January 2016

Ribozymes of the RNA World: Investigating Ancient Enzymes through Modern Self-Cleaving Ribozymes.

Aamir Mir
Purdue University

Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Aamir Mir

Entitled

Ribozymes of the RNA World: Investigating Ancient Enzymes through Modern Self-Cleaving Ribozymes.

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Barbara L. Golden
Chair

Daisuke Kihara

Frederick S. Gimble

Elizabeth J. Tran

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Barbara L. Golden

Approved by: Andrew D. Mesecar 11/29/2016

Head of the Departmental Graduate Program Date

RIBOZYMES OF THE RNA WORLD: INVESTIGATING ANCIENT ENZYMES THROUGH MODERN SELF-CLEAVING RIBOZYMES

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Aamir Mir

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2016

Purdue University

West Lafayette, Indiana

To my mother

ACKNOWLEDGEMENTS

There are many people who I would like to acknowledge who helped me get here. First, I am very grateful to my advisor, Dr. Barbara Golden for patiently guiding me through the ups and downs of my graduate school. Without her support and guidance, none of this would have been possible. I am thankful to her for helping me with research, publications, writing of this thesis, and for other opportunities she provided me with to further my career as a scientist. I would also like to thank the rest of my thesis committee members, Dr. Elizabeth Tran, Dr. Frederick Gimble, and Dr. Daisuke Kihara, for their critical feedback and support during my graduate career. I am also grateful to the Tran lab members (past and present) for their helpful comments, and help with research.

I am especially grateful to Dr. Ji Chen for his help with lab work, and for insightful discussions. Other graduate students of the Golden lab have also been instrumental in my success including Samantha Lee and Rui Gan. I am also thankful to my undergraduate mentees Jaclyn Goodman, Emma Lendy, Kyle Robinson, and others for their contributions to my thesis project. Lastly, I would like to acknowledge the support of my friends and family during my Ph.D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2’ -OH	2’-hydroxyl
RNA	ribonucleic acids
G	guanine
C	cytosine
A	adenine
U	uracil
EDTA	Ethylenediaminetetraacetic acid
Tris-HCl	Tris(hydroxymethyl)aminomethane Hydrochloride
TBE	Tris Borate EDTA
nt	number of nucleotides
WT	wild-type
RzB	SELEX derived hammerhead ribozyme
bp	base pairs
cDNA	complementary deoxyribonucleic acid
At	*Arabidopsis thaliana*
mRNA	messenger RNA
QM/MM	quantum mechanics/molecular mechanics
k_{obs}	observed rate constant
tRNA	transfer RNA
HDV	hepatitis delta virus
snRNA	small nuclear RNA
cryo-EM	cryo-electron microscopy

ABSTRACT

Mir, Aamir. Ph.D., Purdue University, December 2016. Ribozymes of the RNA World: Investigating Ancient Enzymes through Modern Self-Cleaving Ribozymes. Major Professor: Barbara Golden.

Ribozymes are RNAs that can perform catalytic functions. The discovery of ribozymes was an important milestone in science because it not only opened new avenues of research, but also provided insights into the origins of life on Earth. Early life most likely used RNA as catalysts for diverse functions, however, it is unclear how RNA could have performed such diverse catalytic functions. To answer this question, my work has focused on the most abundant natural class of ribozymes in the modern world, called the self-cleaving ribozymes. They catalyze the cleavage of a specific phosphodiester bond in RNA. We investigated the different catalytic strategies used by one such ribozyme, called the hammerhead ribozyme, and uncovered how RNA can utilize tautomerization, divalent metals, and folding to diversify its chemical repertoire. To learn more about the catalytic strategies used by RNA and their biological roles, we developed a method that combines biochemistry of self-cleaving cleaving ribozymes with high-throughput sequencing to discover new self-cleaving ribozymes. New self-cleaving ribozymes will provide insights about their biological roles, and expand the catalytic toolkit available to RNA. Hence, by studying modern self-cleaving ribozymes,

we can learn how ancient ribozymes catalyzed diverse chemical reactions, and the roles they may have played to sustain early life.

CHAPTER 1. THE RNA WORLD HYPOTHESIS

1.1 Perspective

The 'RNA world' hypothesis is a fascinating idea that suggests that in the primordial world, RNA could have existed as an information carrier and as a catalyst (*1*). Such a world may have existed more than four billion years ago, and possibly may have been the key to life's origins. The idea itself originated in the 1960s, when Orgel, Crick and others hypothesized that RNA could be the precursor molecule to all life on Earth (*2, 3*). However, the term 'RNA world hypothesis' was not coined until 1986, when the first ribozymes were discovered (*4, 5*). This illustrates that ribozymes are the backbone that the RNA world hypothesis was built on, and continue to be the most important pieces of evidence for the existence of such a world. The question of how an RNA world would have started in the prebiotic world is important, not only because it attempts to explain how life started on Earth, but also because interest in the RNA world has encouraged researchers to examine the question of what conditions are ideal for life to begin. The answers to these questions are especially important as we look for signs of life in the celestial bodies deep in space.

1.2 Making a case for the existence of the RNA world based on observations from the modern world

Establishing the historical accuracy of the RNA world is nearly impossible, as it occurred over four billion years ago (*6*). Therefore, we have to look for clues in the modern world to support whether an RNA world ever existed. Evolutionary processes almost always leave fossils that paint a picture of the past. In the modern world, there is considerable evidence based on ancient fossils to support that an RNA world existed (*7-9*). We can also establish that RNA came before DNA and protein. However, whether life started with RNA is still being investigated, and evidence for that will be reviewed later.

Perhaps the most important piece of evidence in support of the RNA world is the ribosome. The ribosome is central to life because it is responsible for translating the genetic code. It is where protein synthesis occurs inside the cells. Due to its importance for life, the core of the ribosome is universally conserved across all domains of life (*10*). Therefore, the fact that the ribosome is a ribozyme, is an extremely important piece of evidence in support of the RNA world hypothesis (*11*). The active site of the ribosome is found deep within its conserved core, which is entirely made up of RNA. This is where the peptidyl transferase reaction takes place. It is catalyzed exclusively by the ribosomal RNA (rRNA) (*12*). In fact, there are no proteins found within 18 Å of the active site of the ribosome (*13*). Proteins only appear to be playing a structural role by decorating the outside of the ribosome. It appears that the ribosome must have evolved during the

transition from the RNA world to the protein world, when the genetic code was beginning to take shape (*9*).

It is conceivable that in the RNA world, amino acids, ribonucleotides and other functional compounds acted as coenzymes (*14*). Using *in-vitro* evolution, Connell and Christian showed that a coenzyme A dependent RNA transacylase can be evolved, which significantly enhances the catalytic repertoire available to RNA in the primordial world (*15*). Miller and colleagues even demonstrated a prebiotic synthesis route to coenzyme A, indicating that prebiotic world was capable of forming cofactors to assist RNA catalysis (*16*). Some researchers have speculated that RNA was probably covalently bound to these different cofactors, including amino acids in the RNA world (*17, 18*). Over time, as proteins began to take the dominant catalytic role, the role of RNA diminished to the point where RNA-like cofactors mediate protein-lead catalysis. The ribonucleotide cofactors are thus the fossils of the RNA world. We can observe several of these fossils today, such as, flavin adenine dinucleotide (FAD), nicotinamide mononucleotide (NMN+), nicotinamide adenine dinucleotide (NAD+), nicotinamide adenine dinucleotide phosphate (NADP+) etc. These cofactors contain either a ribose group, phosphate group or an RNA base. This led White to suggest that the adenine ribonucleotide is a vestige of the RNA world (*14*). Many modern cofactors appear to be derived from the adenine ribonucleotide, and it usually only plays a structural role rather than a functional one. Flavin mononucleotide (FMN) is a coenzyme which is very similar to FAD. It lacks the ribose and adenine. Both of these cofactors are seen to be utilized by different enzymes which catalyze very similar reactions (*19*).

There are also several other lines of observational evidence in the modern world that support the existence of the RNA world. RNA is known to fold into complicated and dynamic tertiary structures, which was only thought to be the property of proteins (*20-22*). Discovery of several ribozymes, such as the ribonuclease P, self-splicing introns, and self-cleaving RNAs add further indirect evidence for the presence of the RNA world (*4, 5, 23*). There are also examples of RNA which modulate their function based on the environmental stimuli (*24*). These RNAs are called riboswitches, which are mRNA regulatory elements that bind small molecules and undergo a conformational change which modulates the translation of mRNAs (*25*). There is even a riboswitch that is a ribozyme called the glucosamine-6-phopshate activated (*glmS*) ribozyme (*26*). It shows that, in principle, RNA could not only perform catalysis but could also respond to environmental stresses and provide feedback regulation. Hence, an RNA world, where RNA was the central molecule capable of sustaining life is not such a far-fetched idea.

1.3 Emergence of the early RNA – from prebiotic soup to an RNA polymer

If life started with the RNA world, it is likely that RNA came before any other biomolecule. The first step in the RNA world was the formation of the RNA monomers from the chemicals available in the prebiotic soup (Figure 1.1). A number of experiments have been performed over the years which give us an idea of the chemical compounds most likely to be present in the prebiotic world (*27*). The formose reaction pathway that utilizes formaldehyde is capable of yielding several sugars, which could provide the

simplest pathway to complex sugars such as ribose (*28*). Ritson and Sutherland also uncovered synthesis routes for simple sugars using HCN, UV radiation and cyanometallates (*29*). The famous Miller-Urey experiment provided several compounds including a few amino acids (*30*). Upon re-analysis of the Miller-Urey experiment, many more amino acids were also identified (*31, 32*). Possible synthesis routes towards purine and pyrimidine nucleobases have also been identified (*33, 34*). Interestingly, analysis of the Murchison meteorite revealed a number of amino acids, purines and pyrimidines, thus mirroring the products of the classic prebiotic synthesis experiments (*35, 36*). Hence, the necessary precursors for RNA were present in the prebiotic world. Whether, they were extra-terrestrial or synthesized on Earth is beyond the scope of this discussion.

The precursors available on Earth needed to form activated ribonucleotides to kick-start the RNA world. The complexities this presented were famously referred to as the 'prebiotic chemist's nightmare' (*37*). However, recent progress has been able to unravel most of those complexities. Researchers have demonstrated that under mild conditions, precursors available in the prebiotic soup can be made to form activated ribonucleotides. Powner et al., showed that activated pyrimidine ribonucleotides can be formed from precursors such as cyanamide, cyanoacetylene, glycolaldehyde, glyceraldehyde, and inorganic phosphate (*38*). The role of inorganic phosphate in this synthesis is immensely important, because it provides a pH buffer, acts as a catalyst, and is also a reactant itself. Pasek et al., showed that several phosphate species are plausible when iron-rich meteorites interact with water (*39*). The only problem with this synthesis

route is that it needs specific conditions at every step and therefore is not a one-pot solution. However, central to the activated ribonucleotide synthesis is a systems chemistry approach, where different chemical reactions are interdependent on each other, and where flux and feedback mechanisms are very important (*40*). Such an approach can overcome many of these challenges. Another criticism of the early synthesis of RNA is the chemical heterogeneity present in the prebiotic world. How can a specific molecule like RNA emerge from so many different chemicals? It has been shown that small differences in chemical composition can be amplified to decrease heterogeneity. For example, we know that life depends on levorotatory amino acids and dextrorotatory sugars. However, any synthesis of these compounds usually produces a racemic mixture. But, it appears that perfect racemic mixtures are nearly impossible (*41, 42*). There are always small stochastic differences that get amplified through feedback mechanisms. Lastly, heterogeneity is not all that problematic; RNA is actually tolerable of some chemical heterogeneity. For example, heterogeneity related to 2' -5' versus 3'-5' linkages in RNA can not only allow RNA folding, but also provide benefits by lowering the melting temperature of RNA (*43*). Similarly, the Szostak group also showed that ribose and deoxyribose are both compatible with RNA (*44*). Therefore, the synthesis of activated nucleotides, while complicated, is certainly plausible in the prebiotic world.

The next logical step towards an RNA world would be the formation of RNA polymers from activated nucleotides (Figure 1.1). Activated nucleotides, such as, nucleoside 5' polyphosphates are reactive species which can react to generate an RNA polymer. This non-enzymatic polymerization of RNA has been achieved by using

nucleoside 5'-phosphoramidates including nucleoside 5'-phophorimidazolides (*45, 46*). Cafferty et al., showed the importance of hydrophobic effect in generating long polymers from weakly reactive cyanuric acid and a triaminopyrimidine (*47*). This hydrophobic effect allows RNA bases to stack on top of each other and hold long RNA strands together. RNA polymerization paves the way for life to start by allowing the evolutionary processes to begin.

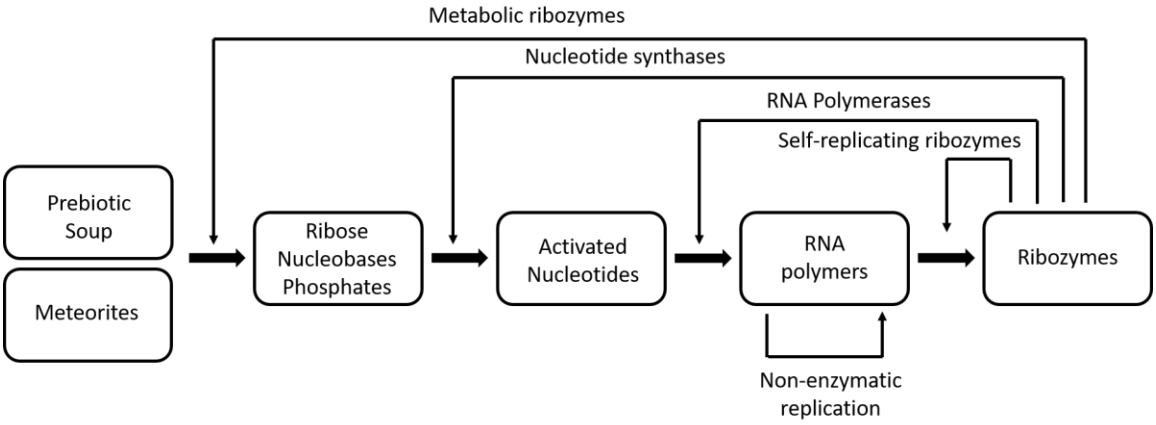

Figure 1.1. Emergence of the RNA world.
The flowchart shows how the RNA world may have proceeded four billion years ago. It shows that both the chemicals available on Earth, and extra-terrestrial chemicals may have been involved in the synthesis of RNA. The synthesis of RNA is expected to have proceeded via non-enzymatic pathways until the early ribozymes evolved. Ribozymes could theoretically then be involved in manufacturing the precursors, RNA monomers, RNA polymers, and possibly could play many other functions.

1.4 Beginnings of Natural Selection - non-enzymatic RNA replication systems

A popular definition of life is any "self-sustaining chemical system capable of Darwinian evolution" (*48*). This is also what NASA refers to as life. At the heart of this

definition is the idea of replication. Replication allows natural selection to act, which in turn allows evolutionary processes to take hold. Initially, in the primordial world, replication had to take place non-enzymatically (Figure 1.1). RNA polymers should be able to make more copies of themselves, and furthermore this replication should be template directed. Several studies have shown that it is possible to replicate short strands of RNA via non-enzymatic polymerization (*49, 50*). The Orgel group discovered that pre-existing a polyC template can direct the synthesis of polyG and vice-versa, and this synthesis is pH-independent to an extent (*51, 52*). This synthesis route did not perform well with all substrates, and suffered from product inhibition. However, it was able to demonstrate, in principle, that non-enzymatic replication of RNA is possible. Recently, a study found that if the RNA strands are immobilized, and if the solution containing the activated ribonucleotides is periodically displaced, RNA strands can be efficiently replicated (*53*). Furthermore, all four nucleotides can be incorporated in the RNA. The temperature appeared to be an important factor in this case. At 0 °C, nucleotides can bind the template more readily and the rate of hydrolysis of the RNA strand is very slow as well. This shows that there may have been a cold start to life, and that RNA synthesis may have occurred on the surface. Non-enzymatic replication mechanisms appear to only produce oligonucleotides of lengths 30-50 nt. Hence, there may have been some mechanisms where longer RNAs were synthesized from template-directed ligation. Orgel and other researchers have demonstrated non-enzymatic template-directed ligation of shorter RNAs (*54-56*). Template-directed ligation is not only capable of producing longer RNAs, but is also more tolerant of non-natural

nucleotides. Some limitations of these approaches include G•U pairing which would affect the fidelity of the replication. Furthermore, long stretches of Cs and Gs can form stable self-structures that would have been hard to replicate. However, these studies still show us a way of how early RNA polymers could have replicated themselves. Further research may overcome the current limitations of these approaches. The primordial world where RNA could copy itself via ligation and replication must have been the starting point for life. The start of evolutionary processes set the stage for the advent of ribozymes, and hence the RNA world.

1.5 The ribozymes of the RNA world – lessons from *in-vitro* evolution studies

Ribozymes are RNAs that are capable of performing catalysis. They are central pieces of evidence in support of the RNA world hypothesis. If the RNA world was able to sustain life, there had to be ribozymes carrying out a variety of functions. These functions include replication, ligation, processing, and metabolism. There is an abundance of research on different types of ribozymes that have been evolved using *in-vitro* selection (Table 1.1). *In-vitro* evolution studies provide an insight into what kind of ribozymes may have evolved during the RNA world. These ribozymes catalyze a number of different chemical reactions such as RNA cleavage, RNA ligation, RNA phosphorylation, RNA capping, alkylation, aminoacylation, aminoacylation, peptide bond formation, etc. (Table 1.1). This is evidence that once RNA had the ability to replicate itself, natural selection may have found ways to evolve these, and more

ribozymes. One area of catalysis where the ribozymes are limited in activity is metabolism (*57*). The RNA world needed to have a constant supply of activated ribonucleotides, and other chemicals. Further research may provide us with efficient ribozymes that can carry out metabolic functions. However, that is not strictly necessary. It has been argued that metabolic function may have been carried out by environmental catalysts (*57*). Studies have shown that reactions analogous to glycolysis, and pentose phosphate pathway are carried out by environmental catalysts. Such catalysts include metal ions, and inorganic phosphates.

1.5.1 RNA Polymerases

While non-enzymatic replication of RNA was important for evolution, an efficient RNA polymerase would have greatly accelerated natural selection. An RNA polymerase ribozyme that uses RNA as a template to produce RNA, is seen as the holy grail of ribozymes for the RNA world. The first such demonstration of RNA polymerization came in 2001, when Johnson et al used the core of the class I ligase to engineer a general RNA polymerase ribozyme called R18 (*58*). They combined engineering with several rounds of *in-vitro* evolution to generate a ribozyme that could perform template directed polymerization, and add 14 nucleotides to an RNA primer. The polymerase was very accurate since it achieved a fidelity rate of 96 %. While this was a big achievement, R18 was unable to synthesize longer templates. Hence, it could not even replicate itself. However, Attwater et al., were able to achieve this feat in 2013 when their ribozyme

called the tC9Y was able to synthesize 206 nt. This success was based on an in-ice selection of ribozyme polymerase activity (*59*). This polymerization activity was also very accurate with a fidelity rate of about 97 %. The tC9Y ribozyme is a significant milestone for the RNA world hypothesis, and key to its success was RNA stabilizing cold conditions. However, there is still one limitation, that it needs to overcome. The RNA synthesized in this case was not the sequence of the ribozyme itself. The ribozyme also suffered from issues, such as, the folding of the newly synthesized RNA, which prevented the ribozyme from reading through the entire template. For now, it seems that the search for the true self-replicating RNA polymerase will continue.

1.5.2 RNA processing ribozymes (ligation and cleavage)

Many ribozymes have been evolved using *in-vitro* evolution approaches that are in involved in RNA processing. Such processing ribozymes include ribozymes that cleave RNA, and ribozymes that can ligate two pieces of RNAs. These activities would have been very important in the RNA world. There is an abundance of self-cleaving ribozymes in nature (reviewed in later chapters, Table 5.1). These include the hammerhead ribozymes, HDV ribozyme, twister ribozymes, and others. Based on bioinformatics, these ribozymes have been found to be widespread in nature (Table 5.1). Furthermore, these self-cleaving ribozymes are the result of convergent evolution since their sequences are not highly conserved. Therefore, it is not surprising that self-cleaving ribozyme are relatively easy to evolve *in-vitro*. The hammerhead ribozyme, in particular,

shows up in most of the *in-vitro* evolution studies (*60, 61*). It is the simplest RNA motif found in nature that can catalyze its own cleavage. Tang and Breaker have found that selecting for Mg^{2+} dependent self-cleaving ribozymes using an *in-vitro* selection approach produces several distinct RNA motifs that can catalyze RNA cleavage reactions (*62*). These ribozymes produce a considerable rate enhancement (up to 10^6) compared to spontaneous RNA cleavage. In addition, other artificial self-cleaving ribozymes have also been found including lead and iron dependent self-cleaving ribozymes (*63, 64*). These self-cleaving ribozymes would have been important players in the RNA world as they could play functional roles in the expression of RNA.

Ligation is another activity that would have been fundamental to the RNA world hypothesis. Ligation combined with RNA replication, and RNA cleavage is capable of producing a very diverse array of RNAs through network effects (*40*). Such network effects have been called the mutually autocatalytic RNA networks. These RNA networks are extremely significant for the RNA world, because they are capable of forming RNAs of different sizes, sequences, and structures. In 1993, Bartel and Szostak were able to evolve RNA ligases that enabled 2'-OH or the 3'-OH of one RNA to attack the phosphate of an adjacent 5'-triphosphate (*65*). These RNAs are already aligned on a template. Rate enhancements found in this case were millions of times faster than the uncatalyzed reaction. It should be noted that natural splicing ribozymes also use a similar mechanism (reviewed in Chapter 2). Later, many other classes of RNA ligases were also discovered including the Class I ribozyme, which was the basis of the first polymerase ribozyme (*66*). That is because it catalyzes a reaction very similar to natural RNA polymerase

enzymes. It catalyzes the formation of the 3'-5' linkage of RNA. Similarly, a short ribozyme ligase was found by Robertson et al., which was also capable of joining non-Watson Crick base pairs (*67*). Many such ligase ribozymes have been found to achieve tremendous rate enhancement, and catalyze multiple turnover reactions (*68*). Furthermore, Vlassov et al., found that in cold conditions fragmented hairpin ribozymes could catalyze the ligation of many different RNAs (*69*). They suggest that freezing environments on Earth could have generated very diverse RNAs through ligation.

While the search for self-replicating RNA polymerase is still going on, a self-replicating ligase ribozyme has already been discovered (*70*). The RNA world could also have relied on something comparable to such a ribozyme. The R3C ligase ribozyme ligates two RNA fragments of itself. The newly formed sequence can serve as a template for more ligations. However, the new RNA was unable to form a good complex with the substrates, which limited exponential amplification. Furthermore, the substrates themselves could anneal to each other. A cross-catalytic system was therefore devised, that uses two different ribozymes which catalyze the ligation of each other's substrates (*71*). This approach is able to yield a strategy of amplifying the sequence of the ligase ribozyme themselves, and forms the basis of mutually autocatalytic network effects.

1.5.3 Other ribozymes

Apart from RNA processing, and RNA replication activities, there are several other ribozymes that might have been useful in an RNA world. Ribozymes capable of

nucleotide synthesis from precursor molecules, and ribozymes that could activate those nucleotides would have come in extremely handy in the primordial world. These ribozymes would have allowed the RNA world to flourish. *In-vitro* selection approaches have produced a few ribozymes which can perform such activities (*72, 73*). For instance, the a15 ribozyme is a nucleotide synthase ribozyme that catalyzes the synthesis of uridine monophosphate from 5-phosphoribosyl 1-pyrophosphate (PRPP) and uracil (*74*). Similarly, the MA ribozyme catalyzes the formation of guanosine monophosphate from PRPP and guanine (*75*). There are other examples of purine and pyrimidine synthase ribozymes in which the glyosidic bond between the ribose and the nucleobase is made. Interestingly, in these studies, purine synthases tend to be much more efficient compared to pyrimidine synthases. The PRPP substrate is an extremely important compound in modern biology. Most of the nucleotide biosynthesis involves PRPP, and it is plausible that it existed in the RNA world.

After the synthesis of nucleotides, they need to be activated before they can be used for polymerization reactions. To activate the nucleotides, the ribozymes of the RNA world should have been able to catalyze the triphosphorylation of the available nucleosides. In this case, the prebioticially plausible starting materials are polyphosphates in addition to nucleosides. One such polyphosphate is the cyclic trimetaphopshate (Tmp), which was used to evolve ribozymes that could triphosphorylate nucleosides at the 5'-OH (*76*). It is important to note that RNAs with 5' -OH were generated from RNA cleavage by a hammerhead ribozyme which can be evolved relatively easily *in-vitro*. Many ribozymes capable of catalyzing the

triphophorylation reactions were obtained, which were about 150 nt long. The triphosphorylation ribozyme 1 (TPR1) was the most efficient ribozyme in that selection. It achieved rate enhancement of approximately 10^7, and reached a completion rate of about 83 % (*76*). These results indicate that a triphosphorylation ribozyme is also possible in the RNA world to generate activated ribonucleotides for polymerization.

Ribozymes discussed so far are the minimal ribozyme types required to sustain a flourishing RNA world. However, other ribozymes may also have existed in the RNA world. *In-vitro* selection approaches have produced several ribozymes capable of performing a diverse set of reactions. They can act on many different bonds including O-P, N-P, C-C, C-N, and C-O, to name a few. The activities performed by these ribozymes are listed in Table 1.1 demonstrating that RNA can perform many different functions if there are selection processes at work.

Table 1.1. A list of ribozymes that were discovered using *in-vitro* evolution. The bonds affected by these ribozymes, the rate enhancements produced, and the references are listed here. For some ribozymes, rate enhancement could not be found. These are labelled nf (not found).

Activity	Bond affected	Rate Enhancement	Reference
RNA cleavage	O-P	~200	(*77, 78*)
Cyclic phosphates hydrolysis	O-P	~50	(*79*)
RNA polymerization	O-P	nf	(*58, 59*)
RNA ligation	O-P	$\sim10^8$	(*65, 80*)
5'-5' RNA ligation	O-P	$\sim10^4$	(*81*)
RNA phosphorylation	O-P	$\sim10^6$	(*82, 83*)
RNA capping	O-P	nf	(*84*)
Amino acid adenylation	O-P	nf	(*85*)
Cofactor synthesis	O-P	nf	(*86*)
Branching reaction	O-P	nf	(*87*)
Diels Alder reaction	C-C	$\sim10^4$	(*88, 89*)
Aldol reaction	C-C	$\sim10^3$	(*90*)
Alcohol oxidation	C-H	$\sim10^7$	(*91*)
Aldehyde reduction	C-H	$\sim10^7$	(*92*)
Nucleotide synthesis	C-N	$\sim10^8$	(*72, 75*)
Peptide bond formation	C-N	$\sim10^6$	(*93*)
Aminoacylation	C-O	$\sim10^5$	(*94, 95*)
Michael reaction	C-S	$\sim10^5$	(*96*)
RNA-protein ligation	N-P	nf	(*97*)

1.6 Role of compartmentalization in the RNA world

At some point during the RNA world, it is highly likely that cell-like compartments began to arise. Today, almost all of biology happens in encapsulated spaces. The benefits of these compartments are obvious. The compartments allow for higher

concentration of reagents through molecular crowding (*98*). Such conditions can also enhance the activity of ribozymes (*99*). Several RNA enzymes require high concentrations of Mg^{2+}, and as Adamala and Szostak demonstrated, a protocell like environment can achieve such high Mg^{2+} concentrations, without any adverse effects (*100*). They demonstrated that the problems associated with fatty acid membrane integrity due to Mg^{2+} can be offset by using citrate ions to chelate the Mg^{2+} ions. Citrate also protects the RNAs from cleavage due to Mg^{2+} ions. Compartments also keep similar sequences in close proximity to each other. They also allow for concentration gradients, including pH gradients, between the inside and the outside of the cell (*101*).

One of the problems associated with the RNA world is that ligation and polymerization would have generated many parasitic RNAs. These RNAs would have been very disruptive to functional RNAs. These RNAs could become templates for replication, and take over the population. Such 'Spiegelman's monsters' can be quite problematic for specific RNA replication (*102, 103*). Compartmentalization can be a very important method for keeping the population resistant to parasitic RNAs. Bansho et al., showed that encapsulating the replicating systems in the water-in-oil emulsions can repress the amplification of the parasitic RNAs, and the entire system seems to be productive longer (*104*). Furthermore, the shorter the protocell, the longer the productive time of the system. It is also important to note that compartmentalization can also aid in the selection of more efficient systems. For instance, once the protocells divide to split the RNA into many groups, it is likely that only the groups with RNAs containing efficient ribozymes would be selected for while least efficient groups will be

selected against. Ichihashi et al., showed that in a cell-like environment the highly replicable mutants of RNA dominate the population (*105*). Hence, compartmentalization is an extremely important process for the beginning of life. It is therefore possible that cell-like structures may have overlapped with the RNA world.

1.7 Objectives of this work

We have seen that ribozymes are hypothesized to be central elements of the RNA world. Furthermore, many *in-vitro* evolution studies have been mentioned that suggest the plausibility of a diverse set of ribozymes in the RNA world. A key question that results from the RNA world is 'how did a simple RNA catalyze such diverse chemical reactions?' To answer such a question, my work has focused on learning from naturally occurring ribozymes, and investigating what catalytic strategies they use for catalysis. More emphasis will be given to the self-cleaving ribozymes since they are the largest class of ribozymes known today. I will explore the known strategies in Chapter 2 for most of the ribozymes. Then, Chapter 3 and 4 will be focused on my work related to investigating the catalytic strategies utilized by the hammerhead ribozyme. Finally, I will describe our approach towards the discovery of more naturally occurring self-cleaving ribozymes in Chapter 5. It will enable us to expand the number of ribozymes available today, and to discover novel catalytic strategies they might use. Hence, this work will shed light on the catalysis performed by RNA in the RNA world.

CHAPTER 2. CATALYTIC STRATEGIES USED BY RIBOZYMES IN THE MODERN WORLD

2.1 Perspective

In Chapter 1, many ribozymes were explored that could have existed in the RNA world. Many of those ribozymes were evolved *in-vitro* using a variety of different selection strategies. To understand how these ribozymes catalyze chemical reactions, and to explore the catalytic strategies used by ribozymes in the RNA world, we need to set our sights at nature. Natural selection is a powerful force that might have facilitated the production of many ribozymes in the RNA world, and many of which might still exist in the modern world. Ribozymes that already exist in nature and have been discovered are a valuable source to study how RNA may have catalyzed such a diverse array of reactions in the RNA world. They can give us an idea of the general strategies that RNA relies on to perform catalysis. In this chapter, the catalytic strategies used by natural ribozymes will be discussed. These include splicing ribozymes, the peptidyl transferase ribozyme, the branching ribozyme, the hydrolase ribozyme, and self-cleaving ribozymes. Particular emphasis will be given to the self-cleaving ribozymes since they are known to be the most abundant ribozymes in nature.

