18:16–23, 2012.

[189] Timothy J Wilson, Yijin Liu, Nan Sheng Li, Qing Dai, Joseph A Piccirilli, and David MJ Lilley. Comparison of the structures and mechanisms of the pistol and hammerhead ribozymes. *J. Am. Chem. Soc.*, 141:7865–7875, 2019.

[190] Timothy J. Wilson and David M.J. Lilley. Do the hairpin and VS ribozymes share a common catalytic mechanism based on general acid-base catalysis? A critical assessment of available experimental data. *RNA*, 17:213–221, 2011.

[191] Juliane K. Soukup. *The Structural and Functional Uniqueness of the glmS Ribozyme*, volume 120, chapter 5, pages 173–193. Academic Press : New York, 2014.

[192] Reed M. Izatt, James J. Christensen, and J. Howard Rytting. Sites and thermodynamic quantities associated with proton and metal ion interaction with ribonucleic acid, deoxyribonucleic acid, and their constituent bases, nucleosides, and and nucleotides. *Chem. Rev.*, 71:439–481, 1971.

[193] Aamir Mir, Ji Chen, Kyle Robinson, Emma Lendy, Jaclyn Goodman, David Neau, and Barbara L. Golden. Two Divalent Metal Ions and Conformational Changes Play Roles in the Hammerhead Ribozyme Cleavage Reaction. *Biochemistry*, 54:6369–6381, 2015. doi: 10.1021/acs.biochem.5b00824. URL `http://pubs.acs.org/doi/10.1021/acs.biochem.5b00824`.

[194] V. D. Sood, T. L. Beattie, and R. A. Collins. Identification of phosphate groups involved in metal binding and tertiary interactions in the core of the Neurospora VS ribozyme. *J. Mol. Biol.*, 282:741–750, 1998.

[195] Yana S. Kovacheva, Svetomir B. Tzokov, Iain A. Murray, and Jane A. Grasby. The role of phosphate groups in the VS ribozyme-substrate interaction. *Nucleic Acids Res.*, 32: 6240–6250, 2004.

[196] Maria T. Panteva, George M. Giambasu, and Darrin M. York. Force Field for $Mg^{2+}$, $Mn^{2+}$, $Zn^{2+}$, and $Cd^{2+}$ Ions that have Balanced Interactions with Nucleic Acids. *J. Phys. Chem. B*, 119:15460–15470, 2015.

[197] Filip Leonarski, Luigi D'Ascenzo, and Pascal Auffinger. $Mg^{2+}$ ions: do they bind to nucleobase nitrogens? *Nucleic Acids Res.*, 45:987–1004, 2017.

[198] Saurja DasGupta, Nikolai B. Suslov, and Joseph A. Piccirilli. Structural Basis for Substrate Helix Remodeling and Cleavage Loop Activation in the Varkud Satellite Ribozyme. *J. Am. Chem. Soc.*, 139:9591–9597, 2017.

[199] Abir Ganguly, Benjamin P. Weissman, Joseph A. Piccirilli, and Darrin M. York. Evidence for a Catalytic Strategy to Promote Nucleophile Activation in Metal-Dependent RNA-Cleaving Ribozymes and 8-17 DNAzyme. *ACS Catal.*, 9(12):10612–10617, 2019.

[200] Fatima D. Jones and Scott A. Strobel. Ionization of a critical adenosine residue in the neurospora Varkud Satellite ribozyme active site. *Biochemistry*, 42:4265–4276, 2003.

[201] M. Duane Smith and Richard A. Collins. Evidence for proton transfer in the rate-limiting step of a fast-cleaving Varkud satellite ribozyme. *Proc. Natl. Acad. Sci. USA*, 104:5818–5823, 2007.

[202] Yaroslav I. Kuzmin, Carla P. Da Costa, Joseph W. Cottrell, and Martha J. Fedor. Role of an active site adenine in hairpin ribozyme catalysis. *J. Mol. Biol.*, 349:989–1010, 2005.

[203] Ian T. Suydam, Stephen D. Levandoski, and Scott A. Strobel. Catalytic Importance of a Protonated Adenosine in the Hairpin Ribozyme Active Site. *Biochemistry*, 49:3723–3732, 2010.

[204] Mathew Vogt, Sahiri Lahiri, Charles G. Hoogstraten, David R. Britt, and Victoria J. DeRose. Coordination Environment of a Site-Bound Metal Ion in the Hammerhead Ribozyme Determined by $^{15}N$ and $^{2}H$ ESEEM Spectroscopy. *J. Am. Chem. Soc.*, 128: 16764–16770, 2006.

[205] Kenneth F. Blount and Olke C. Uhlenbeck. The Structure-Function Dilemma of the Hammerhead Ribozyme. *Annu. Rev. Biophys. Biomol. Struct.*, 34:415–440, 2005.

[206] Jason M Thomas and David M Perrin. Probing general acid catalysis in the hammerhead ribozyme. *J Am Chem Soc*, 131(3):1135–1143, Jan 2009. doi: 10.1021/ja807790e. URL `http://dx.doi.org/10.1021/ja807790e`.