2.2 Splicing ribozymes – group I introns, group II introns, and the spliceosome

RNA splicing involves the excision of the introns from the pre-mRNA, and ligation of the two exons. It is a necessary step in mRNA maturation, and is also responsible for the tremendous diversity in the proteome available to organisms through alternative splicing (*106*). There are several mechanisms in biology that are responsible for splicing, including self-splicing, spliceosomal splicing, and tRNA splicing (*107*). Here, splicing reactions in which RNA is implicated as the catalyst will be discussed, such as, splicing by the spliceosome, and self-splicing in bacteria via group I and group II introns. While prokaryotes and eukaryotes both have some form of self-splicing, the spliceosomal splicing is restricted to eukaryotes (*108*). It is further suggested that group II introns, and the spliceosome shared a common ancestor (*109, 110*). RNA splicing is an important and medically significant event because single nucleotide polymorphisms (SNPs) lead to many diseases due to splicing defects (*106*). More than 90 % of the disease-causing SNPs are located outside of the protein coding sequences. Hence, splicing is an important biological process and RNA catalysis plays a central role in this mechanism.

Group I introns are insertion elements within the genes of several organisms, which can splice themselves out of the transcripts. While the group I introns can self-splice *in-vitro* without the help of any proteins, *in-vivo* they are assisted by intron or host encoded factors (*111*). They are found in the genomes of bacteria, phages, viruses, and rDNA of plants, fungi, and algae (*112*). The secondary structure of group I introns is composed of 10 different stem regions (P1-P10) with several intervening single stranded

areas (Figure 2.1 and ref (*113*)). The two splice sites are located at intron-exon junctions at the 5’ and 3’ end of the group I intron. Base-pairing between the exon and 5’ region of the group I intron makes the P1 region, and determines the 5’ splice site. The 3’ splice site is formed due to base pairing between the 3’ end of the group I intron and the exonic region which makes up the P10 region. The base pairing region between the exons and the group I intron is called the internal guide sequence (IGS). The splicing reaction is comprised of two transesterification steps. The 3’ -OH of an external GTP (αG) bound to the P7 region acts as a nucleophile, and attacks the phosphate at the 5’ splice site. A 3’-5’ phosphodiester linkage is created between αG and the 5’ end of the intron, and the 3’-OH on the upstream exon (reviewed in (*114*)). This also leads to some major conformational changes, so that the αG position is now occupied by the ωG, which is the guanosine located at the 3’ splice site (*115*). This allows the 3’-OH of the upstream exon to attack the phosphate at the 3’ splice site leading to the excision of the intron and ligation of the two exons. Crystal structures of the group I introns including a 3.6 Å structure from phage Twort ribozyme-product complex, and complexes from other states of splicing reveal catalytic and structural roles for at least two divalent metals (Figure 2.1B). There is a G•U wobble involved in the 5’ splice site recognition, and the 3’ splice site is selected based on a three-way interaction between ωG and a G-C base pair (*113, 116*). In a crystal structure that mimics the splicing intermediate, a magnesium ion appears to interact with the 2’ -OH, and the 3’ -O of the guanosine in the ωG binding site (*117*). It also interacts with the pro-R_p oxygen of the scissile phosphate at the 3’ splice site, to stabilize the transition state. Another Mg^{2+} also interacts with this oxygen

for the same purpose, but also makes an additional contact with the 3'-O of the U located at the 3' end of the 5' exon to facilitate a nucleophilic attack (*117*). The 2'-OH of this U also hydrogen bonds with the leaving 3'-OH for the same purpose. These ligand binding sites for the two divalent metals agree with the previous biochemical data (*118*). However, the biochemical data previously had argued for a three-divalent metal ion mechanism (*114, 119*).

Hence, group I ribozymes use a cofactor mediated and divalent metal mediated catalytic strategies to perform the splicing reaction. We also see a role for the 2'-OH in substrate orientation and catalysis. The metals play roles such as transition state stabilization, nucleophile activation, and leaving group stabilization. Furthermore, the consensus catalytic mechanism for the ligation step appears to be a two-metal ion mediated catalytic mechanism where divalent metals are playing an identical role.

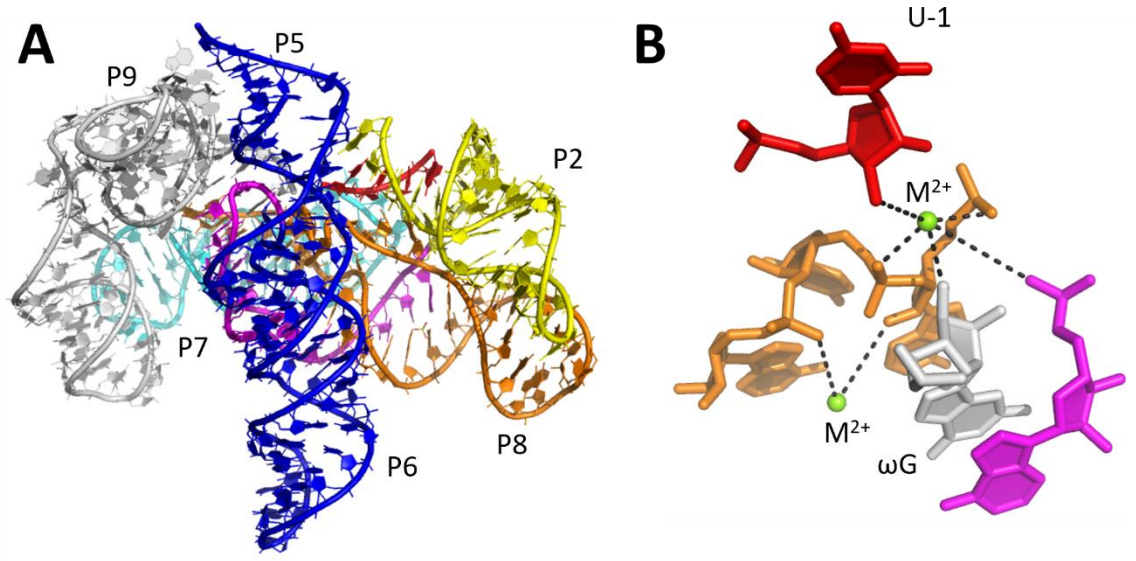

Figure 2.1. Structure of the phage Twort group I ribozyme in complex with its product. **A**. The overall crystal structure of the group I ribozyme product complex is represented as a cartoon. The secondary structural regions are colored separately and labelled. **B.** The active site of the group I ribozyme. The divalent metals are shown as green spheres. The coloring is the same as in A. Potential ligands of the divalent metals are shown using dashed lines. The images were rendered in PyMol using coordinates from PDB ID 1Y0Q(*113*).

The group II introns are also mainly found in bacterial genomes, where they splice themselves out, and get inserted into new locations (*120*). In other words, they are mobile genetic elements. The introns themselves consist of two components: one is the ribozyme that catalyzes splicing, and reverse splicing reactions. The second part consists of the sequence meant for a protein that contains a reverse transcriptase domain. The reverse transcriptase, in addition to reverse transcribing the integrated intron, also helps to enhance the activity of the ribozyme by stabilizing the structure, and tends to be required *in-vivo* (*121*). The secondary structure consists of a central core from which six different domains (D-I to D-VI) originate, and each domain has one or

more base paired regions (ref (*122*) and Figure 2.2A). The active site is located in domain V, where the catalytically important nucleotides AGC, and the AY bulge are found (Figure 2.2B). There are binding sites within these regions for catalytically important Mg^{2+} ions. The active site also contains the branch point nucleotide A (domain VI), the 5' splice site, and the 3' splice site. The reactions catalyzed by the ribozyme include two transesterification steps for splicing the intron. The 2'-OH of the branch point A acts as the nucleophile to attack the phosphate in the 5' splice site (*123*). This leads to an intermediate in which the 5' exon has a 3'-OH, and the intron lariat. Next, the 3'-OH of the 5'-exon is the nucleophile that attacks the 3' splice site phosphate to yield an excised lariat intron, and ligated exons. Again, two Mg^{2+} ions play crucial roles in this catalytic mechanism (*124*). Both the transesterification steps occur in a single or proximal active sites (*125*). This is accomplished by a series of conformational changes that help reorient the active site for the two catalytic steps (*123*). Recent crystal structures indicate that there may be four metal (2 monovalent ions) binding sites. Prior to the first transesterification step, the two divalent metals help coordinate the pro-R_p oxygen of the scissile phosphate at the 5' splice site (*123, 124*). One of the divalent metals also coordinates the leaving 3' oxygen (*126*). The two monovalent ions help in orienting the active site by coordinating several active site nucleotides (*123*). After this splicing reaction, the active site is reoriented such that the 3' splice site is now in the active site. The metals involved in the first step, most likely reform the metal cluster that helps with the second transesterification reaction. This reaction mechanism again shows the importance of divalent metals in RNA catalysis. These divalent metals help

stabilize the transition states as well as the leaving group to facilitate catalysis. There are also roles for the monovalent ions in the orientation of the active site nucleotides. How the nucleophile is activated is still being investigated, however a role for water (specific base) has been predicted based on QM/MM simulations (*127*).

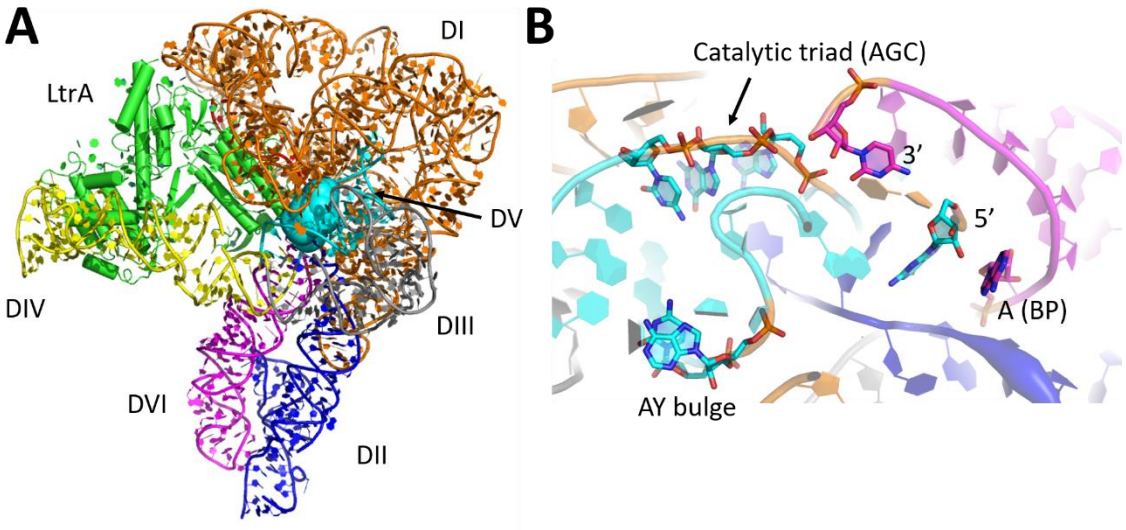

Figure 2.2. The structure of the group II intron and its reverse transcriptase.
A. The overall crystal structure of the group II intron complex. The reverse transcriptase (LtrA protein) is depicted as green cylinders and ribbons, and the RNA is shown as a cartoon of various colors corresponding to the secondary structure regions of group II introns. The spheres indicate the location of the active site and the mRNA is shown in red. B. A close-up view of the active site. **B**. A close-up view of the active site is shown in sticks. The catalytic triad, AY bulge, the branch point adenosine, and the 5' and 3' ends of the introns are labelled. The images were rendered in PyMol using coordinates from PDB ID 5G2X (*122*).

The spliceosome in eukaryotes is thought to have shared a common ancestor with the group II introns (*128*). This is due to the highly similar catalytic transesterification steps that both enzymes use for catalysis. The spliceosome is made up of many subunits including RNA and proteins that work together to remove introns

from the pre-mRNAs (Figure 2.3A and ref (*129*)). The major components of the spliceosome are the five snRNAs called U1, U2, U4, U5 and U6 snRNAs. The spliceosome complex is usually assembled around the splice sites, in a step-wise manner starting with the recruitment of U1 to the 5' splice site. There is extensive evidence for the spliceosome being a ribozyme (*130-133*). The heart of the spliceosome is composed mostly out of RNAs (*129, 134*). Similarly, the Piccirilli group demonstrated that the catalysis is conducted by the U6 snRNA of the spliceosome via a metal meditated mechanism (*135, 136*). Recent progress in the structural biology of the spliceosome provides us with snapshots of the active site of an activated spliceosome (Figure 2.3B). The active site includes the intramolecular stem loop of the U6 snRNA, stem regions of the U2 and U6, and finally the loop region of the U5 snRNA. In addition, the active site also includes four Mg^{2+} ions. Three of the Mg^{2+} ions are observed at the intramolecular stem loop (ISL) of the U6 snRNA, and most likely play a structural role by stabilizing the negative charge generated by the coming together of the phosphate backbone in the loop. In addition, a catalytic Mg^{2+} is located near the oxygen atom of the scissile phosphate. Metal rescue experiments and other evidence indicate that the spliceosome also uses a two-metal ion mechanism for catalysis (Figure 2.3B) (*130, 135*). These metals make at least five contacts with the U6 snRNA ligands during both the branching reaction and the exon ligation reaction. In the branching step, it is proposed that the metals interact with the pro-R_p oxygen of the scissile phosphate, and the leaving oxygen. The two metals are coordinated by the G and U nucleotides of the ISL region of U6 snRNA. In the exon ligation step, G60 and A59 coordinate a Mg^{2+} which interacts with

the leaving group of the 3’ splice site. It is still unclear how the nucleophile is activated since a rescue effect was not observed at that position. However, divalent metals are again seen to be playing key catalytic roles of leaving group and transition state stabilization. The exact positioning of these metals is coordinated by the nucleobases and the backbone of the U6 snRNA.

In summary, all three candidates discussed here are ribozymes that use similar catalytic strategies to achieve effective catalysis. All use divalent metals to either activate the nucleophile or to stabilize the transition state and/or the leaving group. Group I intron also uses cofactor-mediated catalysis. Another common theme is the use of RNA ligands for substrate and metal positioning.

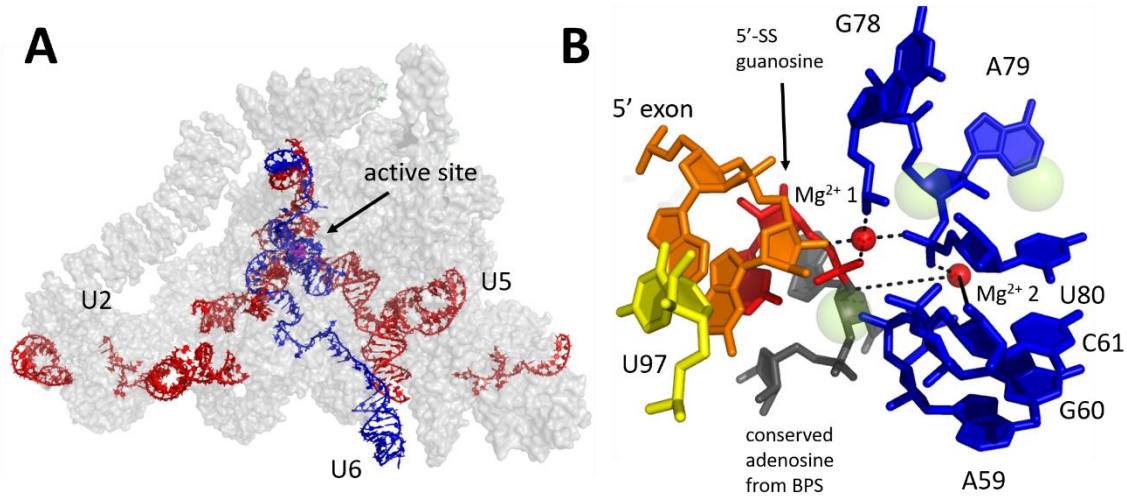

Figure 2.3. Cryo-EM structure of the spliceosome C complex.
A. The overall structure of the spliceosome at catalytic step I is shown. The proteins are shown in transparent surface representation and colored grey. The RNAs are shown as cartoons. U6 is colored blue whereas all the other RNAs are red. The 5' splice site is located at the heart of the spliceosome and is depicted as magenta spheres. **B**. Nucleotides and divalent metals implicated in the splicing reactions are shown close to the 5'-splice site. The divalent metals implicated in catalysis are shown as small red spheres whereas other divalent metals found close to the active site are shown as transparent green spheres. Nucleotides are colored as follows: U6 nucleotides are blue, U5's are yellow, exon nucleotides are orange, 5' splice site is shown as red, and the BPS adenosine is shown in grey. Interactions made by the divalent metals are also shown as black dashed lines. The images were rendered in PyMol using coordinates from PDB ID 5GMK (*129*).

2.3 Peptidyl transferase ribozyme – the ribosome

The ribosome is one of the most important enzymes in biology. It catalyzes the formation of peptide bonds between amino acids called the peptidyl transfer reaction. It is central to the biological process of translation. Without the ribosome, proteins that

control the majority of the processes in biology would not have been possible. The important nature of the ribosome is revealed by the fact that even small defects in the ribosomal pathways lead to a number of diseases, and can also be lethal (*137*). The discovery that ribosome is a ribozyme had profound implications for evolutionary biology, and the RNA world hypothesis (*11, 138*). It is almost universally conserved in biology, and its core is made up of RNA only whereas proteins only play a decorative role (Figure 2.4 and ref (*139*)) (*10*). It suggests that RNA is an ancient entity. It might be a relic of a time when the RNA world was transitioning into the ribonucleoprotein (RNP) world.

The major reaction catalyzed by the ribosome is the peptidyl transfer reaction. However, it also catalyzes the peptidyl hydrolysis reaction at the culmination of translation (*140*). The peptidyl transfer reaction involves a nucleophilic attack by the α-amino group of the aminoacyl tRNA on the carbonyl group of the peptidyl tRNA, leading to the release of the P-site tRNA. Ribosomes can achieve a rate enhancement of about 10^6-10^7 compared to the uncatalyzed reaction. Hydrazinolysis of methyl formate in solution serves as a good model of the peptidyl transfer reaction (*140*). This reaction proceeds through a nucleophilic attack on the carbonyl bond of the methyl formate by the amine group of the hydrazine. This leads to the first intermediate, which is acted on by a general base that deprotonates the positively charged amino group resulting in a second intermediate. This intermediate is resolved when the methoxy cation is released after being protonated. The peptidyl transfer reaction in the ribosome may follow a similar mechanism. However, it is not clear if and when the proton transfer steps occur.

Early crystal structures from *E. coli* led to the belief that the N3 of an active site adenine acts as a general acid-base catalyst (*12*). However, characterization of ribosomes from other species revealed that there may be pH-dependent conformational changes involved in the peptidyl transferase center (PTC) of the ribosome (*141*). Furthermore, substitution at the N3 position of the active site adenine did not hinder the peptide bond formation (*142*). Later, a high resolution crystal structure of the ribosome complexed with tRNAs revealed important catalytic roles for CCA tails of the tRNAs, and about 11 different nucleotides of the 23S rRNA, all residing in the PTC (*143*). Further investigation suggested that the conformation of the PTC differs with and without substrates in the P or A site (*143*). Several mutagenesis studies of the active site nucleotides failed to reveal a dominant effect on the rate of the peptidyl transferase reaction (*140, 144*). These studies began to give rise to the induced fit, and substrate assisted catalytic mechanism (*145*). The crystal structures show a network of hydrogen bonds involving nucleotides in the active site, and water molecules. This hydrogen bonding network probably facilitates catalysis by stabilizing the charged intermediates, and by facilitating proton transfer. This network of hydrogen bonds includes the 2'-OH of the P-site tRNA tail. Mutation of this 2'-OH to 2'-F or 2'-H decreases the rate by about 10^5 fold. Hence, this 2'-OH could be mediating catalysis by orienting the amino group for a nucleophilic attack, and by acting as the general acid-base catalyst. Several different mechanisms have been proposed that govern how the proton shuttling takes place. Studies involving the thermodynamics of this reaction reveal that the major contribution to the rate enhancement comes from the entropic contribution (*146*). The enthalpic

penalty increases in the catalyzed reaction over the uncatalyzed reaction. This argues against a major role for general acid-base involvement, as there is no reduction in the enthalpic factor. The entropic contribution might be coming from the rearrangement of the water molecules within the active site during the reaction, and to a lesser extent from substrate positioning.

In conclusion, while the specifics of how the catalysis proceeds in the peptidyl transfer reaction may not be completely worked out yet, we do know that RNA, whether it's the 23S rRNA or the CCA tails of the tRNA, plays important roles. The RNA-based active site helps position the substrate, and the substrate itself plays an important role in this reaction mechanism. Hence, the ribosome employs several strategies for catalysis, including substrate-assisted catalysis, transition state stabilization, and substrate positioning via an induced fit mechanism.

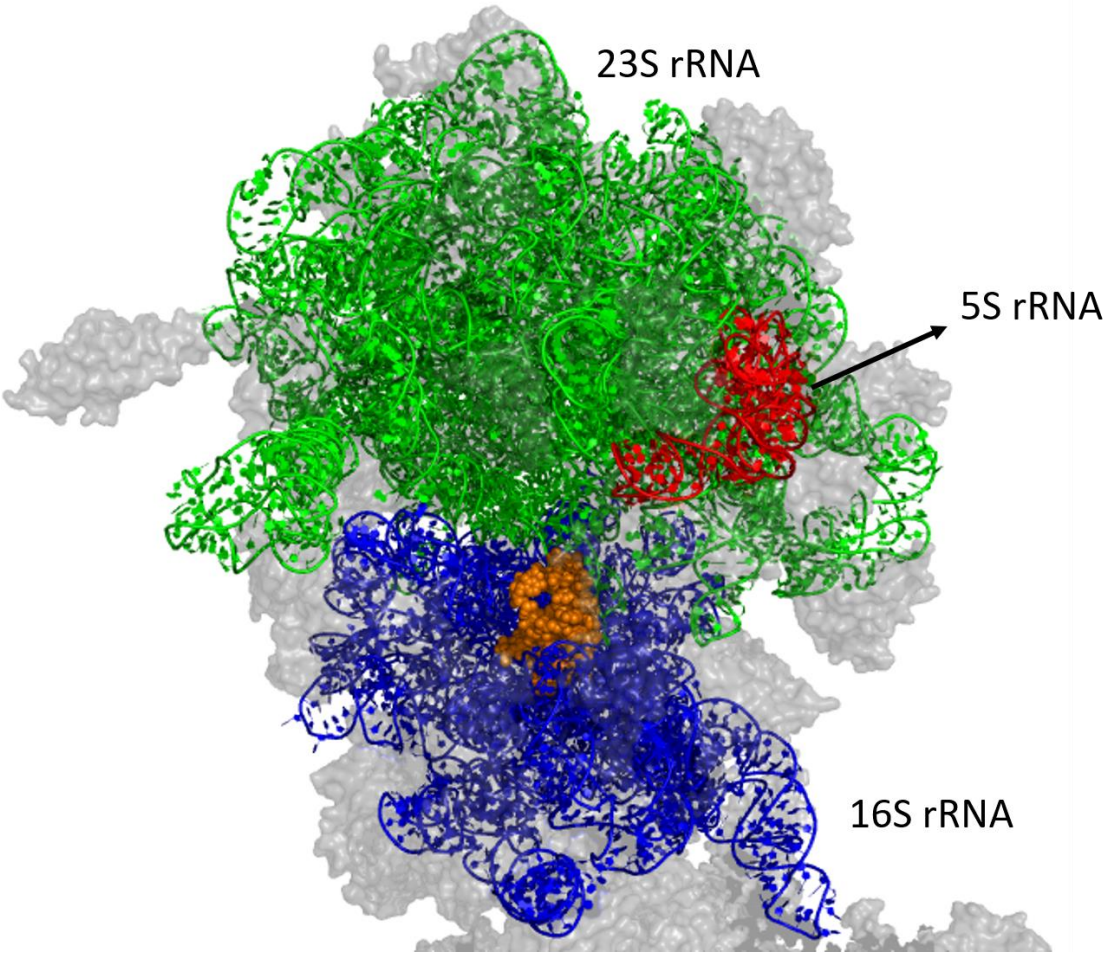

Figure 2.4. The overall crystal structure of a bacterial ribosome.
Proteins are shown using the surface view, and are colored grey. The atoms in the IRES RNA are shown as orange spheres. It occupies the same space that a tRNA normally would. Ribosomal RNAs are shown as cartoons. The image was rendered in PyMol using coordinates from PDB ID 4XEJ (*139*).

2.4 Branching ribozyme (lariat capping) – the GIR1 or LC ribozyme

The GIR1 branching ribozyme is a ~180 nt ribozyme that catalyzes a branching reaction in RNA (*147*). When discovered, the GIR1 ribozyme was thought to be remarkably similar to the group I ribozyme; hence, the name Group I-like ribozymes

(GIR1). However, it does not catalyze the same splicing reaction as the group I intron, and the crystal structure also shows major differences between the folds employed by the group I intron and the GIR1 ribozymes (compare Figure 2.1 to 2.5). In the GIR1 ribozyme, the 2'-OH of an internal nucleotide attacks a phosphate nearby, leading to the formation of a product with a 5'-lariat (2'-5' linkage), and a cleaved RNA with a 3'-OH is released (*148*). For this reason, it is also referred to as the lariat capping (LC) ribozyme (*149*). Naturally it is found in the twin ribozyme introns, which consist of a homing endonuclease gene flanked by the GIR1 and the group I intron ribozyme. The ribozyme does not appear to be very widely distributed, since only a few species are known to have it. However, many artificial examples are known (*150*).

The structure of the GIR1 ribozyme consists of about 11 paired regions (Figure 2.5A and ref (*149*)). The P3-P7-P8-P9 helices are stacked on top of each other to make a domain which is aligned with the P4-P5-P6, and P10-P15 domains. The P2-P2.1 domain sits on top of the other three domains. A few tertiary interactions, and pseudoknots are also involved in the structure. The active site of the GIRI ribozyme resides at the interface of helices P10 and P9, the J5/4 and J9/10 junctions. The 2'-OH of U232 is the nucleophile that attacks the phosphate group of C230. A 2'-5' linkage is formed between U232 and C230. This leads to the formation of a lariat, which contains the nucleotides U232, A231 and C230 (Figure 2.5B), and most likely serves the purpose of protecting the RNA from the activity of 5' exonucleases. Furthermore, the lariat also facilitates the release of mRNAs. The G229 and upstream sequence is released with a 3'-OH after hydrolytic cleavage. This reaction demonstrates the importance of the 2'-OH in RNA

catalysis, since almost every RNA reaction involves the reactive 2'-OH of the ribose. A critical component of the active site is G229, because mutations involving G229 make the ribozyme inactive (*151*). But the crystal structure, which represents the ribozyme-product stage, shows it to be far from the nucleophile, 2'-OH of U232. While the divalent metals were not observed in the active site of the crystal structure like the group I and II introns, they may also be involved in activating the nucleophile and providing stabilization of the intermediates. The catalysis of the GIR1 still needs more work, including high resolution crystal structures with the scissile phosphate modelled in. It would help identify divalent metals in the active site. For now, we can conclude that the GIR1 ribozyme might use catalytic strategies analogous to group I and group II introns for catalysis.

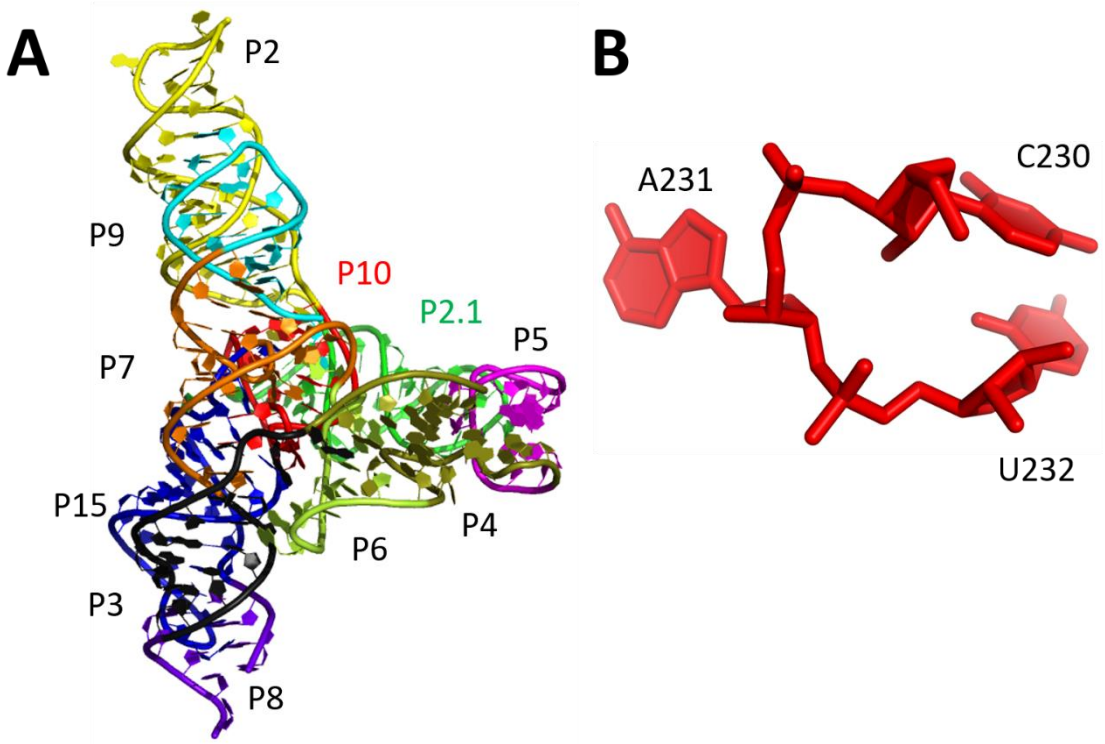

Figure 2.5. Crystal structure of a circularly permuted GIR1 (LC) ribozyme.
The overall crystal structure of GIR1 derived from *Didymium iridis* is shown in **A**, whereas the structure of the 3 nt lariat formed after the branching reaction is shown in **B**. It should be noted that the phosphate connecting U232 and C230 is missing. The stems in the crystal structure are all colored differently and labelled. The color scheme is also consistent between the two pictures. These images were rendered in PyMol using coordinates from PDB file 4P95 (*149*).

2.5 Hydrolase ribozyme (trans-cleaving) – Ribonuclease P

Ribonuclease P (RNase P) is a trans-cleaving enzyme that acts on a variety of RNA substrates (*152*). It is composed of RNA, and one or more proteins. While the protein components are critical for the activity of RNase P, the catalysis is performed by the RNA core (*5, 153*). Therefore, RNase P is a ribozyme, a finding that won Sidney Altman the

Nobel Prize in chemistry in 1989. RNase P was found to be the ribozyme that catalyzes the cleavage of the 5' end of pre-tRNA, and hence functions in the maturation of tRNAs (*154*). Later, however, many other substrates for RNase P were also discovered (*8, 155*). Specifically, RNase P cleaves at the meeting point of a single stranded RNA and double stranded RNA. It is found in all domains of life, suggesting that like ribosome, it is also a living fossil of the RNA world (*156*). Evolutionary studies of the RNase P reveal some very interesting insights into the transition from RNA world to the protein world. The number of protein subunits of RNase P increase from 1 in bacteria to 4 or more in archaea, and finally up to 10 in eukaryotes (*8, 157*). On the other hand, the catalytic activity, and the catalytic role of the RNA subunit of RNase P decreases going up the evolutionary ladder. For instance, the k_{obs} in bacteria is as high as $10min^{-1}$, but it decreases to 0.1 min^{-1} in archaea, and further decreases to 10^{-5} min^{-1} in eukaryotes (*8*). The number of organisms in which the RNA plays a catalytic role in each domain also decreases, from all in bacteria to about 2 in eukaryotes. It shows a wonderful transition from reliance on RNA to protein through evolution.

The structure of the bacterial RNase P's RNA component consists of two domains called S and C. P7 to P14 stem regions make up the S domain, and the rest (P1-P6, P15-P18) make up the S domain (Figure 2.6A and ref (*158*)). The secondary structure packs into a compact structure where the stems in each domain are aligned to each other, and many tertiary interactions including a pseudoknot are observed. The crystal structure of bacterial mature tRNA in complex with RNase P reveals two regions where the substrate contacts the enzyme (Figure 2.6A). The S domain of RNase P interacts with TΨC and D

loops of tRNA, whereas the CCA tail gets inserted into the pseudoknotted portion involving the P6, P15, P16 and P17 stems (*158*). The CCA nucleotides base pair with nucleotides in the L15 loop. The interaction between the highly conserved T and D loops of tRNA, and the conserved portion of the S domain of RNase P, forms the basis of recognition of tRNAs. But RNase P cleaves other substrates as well; however, the recognition mechanism is not properly understood.

The active site of the RNase P is located in the C domain, where the substrate (tRNA), the protein component, and the C domain RNA are found (Figure 2.6B). The active site catalyzes the cleavage of RNA substrates via a hydrolysis reaction that generates two products, which have a 3' -OH and a 5' -phosphate, respectively (*152*). Two divalent metal binding sites have been confirmed to be found near the P4 helix through structural and biochemical means (Figure 2.6B) (*158, 159*). These metals are also critical for the activity of RNase P. A water molecule bound to one of these divalent metals is predicted to be the nucleophile (*160*). A 2'-OH attack by a nucleoside, like in self-cleaving ribozymes would almost inevitably produce RNAs with 2' 3' -cyclic phosphates. Therefore, the active site must shield the 2' -OH groups, such as the 2'-OH at the -1 position. But mutation of this 2'-OH to 2'-H changes the recognition of substrate and metal ion binding properties of RNase P (*161*). While the mechanism of cleavage is not fully understood, the role of divalent metals seems to be crucial. They are predicted to make several interactions with the cleavage site of the substrate, and RNase P nucleobases. They are predicted to stabilize the interaction between the CCA tail and RNase P RNA, and interact with the cleavage site to activate the nucleophile

(*152, 158*). The 2'-OH of U-1 and the exocyclic amine at G+1 are both predicted to have important catalytic functions, and divalent metals may also be coordinating these positions. A divalent metal is also thought to interact with the leaving 3'-OH, and the pro-R_p oxygen to facilitate hydrolysis and transition state stabilization, respectively.

The exact details of the catalytic mechanism seem to be lacking, but there is an important role for the divalent metals in activating the nucleophile which is predicted to be a water molecule, stabilizing the substrate-enzyme complex, and stabilizing interactions with the pro-R_p oxygen and the leaving 3'-OH. Evidence, therefore, points to two divalent metals and possibly nucleobase mediated catalysis like other larger ribozymes.

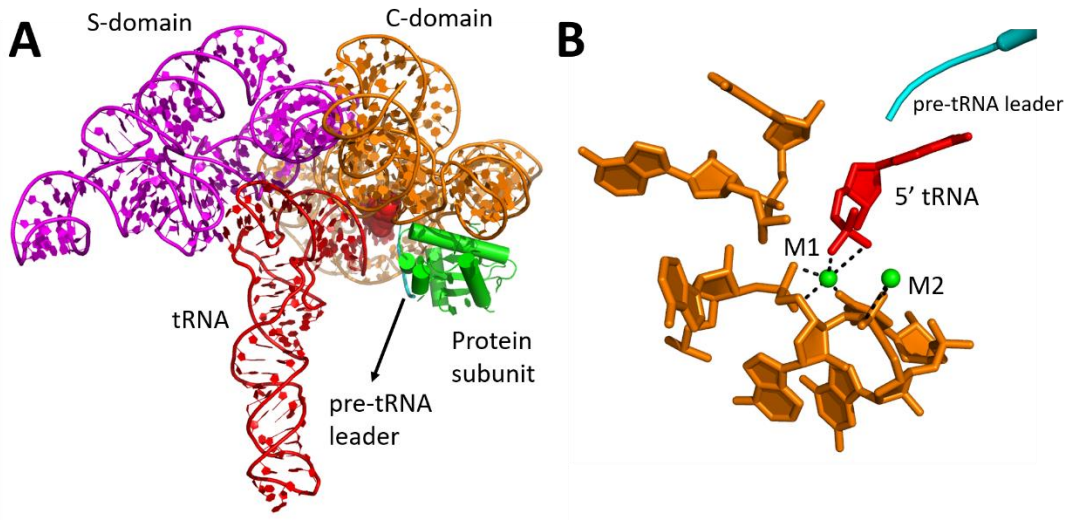

Figure 2.6. The structure of ribonuclease P bound to tRNA.
A. The crystal of the ribonuclease P bound to tRNA is shown. S-domain is colored magenta, C-domain is colored orange, tRNA is colored red, and the protein is shown as green cylinders. The 5' end of the tRNA is shown as green spheres. This is where the active site is located as well. **B.** The active site of the RNase P reveals an important role for two divalent metals, which are shown as green spheres. The predicted ligands of the divalent metals are shown as dashed lines. The images were rendered in PyMol using coordinates from PDB ID 3Q1R (*158*).

2.6 Self-cleaving ribozymes

Self-cleaving ribozymes are the most abundant class of natural ribozymes. Many self-cleaving ribozymes have been discovered including the hammerhead ribozyme, the hairpin ribozyme, the HDV ribozyme, the VS ribozyme, the *glmS* ribozyme, the twister and twister sister ribozymes, the hatchet ribozyme, and the pistol ribozyme. Most of these self-cleaving ribozymes are found in viral or satellite RNAs (*162*). Furthermore, they are also very common in the intergenic regions of other organisms, most likely

through viral insertions. In viruses, these ribozymes are involved in the rolling circle replication mechanism (Figure 2.13A). These are found embedded in the long multimeric RNAs, and self-cleave to produce monomeric RNAs, which are then ligated together. Hence, these ribozymes not only self-cleave, but also catalyze the ligation reaction. One distinct class of self-cleaving ribozyme is the *glmS* ribozyme which functions in gene regulation in bacteria (Figure 2.13B) (*163*). Some other variants of self-cleaving ribozymes are also implicated in gene regulation (Figure 2.13B). Both types of self-cleaving ribozymes will be discussed in the following sections.

The cleavage mechanism usually involves a nucleophilic attack by the 2'-O of a nearby nucleotide, on the scissile phosphate (Figure 2.7). This leads to a pentavalent transition state, which is then resolved into two products. The 5' product has a 2',3'-cyclic phosphate and the 3' product has a 5'-OH. This basic reaction can be accelerated by many different catalytic strategies (Figure 2.15). The RNase A enzyme catalyzes a very similar reaction, and a lot can be learned from the catalytic strategies that it employs (*164*). It uses several different catalytic strategies to achieve extremely efficient catalysis. The rate enhancement achieved by RNase A is close to 3×10^{11} (*165*). The same catalytic strategies are or can be used by self-cleaving ribozyme to perform catalysis (Figure 2.15). Firstly, the in-line geometry at the cleavage site is very important. The angle between the nucleophile, the phosphorus atom, and the leaving oxygen needs to be close to 180 degrees. The structures of ribozymes usually help the active site achieve such a conformation. Second, a general base can enhance the nucleophilicity of the 2'-OH of the adjacent ribose. Third, a proton donor for the 5'-O

can help facilitate the progression from the transition state to the products (general acid catalysis). Fourth, the transition state stabilization can also yield a faster reaction. This can be achieved by interaction of the scissile phosphate with several different players including divalent metals, water, and other nucleotides. RNase A uses all these strategies (*164, 166*). His 12 deprotonates the nucleophile, Lys 41 provides the transition state stabilization of the pro-R_p oxygen, and His 119 provides the proton to the 5'-O. While RNA lacks the catalytic repertoire of proteins, self-cleaving ribozymes can use remarkable strategies to achieve efficient catalysis. They can use some or all the above-mentioned strategies for rate enhancement, as will be discussed below.

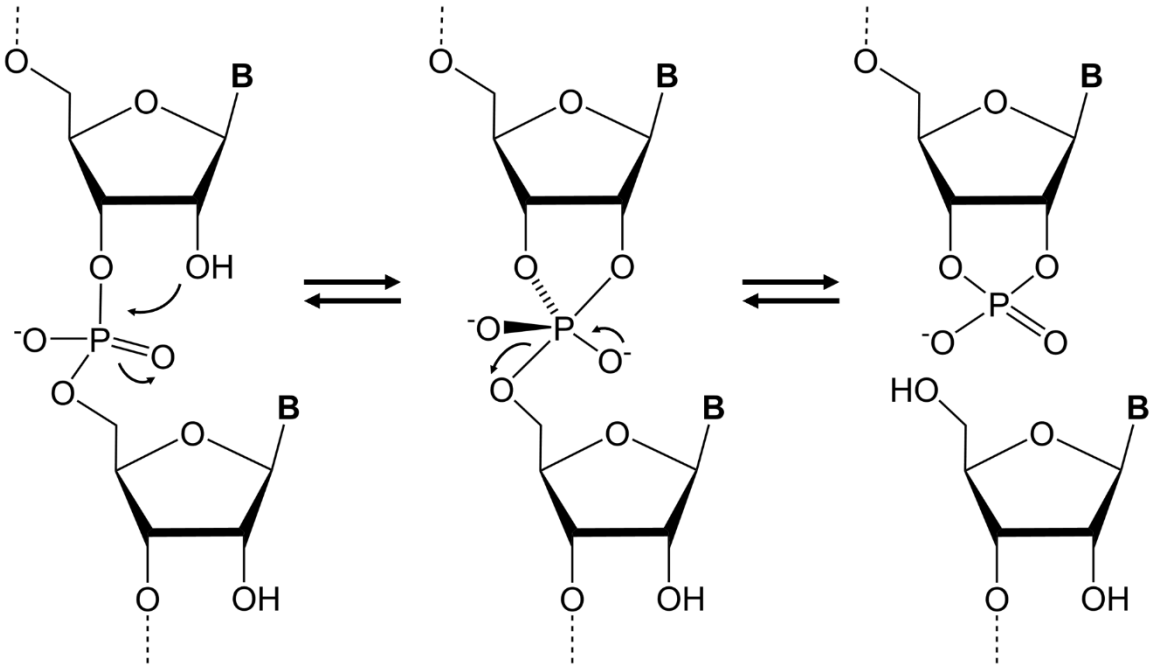

Figure 2.7. Cleavage of a phosphodiester bond in RNA.
2'-OH is the nucleophile that attacks the scissile phosphate leading to a pentavalent intermediate. The intermediate is resolved when the 5'-O accepts a proton leading to two products: a 5' product with a 2', 3' -cyclic phosphate and a 3' product with a 5' -OH.

2.6.1 Self-cleaving ribozymes derived from viral RNAs

The earliest self-cleaving ribozymes were all discovered in viral RNAs. Specifically, they were found in the satellite RNAs from plants. These include the hammerhead ribozyme, the HDV ribozyme, the hairpin ribozyme, and the VS ribozymes. Some recently discovered ribozymes in bacteria, including the hatchet ribozyme, the pistol ribozyme, the twister and twister sister ribozymes are predominantly located in the non-coding regions of the genomes. It is highly likely that these originated from viral sequences as well, since they were discovered due to their genetic association to the hammerhead ribozyme. However, these ribozymes can also evolve independently. The following discussion will primarily focus on the catalytic strategies used by these ribozymes. The hammerhead ribozyme will be reviewed in the next two chapters.

2.6.1.1 The hairpin ribozyme

Along with the hammerhead ribozyme, hairpin ribozyme was one of the first ribozymes to be discovered. It was found in the (-) strand of the tobacco ringspot virus satellite RNA, where it functions in the rolling circle replication (*167*). Like the hammerhead ribozyme, it also catalyzes the site-specific and reversible cleavage of RNA. Unlike other self-cleaving ribozymes, it is particularly competent at ligating the cleaved version of itself (*168*). The hairpin ribozyme produces a rate enhancement of 10^7-10^9

compared to spontaneous RNA cleavage (*169*). The structure of the hairpin ribozyme consists of a four-way helical junction, and the adjacent arms A and B are characterized by two internal loops (Figure 2.8A and ref (*170*)). The full-length hairpin differs from the minimal hairpin ribozyme in the sense that helices C and D are missing in the minimal version. The cleavage site is in the A loop (Figure 2.8). Either due to the complexity of the structure of this ribozyme or other reasons, the hairpin motif is not abundant in nature. Only three examples of the hairpin ribozyme are known today (*171*).

A significant body of experimental evidence suggests that the hairpin cleavage reaction is catalyzed by a general acid-general base mechanism (*172*). The two nucleotides responsible for these roles are G8 and A38 respectively. G8 is located in loop A, and A38 in loop B. The crystal structures of the hairpin ribozyme show that loop A and B are very close together (Figure 2.8A). The docking of loops A and B is bound to generate repulsion between the similarly charged backbones of RNA. However, this interaction is stabilized by positively charged divalent metals and base pairing between loop A and loop B. This interaction creates the active site of the hairpin ribozyme where G8 and A38 are near the cleavage site (Figure 2.8B). The transition state structure shows A38 to be interacting with the 5'-O, and G8 to be interacting with the 2'-O (*170*). Furthermore, the in-line geometry is preserved in the transition state. The crystal structure suggests that N1 of G8 plays the role of the general base, whereas A38 acts as the general acid in the hairpin ribozyme catalysis. N1 of G8, which needs to be deprotonated abstracts the proton from the nucleophile, whereas the protonated N1 of A38 provides the proton to the 5'-O. The pH-rate profiles also agree with this

interpretation. The rate increases with pH until about pH 7, and then stays constant. This pH rate profile is to be expected if the general base has a pK_a of greater than 9.6, which is consistent with the pK_a of guanine in solution, and the general acid pK_a to be around 6. However, the pK_a of unperturbed adenine in solution is close to 3.5. There is significant evidence that A38 is the general acid including a 5'-PS substitution at the 5'-O position. 5'-PS nullifies the requirement for a general acid, and it was shown to rescue the decrease in the hairpin rate of reaction resulting from mutation of A38 to a purine nucleobase (*172*). The question still remains regarding the pK_a of A38 in the context of the tertiary structure. Raman crystallography measured the microscopic pK_a of A38 in the hairpin ribozyme to be 5.46 (*173*). This is one of the strategies RNA can use to alter its chemical reactivity. It is possible that the active site interactions of A38 enable it to alter its pK_a. The proximity of the N1 of A38 to the negatively charged scissile phosphate could potentially shift the pK_a upwards.

In summary, the catalytic strategies employed by the hairpin ribozyme appear to be general acid-general base catalysis coupled with in-line geometry at the cleavage site. The active site environment also creates the necessary conditions to shift the pK_a of the general acid towards neutrality for more efficient catalysis.

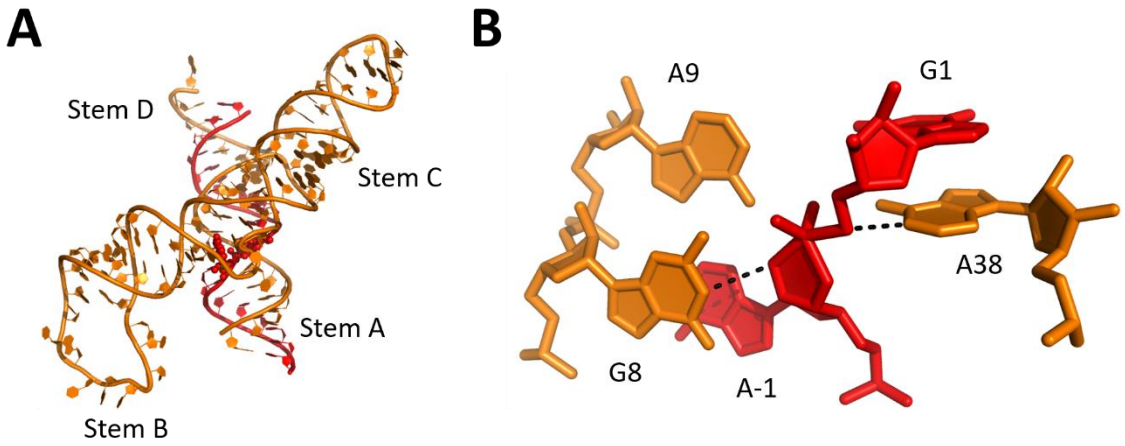

Figure 2.8. Transition state analog structure of the hairpin ribozyme.
A. The overall crystal structure is shown in a cartoon representation. The ribozyme strand and substrate strand are colored orange and red respectively. The cleavage site dinucleotides are shown as spheres. The four helices in the structure are labelled. **B.** The active site of the hairpin ribozyme during the transition state is shown. Important nucleotides and interactions in the active site are labelled. The images were rendered in PyMol using coordinates from PDB ID 1M5O (*170*).

2.6.1.2 The HDV ribozyme

The hepatitis delta virus (HDV) ribozyme was discovered in the genomic and anti-genomic RNA strands of the hepatitis delta virus (*174*). It functions in the rolling circle replication of the hepatitis delta virus. The HDV ribozyme was later found in many different organisms using conserved secondary structure searches (*175, 176*). One interesting observation from these studies came in *Anopheles gambiae.* It was found that the expression of HDV-like ribozymes in *A. gambiae* is also dependent on the developmental stage of the organism. It suggested a possible role for the HDV ribozyme,

and quite possibly other ribozymes such as the hammerhead ribozyme in gene regulation during development. The HDV ribozyme is about 85 nt long and appears to adopt a double pseudoknot structure which contains five different base paired regions (Figure 2.9A and ref (*177*)). The cleavage site is positioned in the P 1.1 region at the 5' end of the ribozyme. The HDV is a unique ribozyme in the sense that it does not require significant interactions with the substrate, and can cleave almost any substrate that is attached to it. It is for this reason that it has found uses in molecular biology for uniform transcript production (*178*).

From the very beginning, the HDV ribozyme appeared to be slightly different compared to other self-cleaving ribozymes. A lot of evidence pointed to a possible role of a divalent metal in the active site (*179-181*). The evidence included the need for a divalent metal at physiological conditions. The early crystal structures of the HDV ribozyme established a crucial role for the C75 nucleotide in catalysis (*182*). The N3 position of the C75 appeared to interact with the 5'-O, possibly suggesting a general acid role. C75 is so important, that mutating it to a U can inactivate the ribozyme (*183*). However, further details about the active site were not possible until a recent crystal structure (*177*). In addition to a catalytic role, C75 also appears to be important for the orientation of the active site. It makes several interactions in the active sites including hydrogen bonding with the scissile phosphate (Figure 2.9B). Furthermore, the pK_a of C75 was also found to be shifted towards neutrality using Raman crystallography (*184*). In this case, the hydrogen bonding interaction between N4 and the negatively charged pro-R_p oxygen serves to stabilize the protonated form of C75 leading to a shift in the pK_a

towards neutrality. The role of C75 as a general acid was confirmed when the 5'-O was substituted with the sulfur atom, which abolished the need for a proton donor (*185*). In this 5'-PS HDV, a C75U mutation, which would otherwise make the ribozyme inert, had minimal effect on the rate of cleavage reaction. The crystal structure also shows a Mg^{2+} within the active site, and it appears to interact with the 2'-O and the pro-R_p oxygen of the scissile phosphate (Figure 2.9B). This observation pointed to a role of Mg^{2+} as a Lewis acid by activating the nucleophile directly, or as a Bronsted base that deprotonates the 2'-OH. In reality, Mg^{2+} could be serving as both a Lewis acid and as a Bronsted base. In addition, Mg^{2+} could also stabilize the negative charge on the scissile phosphate, thus reducing the activation energy required to reach the transition state. The key nucleotides positioning the Mg^{2+} for this role are the G25•U20 reverse wobble base pair (*186*). When this G•U is mutated to an A•C base pair, the Mg^{2+} binding site is disrupted. The pH-rate profile of this mutant resembles the pH-rate profile of the HDV ribozyme in the absence of any divalent metals, with pK_as of 5.3 and 7.3. This confirmed that there is Mg^{2+} binding site in the active site. The pK_as correspond to C75 (7.3), and either an unknown base or acid denaturation of the ribozyme (5.3). In the presence of Mg^{2+}, the HDV ribozyme's pH-rate profile only shows a single pK_a around 6.1, which can be attributed to C75. Since the majority of the catalytic power of the HDV ribozyme comes from the C75 nucleotide, the HDV appears to have a 'hybrid engine' (*187*). It can function in absence or presence of Mg^{2+}.

The HDV catalysis demonstrates a direct role for divalent metals in catalysis, which was previously only seen in large ribozymes such as RNase P. HDV appears to use

many different strategies to produce a considerable rate enhancement of ~10^7 compared to the uncatalyzed cleavage. It uses a divalent metal to activate the nucleophile directly. C75 not only acts as a general acid but also positions the active site, and in doing so, shifts its pK_a upwards for more efficient catalysis. Both, C75 and Mg^{2+} also help lower the negative charge at the scissile phosphate. Lastly the cleavage site also adopts an in-line geometry needed for efficient catalysis. Hence, the HDV ribozyme is a good model RNA that can use almost all the catalytic strategies that a similar protein enzyme can. It also demonstrates the versatility of RNA catalysis because it can switch between different catalytic strategies based on the available conditions.

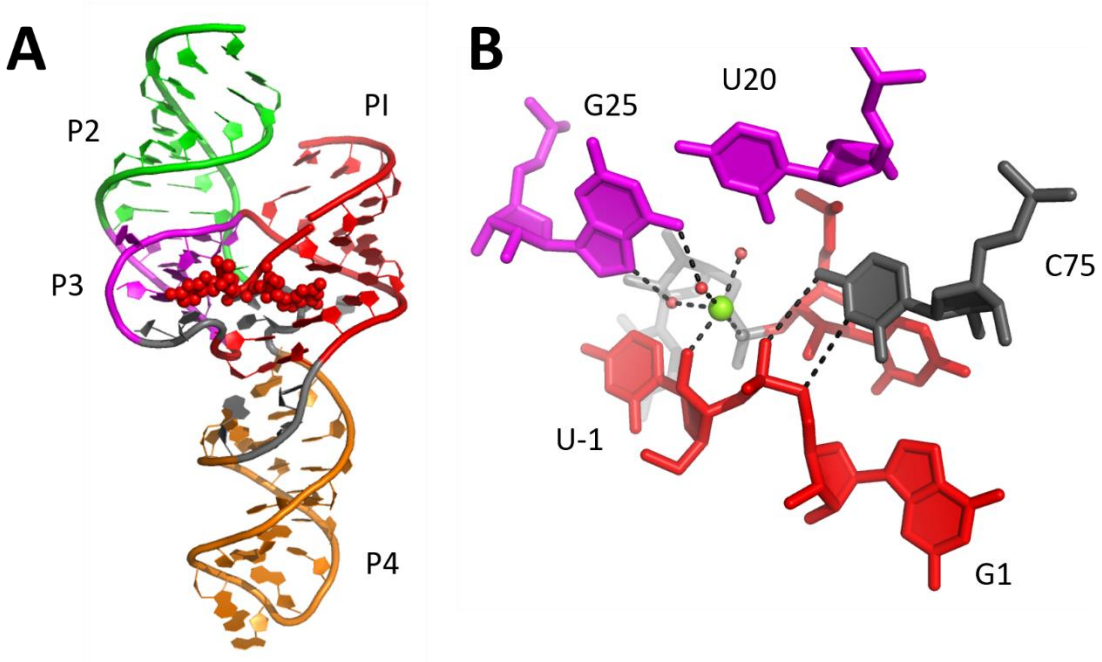

Figure 2.9. Structure of the HDV ribozyme.
A. The crystal structure of the HDV ribozyme is shown as cartoon. The paired regions are labelled and colored differently. Cleavage site dinucleotide atoms are denoted by red spheres. **B.** Active site of the HDV ribozyme. It was modelled using the active site of the hammerhead ribozyme. Nucleotides predicted to be important for catalysis are shown and labelled. The active site Mg^{2+} is shown as a green sphere. It is bound to three water molecules which are shown as red spheres. The interactions made by the divalent metal and the catalytic nucleotides are shown as dashed lines. The images were rendered in PyMol using coordinates from PDB ID 3NKB (*177*).

2.6.1.3 The VS ribozyme

The Varkud Satellite is another self-cleaving ribozyme whose catalytic properties depend on nucleobases (*188, 189*). Like the hairpin ribozyme it also undergoes a general acid-general base mechanism. The VS ribozyme can be isolated from the mitochondria of many species of *Neurospora*, where it functions in the rolling circle replication mechanism (*190*). It is the largest (by size) class of self-cleaving ribozymes

known (~150 nt) (*191*). The VS motif that was crystallized contains a dimer of seven helical segments that are joined by a three-way helical junction. The cleavage site is found in an internal loop within helix 1. The minimal version of this ribozyme is a monomer with six helical segments (I to VI) (Figure 2.10A and ref (*192*)).

The crystal structure of the VS ribozyme revealed a domain swapped dimer with two active sites, similar to our crystal structure of the RzB hammerhead ribozyme (Chapters 3 & 4). Helices I and VI contain the two indispensable catalytic nucleotides for VS ribozyme. In the crystal structure, the internal loop of helix I and the minor groove of helix VI interact to form the active site (Figure 2.10B). Hence, no Watson-Crick base pairing is observed between the substrate and the ribozyme, which is quite unlike other self-cleaving ribozymes. In the active site, the N1 of G638 is seen to be hydrogen bonding with 2'-OH of G620 (Figure 2.10B). Moreover, the A756 is found near the 5'-O. Like the hairpin ribozyme, this suggests a general acid-general base mechanism where G638 serves as the general base and A756 as the general acid. The biochemical data largely agrees with this catalytic model (*189, 193*). The VS ribozyme has a classic bell shaped pH-rate profile. If simulated, such a pH-rate profile can result from the pK_a values of ~5 and ~8.5. The 5'-PS substrate which abolishes the need for a leaving group only rescues the activity of a mutation at the 756 position, suggesting A756 is indeed the general acid in the catalytic mechanism (*193*). Finally, the observed pH rate profile suggests that the pK_as of both the G638, and A756 are shifted towards neutrality. The mechanisms behind these pK_a shifts are unclear, but electrostatics within the active site are the usual culprits for such pK_a shifts.

A final observation from the VS ribozyme is that the G620 and A621 nucleotides, which make up the cleavage site, adopt a so-called 'splayed conformation' like other self-cleaving ribozymes, which usually results in the in-line geometry of the nucleophile, the phosphate and the leaving group. However, in the crystal structure, this angle called τ is close to 97° (*192*). The angle, although more than a traditional τ angle of 70°, does not produce the required in-line geometry. While the solution conformation might be different, this does indicate that the in-line geometry might not be as important as other catalytic strategies for cleavage mechanisms (*164*). Hence, the primary catalytic strategy employed by the VS ribozyme is the nucleobase mediated general acid-general base catalysis, where the pK_as of the nucleobases are shifted due to the local environment of the active site.

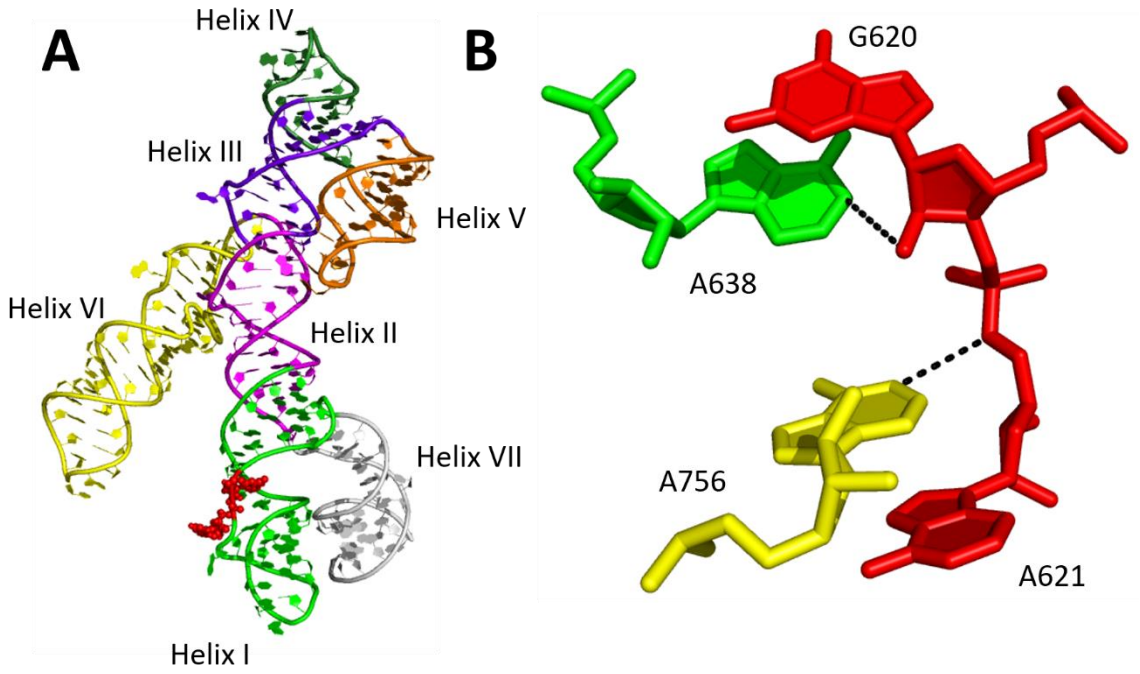

Figure 2.10. Structure of the G638A mutant of the VS ribozyme.
A. The overall crystal structure of the VS ribozyme is shown. Only a single monomer is shown. The ribozyme crystallized as a dimer. The seven helical segments are labelled and colored differently. The cleavage site dinucleotide is depicted as red spheres. **B.** The active site of the VS ribozyme dimer. The nucleotides involved in catalysis are labelled and the interactions between the general base A638 (G in wt) and general acid A756 are shown. The A756 in the active site is from the second monomer. The images were rendered in PyMol using coordinates from PDB ID 4R4V (*192*).