[207] Erica A. Frankel, Christopher A. Strulson, Christine D. Keating, and Philip C. Bevilacqua. Cooperative Interactions in the Hammerhead Ribozyme Drive pKa Shifting of G12 and Its Stacked Base C17. *Biochemistry*, 56:2537–2548, 2017. doi: 10.1021/acs.biochem.7b00174.

[208] Sandro Neuner, Christoph Falschlunger, Elisabeth Fuchs, Maximilian Himmelstoss, Aiming Ren, Dinshaw J. Patel, and Ronald Micura. Atom-Specific Mutagenesis Reveals Structural and Catalytic Roles for an Active-Site Adenosine and Hydrated $Mg^{2+}$ in Pistol Ribozymes. *Angew. Chem. Int. Ed.*, 56:15954–15958, 2017.

[209] Tai-Sung Lee, Carlos Silva-Lopez, Monika Martick, William G. Scott, and Darrin M. York. Insight into the role of $Mg^{2+}$ in hammerhead ribozyme catalysis from x-ray crystallography and molecular dynamics simulation. *J. Chem. Theory Comput.*, 3:325–327, 2007.

[210] Tai-Sung Lee, Carlos Silva Lopez, George M. Giambaşu, Monika Martick, William G. Scott, and Darrin M. York. Role of $Mg^{2+}$ in hammerhead ribozyme catalysis from molecular simulation. *J. Am. Chem. Soc.*, 130(10):3053–3064, 2008.

[211] Tai-Sung Lee and Darrin M York. Computational mutagenesis studies of hammerhead ribozyme catalysis. *J. Am. Chem. Soc.*, 132(38):13505–13518, 2010. doi: 10.1021/ja105956u. URL `http://dx.doi.org/10.1021/ja105956u`.

[212] Kin-Yiu Wong, Tai-Sung Lee, and Darrin M. York. Active participation of the $Mg^{2+}$ ion in the reaction coordinate of RNA self-cleavage catalyzed by the hammerhead ribozyme. *J. Chem. Theory Comput.*, 7(1):1–3, 2011.

[213] Yuka Nakamatsu, Masaki Warashina, Tomoko Kuwabara, Yoshiyuki Tanaka, Koichi Yoshinari, and Kazunari Taira. Significant activity of a modified ribozyme with N7-deazaguanine at G[sub 10.1]: the double-metal-ion mechanism of catalysis in reactions catalysed by hammerhead ribozymes. *Genes Cells*, 5:603–612, 2000.

[214] Alessio Peracchi, Leonid Beigelman, Nassim Usman, and Daniel Herschlag. Rescue of abasic hammerhead ribozymes by exogenous addition of specific bases. *Proc. Natl. Acad. Sci. USA*, 93:11522–11527, 1996.

[215] Po-Jung Jimmy Huang and Juewen Liu. Rational evolution of $Cd^{2+}$-specific DNAzymes with phosphorothioate modified cleavage junction and $Cd^{2+}$ sensing. *Nucleic Acids Res.*, 43(12):6125–6133, 2015.

[216] Alessio Peracchi, Maria Bonaccio, and Marcello Clerici. A mutational analysis of the 8–17 deoxyribozyme core. *J. Mol. Biol.*, 352:783–794, 2005.

[217] Kimberly A. Harris, Christina E. Lünse, Sanshu Li, Kenneth I. Brewer, and Ronald R. Breaker. Biochemical analysis of pistol self-cleaving ribozymes. *RNA*, 21:1852–1858, 2015. doi: 10.1261/rna.052514.115. URL `http://www.rnajournal.org/cgi/doi/10.1261/rna.052514.115`.

[218] Matthew W L Lau and Adrian R Ferré-D'Amaré. An *in vitro* evolved *glmS* ribozyme has the wild-type fold but loses coenzyme dependence. *Nat. Chem. Biol.*, 9:805–812, 2013.

# Appendix A
# Supplemental Media

## A.1 TS ribozyme: Mg Acid Model

3-D structure file in PDB format (TSrz_Mg_Acid_Model.pdb) representative of the Mg Acid Model derived from molecular dynamics simulation of the TS ribozyme.

## A.2 TS ribozyme: C7 Acid Model

3-D structure file in PDB format (TSrz_C7_Acid_Model.pdb) representative of the C7 Acid Model derived from molecular dynamics simulation of the TS ribozyme.

## A.3 Twister ribozyme: Summary Movie

Movie (Twister_Summary_Movie.mp4) summarizing the key results and models obtained by applying the Computational RNA Enzymology approach to the twister ribozyme.

## A.4 L-platform/L-scaffold: PyMOL Session

Zip Archive (Lplatform_for_Pymol.zip) including a PyMOL script (L_platform.pml) and collection of PDB files comparing the active site L-platform/L-scaffold motifs for the following ribozymes: Twister (twister_Lplatform.pdb), Hairpin (hairpin_Lplatform.pdb), VS (VS_Lplatform.pdb), Hammerhead (hammerhead_Lplatform.pdb), 8-17 DNAzyme (DNAzyme_Lplatform.pdb), and Pistol (pistol_Lplatform.pdb).