2.6.1.4 Twister and twister sister ribozymes

The twister ribozyme was discovered in 2014, after about a decade passed by without a discovery of a new self-cleaving ribozyme (*194*). Over 2700 different twister-like sequences were found in bacteria (mostly *Clostridia* genus), and some eukaryotic organisms. The twister ribozyme is associated very closely to the hammerhead ribozyme, and could even be functionally equivalent to the hammerhead ribozyme. In

fact, twister ribozyme is found within 6 kb of the hammerhead ribozyme in many instances, and that was an important reason that led to its discovery. The conserved secondary structure of the twister ribozyme consists of a stem loop like structure (Figure 2.11A and ref (*195*)). Within the stem-loop, there are two additional loops. Loop L1 makes a tertiary contact (pseudoknot) with the L4, and L4 makes an additional pseudoknot with L2. The cleavage site is found in L1. Like the hammerhead ribozyme, twister is also circularly permuted. It can be divided into three different classes depending on which stem contains the opening of the circle: P1, P3 and P5 (*194*).

Despite the recent discovery of the twister ribozyme, several crystal structures of the twister ribozyme have been solved (*195-197*). Discussed here is the structure obtained by the Lilley group, and the same numbering system is also used (*195, 198*). The crystal structure shows how loops L4 and L1 are brought together to make the active site (Figure 2.11). G33, the presumed general base in this catalytic mechanism is positioned in stem P2, whereas the general acid A1 is found immediately 3' of the scissile phosphate. A1 and U-1 also make up the cleavage site of this ribozyme. The pH-rate profile of the twister ribozyme is bell-shaped with pK_a values of 6.9 and 9.5 (*198*). The higher pK_a value can be attributed to the general base, G33. When G33 is mutated to an A, the pH rate profile still resembles a bell curve but become sharper and shifts to a lower pH value. This is consistent with adenine replacing guanine as a general base. The crystal structure, however is not in full agreement with this data. In the crystal structure, the N1 of G33 does not seem to be interacting with the 2'-OH (Figure 2.11B). This could be due to the absence of a 2'-OH in the inhibited complex used for

crystallization, or it could be an artifact of the crystallized form of the twister RNA. Instead G33, through its N3 position appears to make a hydrogen bond with the pro-R_p oxygen of the scissile phosphate, suggesting a role in the transition state stabilization during the cleavage mechanism (*195*). We observed a similar feature in the RzB crystal structure where 2'-OH of G8 appears to hydrogen bond with the pro-R_p oxygen of the scissile phosphate. This is a new catalytic strategy employed by the self-cleaving ribozymes. Wilson and colleagues argue that A1 functions as the general acid (*198*). In the crystal structure, N3 of A1 appears to be within interacting distance of the 5'O (Figure 2.11B). A shifted pK_a of A1 could certainly be the general acid. However, N3 has even lower pK_a than the N1 nitrogen. NMR studies have found the pK_a of A1 to be around 5.2, and mutational studies find significant rate loss due to substitutions at positions N3 and N1 (*199*). N3C substitution virtually makes the ribozyme inert, whereas N1C still retains activity. In addition, an N7C substitution, which increases the pK_a of both N3 and N1 positions shifts the pH-rate profile upwards. These data do indicate a role for A1 in the general acid catalysis, however the pK_a of A1 needs to shift further towards neutrality to agree with the pH-rate profile. Additionally, the twister ribozyme does not appear to adopt the in-line geometry at the cleavage site. Only one crystal structure shows such a conformation, but that may not be the active conformation (*197*). Hence, further investigation is necessary to reach a consensus on the catalytic mechanism of the twister ribozyme. In summary, the twister ribozyme uses a general acid-general base mechanism for catalysis, and might utilize a nucleobase for transition state stabilization.

The twister sister ribozyme was also discovered due to its association with the twister and the hammerhead ribozymes within the bacterial genomes (*176*). The secondary structure of the twister sister ribozyme looks remarkably similar to the twister ribozyme, hence the name twister sister. It consists of a stem-loop with two additional loops. The cleavage site is still located in L1, however the cleavage site location within the loop is different (*176*). In addition, there appear to be no pseudoknots in its tertiary structure, unlike the twister ribozyme. Early evidence suggests that it might share similar catalytic properties to the twister ribozyme due to a similar pH-rate profile and divalent metal dependence. However, the maximum rate achieved by twister sister ribozyme is considerably lower than the twister ribozyme. Hence, twister sister might just be a degenerate version of the twister ribozyme. Further investigation into the twister sister will provide further details about its catalytic properties.

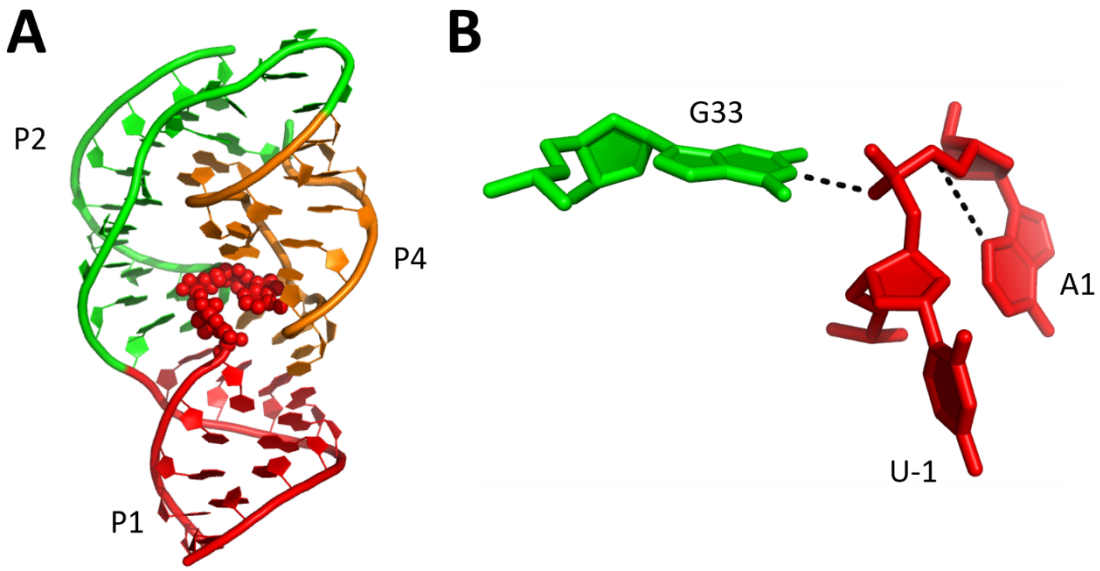

Figure 2.11. The inhibited pre-cleavage structure of the twister ribozyme.
A. The overall crystal structure of the *O. stavia* twister ribozyme is shown as a cartoon. The secondary structure elements are labelled and colored separately. The cleavage site dinucleotide is shown as spheres. **B.** The active site of the twister ribozyme is shown. The nucleotides presumed to be involved in catalysis are labelled and some of the important interactions are shown as dashed lines. The images were rendered in PyMol using coordinates from PDB ID 4OJI (*195*).

2.6.1.5 The pistol ribozyme

The pistol ribozyme was also discovered by the Breaker group due to its genetic association with other self-cleaving ribozymes (*176*). Despite its recent discovery, there is already a crystal structure and some early characterization of its biochemical properties (*200, 201*). The structure of the pistol ribozyme is composed of a hairpin (P1), which is connected to two base-paired stems (P2 and P3) (Figure 2.12A and ref (*200*)). The P2 and P3 are separated by an internal loop, which is where the cleavage site is located. It was also revealed that there is a pseudoknot interaction between the internal

loop and the loop in P1. The crystal structure revealed that six base pairs are involved in this pseudoknot. The pseudoknot brings the six conserved nucleotides (AAG and GCG) within the internal loop close to the cleavage site dinucleotides G-U (Figure 2.12B). This makes up the local environment of the active site. The pistol active site also displays an in-line trajectory at the cleavage site due to a 'splayed' conformation. The closest nucleotides to the cleavage site include the G40, and A32 (Figure 2.12B). N3 of A32 is positioned close to the 5'-O, and N1 of G40 is within 3.4 Å of the 2'-OH. Hence, G40 could act as a general base. The 2'-OH of A32 is also close to the 5'-O, however, hydroxyl group of the ribose has a considerably higher pK_a. Furthermore, a shifted pK_a of 4.7 was also found for A32 using NMR spectroscopy, supporting a general acid role for A32 (*200*). There was also a hydrated Mg^{2+} ion bound to the G33 nucleotide, and is 4 Å away from the pro-R_p oxygen. Lastly, mutations involving a substitution at the N1 position of G40 and N3 position of A32 had a substantial effect on catalysis.

Hence, the crystal structure gives rise to an early model of catalysis where N1 of G40 deprotonates the nucleophile, the N3 of A32 provides a proton to the leaving 5' oxygen, and a divalent metal possibly interacts with the pro-R_p oxygen of the scissile phosphate to stabilize the transition state. In addition, the cleavage site also adopts an in-line geometry. All these strategies give it a k_{obs} of approximately 10 min^{-1} in physiological conditions. The similarities with the RzB hammerhead ribozyme are quite noteworthy, and can also signal the need for more in-depth investigation of the catalytic mechanism of this ribozyme.

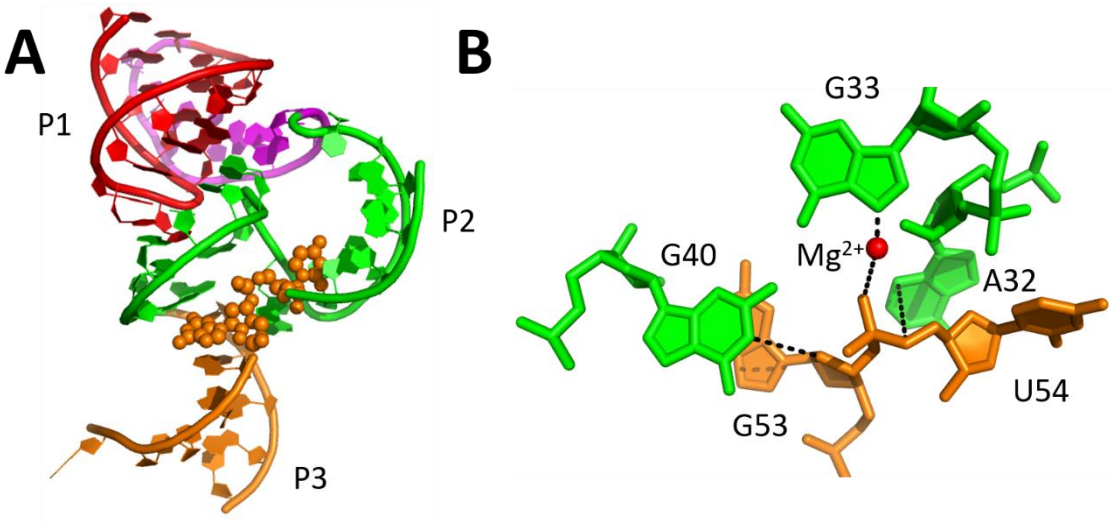

Figure 2.12. Inhibited pre-cleavage structure of the pistol ribozyme.
A. The crystal structure of the pistol ribozyme is shown as a cartoon. The secondary structure regions P1-P3 are colored separately. The cleavage site dinucleotides are shown as spheres. **B.** The active site of the pistol ribozyme is shown. The catalytic nucleobases and the divalent metals are labelled and some of the probable interactions are shown as dashed lines. The images were rendered in PyMol using coordinates from PDB ID 5K7C (*200*).

2.6.1.6 The hatchet ribozyme

The hatchet ribozyme was another product of the 'discovery by genetic association' approach. There are about 159 known examples of the hatchet ribozyme in all the genetic data available, however, rather surprisingly it is not found in any of the completely sequenced genomes (*176*). The environmental sequences it is found in may be bacterial and/or archaeal species, which have not been identified yet. The secondary structure consists of four stem regions named P1 through P4. P1 is a hairpin connected to P2 by a conserved CAA linker. There are two internal loops separating P2, P3 and P4. Like the HDV ribozyme, the cleavage site is located at the 5' end of the ribozyme just

before the start of P1. There is no crystal structure of the hatchet ribozyme at present. Early kinetic analysis of the hatchet ribozyme shows a maximum cleavage rate of ~4 min^{-1}, which is relatively low compared to other self-cleaving ribozymes, likely due to a more open structure as it doesn't appear to have any tertiary interactions in its structure (*202*). However, the importance upstream and downstream sequences of the hatchet ribozyme should not be discounted, as they could potentially form tertiary contacts needed to accelerate the rate of reaction. The maximum cleavage occurs at pH 7.5 and 10 mM Mg^{2+}. The pH-rate profile resembles that of the hammerhead ribozyme with a single apparent ionizable group at pH 7.5. However, more experimental details are needed to create a working model for the catalytic mechanism of the hatchet ribozyme.

2.6.2 Self-cleaving ribozymes found in pre-mRNAs

Up until 2004, it was thought that self-cleaving ribozymes are only found in the viral genomes, and that they do not have any functional relevance in higher organisms. This was because all the self-cleaving ribozymes discovered prior to the *glmS* ribozyme discovery were only discovered in viral genomes, where their only function appeared to be in the rolling circle replication mechanism (Figure 2.13A). However, that all changed with the discovery of the *glmS* ribozyme, which is involved in gene-regulation (Figure 2.13B) (*203*). The manner of the *glmS*'s discovery via conserved motif searching opened the floodgates for other self-cleaving ribozymes. Several groups developed software to search for the existing self-cleaving ribozyme based on consensus secondary structures.,

which led to the discovery of more self-cleaving ribozymes in higher organisms with potential roles in gene expression (Figure 2.13B). The majority are still found in the intergenic regions, however several have been found in intronic sequences, and untranslated regions of the messenger RNAs (*204-206*). A few examples will be discussed below.

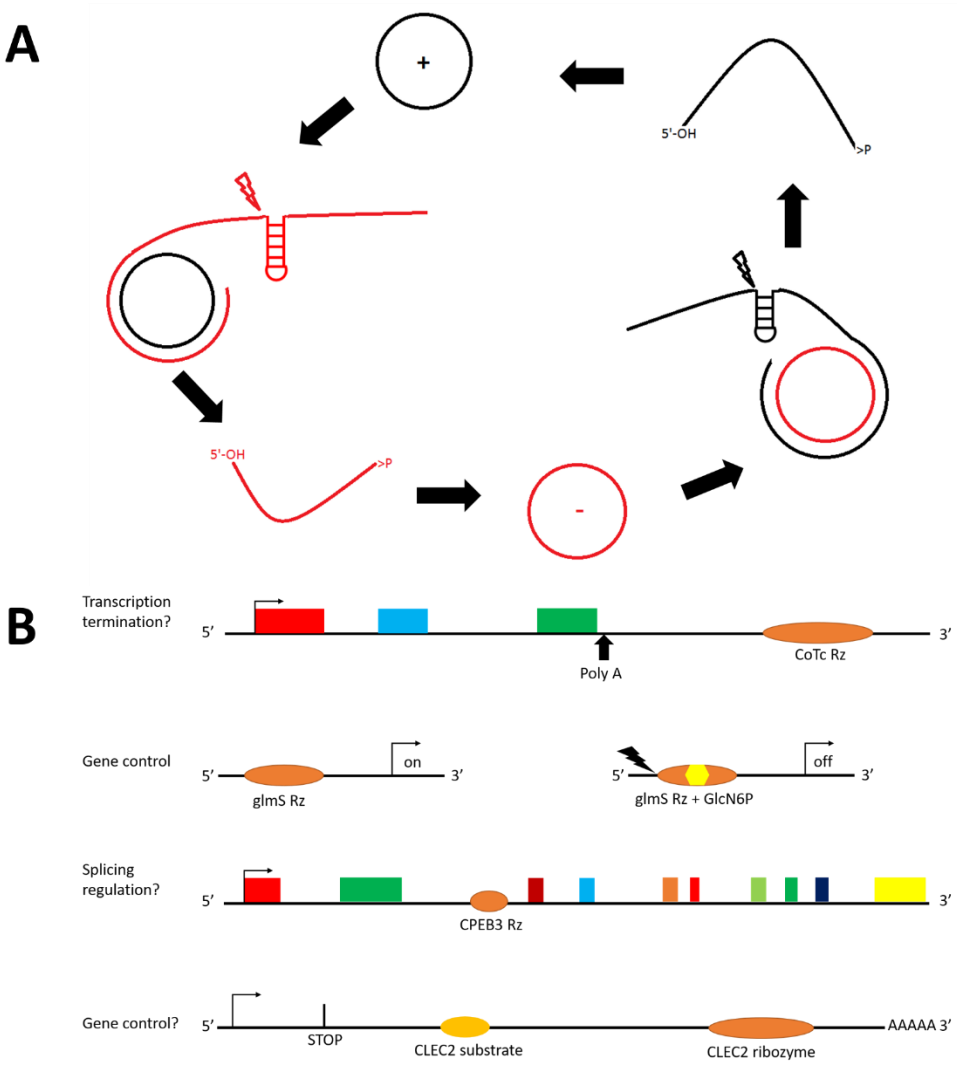

Figure 2.13. Biological roles of self-cleaving ribozymes.
A. The majority of the self-cleaving ribozymes function in the rolling-circle replication in satellite RNAs from plant viruses. The genomic (+) and anitgenomic (+) strands are replicated, cleaved by ribozymes, and then ligated by ribozymes. **B.** The roles of four self-cleaving ribozymes found in mRNAs are shown. CoTc ribozyme is found downstream of the polyA tail in the human beta globin gene and cleaves in the presence of GTP, the *glmS* ribozyme is found in the 5'-UTR region of the *glmS* mRNA and cleaves only in the presence of GlcN6P, CPEB3 ribozyme is located in the second intron of CPEB3 gene in humans, and CLEC2 ribozyme is located in the 3'-UTR region of C-type lectin II gene in mammals. CLEC2 ribozyme is a discontinuous ribozyme because the substrate and ribozyme strands are located away from each other. Apart from the *glmS* ribozyme, none of the other functions of the ribozymes have been confirmed.

2.6.2.1 The *glmS* ribozyme

The rise of bioinformatics search methods together with availability of sequenced genomes led to the discovery of several conserved RNA motifs. These included the riboswitch RNAs that respond to environmental stimuli by undergoing conformational changes upon binding small molecules, leading to changes in the expression of mRNAs (*24*). A similar search for conserved RNA sequences in bacteria revealed a conserved sequence in the 5'-UTR regions of the *glmS* genes in gram-positive bacteria (*203*). This gene encodes for glucoseamine-6-phosphate synthetase (GlmS). GlmS enzyme catalyzes the formation of glucoseamine-6-phosphate (GlcN6P) from fructose-6-phopshate and glutamine (*207*). This is the second step in the hexoseamine pathway (*208*). Unlike other riboswitches, the conserved element within the 5'-UTR region did not change conformation in the presence of the reaction products, but instead it cleaved the mRNA (*203*). Subsequent investigation found that this was actually a self-cleaving ribozyme due to the production of RNAs with 2',3' -cyclic phosphate and 5' -hydroxyl (*26*). It turns out that under sufficiently high concentration of GlcN6P, the 5'-UTR region undergoes cleavage leading to formation of a 5'-OH on the mRNA. An exposed hydroxyl on the 5' site of the mRNA is a signal for mRNA decay, as RNase J1 degrades such mRNAs in bacteria (*209*). Loss of *glmS* ribozyme activity leads to sporulation defects in certain bacteria (*210*). It is the first phenotype observed that is controlled by a self-cleaving ribozyme.

The secondary structure of the *glmS* ribozyme is composed of four stem loop regions called P1 through P4. P2 and P4 contain internal loops whereas P3 and P1 have end loops. The cleavage site is located on the 5' end of the ribozyme (Figure 2.14A and ref (*211*)). There are three pseudoknots in the structure of the *glmS* ribozyme (Figure 2.14A). The 5' region of the ribozyme makes a pseudoknot interaction with P2.2. A second pseudoknot is formed when the linker region between P2 and P1 base-pairs with the P2.1 region. The last pseudoknot is the interaction between the 3' end of the ribozyme with P3.1. This triple pseudoknotted structure makes the *glmS* ribozyme highly compact and reactive. Rate enhancement of up to 10^7 can be achieved under the right experimental conditions. Unlike other riboswitch domains, GlcN6P does not induce a conformational change in the structure of the *glmS* ribozymes. In the crystal structures the active site structure is pre-organized prior to GlcN6P binding (*211*). Instead, it acts as a coenzyme for this catalysis. This is the first and the only known cofactor based catalytic strategy employed by self-cleaving ribozymes, and has important repercussions for the RNA world hypothesis, as discussed in Chapter 1.

GlcN6P is thought to be extremely important for the cleavage mechanism of the *glmS* ribozyme (*212*). Using several different GlcN6P-like activators, it was found that the amine group is vital for catalysis, indicating a role for the GlcN6P in the catalytic mechanism. In the crystal structures, the ammonium group is able to hydrogen bond with the 5'-leaving oxygen (Figure 2.14B). Furthermore, there is a linear correlation between the microscopic pK_as of different cofactor analogs and the apparent pK_a of the ribozyme (*212*). This, together with the observations from the crystal structures,

implicates the GlcN6P in general acid catalysis. In addition, the hydroxyl group on the C1 position can hydrogen bond with the pro-R_p oxygen of the scissile phosphate providing the necessary stabilization during the transition state. It has also been suggested that the GlcN6P may also participate in general base catalysis. The reason being that the ribozyme is almost completely inactive without it, despite a pre-organized active site. The crystal structures also show two well-ordered water molecules in the active site which could, theoretically be activating the nucleophile (*211*). However, the general base catalysis role for GlcN6P is far from conclusive. The G40 nucleotide, the closest nucleotide to the 2'-OH of A-1 in crystal structure, may act as the general base. The N1 of G40 is within 3 Å of the nucleophile, indicating that it may be responsible for the deprotonation of the 2'-OH of A-1. This role is supported by the fact that when G40 is mutated to A40, the ribozyme becomes completely inactive, showing its importance for catalysis (*213*). The G40A mutant also folds into the same tertiary structure as the WT *glmS* ribozyme, ruling out major folding defects in the mutant ribozyme. G40 could also be involved in orienting the active site during the transition state, since it interacts with 2'-OH and the pro-S oxygen in the cleavage site (Figure 2.14B). It has also been suggested that the G40 nucleobase may be assisted by a Mg^{2+} for general base catalysis similar to the RzB hammerhead ribozyme (Chapter 3) (*214*).

The *glmS* ribozyme is another self-cleaving ribozyme that uses the general acid-base catalysis. The role of general acid is most likely played by its cofactor, GlcN6P. The cofactor also helps stabilize the transition site. However, activation of the 2'-OH is still unclear. N1 of G40 (also referred to as G33), a water molecule, or GlcN6P are the likely

culprits. The *glmS* ribozyme is an important discovery in the world of ribozymes because it provided us with a new strategy that RNA could utilize to self-cleave.

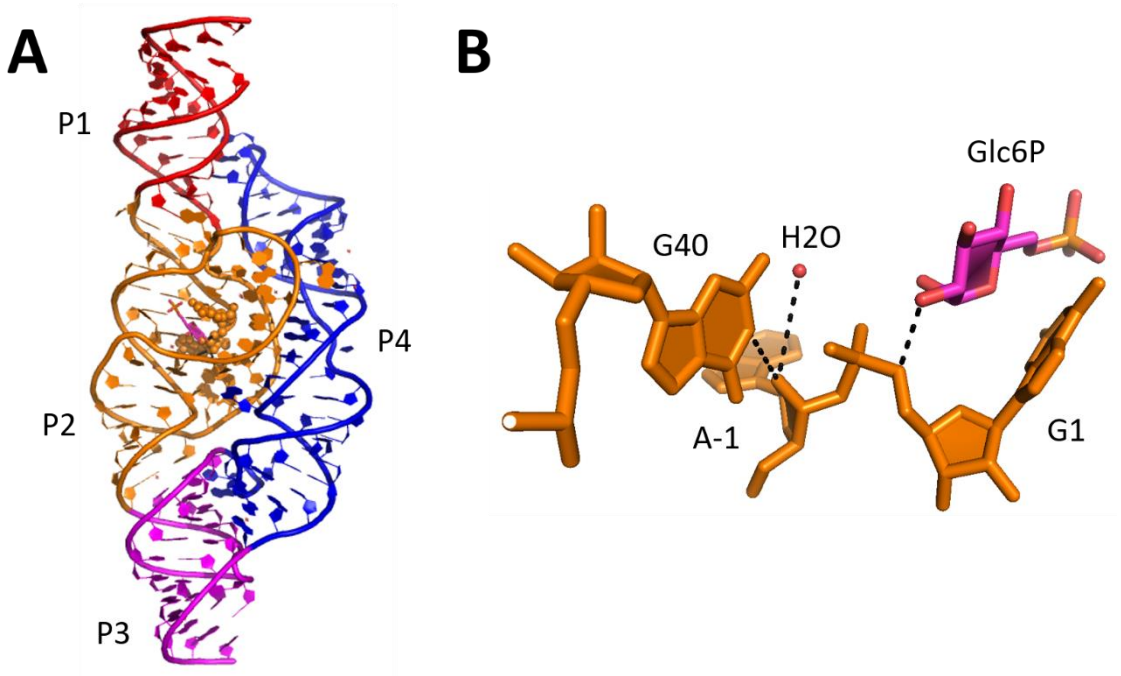

Figure 2.14. Pre-cleavage structure of the *glmS* ribozyme.
A. The overall crystal structure of the *glmS* ribozyme is depicted as a cartoon. The different secondary structure regions are colored differently, and the cleavage site dinucleotides are shown as small spheres. **B.** The active site of the *glmS* ribozyme is shown. The important players in the catalysis of *glmS* ribozyme are labelled. The images were rendered in PyMol using coordinates from PDB ID 2H0Z (*211*).

2.6.2.2 Other self-cleaving ribozymes: CPEB3, Vg1, CoTc, and CLEC2 ribozymes

Apart from the *glmS* ribozyme, several other self-cleaving ribozymes have been found that are a part of the mRNA either in the pre- or mature form. These include the CPEB3 ribozyme, the CLEC2 ribozyme, the CoTc ribozyme, and the Vg1 ribozymes. The

CPEB3 ribozyme and the CLEC2 ribozyme are variations of the HDV ribozyme, and the hammerhead ribozyme respectively, whereas the CoTc ribozyme and Vg1 ribozyme are not very well characterized. Therefore, the discussion here will focus on the function rather than the catalytic mechanism of these ribozymes.

The HDV-like mammalian cytoplasmic polyadenylation element-binding protein 3 (CPEB3) ribozyme was discovered via an *in-vitro* selection strategy, which was modified to find ribozymes in the human genome (*205*). It was found to be embedded in the second intron of the CPEB3 gene. CPEB3 protein is thought to be involved in the regulation of mRNA polyadenylation. Later, CPEB3 ribozyme was found to be highly conserved in mammalian species suggesting that it evolved relatively recently (*175*). Initial characterization suggested that CPEB3 is a slow reacting ribozyme with a rate constant of approximately 0.01 min^{-1}, perhaps to prevent interference with normal splicing mechanisms. However, the Bevilacqua group found that CPEB3 itself is intrinsically fast reacting (*215*). They demonstrated the importance of the upstream sequence to the HDV ribozyme for fast kinetics. There is also speculation that a trans-acting factor, such as a ligand could upregulate the cleavage rate *in-vivo*, resulting in premature termination of the mRNA, and leading to downregulation of gene expression or instead a shortened mRNA (*216*). However, evidence for these roles is not available yet.

The CLEC2 ribozyme is an interesting self-cleaving ribozyme, which resembles a trans-acting hammerhead ribozyme. It was discovered in the 3'-UTR region of the C-type lectin II (Clec2) genes in rodents (*204*). The loop in the stem I region of the hammerhead

ribozyme consists of hundreds of nucleotides. Hence, it was called the discontinuous hammerhead ribozyme. The self-cleavage of the CLEC2 ribozyme has been demonstrated *in-vitro* and *in-vivo*. Cleavage *in-vivo* was also shown to downregulate protein expression, cleavage results in the removal of the poly-A tail of the mRNA, which is essential for efficient translation. The precise mechanism of gene regulation by the CLEC2 ribozyme is still unclear. The self-cleavage activity of the CLEC2 ribozyme must be regulated to fine-tune the expression of the protein, and it could be done through the folding of the ribozyme. Since, the CLEC2 ribozyme results from the interaction between two domains located far from each other, certain conditions within the cell may enable these domains to come together resulting in cleavage of the mRNA.

The CoTc (co-transcriptional cleavage) ribozyme was discovered when researchers studying the transcription termination of the human β-globin gene found that the pre-mRNA was being co-transcriptionally cleaved (*206*). They went on to characterize the minimum core of the ribozyme, which was also found to be highly conserved in the 3' regions of β-globin genes in primates. So, the CoTc ribozyme may be involved in transcription termination in these organisms. However, subsequent evidence in support of a CoTc ribozyme, and its characterization is lacking.

Lastly, the Vg1 ribozyme was found in the 3'-UTR region of Vg1 and β-actin genes (*217*). Vg1 only cleaves in the presence of Manganese or Cadmium ions. It is the shortest motif capable of undergoing cleavage, since it only requires a stem composed of UUU and GAAA. *In-vitro* assays demonstrated cleavage, which was dependent not only on pH and divalent metals (Mn^{2+} and Cd^{2+}), but also on the RNA concentration. However, *in-*

vivo cleavage activity was found to be lacking. There was no physiological difference between mRNAs with or without the Vg1 minimum motif. It is perhaps not very surprising, given that a stem of four base pairs involving mostly A-U base pairing, is unlikely in the context of a much longer RNA.

These ribozymes are important because they indicate that self-cleaving ribozymes might play important functional roles in gene regulation. This is especially true for the CLEC2 and CPEB3 ribozymes. It might be the case that other self-cleaving ribozymes have similar roles in nature, but just have not been discovered yet. This was part of the reason why we are pursuing self-cleaving ribozymes using a functional genomics approach (Chapter 5).

2.7 Lessons from existing natural ribozymes, and the need to explore more

Natural self-cleaving ribozymes employ different catalytic strategies to cleave a phosphodiester bond, and they are summarized in Figure 2.15. We have seen that self-cleaving ribozymes tend to activate the nucleophile (2'-OH) using different approaches. A common strategy is to use a nucleobase, to deprotonate the nucleophile, and to protonate the leaving 5'-O. This strategy called the general acid-general base strategy must overcome the relatively extreme pK_as on the functional groups of RNA nucleobases. Efficient proton transfer at physiological conditions requires that the pK_as of proton donors and acceptors be close to neutral pH. We have seen that RNA uses its tertiary structure to position nucleobases, such as cytosines, and adenines close to

negatively-charged phosphate groups. It allows the RNA to stabilize the protonated forms of nucleobases, which shifts the pK_as towards the basic pH. We witnessed this phenomenon in several ribozymes like the HDV, pistol, twister, and the hairpin ribozymes. On the other hand, nucleobases like guanines and uracils tend to have pK_as on the basic side. To shift the pK_as towards neutrality, RNA folding places the deprotonated nucleobases close to positively charges groups, such as divalent metals, and monovalent metal ions. Such a strategy can be seen in the hammerhead ribozyme (Chapter 3). Some ribozymes might employ hydrated divalent metals as general acids or general bases. The hammerhead ribozyme might be using this strategy for general acid catalysis (Chapter 4). The *glmS* ribozyme is thought to use a cofactor for general acid catalysis. The HDV ribozyme also uses a divalent metal to directly activate the nucleophile (Lewis acid). Usage of Lewis acid catalysis to stabilize the leaving 5' – O has not been observed in the self-cleaving ribozymes so far.

Specific acid or specific base catalysis can also be employed by self-cleaving ribozymes, as some studies suggest for the *glmS* ribozyme (specific base). It is also possible that many self-cleaving ribozymes use multiple strategies for the same purpose. For example, the HDV ribozyme can function in the absence of divalent metals, and it is likely that a specific base provides the nucleophile activation in such conditions. Additionally, the *glmS* ribozyme's crystal structure also showed a water molecule close to the nucleophile. The transition state in the self-cleavage reaction results in a huge build-up of negative charge on the non-bridging oxygens. Self-cleaving ribozymes use hydrogen bonding via nucleobases or cofactors to stabilize the transition state. There

are also examples of divalent metals stabilizing the negative charge on the non-bridging oxygens (hammerhead, HDV and pistol). Finally, many self-cleaving ribozymes also adopt the so-called 'splayed conformation' to achieve an in-line geometry between 2'-OH, P and 5'-O. However, the rate-enhancement from this strategy is only expected to be 10-20 fold (*218*). It is perhaps for this reason why ribozymes such as the twister and the VS ribozymes do not adopt in-line geometries in their respective cleavage sites.

Larger ribozymes also use several different catalytic strategies that take advantage of RNA structure to position the nucleophiles, leaving groups, divalent and monovalent metal ions to catalyze different chemical reactions. The nucleophiles are the chemical moieties on either endogenous nucleobases of RNA or exogenous molecules. The larger ribozymes also tend to rely heavily on divalent metals for catalytic purposes. Several large ribozymes use divalent metals for transition state, and leaving group stabilization. There are also predicted roles for active site waters in nucleophile activation. Several of the larger ribozymes also work together with proteins to achieve efficient catalysis.

Some of the catalytic strategies used by larger ribozymes have not been observed in self-cleaving ribozymes primarily because self-cleaving ribozymes tend to be small ribozymes. However, it is likely that many larger self-cleaving ribozymes which may use such catalytic strategies have not been discovered yet. Furthermore, there may be other unknown catalytic strategies that RNA could potentially use, which can be studied by finding novel self-cleaving ribozymes. Such catalytic strategies may also be tied to their roles in gene regulation of mRNAs or non-coding RNAs. For example, one can imagine a self-cleaving ribozyme that only cleaves in response to binding of amino acids or

peptides. Such hypothetical ribozymes might be of significance in the regulation of metabolic pathways related to amino acid synthesis. Similarly, in the absence of RNA chaperones and helicases, RNA folding is likely to change, and this could potentially lead to ribozyme-mediated cleavage of mRNA or non-coding RNA followed by degradation. Hence, it is important to find new approaches of discovering novel self- or trans-cleaving ribozymes, which was the motivation for Chapter 5. In addition to discovering new ribozymes, structural and biochemical characterization of existing self-cleaving ribozymes and other natural ribozymes is also crucial. Many new catalytic strategies and their relevance to ribozyme's biological roles will only become apparent through such in-depth studies. This is one of the reasons why the next two chapters are dedicated to the in-depth analysis of the catalytic mechanism of the hammerhead ribozyme.

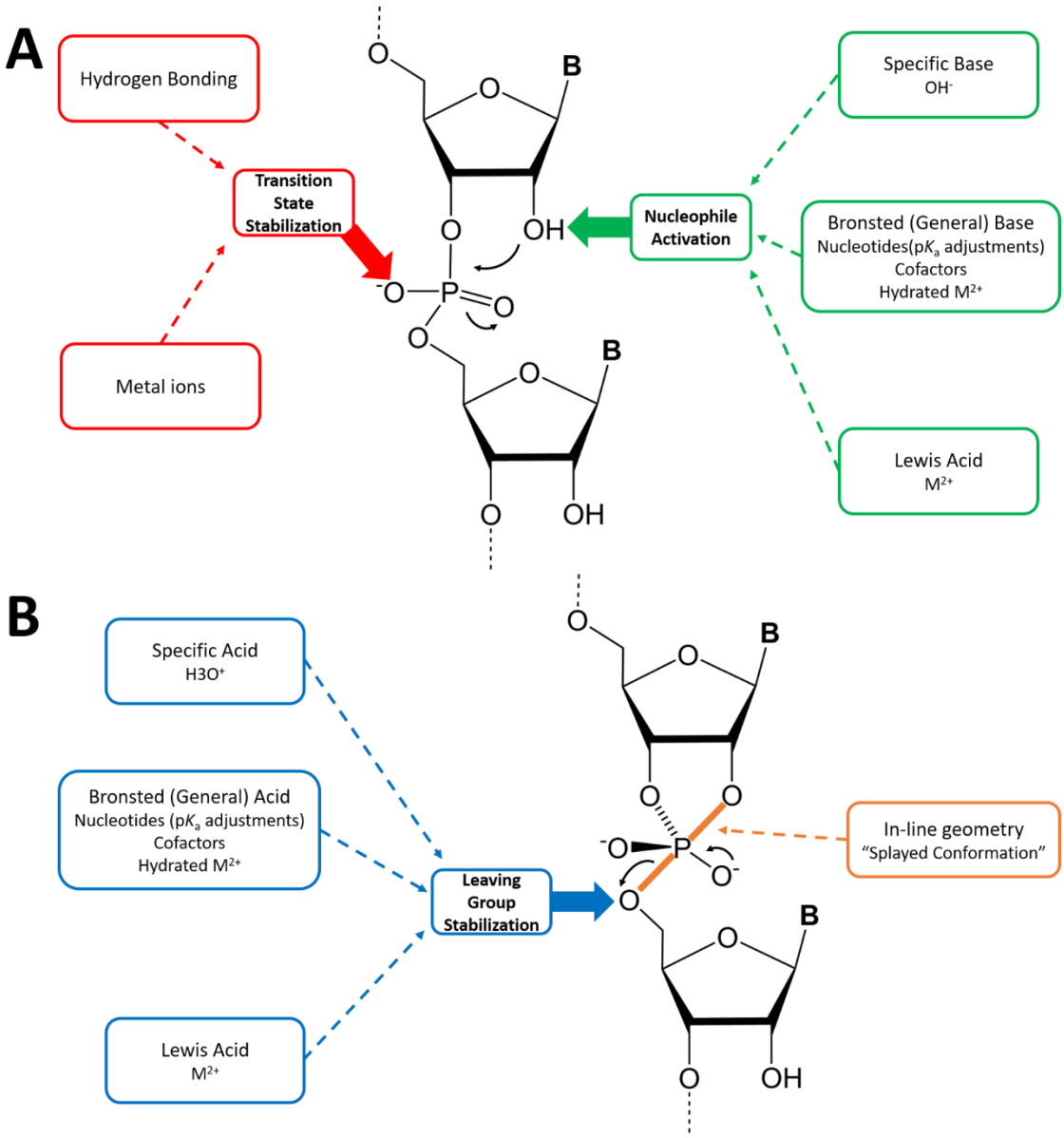

Figure 2.15. Catalytic strategies employed by self-cleaving ribozymes.
Four major categories are described as observed in natural self-cleaving ribozyme. Nucleophile activation and charge stabilization of the non-bridging oxygen is described in **A**, whereas, leaving group stabilization and in-line geometry is described in **B**. See test for details.

CHAPTER 3. TWO DIVALENT METALS IONS AND CONFORMATIONAL CHANGES PLAY ROLES IN THE HAMMERHEAD RIBOZYME CLEAVAGE REACTION

3.1 Declaration of collaborative work

The work described in this chapter was the result of collaborative efforts of Aamir Mir, Ji Chen, Emma Lendy, Kyle Robinson, Jaclyn Goodman, and David Neau, under Barbara Golden's supervision. Ji Chen crystallized, collected X-ray diffraction data (together with David Neau) and solved all the crystal structures of the WT RzB. Aamir Mir was assisted by Emma Lendy, Jaclyn Goodman, and Kyle Robinson to prepare and crystallize the G12A mutants of the RzB ribozyme. Aamir Mir and David Neau collected the diffraction data, and Aamir Mir solved all the crystal structures of the hammerhead ribozymes. Kinetic analyses described in this chapter were performed by Aamir Mir, Emma Lendy, and Kyle Robinson. Aamir Mir prepared the figures containing kinetic analyses, and Barbara Golden, Aamir Mir, and Ji Chen prepared figures containing structural data. The chapter was written and edited by Barbara Golden, Aamir Mir, and Ji Chen. Dan Hershlag, and Victoria DeRose provided critical insights in the preparation of this chapter. The text, and figures were reprinted and adapted with permission from reference (*219*).

3.2 Introduction

In 1987, Olke Uhlenbeck first reported the existence of a small, self-cleaving RNA known as the hammerhead ribozyme (*220*). This compact RNA catalyst was composed of three short base-paired stems and a cluster of core nucleotides. Although these minimal hammerhead ribozymes are functional, it was eventually discovered that maximal activity requires some form of tertiary interaction between stems I and II (Figure 3.1A). Using the known conserved elements of the hammerhead ribozyme, several groups have used bioinformatics and secondary structure predictions to search for hammerhead-like sequences in the available genomes. Now, thousands of hammerhead ribozyme sequences have been identified in this spanning all domains of life (*221-224*). While many viroids use hammerhead ribozyme sequences to processes their RNA genomes, the physiological functions of the majority of these hammerhead-like sequences are still unclear.

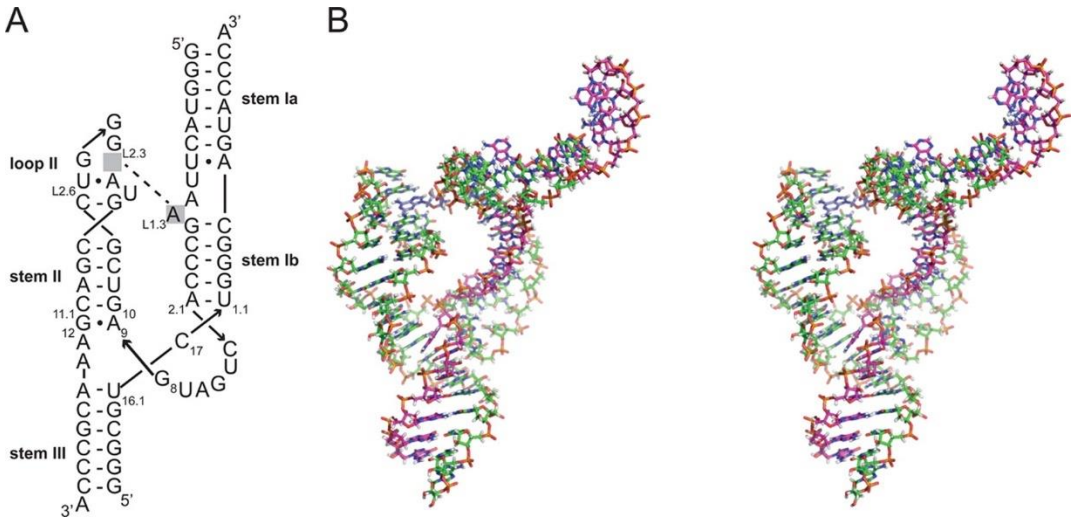

Figure 3.1. Secondary structure and tertiary structure of the hammerhead ribozyme. **A.** The RzB hammerhead ribozyme folds into three helical domains, stem I, stem II, and stem III. Stem I is interrupted by an internal loop, and stem II is capped by loop II. A single adenosine base, A(L1.3), from stem I makes a tertiary contact within loop II. The numbering system used here was described by Hertel et al. **B.** The tertiary structure of the RzB ribozyme (wall-eyed stereo).

Although the hammerhead ribozyme was the first small nucleolytic ribozyme to be discovered, its catalytic mechanism remains elusive. Most small nucleolytic ribozymes often cleave their target RNAs by activating a 2′-hydroxyl for nucleophilic attack at the phosphate of the adjacent nucleotide (Figure 3.2). The 5′-hydroxyl of the following nucleotide is released, generating 2′,3′-cyclic phosphate and 5′-hydroxyl termini on the products. This reaction requires deprotonation of the 2′-hydroxyl that serves as the nucleophile and protonation of the 5′-hydroxyl leaving group. The mechanism by which this reaction occurs, however, varies from ribozyme to ribozyme.

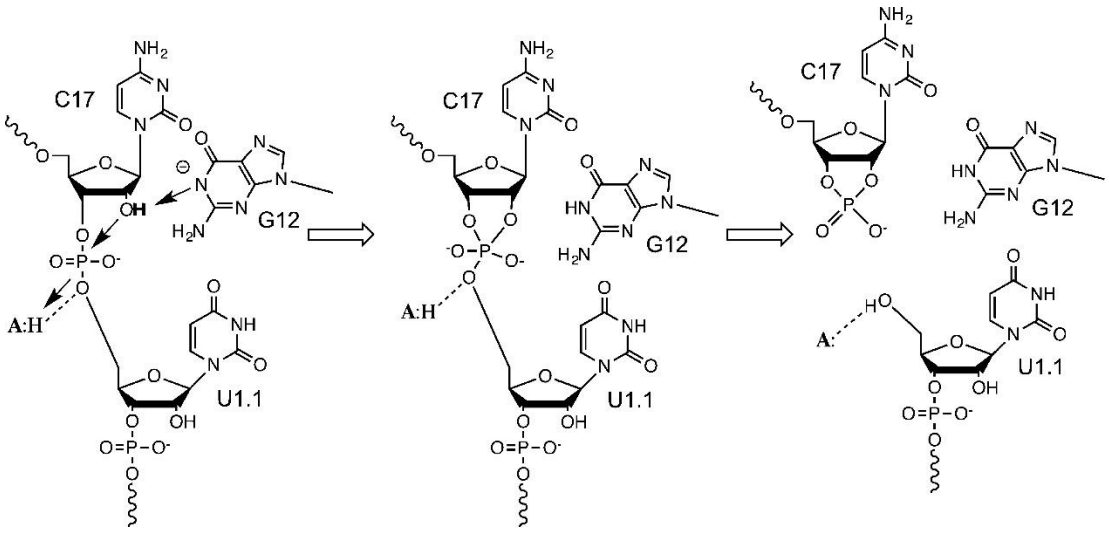

Figure 3.2. Cleavage reaction of the hammerhead ribozyme.
Small nucleolytic ribozymes cleave their substrates by activating a 2'-O for attack at the phosphate of the following nucleotide and release of the 5'-OH. G12 appears to serve as a general base, accepting a proton from the 2'-OH of the nucleophile. A general acid (A:H) could facilitate catalysis by donating a proton to the 5'-O leaving group

Small nucleolytic ribozymes can be divided into two distinct classes based on their ability to react in the presence of $[Co(NH_3)_6]^{3+}$ (*187*). $[Co(NH_3)_6]^{3+}$ is a trivalent ion with nonexchangeable ligands that mimics hydrated Mg^{2+} ions. Unlike $[Mg(H_2O)_6]^{2+}$, $[Co(NH_3)_6]^{3+}$ cannot coordinate directly to ligands from the RNA, and $[Co(NH_3)_6]^{3+}$ does not readily participate in proton transfer reactions because the pK_a of the amine ligands of $[Co(NH_3)_6]^{3+}$ is much higher than that of a Mg^{2+}-bound water (*225*). Class I ribozymes, such as the hairpin ribozyme, glmS ribozyme, and twister ribozyme can react in the presence of $[Co(NH_3)_6]^{3+}$ alone (*226-230*). Consistent with this lack of dependence on Mg^{2+}, none of these ribozymes appear to use metal ion-mediated catalysis in the reaction mechanism. Class I ribozymes use nucleobase or cofactors to perform general

acid or general base catalysis. In many class I ribozymes, the pK_a of the general acid is shifted by the active site environment toward neutrality, increasing the population of the catalytically active protonation state (*173, 184, 231, 232*).

In contrast, class II ribozymes, such as the hammerhead and hepatitis delta virus (HDV) ribozymes, do not have high activity in the presence of $[Co(NH_3)_6]^{3+}$ alone, and their Mg^{2+}-dependent reaction is strongly inhibited by the presence of $[Co(NH_3)_6]^{3+}$ (*233-235*). Recent crystallographic studies have demonstrated that the hepatitis delta virus ribozyme binds a Mg^{2+}ion within its active site (*177*). Mutational analyses and simulations suggest that this Mg^{2+} ion participates in the cleavage reaction (*186, 236*). On the basis of the response of the hammerhead ribozyme to $[Co(NH_3)_6]^{3+}$, this ribozyme would also be predicted to use Mg^{2+} ions in its catalytic mechanism (*234*).

A Mg^{2+}-independent mechanism for the hammerhead ribozyme has been proposed based on cross-linking studies, kinetic analyses, and the crystal structure of the *Schistosoma mansoni* hammerhead ribozyme (*237-240*). In this mechanism, the nucleobase of G12 serves as a general base to abstract the proton from the 2'-hydroxyl that serves as the nucleophile. The N1 of G12 would therefore need to be deprotonated, and negatively charged, prior to cleavage (Figure 3.1). The general acid was proposed to be 2'-hydroxyl of G8 which is observed to be hydrogen bonded to the 5'-hydoxyl leaving group in crystal structures.

This proposed mechanism is problematic for several reasons. The pK_a's of N1 of G12 and 2'-OH of G8 are ~9 and ~13 respectively if they are not shifted by the active site environment. These pK_a values are far from neutrality, and, unless their pK_a's are

shifted by the active site, are not optimized for general acid–base catalysis near neutrality as the fraction of ribozyme in the active state would be very small (*241*). It is not clear what elements within the hammerhead ribozyme would serve to shift the pK_a's of the proposed general acid and general base.

There is also considerable evidence from solution biochemical experiments that a divalent metal ion, such as Mg^{2+}, is involved in the cleavage reaction. A direct interaction between a metal ion and the pro-R_p oxygen of the scissile phosphate has been proposed based on metal rescue of phosphorothioate-substituted substrates (*242-245*). While the hammerhead ribozyme is capable of reaction in the absence of divalent metal ions, (*246*) very high concentrations of monovalent ions are required to promote cleavage under these conditions, and the reaction appears to proceed through an alternate channel (*234, 243*).

Divalent metals have been shown to produce considerable rate enhancements when compared to reactions in their absence, (*247*) and the rate constant of the reaction appears to be affected by the identities of the divalent metals used (*248*). For example, reactions in Mn^{2+} proceed faster than Zn^{2+} or Mg^{2+}. Crystal structures and electron paramagnetic resonance experiments have revealed a high affinity Mn^{2+} binding site at the A9–G10.1 region; (*249-251*) however based on crystal structures, this ion cannot interact directly with the cleavage site. York and colleagues have performed molecular dynamics simulations that suggest prior to catalysis a metal ion moves from the position near G10.1 observed in the crystal structure to the scissile phosphate where it is well-positioned to participate in catalysis (*252-254*). Together,

these studies point to significant roles for divalent metal ions in the cleavage reaction of the hammerhead ribozyme under cellular conditions.

To address these open questions in what may be the most heavily investigated of the known small nucleolytic ribozyme, we have reinvestigated the crystal structure of the hammerhead ribozyme using an artificial hammerhead ribozyme, RzB, generated by the Burke laboratory (*255*). Compared to the *S. mansoni* and satellite tobacco ringspot virus hammerhead ribozymes previously studied, (*238, 240, 256*) this ribozyme has a different tertiary interaction between stems I and II, crystallizes under different conditions, and has distinctly different crystal packing interactions. In spite of these differences, the active site structure is not significantly different than the previously determined structures.

We therefore examined the activity, metal specificity, and the pH–rate profiles of ribozymes with various mutations. We found that the solution biochemical results were not fully consistent with the crystal structures. We identify two divalent ion binding sites which have the potential to directly contribute to metal-mediated catalysis in the hammerhead ribozyme. Taken together, these data suggest that the ground state of the hammerhead ribozyme observed in crystal structures does not represent a catalytically active conformation, but rather a precatalytic state that must rearrange, at least subtly prior to catalysis (*245, 252*).

3.3 Materials and Methods

3.3.1 RNA Synthesis

The two trans-acting versions of the hammerhead ribozyme were designed based on the previously reported sequences of *S. mansoni* ribozyme (*238*) and SELEX derived RzB (*255*). To make the ribozyme strand, the ribozyme sequence was placed between a T7 promotor and an EarI restriction site, which was then inserted into pUC-19 and amplified by *E. coli* DH5α cells. DNA plasmid was obtained using QIAGEN Mega kit. *In vitro* transcription and RNA purification by denaturing PAGE were performed as previously described (*257*). The RNA sample was concentrated using an Amicon Ultra centrifugal filter and stored at -20 °C before use.

For crystallization, a 20-nt inhibitor strand carrying a 2'-deoxy or 2'-amino modification at the C17 position was purchased from Thermo Scientific, deprotected and further purified by denaturing PAGE gel as previously described. For solution kinetic experiments, the substrate used was a chemically synthesized, DY-547 labeled oligomer containing an all ribose backbone.

3.3.2 RNA Crystallization

To refold the hammerhead ribozyme, equal molar concentrations of ribozyme strand and the inhibitor strand were mixed in a buffer containing 10 mM potassium

cacodylate at pH 6.0). For transition state mimic, the inhibitor strand was used in two pieces: 7 and 13 nt long RNA pieces of the substrate strand. The complex was heated at 90 °C for 1 min, cooled to room temperature for 10 min. $MgCl_2$ was then added to the complex to achieve a concentration of 10 mM. Following addition of Mg^{2+}, the complex was heated at 50°C for 5 min and cooled to room temperature for 20 min prior to use. For transition state mimic 2 mM ammonium vanadate was also added prior to incubation at 50 ^{o}C. The final concentration of the refolded complex was 5.5 mg/mL.

The ribozyme-inhibitor complex was crystallized by vapor diffusion using hanging drops. Drops were created by mixing 1 or 2 µL of the refolded complex with 1 µL of a reservoir solution containing 33 % MPD, 0.4 M potassium chloride, 50 mM potassium acetate pH 5.0, 0.5mM spermine, and for transition state mimic, 2 mM ammonium vanadate The drops were equilibrated against 1mL of reservoir solution in a Linbro plate (Hampton Research) at room temperature. Ribozyme crystals started to appear in ~4 weeks, and grew to the size of ~0.2 mm x 0.1 mm x 0.05 mm in the next few weeks. Prior to data collection, crystals were soaked into a solution containing 33% MPD, 50 mM buffer (potassium acetate pH 5.0 or Tris-HCl pH 8.0) 5% isopropanol, 0.4 M potassium chloride, 1 mM spermine and 50 mM $MgCl_2$ for 0.5-1 h. Crystals were then flash-frozen and stored in liquid nitrogen.

Diffraction data were collected at LS-CAT beamline 21-ID-F and 21-ID-D at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL). Data were processed by HKL2000 (*258*) in space group of $C222_1$.

3.3.3 Structure determination and refinement

The search model used for molecular replacement was derived from the crystal structure of *S. mansoni* hammerhead ribozyme (PDB code: 3ZD5), from which the entire stem I, the stem II loop, and active site residues including C3, G8, A9, G12, C17 and C1.1 were removed. The phase was determined using Phaser in Phenix (*259*). The asymmetric unit contains a single ribozyme and the best solution had a TFZ score of 8.8 a LLG of 222. The initial model was improved by the Autobuild routine implemented in Phenix. The initial model was improved by several rounds of manual building and refinement using Coot (*260*) and Phenix. Waters and Mg^{2+} ions were manually modeled using Coot. Variants of the initial structure solution were refined using a constant set of test set of reflections used to calculate R_{free} in the initial structure.

3.3.4 Kinetic Assays

Single turnover reactions were performed under various conditions to obtain observed rate constants (k_{obs}). Ribozyme (2 µM) was mixed with 50 nM substrate in 50 mM buffer adjusted to a desired pH. Buffers used were Tris-HCl (pH 7.5, 8 and 8.5), MOPS (pH 7.0), MES (pH 5.5, 6.0 and 6.5) and Potassium Acetate (pH 4.5 and 5.0). For reactions in 1 M NaCl, 100 mM EDTA was also added to the reaction mixture. The mixture was then heated at 90 ^{o}C for 2 min, cooled to room temperature for 10 min and equilibrated at 37 ^{o}C for 2 min. Reactions were initiated using 5 µL of initiation buffers

which contained the appropriate divalent metal ion ($MgCl_2$, $MnCl_2$ or $Zn(O_2CCH_3)_2$) at a desired concentration and 50 mM buffer. At various time intervals, 5 µL aliquots of reaction were thoroughly mixed, on ice, with 5-20 µL of the quenching buffer (25 mM EDTA, 0.5x TBE and 90% Formamide) to ensure efficient quenching. The progress of the reaction was analyzed by separating the cleaved and uncleaved substrate on denaturing (7 M urea) 10% acrylamide gels, detecting the emission of DY-547 on Typhoon 8600 variable mode imager (Molecular Dynamics) and quantifying bands using ImageQuant 5.1 (Molecular Dynamics). Fraction cleaved (f) values were plotted against time (t) on KaleidaGraph (Synergy Software) and data were well fit to a single exponential equation:

$$f = A + B[e^{(-k_{obs}t)}] \qquad (1)$$

where A is the fraction cleaved at completion, A+B is the burst fraction and k_{obs} is the first order rate constant. For very slow reactions, A was artificially set at 0.9.

To find apparent pK_a values, k_{obs} obtained from equation 1 were plotted as a function of pH and the data were fit to equation 2 or 3 as appropriate:

$$k_{obs} = k_{max}/[1 + 10^{(pKa-pH)}] \qquad (2)$$

$$k_{obs} = k_{max}/[1 + 10^{(pKa1-pH)} + 10^{(pH-pKa2)} + 10^{(pKa1-pKa2)}]$$

Finally, plotted data from metal titration experiments were fit to equation 3:

$$k_{obs} = k_{max}/[1 + (K_{D,M^{2+}}/[M^{2+}])^{(n_{Hill})}] \quad (3)$$

where k_{max} is the maximal rate constant, $K_{D,M^{2+}}$ is the apparent dissociation constant of

divalent metal, and n_{Hill} is the apparent Hill coefficient for divalent metal binding to the ribozyme.

3.4 Results

3.4.1 Crystal structure of RzB hammerhead ribozyme

To better explore the structural biology of the hammerhead ribozyme, we sought to identify and characterize a crystal form that differed significantly from the previously described crystals of the ribozymes from *S. mansoni* (*238, 240*) and satellite tobacco ringspot virus (*256*). We selected RzB, an artificially evolved variant of the hammerhead ribozyme with an alternate tertiary contact between stems I and II (*255*). Several variants were screened to identify RNAs that would readily crystallize, and the sequence used in this study is shown in Figure 3.1. To prevent cleavage of the substrate RNA strand during crystallization and data collection, C17 was replaced with a deoxynucleotide, a substitution that deletes the nucleophile. Crystallization screens were performed under relatively low monovalent salt conditions that would allow Mg^{2+} ions and other divalent metals to interact with the ribozyme. The structures described here have been determined at a resolution of 2.7−3.1 Å. Data collection and refinement statistics can be found in Table 3.1, and composite omit maps are shown in Figure 3.3.

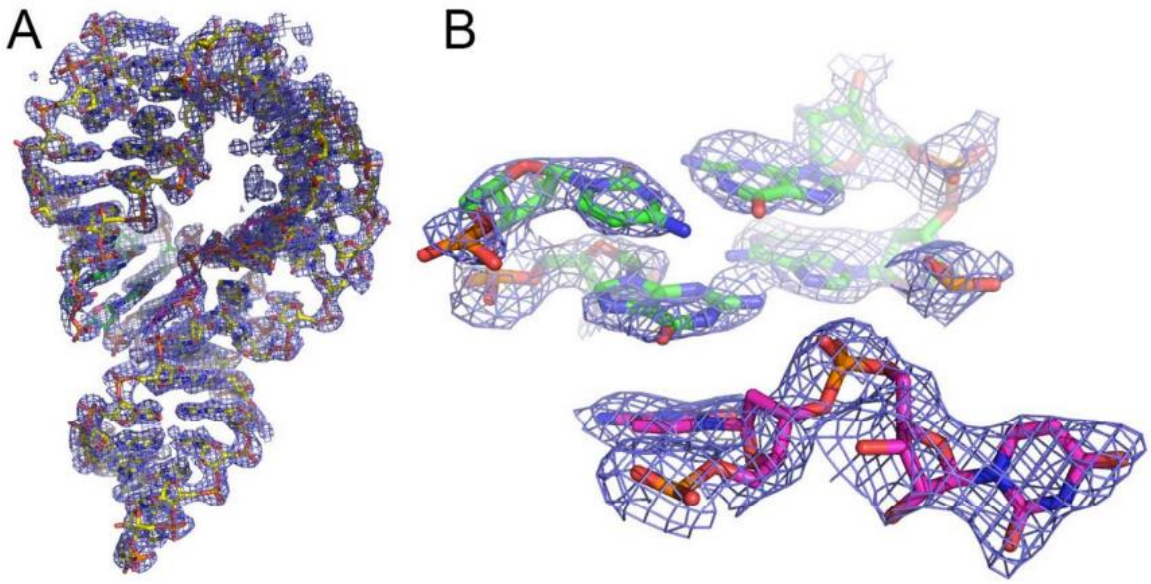

Figure 3.3. Composite 2Fo-Fc omit maps contoured at 1.2 σ.
A. Stems Ib, II and III of the ribozyme. Stem Ia is out of the plane in this image and is therefore not visible. **B.** The active site of the ribozyme, including A9, G10.1, C11.1, G12 (green) and the cleavage site dinucleotide (pink)

Table 3.1. Crystallographic data collection and refinement statistics.

	WT RzB			**G12A**		
data collection	**pH 5.0 Mg^{2+}**	**pH 5.0 Mn^{2+}**	**pH 8.0 Mn^{2+}**	**pH 5.0 Zn^{2+}**	**pH 4.0 Mn^{2+}**	**pH 5.0 Mg^{2+}**
PDB id	5DQK	5DI4	5DI2	5DH8	5DH7	5DH6
cell dimensions *a*, *b*, *c* (Å)	79.76, 86.07, 103.3	80.40, 86.05, 104.21	79.52, 85.88, 103.72	80.94, 86.02, 103.37	81.3, 85.8, 103.5	81.41, 85.96, 102.97
wavelength (Å)	0.97872	1.5011	1.5011	1.181	1.8923	0.976
resolution (Å)	42.0–2.70 (2.75–	50.0–2.94 (2.99–	50.0–3.00 (3.05–	*50.0–3.10 (3.15–*	50.0–3.08 (3.13–	50.0–2.80 (2.85–
R_{merge} (%)	5.9 (78.5)	7.6 (37.9)	8.9 (33.7)	11.9 (28.6)	13.5 (25.2)	7.8 (65.5)
$I/\sigma(I)$	28 (2.5)	36.9 (7.7)	34.4 (14.1)	49.4 (22.4)	26.5 (6.0)	61.7 (6.3)
no. of reflections	9951	14606	13895	12541	12239	9318
redundancy	8.1 (8.3)	14.2 (14.5)	13.6 (14.6)	13.3 (14.6)	4.4 (3.0)	14.3 (15.0)
completeness (%)	99.7 (100)	99.8 (100.0)	99.8 (100.0)	99.5(95.6)	92.9 (70.2)	99.5 (96.7)
Refinement						
resolution (Å) (last shell)	26.1–2.71	28.2–2.95	39.8–2.99	28.3–3.30	40.0–3.06	28.4–2.78
	(2.85–2.71)	(3.05–2.95)	(3.10–2.99)	(3.47–3.29)	(3.19–3.06)	(2.93–2.78)
no. of reflections	9627	14432	13759	10413	12124	9231
no. of reflections in the test	968	1446	1385	1048	1206	920
R_{work}/R_{free}(%) (last shell)	22.42/27.78	20.94/23.30	22.10/25.99	21.68/25.18	20.20/25.37	21.08/24.16
	(36.14/42.27)	(30.54/29.65)	(25.85/32.97)	(21.97/30.91)	(26.49/33.52)	(35.27/36.02)
no. of atoms	1479	1466	1462	1467	1465	1467
RNA	1455	1455	1455	1454	1454	1454
ions	5	11	7	13	11	5
water	19	0	0	0	0	8
average *B*-factor	27.1	76.30	73.37	74.40	86.62	56.1
root-mean-square deviation						
bond length (Å)	0.008	0.001	0.001	0.001	0.001	0.002
bond angles (deg)	1.378	0.388	0.373	0.443	0.376	0.470
coordinate error (Å)	0.45	0.32	0.38	0.33	0.34	0.37

The RNA crystallizes as a domain-swapped dimer in which one strand from stem Ia forms intermolecular base pairs with a second molecule within the unit cell (Figure 3.4). The two monomers are related by crystallographic symmetry. Although the RNA strands are swapped to form the dimer, a monomer model can be generated by switching the crossed over strands (Figure 3.3). In spite of the intermolecular crystal contact, the tertiary contact between stem I and stem II appears to be intact (Figure 3.5). Loop II forms a pocket in which an adenosine from stem I can dock. The U at position L2.1 and A at position L2.5 form a Hoogsteen base pair. The A at position L1.3 in stem I is extruded and stacks upon the Hoogsteen base pair from loop II and the Watson−Crick face of L1.3 hydrogen bonds to the minor groove face of the first G in loop II. Overall, the topology of the three guanosines in loop II resembles that of a GNRA tetraloop with a long-range tertiary interaction taking the place of the adenosine normally found at the last position of the tetraloop (Figure 3.5).

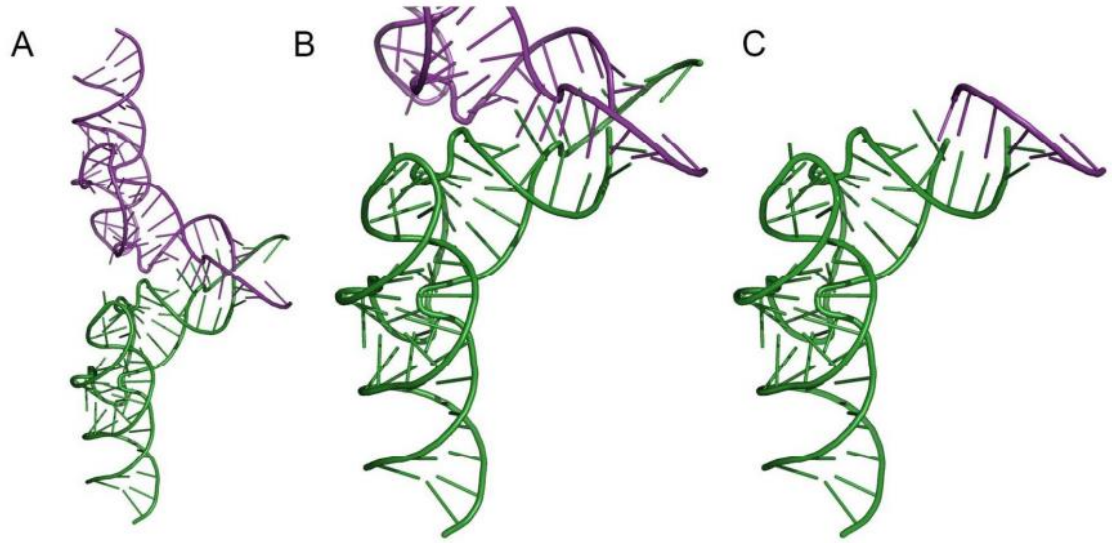

Figure 3.4. Crystal contacts involve intermolecular base pairs.
A. In crystals of the RzB ribozyme, an intermolecular dimer is formed. Stem Ia is disrupted, allowing base pairing between two adjacent molecules. The green molecule and purple molecule are related by crystallographic symmetry. **B.** Close up view of the packing interaction. **C.** A structure for the 'monomeric' ribozyme can be generated using appropriate nucleotides from green and purple molecules to generate an intact stem Ia. In the resulting model for the monomer, the termini of the green and purple fragments are close enough to be linked.

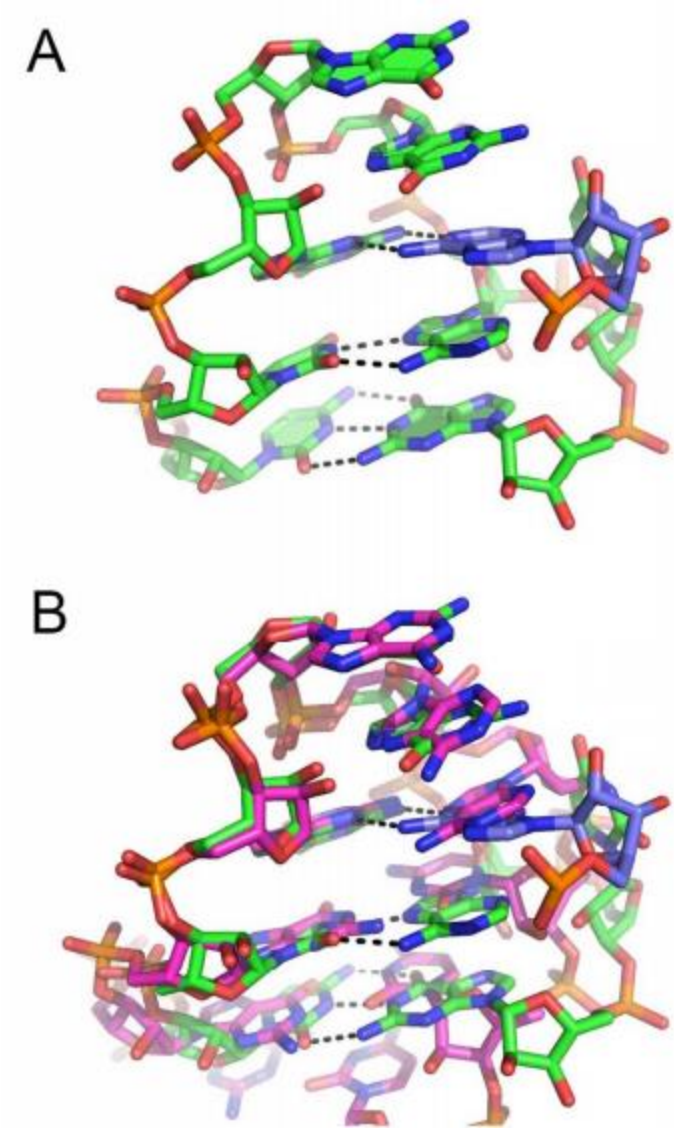

Figure 3.5. Tertiary contact between stem I and loop 2.
A. A single adenosine at position L1.3 (blue) is extruded from stem I and docks within loop II (green). Recognition is mediated through hydrogen bonding and stabilized by stacking interaction. **B.** Loop II resembles a GNRA tetraloop. The structure of a GNRA tetraloop (pink) was superposed on loop II revealing a similarity in the overall architecture of the loops. The adenosine from stem I takes the place of the 'A' in the GNRA loop. The orientation of the adenosine is reversed, however.

The conformation of the active site of hammerhead ribozyme depends critically on the tertiary interaction between stem I and stem II. The active site of the RzB

hammerhead ribozyme is strikingly similar to that observed in the *S. mansoni* hammerhead ribozyme and differs significantly from the structure of the minimal hammerhead ribozyme. This suggests that the active site structure observed in crystals of the RzB hammerhead ribozyme is native in spite of the base pairing formed across the molecules in the crystals (Figure 3.6). The structural similarities between the *S. mansoni* and RzB ribozymes remain in spite of the differences in sequence, in crystal contacts, and in crystallization conditions. We therefore used these crystals and this RNA sequence to explore the roles of divalent metal ions and individual nucleotides to hammerhead ribozyme catalysis.

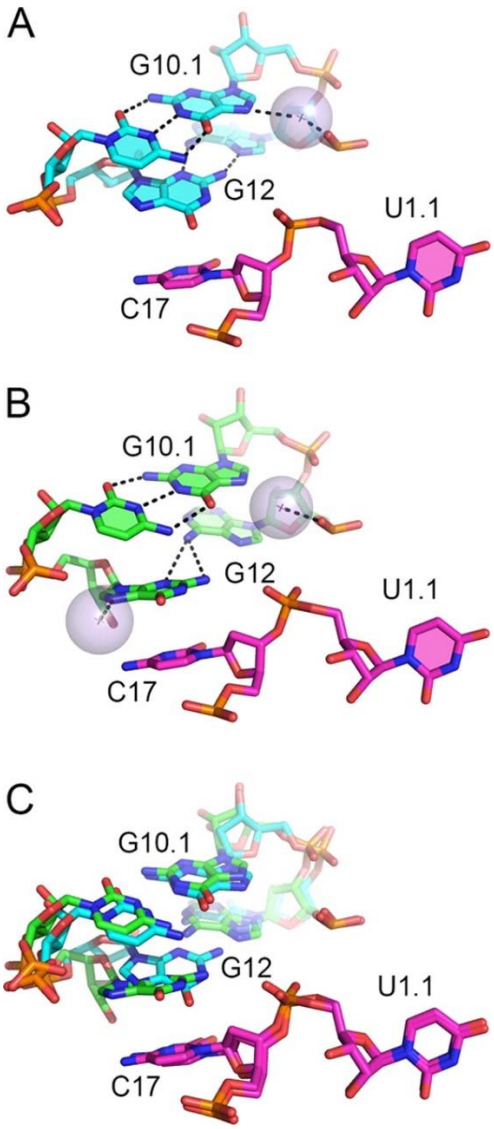

Figure 3.6. A pH-dependent conformational rearrangement of the active site. **A.** At pH 5.0, G12 is observed to be involved in a sheared base pair with A9. A single Mn^{2+} ion is observed in the active site near G10.1. **B.** At pH 8.0, G12 has changed position to stack upon C17. Two long, ~3.5 Å, potential hydrogen bonds link G12 and A9. Two Mn^{2+} ions are observed in the active site at pH 8.0. The first, near G10.1 as seen at pH 5.0. The second is observed at the Hoogsteen face of G12. **C.** The structures of the RzB hammerhead ribozyme at pH 5.0 and 8.0 were superposed based on the position of the cleavage site dinucleotide (pink). Active site nucleotides are shown in cyan (pH 5.0) and green (pH 8.0). The largest difference in the two structures is the position of G12. Movement of G12 away from the scissile phosphate and the recruitment of a metal ion (shown in panel B) to G12 suggest that G12 may be deprotonated and negatively charged at pH 8.0.

3.4.2 The G12A mutation is disruptive to catalysis.

We next investigated the proposed general base, G12. Mutation of the highly conserved G12 nucleotide to an adenosine severely reduced the kobs of the reaction in 10 mM Mg^{2+} at pH 6.5 (Table 3.2). The rate constant for the reaction of the WT ribozyme was measured to be ∼1.4 min−1. In contrast, the kobs measured for G12A under similar conditions was ∼0.00011 min−1, which represents about a 13 000-fold decrease in the reaction rate. Such a low reaction rate is consistent with previous studies showing the importance of G12 for the cleavage reaction (*237, 256, 261*).

Table 3.2. Observed rate constants for WT RzB hammerhead ribozyme and its mutants. Reactions were performed at 37 ºC under single turnover conditions in 5 mM Mg^{2+} and 50 mM MES pH 6.5. k_{rel} represents the ratio of the observed rate constant to that of the WT RzB hammerhead ribozyme.

	k_{obs} (min^{-1})	k_{rel}
WT	1.39 ± 0.04	NA
G12A	0.00011 ± 0.00003	0.00008
G10.1C/C11.1G	0.00132 ± 0.00008	0.001
G10.1A/C11.1U	0.078	0.06

To confirm that the G12A mutation does not cause significant structural changes in the RzB ribozyme, we determined the crystal structure of the G12A mutant ribozyme (Figure 3.7). As previously observed, there are minimal structural differences between the WT ribozyme and G12A mutant (*256, 261*). The ribozyme was crystallized with a deoxynucleotide at C17, and therefore an exact measurement between the N1 of the

purine at position 12 and the 2'-hydroxyl of C17 is impossible. The distance between C2' of C17 and the N1 of A12 in the G12A crystal structures is 4.0 Å, while the equivalent distance is 4.2 Å for the N1 of G12 in the WT ribozyme. The distance of the proposed general acid, the 2'-OH of G8 to the 5'-O leaving group on U1.1, remains unchanged at 3.2 Å in both the G12A mutant and WT ribozymes. Thus, the A12 mutation does not appear to cause significant structural changes in the ground state conformation observed in the crystal structure.

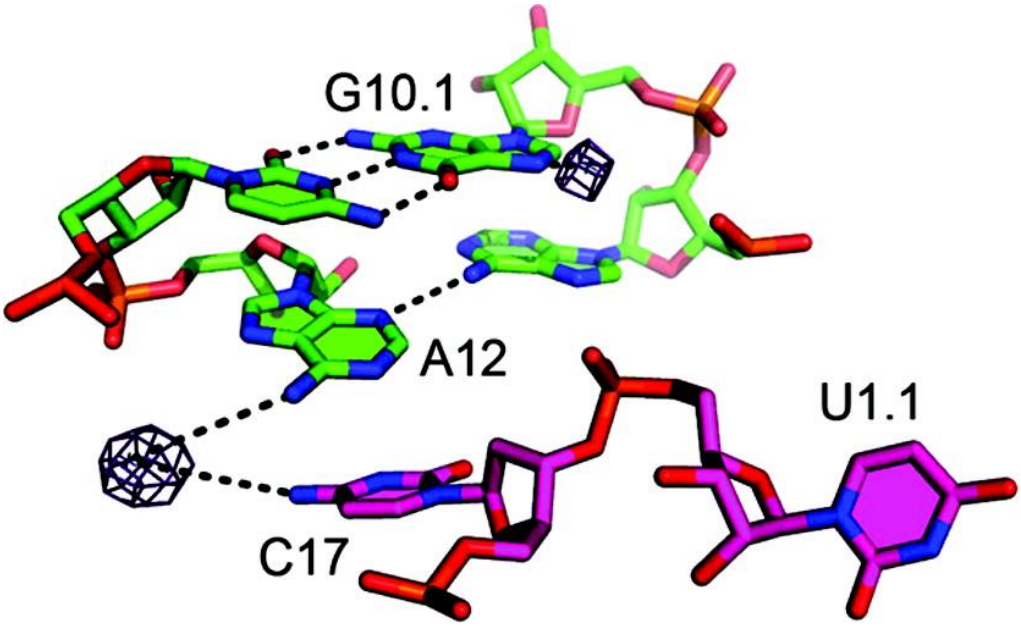

Figure 3.7. Active site of the G12A mutant hammerhead ribozyme.
The structure of the G12A mutant ribozyme is very similar to that of the WT ribozyme. Due to the loss of the N2 exocyclic amine at position A12 is displaced slightly from the position occupied by G12 in the WT ribozyme. There is a 3.1 Å hydrogen bond between the N3 of A12 and the N6 of A9. Crystals of the G12A mutant ribozyme were soaked in 10 mM Zn(O2CCH3)2 prior to freezing, and anomalous diffraction data were collected using X-rays at a wavelength of 1.181 Å. An anomalous difference density map, shown here contoured at 4σ, was used to identify Zn^{2+} binding sites within the ribozyme. A clear peak for Zn^{2+} is observed near A12. The distance between the N6 of A12 and the Zn^{2+} ion is 4.5 Å, and the distance between the N4 of C17 and the Zn^{2+} ion is 5.1 Å. In addition, a Zn^{2+} ion occupies the divalent ion binding pocket near G10.1.

It has been reported previously that the observed rate at which WT ribozyme cleaves increases log−linearly with respect to pH (*237, 247*). A similar pH−rate profile was obtained for the WT ribozyme in 5 mM Mg^{2+} (Figure 3.8). The observed rate constant is observed to be log−linear from pH 4.5 through ∼ pH 8.0. Such a profile is consistent with a theoretical pH−rate profile that would arise from a general base having a pK_a of ∼8 (N1 of G12), and thus is largely consistent with the mechanism proposed by Scott (*238*). The change in slope at pH 8.0 suggests that the N1 of G12 may have a pK_a that is lowered toward neutrality in the context of the folded hammerhead ribozyme as suggested previously (*262*). Above pH 8.5, the rate constant decreases, and it is not clear if this is due to deprotonation of a general acid, such as the 2'-OH of G8 or a water molecule, being deprotonated or if this is due to denaturation of the RNA.

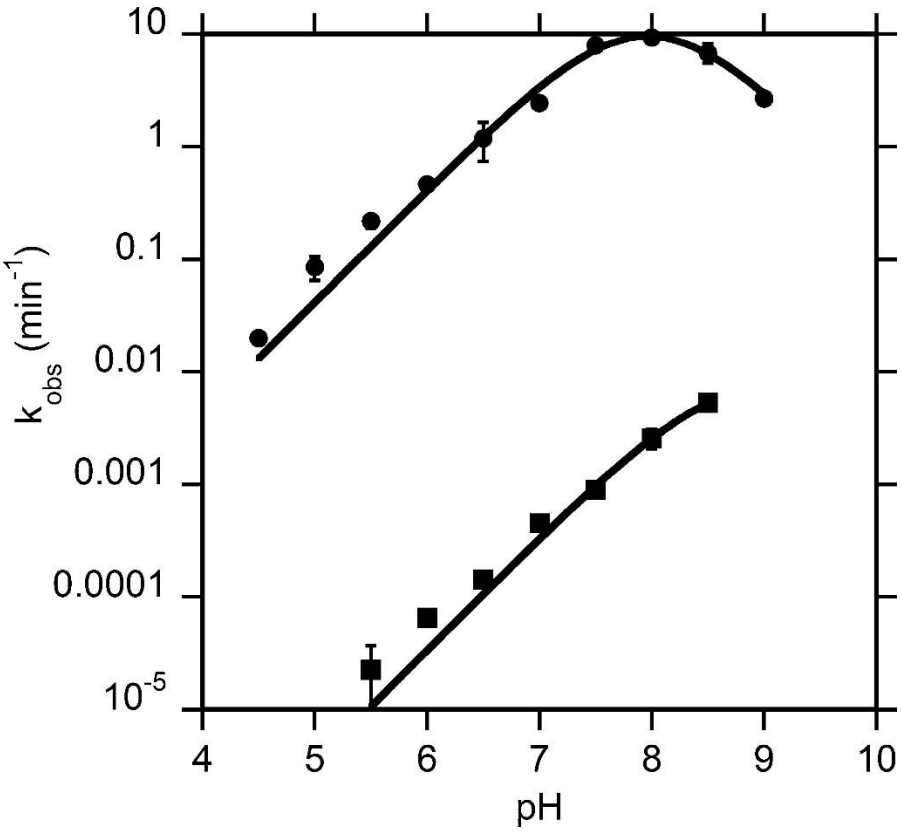

Figure 3.8. pH–rate profile of WT hammerhead ribozyme (circles) is compared to pH–rate profile of G12A (squares).
Single turnover reactions were performed in 50 mM appropriate buffer. WT ribozyme reactions were conducted using a 5 mM Mg^{2+} concentration, and G12A mutant ribozyme reactions were performed at 50 mM Mg^{2+}. Data were fit to eq 3 or to a line as appropriate. Error bars represent standard deviations from at least three replicates.

In the G12A mutant, however, the pH−rate profile would be expected to be significantly different than that observed for the wild type ribozyme. Adenosine in solution has a pK_a of ∼3.5. A general base with that pK_a and a general acid with a pK_a ≈ 13 would be expected to give rise to a pH−rate profile that is flat in the measurable pH range (Figure 3.9). We characterized the pH−rate profile of the G12A mutant using 50 mM Mg^{2+} where the reaction is faster, allowing more accurate measurements to be

made (Figure 3.8). As with the WT ribozyme, the pH−rate profile of the G12A mutant is log−linear; however, a change in the slope at high pH is not observed with this mutant (Figure 3.8). As the G12A mutant replaces a potential general base with a pK_a of ~9 with a potential general base with a pK_a of around 4, these data suggest either that the pH−rate profile of the hammerhead ribozyme does not reflect the protonation state of G12 or that the mechanism of the G12A mutant must be revisited.

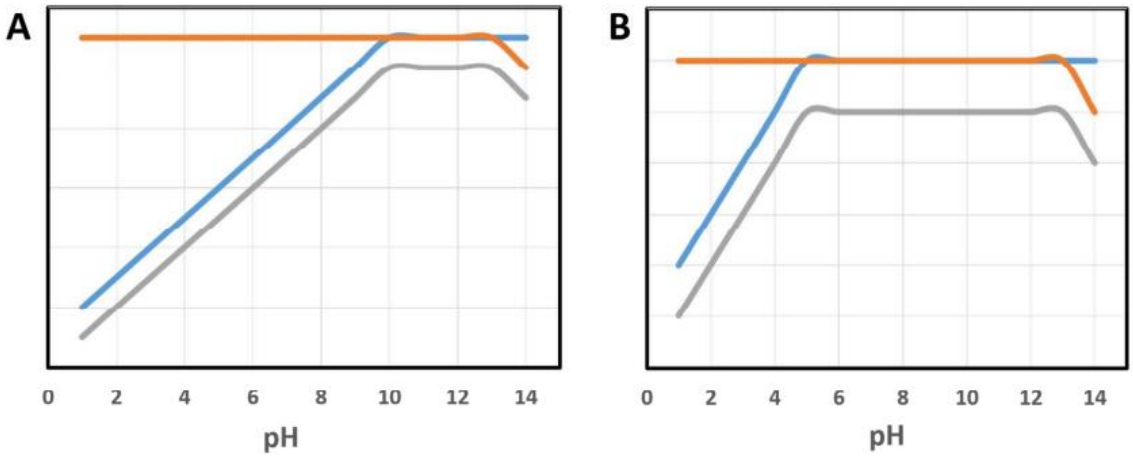

Figure 3.9. Simulated pH-rate profiles of WT RzB and G12A RzB.
The individual fractions of functional general acid, fA (orange) and functional general base, fB (blue) give rise to the pHrate profile which is shown in grey. For wild-type hammerhead ribozyme **(A)**, the pK_a s of general acid and general base were assumed to be ~13 and ~9 respectively, whereas for G12A mutant (**B**), the pK_a of general base was assumed to be ~4.

3.4.3 A pH-dependent conformational switch in the hammerhead ribozyme active site

The hammerhead ribozyme exhibits maximal activity at high pH. We therefore compared crystal structures of the hammerhead ribozyme at pH 5.0 and pH 8.0 (Figure

3.6). The WT ribozyme exhibits maximal activity at pH 8.0 with the observed rate constant about 3 orders of magnitude higher in pH 8.0 than in pH 5.0. In both structures, $MnCl_2$ was included in the soaking buffer to allow unambiguous identification of divalent metal binding sites using anomalous diffraction.

A significant rearrangement of the active site of the hammerhead ribozyme is observed at pH 8.0. The observed changes in conformation are centered around G12 (Figure 3.6C). At pH 5.0, G12 makes a sheared base pair with A9 with two hydrogen bonds formed between the two bases. When the pH is shifted to pH 8.0, this base pair is disrupted and G12 shifts 2−3 Å away from A9 to stack upon C17. The only other significant shift is at nucleotide A9, which shifts in the same direction but only ∼1−1.5 Å. In these crystal structures, we are missing the 2'-OH nucleophile, and therefore it is difficult to determine how the distance from G12 to the nucleophile changes. However, the distance from the N1 of G12 to the phosphorus atom of the scissile phosphate can be readily measured. At pH 5.0, this distance is 5.4 Å, and at pH 8.0, this distance is 5.9 Å.

The locations of Mn^{2+} ions can be unambiguously identified in these structures through anomalous scattering experiments. At both pH 5.0 and pH 8.0, an Mn^{2+} ion is observed near G10.1 as previously described (*251*). An additional Mn^{2+} ion is observed to bind near the Hoogsteen face of G12, but only at pH 8.0 (Figure 3.6 and Figure 3.10). The position of this Mn^{2+} ion suggests that the coordination to N7 of G12 is direct, while coordinate of O6 of the metal ion of G12 is outer sphere and mediated by a metal-bound water molecule. If a Mg^{2+} ion were to bind at this site, the ligands would be

expected to be slightly different as Mg^{2+} does not usually coordinate directly to the N7 nitrogen.

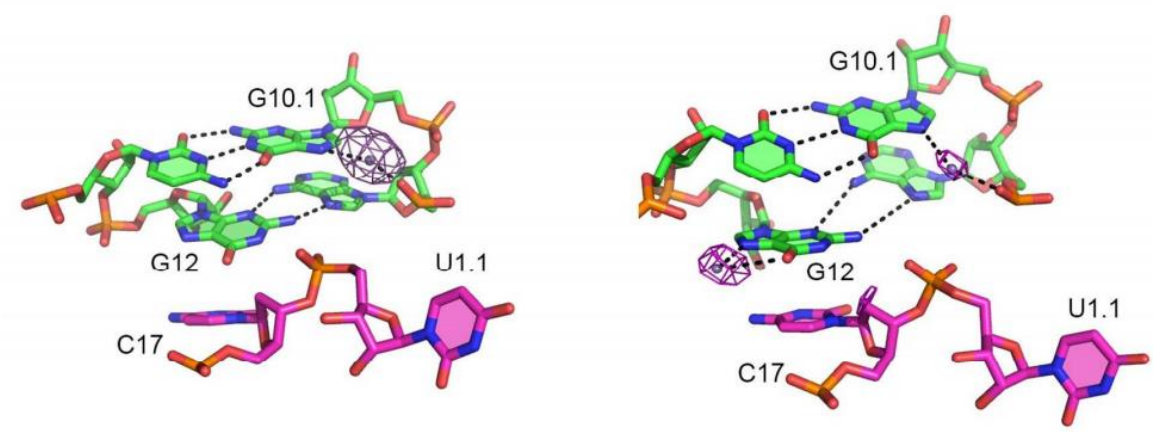

Figure 3.10. Anomalous diffraction maps of Mn^{2+} ions.
Mn^{2+} anomalous diffraction maps at pH 5.0 (left) and pH 8.0 (left). Maps are contoured at 4σ.

It is impossible to tell from a static crystal structure if the observed change in conformation is due to a change in the ionization state of a functional group on the ribozyme or which functional group is ionizing. However, there is striking change in conformation of G12 and a small, but significant movement away from the negatively charged scissile phosphate. In addition, a positively charged metal ion is only observed to associate with G12 at pH 8.0. Both of these observations suggest that G12 may be deprotonated and negatively charged at pH 8.0, consistent with the inflection point observed in the pH−rate profile. Curiously, at high pH, G12 is in a position similar to that of A12 in the G12A mutant. If G12 is, in fact, deprotonated at pH 8.0, this similarity may

reflect the conversion of the hydrogen bond donor at N1 of G12 into a hydrogen bond acceptor.

3.4.4 The G12A mutation exhibits a metal ion specificity switch.

Several studies have shown that the divalent metals influence the observed rate constants in tertiary-stabilized hammerhead ribozyme (*247, 248*). In an attempt to understand how divalent metals impact the reaction, metal ion titration studies were performed using both WT and G12A mutant ribozymes. For this study, we focused on the response of the hammerhead ribozyme to Mg^{2+}, Mn^{2+}, and Zn^{2+} ions. The pH was kept low (6.5) to avoid complications arising from lower solubility of divalent metals at higher pH conditions. For the wild-type RzB ribozyme, the kobs values only show minor differences between Mn^{2+}, Mg^{2+}, or Zn^{2+} (Figure 3.11). Reactions in Zn^{2+} and Mn^{2+} appear slightly faster when compared to Mg^{2+} as previously observed; (*247*) however the observed kmax values in the different metal ions are within 2-fold of each other (Table 3.3). Likewise, the reactions in the various metal ions exhibit only small differences between the apparent K_d's and Hill coefficients.

In contrast, the G12A mutant of RzB ribozyme shows a very different preference for divalent metals than the wild-type ribozyme. The observed rate constant was significantly higher in Zn^{2+} as compared to Mn^{2+} or Mg^{2+} (Figure 3.11B). kmax of the RzB G12A mutant in the presence of Zn^{2+} was ∼0.10 min−1, which is 10-fold higher than that observed in Mn^{2+} and about 50-fold higher than that observed in Mg^{2+}. The trends in

apparent dissociation constants and Hill coefficients remain similar to those observed in the wild-type RzB ribozyme. In contrast to the 13 000-fold difference in reactivity between the WT and G12A mutant ribozymes in Mg^{2+} ion, there is only a 170-fold difference in kmax between the WT and G12A mutant ribozymes in Zn^{2+} ion (Table 3.3).

Metal specificity switches can result from binding of metal ions at locations distant from the active site. We therefore repeated these experiments using the *S. mansoni* hammerhead ribozyme. This ribozyme has different sequences in the peripheral helices, but the core nucleotides in the active site are invariant. It was not possible to saturate the reaction of the WT *S. mansoni* hammerhead ribozyme with Mg^{2+} precluding an accurate description of the kinetic parameters for the reaction in this ion. The observed rate constants in all three metal ions, however, are similar (Figure 3.11, Table 3.3). We then examined the metal ion dependence of the *S. mansoni* G12A mutant. This ribozyme displays the same preference for the divalent metals as the RzB G12A mutant ribozyme (Figure 3.11, Table 3.3). Zn^{2+} significantly enhances the reactivity over that observed in Mn^{2+} or Mg^{2+}. As with the WT *S. mansoni* ribozyme, saturation was not observed in the Mg^{2+}-containing reactions. These experiments demonstrate that replacing G12 with an adenosine switches the divalent metal binding preference in at least two hammerhead ribozyme sequences. As the metal ion specificity switch is observed in different ribozymes with different peripheral elements, the switch in divalent metal specificity is likely a conserved property of the catalytic core nucleotides in the active site of the G12A mutant.

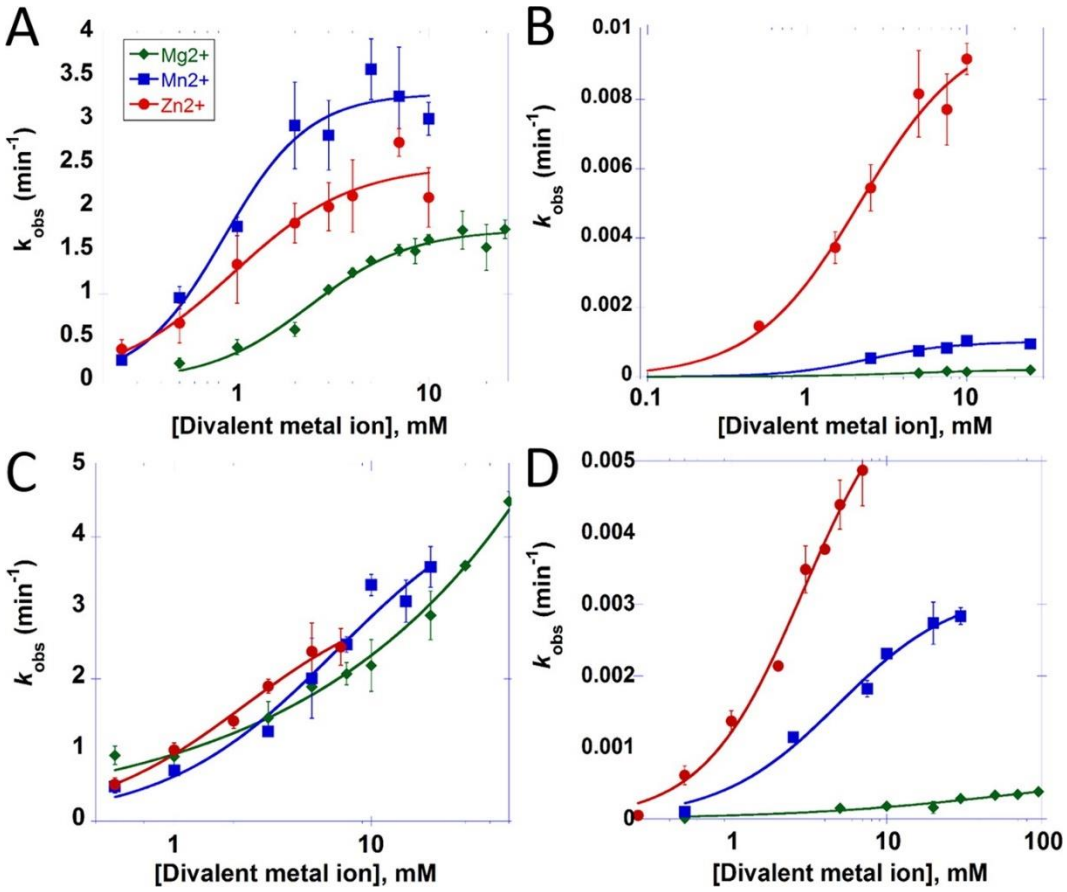

Figure 3.11. G12A mutant shows a metal ion specificity switch.
Single turnover reactions were performed in Mg^{2+} (green ◊), Mn^{2+} (blue □), or Zn^{2+} (red ○). Reactions contained 2 µM ribozyme, 50 nM fluorescently labeled substrate, and 50 mM MES pH 6.5. The reactions were performed using RzB hammerhead ribozyme **(A)** or the G12A mutant **(B)**. Data from these graphs were fit to eq 4, and the resulting parameters are given in Table 3. Measurements at higher Zn^{2+} concentrations were not possible due to the loss of Zn^{2+} solubility. Error bars represent standard deviation from three replicates. The *S. mansoni* hammerhead ribozyme was characterized under the same conditions in panels **(C)** (WT) and **(D)** (G12A).

Table 3.3. Kinetic parameters for the hammerhead ribozymes cleavage reactions in various divalent metal ions.
Reactions were performed in 50 mM MES pH 6.5 and various concentrations of the indicated divalent metal ions.

	WT			G12A		
	Zn^{2+}	Mn^{2+}	Mg^{2+}	Zn^{2+}	Mn^{2+}	Mg^{2+}
RzB						
k_{max} (min^{-1})	1.7 ± 0.1	3.3 ± 0.2	2.5 ± 0.3	0.010 ± 0.001	0.0010 ± 0.0001	0.0002 ± 0.0001
K_D (mM)	2.4 ± 0.2	0.9 ± 0.1	1.0 ± 0.2	2.1 ± 0.6	2.4 ± 0.5	4.0 ± 3
n_{Hill}	1.7 ± 0.3	2.0 ± 0.5	1.4 ± 0.4	1.3 ± 0.3	1.7 ± 0.8	1.3 ± 3
S. mansoni						
k_{max} (min^{-1})	3.2 ± 0.5	5 ± 1		0.007 ± 0.001	0.0023 ± 0.0003	0.0005 ± 0.0001
K_D (mM)	2.3 ± 0.8	7 ± 4		2.9 ± 0.9	2.5 ± 1.0	19 ± 9
n_{Hill}	1.1 ± 0.2	1.0 ± 0.3		1.4 ± 0.3	2.2 ± 2	0.7 ± 0.2

3.4.5 A Zn^{2+} ion binding pocket exists near A12

To characterize the metal ion binding properties of the G12A mutant ribozyme, we soaked crystals of this ribozyme in stabilization buffer containing Zn^{2+} and collected anomalous diffraction data from the Zn^{2+} soaked crystals. A >5σ peak is observed near A12 in the Zn^{2+} anomalous density map (Figure 3.7). Potential ligands to this Zn^{2+} ion are the exocyclic amine of A12 (4.5 Å distant), the pro-S_P oxygen of U16.1 (4.1 Å distant), and the exocyclic amino of C17 (5.1 Å distant). These distances are consistent with outer sphere interactions between the ligands from the RNA and the water molecules bound to the hydrated Zn^{2+} ion. Mn^{2+} is observed at the same site in an anomalous difference map obtained using an Mn^{2+} soaked crystal. However, the peak from the Mn^{2+} ion is

only ~3.5σ in height, consistent with lower occupancy and weaker rescue of the G12A mutant ribozyme with Mn^{2+}.

3.4.6 There is a divalent metal ion binding pocket at G10.1

A divalent metal ion has been previously observed at the G10.1−C11 base pair in the crystal structure of the *S. mansoni* hammerhead ribozyme. There is significant evidence to suggest that this metal ion participates in catalysis and interacts with the scissile phosphate during the cleavage reaction (*242, 243*). In the RzB ribozyme, we are also able to observe electron density at the Hoogsteen face of G10.1 that appears to be a Mg^{2+} ion. To confirm metal ion binding at this position, we soaked crystals in Mn^{2+} and calculated anomalous difference electron density maps. At pH 5.0 Mn^{2+} ion at G10.1 gives rise to a > 5σ peak (Figure 3.6). As has been observed previously with the *S. mansoni* hammerhead ribozyme, (*251*) this ion is 4−5 Å distant from the scissile phosphate in the ground state crystal structure.

To characterize the role of the metal ion bound near G10.1, double mutants were generated that preserved the base-pairing potential of nucleotides 11.1 and 10. The G10.1C−C11G double mutant swaps the positions of the purine and the pyrimidine in the base pair. Thus, the N7 on guanosine is moved to the opposing strand of RNA, and the exocyclic amine of the cytosine occupies the major groove near the metal binding site. Secondary structure predictions suggest that this double mutant retains the ability to fold into the wild-type secondary structure. The activity of this mutant is 1050-fold

lower relative to that of the wild-type ribozyme (Table 3.2). The G10.1A−C11U double mutant retains the purine and the N7 at the 10.1 position but replaces the keto group of the guanosine with the exocyclic amine of an adenosine. This double mutant is also predicted to retain the secondary structure of the wildtype ribozyme. The A10.1−U11 double mutant is only 18-fold less active than the WT ribozyme. These results are consistent with previous studies of the minimal hammerhead ribozyme that probed base pairing between G10.1 and C11 and the role of the N7 of G10.1 (*263*).

The mutagenesis experiments suggest that the metal ion binding site at the G10.1−C11 base pair is important for catalysis, in spite of being 4−5 Å distant from the cleavage site dinucleotide. Loss of the N7 on the purine is expected to be highly disruptive to metal ion binding, while replacement of the keto oxygen on the guanosine with the exocyclic amino group of adenosine is expected to be modestly disruptive.

3.5 Discussion

The hammerhead ribozyme has been extensively characterized since it was first discovered in 1987. Early studies focused on a minimal ribozyme design containing only the core nucleotides in the active site and three small stems. This molecule was systematically studied using a variety of single atom substitutions and unnatural nucleotides (*264*). The crystal structure of the minimal hammerhead ribozyme was also the very first ribozyme to be characterized by X-ray crystallography (*265, 266*). Surprisingly at the time, the crystal structure of the minimal hammerhead ribozyme was

not supported by the extensive biochemical data (reviewed in ref (*264*)), suggesting that the structure was not in a catalytically active conformation. Eventually, a tertiary contact between stem I and stem II of the hammerhead ribozyme was discovered to exist in naturally occurring hammerhead ribozymes. When the hammerhead ribozyme is extended to include this tertiary contact, the reactivity of the ribozyme is greatly increased (*255, 267-269*). As predicted by the biochemical data, this tertiary contact is necessary to reorganize the conformation of the active site, and the crystal structure of the tertiary-contact stabilized ribozyme revealed a distinctly different conformation in the active site. The structure of the active site in the extended hammerhead is in agreement with most of the available biochemical data (*270, 271*). Crystal structures, however, only reveal the structure of a trapped ground state, and evidence is accumulating that suggests the active site of the hammerhead must further rearrange prior to catalysis or during the reaction pathway.

3.5.1 The curious case of G12

In all the crystal structures of the extended hammerhead, G12 is pointed toward the position that would be occupied by 2'-hydroxyl of U16.1, the nucleophile in the cleavage reaction. G12 has been proposed to serve as the general base, capable of accepting a proton from the nucleophile when G12 is in its deprotonated, negatively charged, state. Guanosine at position 12 is completely conserved in all known hammerhead ribozymes, and substitution with adenosine or unnatural purine

analogues, including 2,6-diaminopurine, inosine, and 2-aminopurine, is more detrimental to catalysis than would be predicted based on the position of G12 within the active site and its function as a general base (*237, 261*).

G12 and A9 form a sheared base pair, and there is a hydrogen bond between the exocyclic amine of G12 and the N7 of A9. Surprisingly, substitution of G12 with an inosine, which lacks only the exocyclic amine, is severely disruptive to catalysis (*261, 272*). Inosine has a pK_a similar to that of guanosine (9.6 for guanosine vs 8.7 for inosine), and an inosine at position 12 would be expected to largely retain function. The nucleobases 2-aminopurine or 2,6-diaminopurine retain the exocyclic amine at position 2, but either lacks a functional group at position 6 or substitutes an exocyclic amine for the keto oxygen of guanine. Like adenosine, 2-aminopurine and 2,6-diaminopurine nucleobases severely reduce cleavage activity when introduced at position 12. Comparing the activity for all of these nucleobase substitutions suggests that maximal activity requires both an exocyclic amine at position 2 and either a keto oxygen or an exocyclic amine at position 6. This requirement is not supported by the crystal structures as the keto oxygen of G12 points toward solvent and is not engaged in any hydrogen bonds or direct metal ion interactions (Figure 3.6).

More curiously, all of these purine analogues retain the ability to perform proton transfer reactions and would be expected to support function in the hammerhead ribozyme cleavage reaction. The pK_a's of the purines used in these studies ranges from ~4 to ~9. Yet, the pH–rate profile for ribozymes with purine substitutions at position 12 are all very similar to a log linear profile from ~pH 4 to ~pH 8.5. If G12 were involved in

a proton transfer reaction, the activity of the ribozyme would be expected to depend on the protonation state of G12 or the protonation state of the purine introduced at position 12. This is clearly not evident in the pH–rate profile (*237*). This suggests either that the pH–rate profile does not reflect the protonation state of G12 or that the path to the cleavage reaction is more complex than expected based on the ground state crystal structure. Thus, G12 likely plays key roles in organization of the active site that have not yet been fully elucidated by the structures observed in the available crystals.

3.5.2 G12 positions an active site metal ion

Here, we present data that suggest that the purine at position 12 plays critical roles in positioning an active site metal ion. As observed in previous studies, the activity of the G12A mutant ribozyme is over 4 orders of magnitude slower than the WT ribozyme in the presence of Mg^{2+} (*261*). However, the maximal rate of the G12A ribozyme is only 2 orders of magnitude slower in the presence of the Zn^{2+} ion (Figure 3.7, Table 3.3). The interaction between A12 in the mutant ribozyme and the rescuing Zn^{2+} ion could be direct, or it could result from a long-range effect from Zn^{2+} binding elsewhere in the ribozyme. The Zn^{2+} rescue of the G12A mutant is not dependent on the identity of the hammerhead ribozyme as it is observed both in the RzB ribozyme and in the *S. mansoni* ribozyme. Thus, the rescuing Zn^{2+} likely binds to an active site nucleotide present in the catalytic core common to both ribozymes. In addition, we are able to observe a Zn^{2+} ion interacting with the Hoogsteen face of A12 in the mutant ribozyme

(Figure 3.8). Both monovalent and divalent ions have been observed to interact with the Hoogsteen face of G12 in crystal structures described here and elsewhere (*240, 251*).

What role does the metal ion at the Hoogsteen face of position 12 play? One possibility is that this metal ion serves solely to lower the pK_a of G12 in the WT ribozyme through electrostatic effects. However, the Zn^{2+} rescue of the G12A mutant argues against this. Just as metal ion binding at the Hoogsteen face of G12 could shift the pK_a of G12 down toward neutrality, metal ion binding at A12 in the mutant ribozyme would be expected to shift the pK_a of A12 down. But this shift would push the pK_a of A12 below ~4 and away from neutrality. While this positively charged metal ion could stabilize the negative charge at the N1 in the deprotonated state of G12, it should destabilize protonation of A12 due to juxtaposition of multiple positive charges (Figure 3.12). Thus, binding of metal ion at the Hoogsteen face would be expected to disfavor proton transfer if A12 were serving as the general base in this mutant, and the only role for Zn^{2+} were to lower the pK_a of the purine nucleobase (Figure 3.12). In contrast to the inhibition expected in this scenario, however, we observe that Zn^{2+} rescues the G12A mutant to within 2 orders of magnitude of the WT ribozyme. This suggests that the rescuing Zn^{2+} enhances the reactivity of the G12A mutant through a different mechanism.

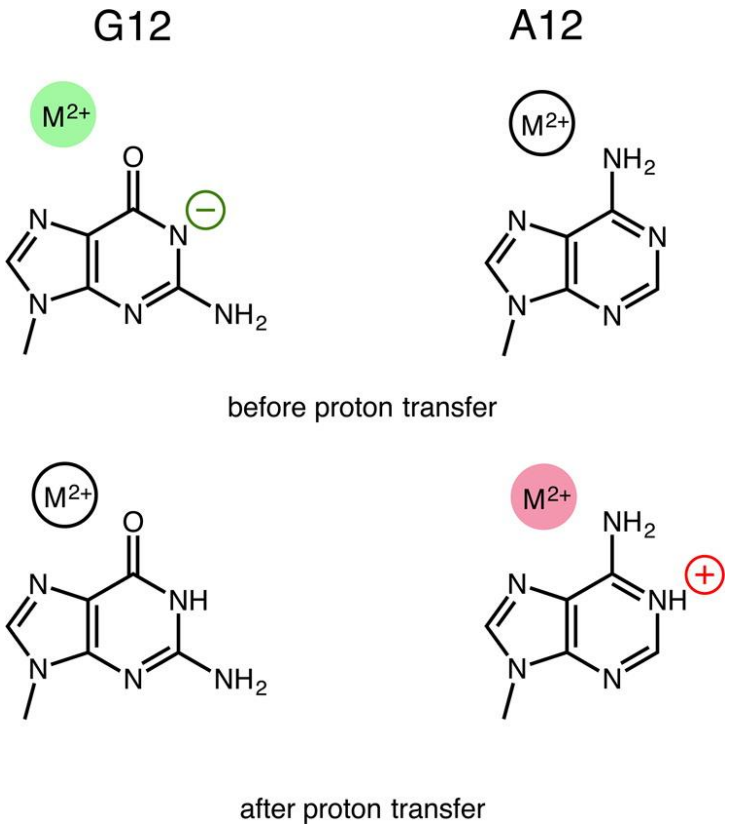

Figure 3.12. Metal ion activation of G12.
Coordination of a divalent metal ion to G12 is predicted to lower the pK_a of the nucleobase from ~9 toward neutrality because the positively charged metal ion can stabilize the negative charge on the guanine base. In contrast, metal ion binding to adenine is predicted to shift the pK_a of the adenine nucleobase away from neutrality as the metal ion would destabilize the positively charged, protonated form of the nucleobase

The location of the rescuing metal ion observed in Zn^{2+} anomalous difference maps of the G12A ribozyme is also problematic. This Zn^{2+} has three potential ligands that are atoms within the covalent structure of the ribozyme. However, all of these atoms are >4.5 Å distant from the metal ion, suggesting the metal ion only interacts with the RNA through its water shell (Figure 3.7). Most metal ion specificity switches

result from metal ions coordinating directly to atoms within the covalent structure of RNA. In addition, the metal ion is at the periphery of the RNA, distant from most of the active site and the cleavage site (Figure 3.7). Combined, the specificity switch and the location of the Zn^{2+} ion near A12 suggest that the hammerhead ribozyme undergoes a conformational change prior to or during the cleavage reaction. This proposed conformational change could serve to bring A12 into direct contact with the Zn^{2+} ion and give rise to the observed metal specificity switch.

3.5.3 Zn^{2+} could induce tautomerization of A12.

Zn^{2+} ions, while commonly found at the Hoogsteen face of guanosine nucleotides, are not observed to interact with adenosine nucleotides. The only difference between the Hoogsteen face of G and the Hoogsteen face of A is that the keto group at the six position of guanine is replaced with the exocyclic amine at the six position of adenine. Metal specificity switches typically result from a direct contact between a metal ion and the ligand being probed. The observed metal specificity switch would therefore suggest that the Zn^{2+} ion at the Hoogsteen face of A12 becomes directly coordinated to the N6 of A12 in the G12A mutant. Chemically, this is problematic because aromatic exocyclic amine groups are not particularly good metal ligands for Zn^{2+} (Figure 3.13). In addition, as previously discussed, a direct interaction between Zn^{2+} and A12 is expected to lower the pK_a and impair, not improve, the ability of A12 to serve as a general base (Figure 3.12).

We propose a mechanism in which A12 tautomerizes during the cleavage reaction of the hammerhead ribozyme (Figure 3.13). Nucleobases have the ability to adopt tautomeric structures, especially when constrained in the context of protein or RNA active sites. Xanthine, a nucleobase not typically found in RNA, is observed to be in the rare tautomeric species when bound in the binding pocket of the purine riboswitch (*273*). Rare tautomers of naturally occurring DNA nucleobases also appear to be trapped within the active site of polymerases (*274, 275*).

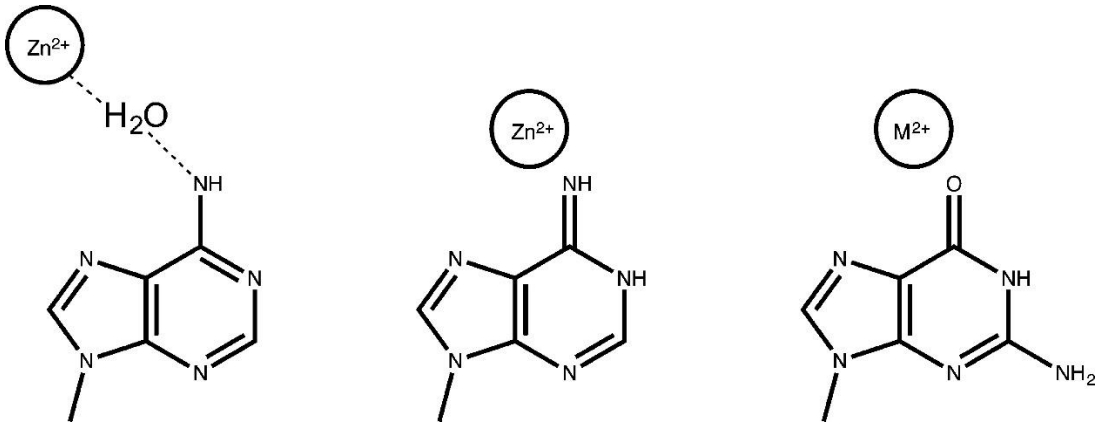

Figure 3.13. A conformational change is needed to stabilize the rare imino tautomeric form of A12.
In the crystal (left), a Zn^{2+} ion is observed to coordinate the N6 of A12 through its water ligand. Divalent metal ions are observed to stabilize tautomeric forms of adenine; however, this involves direct coordination of the metal ion to the nucleobase. Zn^{2+} rescue of the G12A mutant suggests that the Zn^{2+} ion could become directly coordinated to the N6 of A12, and the pH–rate profile of the G12A mutant is consistent with a model in which the reactive species is the rare imino tautomer (middle). In the WT ribozyme (right), a Mg^{2+} likely similarly coordinates to G12 to help lower the pK_a of G12 and organize the active site for catalysis.

The imino-N1H tautomeric species of adenosine has chemical properties that could explain how the N6 of A12 could directly interact with the Zn^{2+} ion, and why the pH−rate profile of the G12A mutant differs so little from that of the WT ribozyme. If A12 adopted the imino-N1H tautomeric form in the context of the hammerhead ribozyme active site, N6 would become more electron-rich and therefore a better inner-sphere ligand to Zn^{2+}. Computational studies predict that interactions between divalent metal ions are strengthened in the rare, imino-NH1 species of adenine and, likewise, that the rare tautomeric species would be stabilized by the presence of the metal ion (*276*). Thus, formation of the rare tautomer of adenine and stabilization of the same by Zn^{2+} coordinate could explain the observed metal specificity switch. This mechanism would predict that a Mg^{2+} ion is bound at a similar position to G12 in the WT ribozyme, leading to reduction in the pK_a and activation of the general base.

Tautomerization of A12 in the G12A mutation could also explain the pH−rate profile of the G12A mutant (Figure 3.8). If the dominant, amino, tautomer of A is functioning in the reaction, we would expect to see a different pH−rate profile in the WT and G12A mutant ribozyme reactions as the pK_a's of A and G differ significantly. However, the imino-N1H form of adenosine resembles guanosine in that N1 is protonated in its neutral form and the pK_a of this species would be predicted to be closer to that of guanine than that of adenine. If the minor, imino, tautomer of adenosine were to serve as a general base, the pH−rate profile for the G12A mutant ribozyme would therefore be expected to have the same overall shape as the WT ribozyme, as we observe. In addition, the same driving forces that facilitate

deprotonation of the N1 of guanosine could facilitate deprotonation of the imino-N1 tautomeric form of A12. Such a mechanism would help to explain the increased reactivity of 2,6-diaminopurine relative to 2-aminopurine at position 12 as a metal ion can interact with and stabilize the N6 imino-N1H tautomer of 2,6-diaminopurine nucleobase, but not the 2'-aminopurine nucleobase. We would predict that if G12 were replaced with a 2,6-diaminopurine, Zn^{2+} could further enhance the cleavage activity as this nucleobase possesses the N2 exocyclic amine lacking from adenine.

3.5.4 Active site rearrangements during the hammerhead ribozyme cleavage reaction.

There are significant indications presented here and elsewhere (*245, 252, 277*) that the ground-state structure of the hammerhead ribozyme observed in crystals must undergo a significant conformational change prior to or during the cleavage reaction. We are able to observe a striking change in the position of G12 at high pH where the hammerhead ribozyme is the most active. We cannot determine if the change in the conformation of the ribozyme is due to deprotonation of G12. However, the change in position of G12 is consistent with deprotonation of the nucleobase as G12 is observed to move away from the scissile phosphate, yet remain in position to deprotonate the nucleophile. Such a change in conformation in the active site has not previously been observed.

The location of the two active site metal ions in the available crystal structures suggests that significant additional conformational changes are likely needed to occur

prior to or during the cleavage reaction. The metal ion near G12 appears to be critical for ribozyme function, yet at maximum, this ion has a single inner-sphere ligand to the ribozyme and is located at the periphery of the active site. The metal ion at G10.1 is also distal to the active site, yet significant evidence exists that it contacts the scissile phosphate prior to the cleavage reaction, and this has not been observed in any crystal structure. Wang et al., elegantly linked Cd^{2+} rescue of phosphorothioate-substitutions at A9 and the pro-R_p oxygen of the scissile phosphate to the N7 of G10.1 in the context of a minimal hammerhead ribozyme that lacks the tertiary contact between stems I and II (*242*). These data suggest that the metal binding pocket near G10.1 must contract and bring the metal ion into proximity of the scissile phosphate. Ward et al. demonstrated that this metal ion interacts with the pro-R_p oxygen of the scissile phosphate in the ground state of the *S. mansoni* ribozyme. We are unable to observe metal binding at this position in any of the anomalous diffraction experiments described here. These data suggest that all of the crystal structures, including those presented here, represent an open structure that must subtly rearrange during the cleavage reaction.

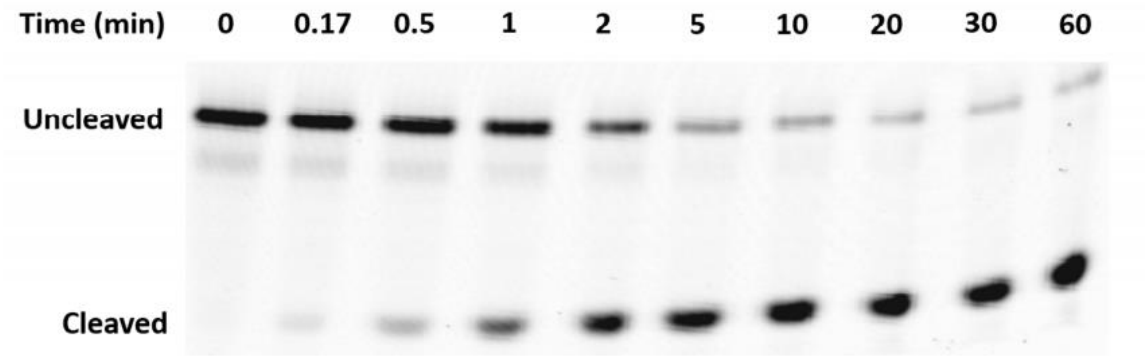

Figure 3.14. Sample reaction time course.
Reactions contained 2 µM ribozyme, 50 nM fluorescently labeled substrate, 50 mM MES pH 6.5 and 2 mM MgCl2. The reaction was performed at 37 °C and initiated by the addition of Mg^{2+} ion. Cleavage products were separated by PAGE and imaged using a Typhoon FLA 9500 imager.

3.6 Conclusions: a new strategy in RNA catalysis

Over 25 years after the discovery of the hammerhead ribozyme, we still do not fully understand how the hammerhead ribozyme performs its cleavage reaction. Most of the known small nucleolytic ribozymes appear to be largely preordered for catalysis. The hammerhead, however, seems to predominantly exist in an open and precatalytic state, and its cleavage mechanism appears to involve significant molecular motions. In addition, there are at least two distinct divalent metal ions involved in the cleavage reaction. One, near G10.1, is well characterized and predicted biochemically to interact directly with the cleavage site and likely participates in the cleavage reaction. The second, observed here to be near G12, is predicted to be important for organization of the active site as it proceeds through the cleavage reaction and likely tunes the pK_a of G12.

Our studies of the G12A mutant suggest that rare nucleobase tautomers could potentially play key roles in ribozyme reaction mechanisms. While we observe this phenomenon in a mutant ribozyme, native, wild-type ribozymes could also use this mechanism. Tautomeric species may also be key intermediates in and facilitate nucleobase protonation/deprotonation reactions in the nucleolytic ribozymes (*278*). Interpretation of pH dependence is often complicated due to the principle of kinetic equivalence, in which different reaction mechanisms can give rise to the same pH−rate profile. Kinetic equivalence makes it difficult to unambiguously identify the general acid and general base in enzymatic reactions. Tautomerization of nucleobases has the potential to add further complexity to such interpretation as seen here, where mutations and substitutions of the general base do not result in predictable changes in the pH−rate profile of the hammerhead ribozyme. Many of the small nucleolytic ribozymes, including the hairpin, glmS, VS, and twister ribozymes, employ conserved guanosine nucleotides that are well positioned to serve as proton donors and acceptors (*279*). Formation of tautomeric species, and the accompanying changes in pK_a, have the potential to complicate analysis of these ribozyme reactions. Thus, we still have much to learn about the mechanisms by which ribozymes can catalyze chemical reactions under biological conditions.

CHAPTER 4. TWO ACTIVE SITE DIVALENT IONS IN THE CRYSTAL STRUCTURE OF THE HAMMERHEAD RIBOZYME BOUND TO A TRANSITION STATE ANALOGUE

4.1 Declaration of collaborative work

Aamir Mir and Barbara Golden contributed to this chapter. Under the supervision of Barbara Golden, Aamir Mir crystallized the transition state analog RzB ribozyme, collected diffraction data, and solved the crystal structures. Barbara Golden prepared the figures containing RzB structures, and Aamir Mir and Barbara Golden wrote, and edited the chapter. Victoria DeRose and Dan Herschlag provided critical feedback. This chapter is reprinted and adapted with permission from reference (*280*). Copyright (2016) American Chemical Society.

4.2 Introduction

The hammerhead ribozyme was the first of the small nucleolytic ribozymes to be described (*220*) and the first to have its crystal structure determined (*265, 266*). While these initial structures were found to be trapped in an inactive conformation, (*264, 281*) structures of the extended hammerhead ribozymes inhibited by placement of a deoxy or methoxy modification at the cleavage site were largely in agreement with solution

biochemical experiments (*238, 240, 270*). Nevertheless, a significant body of data that cannot be reconciled with the crystal structures of the trapped, inhibited versions of the extended ribozyme previously studied remains (*219, 242, 245*). We therefore sought to obtain a crystal structure of the hammerhead ribozyme bound to a transition state analogue.

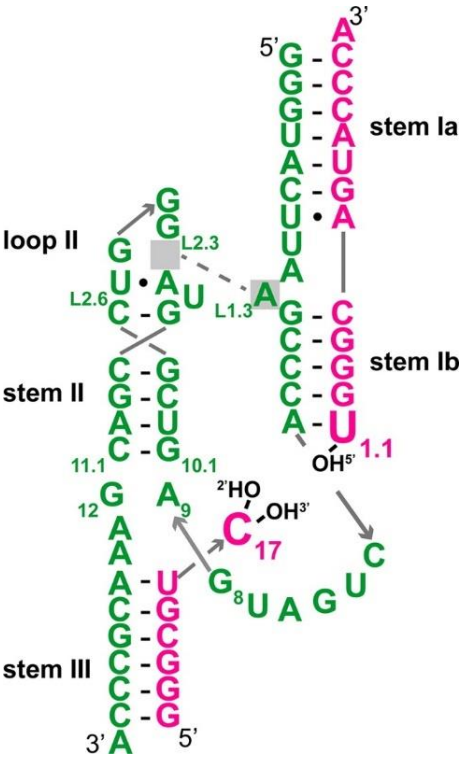

Figure 4.1. Secondary structure of the RNA used to determine the structure of the transition state analogue.
The ribozyme strand is colored green. The strand corresponding to the substrate (pink) is synthesized in two fragments. The free 2'-, 3'-, and 5'-hydroxyls at the cleavage site bind and position the vanadate ion.

4.3 Materials and Methods

4.3.1 RNA Synthesis

The molecule described here was designed based on the previously crystallized RzB hammerhead ribozyme (*219*). The ribozyme was assembled from three pieces of RNA. The first piece of RNA is formally the ribozyme strand (green in Figure 4.1). The RNA that is formally the substrate strand and that would be cleaved if synthesized as a single RNA was divided into two fragments. The first fragment consists of the first 7 nucleotides of the substrate strand through and including C17. The second fragment was 13 nucleotides in length and first nucleotide of this fragment is U1.1. C17 has free 3'- and 5'-hydroxyl groups and U1.1 has a free 5'-hydroxyl. These are shown in pink in Figure 1. Thus, when the three strands are annealed, the phosphate group at the cleavage site is absent.

The ribozyme strand was synthesized by in vitro transcription as described previously (*257*). Briefly, a DNA template was designed which contained a T7 promoter, the ribozyme sequence of RzB and an Ear I restriction site. The DNA was then cloned into pUC-19 vector and amplified in DH5α cells. Following plasmid purification and cleavage by Ear I, the DNA was used as a template for in vitro transcription using T7 RNA polymerase. The ribozyme strand was then purified using a PAGE gel. The two strands that contain nucleotides that would make up the substrate were purchased from

Dharmacon (GE Life Sciences). They were deprotected and desalted according to manufacturer's guidelines prior to use.

4.3.2 RNA Crystallization

The ribozyme was reconstituted by mixing 0.226 mM of the ribozyme strand and 0.235 mM of each substrate strand in 10 mM potassium cacodylate pH 6.0. To ensure proper refolding of the transition state analog, the mixture was heated to 90 o C for 1 min, and cooled to room temperature for 10 min. Next 10 mM $MgCl_2$ and 2 mM NH_4VO_3 were added to the mixture and the RNA was incubated at 50 °C for 5 min. The mixture was cooled again for 10 min at room temperature prior to use.

To crystallize the RNA, a 1 µL drop of the refolded ribozyme was mixed with a 1 µL drop of the reservoir solution (32 % 2-methyl-2,4-pentanediol, 0.9 M potassium chloride, 50 mM Tris-HCL pH 8.2, 0.5 mM spermine and 2 mM NH4VO3). These drops were equilibrated by vapor diffusion against 1 mL of reservoir solution. Crystals were obtained in about 4 weeks. Crystals were soaked in a cryostabilization buffer containing reservoir solution, 5 % isopropanol and either 50 mM of Mg^{2+} or 10 mM Mn^{2+}. These crystals were then flash cooled and stored in liquid nitrogen.

4.3.3 Crystal structure determination

Data collection was performed at LS-CAT beamlines 21-ID-F and 21-ID-D at the Advanced Photon Source in Argonne national lab. Data was collected at 12.7 keV and

8.26 keV for crystals soaked in Mg^{2+} and Mn^{2+} respectively. Data were processed and indexed in HKL2000 in the space group C2221. We used PHENIX (*259*) for phase determination and refinement. The phase was determined using RzB WT structure (PDB id: 5DI2) as search model in Phaser-MR module provided in PHENIX. The successful MR solution had TFZ score of 14.4 and LLZ of 1341.133. The initial structure was then improved by several rounds of manual building in Coot (*260*) and refinement in Phenix.

4.4 Results and Discussion

We started with the sequence of the artificial hammerhead RzB ribozyme (*255*). To obtain the crystal structure of the transition state analogue, we mixed the three RNA strands shown in Figure 4.1. The resulting complex lacks the scissile phosphate but retains the 2'-, 3'-, and 5'-hydroxyls at the cleavage site. This complex was cocrystallized with NH4VO3. Details are provided in the methods and in Table 4.1, and electron density for the active site is shown in Figure 4.2. We obtained crystal structures for crystals soaked in Mg^{2+} or Mn^{2+} ions at 3.0 and 3.2 Å resolution, respectively. Both of these divalent ions support hammerhead ribozyme activity (*247, 248*).

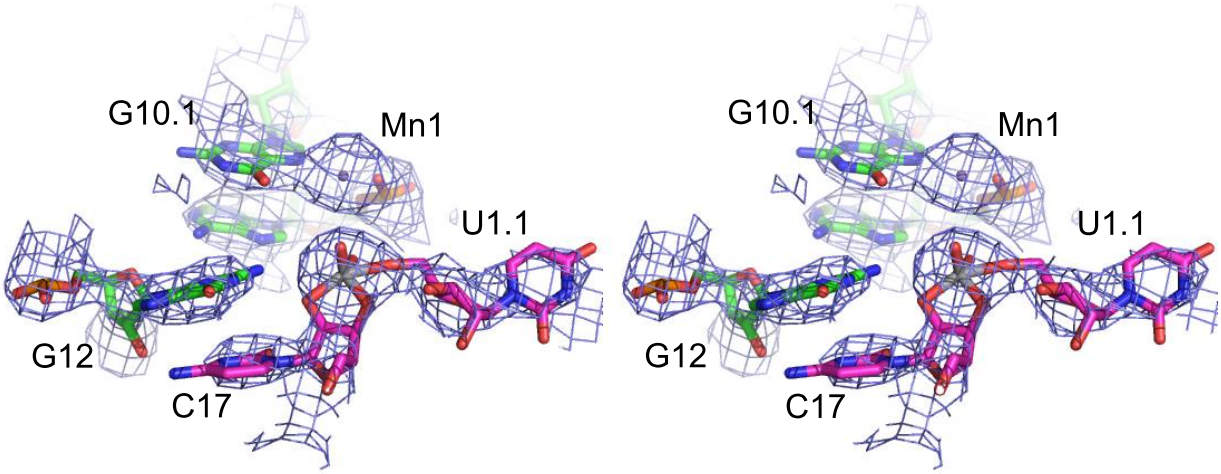

Figure 4.2. Electron density for the hammerhead ribozyme active site.
Crystals of the hammerhead ribozyme bound to a transition state analog were grown in Mg^{2+} containing buffer and then soaked into a cryostabilization buffer containing Mn^{2+}. The $2F_o$-F_c map is shown contoured in blue at 1.5 σ. The image is rendered as wall-eyed stereo using PyMol

Table 4.1. Data collection and refinement statistics of the transition state analog structure of the hammerhead ribozyme.

Data Collection	**Mg^{2+} (pdb id-5EAO)**	**Mn^{2+} (pdb id-5EAQ)**
cell dimensions (Å)	80.0, 87.0, 105	81.3, 87.4, 106
Wavelength (Å)	0.979	1.501
Resolution (Å) (last shell)	50.0-3.00 (3.05-3.00)	50.0-2.95 (3.00-2.95)
R_{merge} (%)	14.2 (55.5)	19.6 (63.6)
I/σ(I)	51.6 (5.7)	139.7 (19.4)
no. of reflections	14333	12773
redundancy	12.5 (12.6)	6.3 (5.0)
completeness (%)	99.4 (97.8)	99.4 (97.2)
Refinement		
Resolution (Å) (last shell)	28.3-3.00 (2.99-3.09)	27.4-3.20 (3.35-3.20)
no. of reflections	14229	11834
no. of reflections in the test set	1405	1180
R_{work}/R_{free} (%) (last shell)	22.7/26.9 (30.4/35.3)	24.6/28.6 (28.6/35.9)
no. of atoms		
RNA	1456	1456
ions	8	7
average B-factor	39.9	76.1
rmsd bond-lengths, Å	0.006	0.007
RMSD bond angles, °	1.154	1.488
Coordinate error (Å)	0.34	0.36

4.4.1 Vanadate was incorporated at the scissile phosphate position

We first examined the structure for evidence that vanadate was indeed present at the position corresponding to the scissile phosphate. Anomalous difference maps

calculated with Mg^{2+}- or Mn^{2+}-soaked crystals reveal a 4−5σ peak at the cleavage site centered on the position normally occupied by the scissile phosphate (Figure 4.3). While the phosphorus atoms in the RNA backbone can potentially give rise to an anomalous signal, there is no density in the anomalous difference map for any of the backbone phosphate groups. We therefore attribute the peak in the anomalous map to vanadium. Distances between the 2'-, 3'-, and 5'-oxygens and the vanadium atom are consistent with direct coordination between the oxygen atoms and the vanadium atom. Finally, the structure of the active site is distinctly different than that observed for the previous structures containing a tetravalent phosphate at the active site (Figure 4.4). The description of the active site described here will focus on the structure of the ribozyme in Mn^{2+} ion as locations of the divalent metal ion binding sites can be unambiguously characterized using anomalous diffraction.

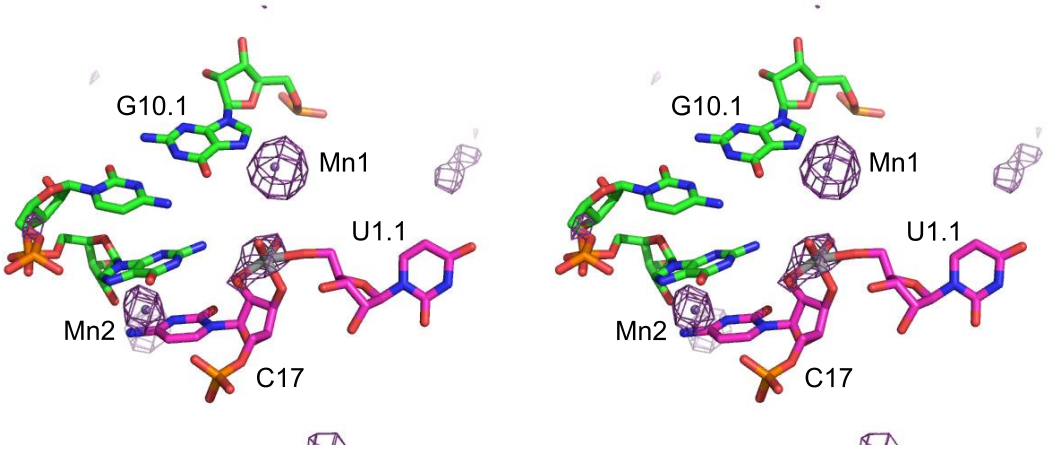

Figure 4.3. Anomalous diffraction from Mn^{2+} -soaked crystal.
Crystals of the hammerhead ribozyme bound to a transition state analog were grown in Mg^{2+} containing buffer and then soaked into a cryostabilization buffer containing Mn^{2+}. An anomalous diffraction map, contoured here at 3 σ, was calculated to determine the positions of the Mn^{2+} ions and vanadate bound to the ribozyme. The image is rendered as wall-eyed stereo using PyMol.

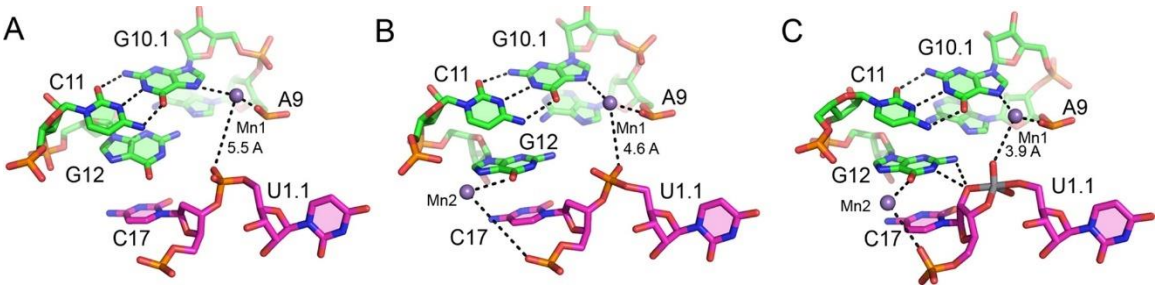

Figure 4.4. Rearrangements in the active site of the hammerhead ribozyme. Nucleotides in the ribozyme active site are colored green; the cleavage site dinucleotide is colored pink, and the Mn^{2+} ions in the active site are colored purple. The three structures shown represent the ribozyme inhibited by a deoxy substitution at C17 at pH 5.0 **(A)** and pH 8.0 **(B)** and the transition state analogue at pH 8.2 reported here **(C)**. The Protein Data Bank entries for these structures are 5DI4, 5DI2, and 5EAQ respectively. Dashed lines represent base-pairing interactions and the distances listed in Table 1. **A.** At pH 5.0, G12 forms a sheared base pair with A9 and an active site Mn^{2+}, Mn1, is 5.5 Å from the pro-R_p oxygen of the scissile phosphate. **B.** When the pH is increased to 8.0, the sheared base pair between G12 and A9 is broken, allowing G12 to stack upon C17. A second Mn^{2+} ion, Mn2, is observed to interact with the Hoogsteen face of G12, and Mn1 is observed to be 4.6 Å from the pro-R_p oxygen of the scissile phosphate. **C.** In the structure of the transition state analogue, G12 is positioned such that both N1 and N2 can hydrogen bond with the nucleophile. Mn2 is observed to directly ligand the O6 keto oxygen of G12, and Mn1 is 3.9 Å from the position of the pro-S_P oxygen of the scissile phosphate.

4.4.2 The G10.1 divalent metal (Mn1)

The first major difference between the inhibited structures and the transition state analogue involves the G10.1-C11 base pair. A significant body of work suggests that there is an active site divalent metal ion whose ligands include N7 of G10.1, the pro-R_p oxygen of A9, and the pro-R_p oxygen of the scissile phosphate (*242, 243*). This metal ion, which we call Mn1, is observed in crystals to bind near N7 of G10.1 and the pro-R_p oxygen of A9; however, it is always distant from the cleavage site (*219, 240, 251*).

For example, in the crystal structure of the RzB ribozyme at pH 5.0, Mn1 is 4.6 Å from the pro-R_P oxygen of the scissile phosphate.

In the crystal structure of the transition state analogue, the active site has, in fact, undergone a small but significant conformational change that results in rotation of the oxygens on the scissile phosphate. As a result, Mn1 is now observed to be only 3.9 Å from the oxygen atom from the vanadate that would correspond to the pro-S_P oxygen of the scissile phosphate (Figure 4.4C and Table 4.2). The oxygen atom corresponding to the *pro-R_P* oxygen of the scissile phosphate is 5.3 Å distant. While the position of Mn1 in the crystal structure of the transition state analogue still does not allow direct interactions with all of the predicted ligands to occur simultaneously, there is net movement of the metal ion toward the scissile phosphate providing better agreement with the available solution biochemical data (*242, 243, 245*). It is possible that there are additional changes in the active site structure that occur during the cleavage reaction that allow all three predicted ligands to interact with Mn1. Alternately, it is possible that vanadate is not a sufficiently good mimic of the pentavalent phosphate intermediate to observe these interactions in what is still a trapped and unreactive crystal structure.

Table 4.2. Atomic distances (Å) in the hammerhead ribozyme active sites. Distances were obtained from structures of PDB entries 5DI4 (pH 5.0), 5DI2 (pH 8.0), 5EAQ (TS).

		pH 5.0	pH 8.0	TS
G12 N2	C17 2'-OH			3.0
G12 N1	C17 2'-OH			2.9
G10.1 N7	Mn1	3.2	3.3	2.7
A9 *pro-R_p* O	Mn1	1.9	2.6	2.2
U1.1 *pro-R_p* O	Mn1	5.5	4.6	5.3
U1.1 *pro-R_p* O				3.9
G12 O6	Mn2		3.3	2.6
C17 *pro-R_p* O	Mn2		6.2	3.8
U1.1 *pro-R_p* O	G8 2'-OH			2.7

4.4.3 G12 undergoes significant conformational changes

The second major difference between the inhibited structures and the structure of the transition state analogue centers around G12. The nucleobase of G12 appears well placed to serve as the general base in the cleavage reaction. In all of the inhibited structures, G12 is positioned such that N1 can hydrogen bond with the 2'-hydroxyl of C17, the nucleophile in the cleavage reaction. In the structure of the transition state analogue, the position of G12 is distinctly different from that observed previously. There are three major changes. First, G12 has shifted slightly such that the 2'-oxygen of C17 is positioned by a bifurcated bond involving both N1 and exocyclic N2 of G12 (Figure 4.4 and Table 4.2). Second, a metal ion, here termed Mn2, is observed to interact directly with O6 of G12 and indirectly with the pro-R_p oxygen of C17. Third, the base pairing

between G12 and A9 has been broken, allowing G12 to move toward the cleavage site. Interestingly, the position of G12 in this transition state analogue is similar to that previously observed at high pH and in structures of the G12A mutants (*219, 256, 261*).

The conformation of G12 as observed in the structure of the transition state analogue is supported by a significant body of solution biochemical studies. While the sheared base pair between G12 and A9 is observed in structures of the inhibited ribozyme, replacement of A9 with a purine nucleobase, which cannot base pair with G12, is only modestly detrimental to catalysis (*282, 283*). This suggests this base-paring interaction is not critical for ribozyme activity and could, in fact, be inhibitory. The exocyclic amine of G12, however, does seem to play an important role in catalysis as there is a significant difference in activity when G12 is replaced with inosine, a nucleobase that resembles guanosine but lacks the N2 exocyclic amine (*237, 282, 283*). These data point to a significant role for the exocyclic amine of G12. The hydrogen bond between N2 of G12 and the 2'- oxygen of C17 could play critical roles in catalysis. This interaction likely serves to help position the nucleophile, and it could also facilitate deprotonation of the 2'-hydroxyl of C17.

4.4.4 G12 divalent metal (Mn2) may be involved shifting the pK_a of G12

While monovalent and divalent ions have previously been observed to interact with the Hoogsteen face of G12 at the N7 and O6 positions, the interaction with O6 was water-mediated and the metal ions were distant from the active site (*219, 240*). A new

divalent metal ion binding mode at G12 is observed in the transition state analogue. In this structure, Mn2 is observed to directly coordinate to O6 of G12 (Figure 4.4 and Table 4.2). Anomalous scattering experiments clearly demonstrate Mn^{2+} binding at this position, and a 3σ peak is observed at this position in the Fo − Fc map from the crystal containing Mg^{2+}. The importance of this divalent ion has been demonstrated using a G12A mutant of the hammerhead ribozyme. When the keto oxygen of G12 is replaced with the exocyclic amine of A12, a metal specificity switch is observed, with the G12A mutant exhibiting a clear preference for Zn^{2+} ion (*219*). This suggests that there is a direct interaction between the Zn^{2+} ion and the exocyclic amine of A12 and, by analogy, an interaction between a divalent metal ion and O6 of G12 in the wild-type ribozyme. The pro-R_p oxygen of C17 also interacts with this metal ion, likely through an outer-sphere, water-mediated interaction consistent with the lack of a thio effect at this phosphate (*284*).

It has been suggested that the pK_a of G12 is shifted from a value of ∼9.4 observed for free guanosine to a value of ∼8.0 in the active site of the hammerhead ribozyme (*219, 262*). Coordination of the Mn^{2+} ion at O6 has the potential to lower the pK_a of G12 either through electrostatic effects or through favoring the formation of tautomeric species that help distribute the negative charge on N1.

4.4.5 A water bound to G10.1 divalent metal (Mn1) may be serving as a general acid

How, or even if, the ribozyme activates the 5'-oxygen leaving group remains unclear. In most structures of the inhibited ribozyme, (*219, 238, 240*) the 2'-hydroxyl of G8 interacts with the 5'- hydroxyl and was proposed to serve as a general acid in the cleavage reaction. Solution biochemical experiments clearly demonstrate that the 2'-hydroxyl of G8 plays a significant role in the catalytic reaction (*285, 286*). In the crystal structure of RzB at pH 8.0 and in the transition state analogue, (*219*) however, the 5'-oxygen leaving group is pointed away from the 2'-hydroxyl of G8 (Figure 4.5). In these structures, the 2'-hydroxyl of G8 is engaged in a hydrogen bond with the vanadate oxygen that would correspond to the pro-R_p oxygen of the scissile phosphate. Thus, it is possible that the 2'-hydroxyl of G8 plays a role other than serving as the general acid. The structure of the transition state analogue suggests that the 2'-OH of G8 could help organize the active site by positioning the scissile phosphate and dissipating any negative charge that builds up on the nonbridging oxygen.

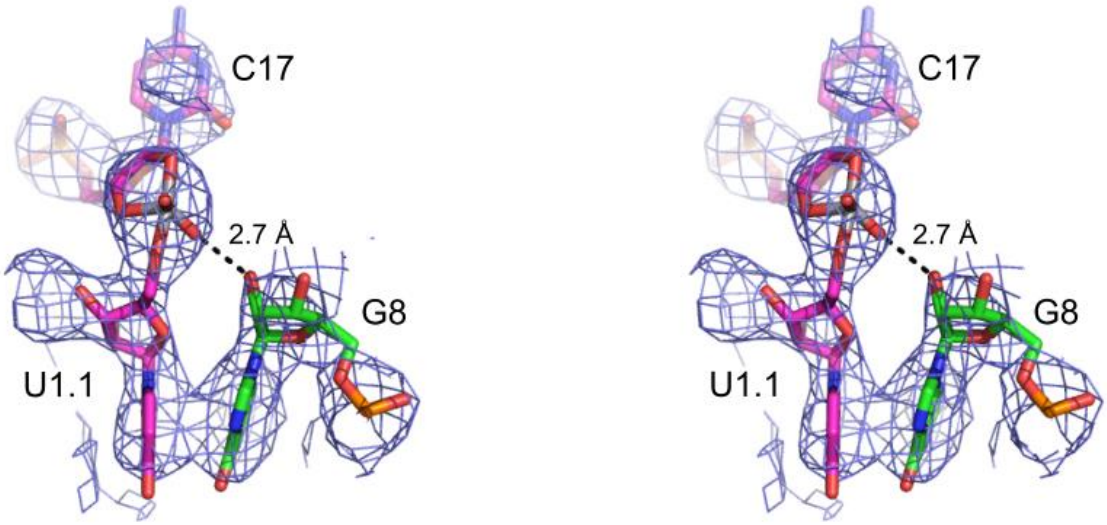

Figure 4.5. Electron density around G8.
The $2F_o$-F_c map is contoured at 1.2 σ. The image is rendered as wall-eyed stereo using PyMol.

It remains possible that a Mn^{2+} bound water serves as the general acid in the cleavage reaction. Mn1 remains ~3 Å from a position that would allow direct coordination to the pro-R_p oxygen of the vanadate, while solution biochemical experiments suggest it becomes directly coordinated to this atom. If Mn1 does move to coordinate the pro-R_p oxygen of the scissile phosphate as predicted, the 5'-hydroxyl leaving group is in a conformation that would be within hydrogen bonding distance of the water molecules coordinated to the divalent metal ion. Thomas and Perrin demonstrated that a ribozyme modified to contain a 5'-sulfur leaving group, a hyperactive substrate, was insensitive to metal ion substitutions and retained activity even in the absence of divalent metal ions (*286*). This observation is consistent with an active site divalent metal ion that is important for stabilization of the 5'-O leaving group.

4.5 A revisited catalytic mechanism of the hammerhead ribozyme.

We propose a mechanism for the hammerhead ribozyme shown in Figure 4.6. In this model, G12, activated by a divalent metal ion, serves as the general base. A second divalent metal ion, coordinated to N7 of G10.1, the pro-R_p oxygen of A9, and the pro-R_p oxygen of the scissile phosphate, is positioned such that one of its water ligands can serve as the general acid. A general base with a pK_a of 8−9 and a Mg^{2+}- bound water (pK_a ∼ 11.4) (*287*) would be consistent with the observed pH−rate profiles for this ribozyme (*219, 247*). While there are significant experimental data available to support this model, it will clearly require further validation, both to determine if G12 is, in fact, a general base and to learn if and how the hammerhead ribozyme stabilizes the leaving group.

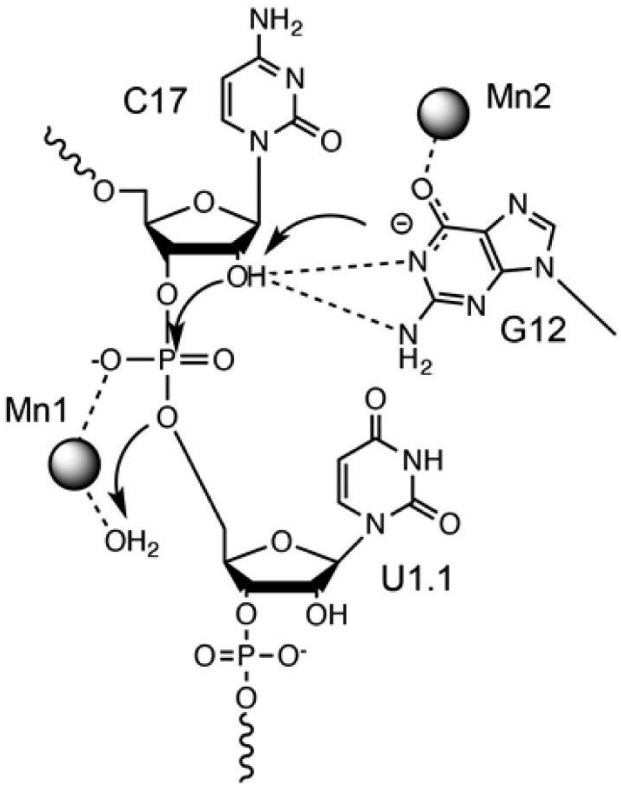

Figure 4.6. Proposed mechanism for general acid/general base catalysis by the hammerhead ribozyme.
G12 serves as the general base. Mn2 binds at O6 of G12 where it can lower the pK_a of G12 and help to organize the active site. Mn1 becomes positioned near the 5′-oxygen leaving group, positioning one of its bound water molecules such that it could serve as a general acid.

4.6 Concluding remarks and future studies regarding the hammerhead ribozyme investigation

The hammerhead ribozyme has been one of the most well-studied self-cleaving ribozymes. Using studies mentioned in Chapters 3 and 4, together with investigations done by other groups, we now have a clearer picture of the hammerhead ribozyme catalysis. The hammerhead ribozyme uses a general acid-general base mechanism for

RNA cleavage. However, it undergoes conformational changes which are dependent on the pH of the solution which brings the ribozyme to an active conformation. In this conformation, the hammerhead ribozyme uses a bifurcated interaction between the deprotonated G12 (via N1 and -NH2) and the 2'-OH of C17 to activate and deprotonate the nucleophile. The deprotonated G12 is stabilized by a divalent metal as it directly coordinates the O6 of G12. Stabilization of deprotonated G12 leads to shifting of the pK_a of N1 of G12 to ~8. The transition state stabilization is provided by the 2'-OH of G8 as it hydrogen bonds with pro-R_p O of the scissile phosphate in the active conformation, and divalent metal(s) interaction with pro-R_p and/or pro-S_p oxygens. The G10.1 divalent metal interacts with pro-R_p oxygen in the ground-state based on biochemical data, and pro-S_p oxygen in the transition state. The leaving group stabilization is most likely provided by a water molecule bound to G10.1/A9 divalent metal. The hammerhead ribozyme also adopts the in-line geometry needed to accelerate RNA-cleavage reaction. Using these catalytic strategies, the hammerhead ribozymes is able to cleave RNA very efficiently.

However, the hammerhead ribozyme catalysis still needs more work. An unpublished crystal structure of the RzB hammerhead ribozyme, where the nucleophile is replaced with 2'-NH2 group, shows at least two divalent metals in the active site, but at different positions. The G10.1 divalent metal is at the regular binding site, but the second divalent metal is directly bound to the pro-S_p oxygen. There is electron density close to O6 of G12, but it is not due to anomalous diffraction. Therefore, it could either be a monovalent ion, K^+ or a Mg^{2+} ion. While the 2'-NH2 may create an unnatural

divalent metal binding site, it is however possible that the hammerhead ribozyme could use an additional divalent/monovalent for cleavage reaction. We also, indirectly, observed an imino tautomerized A12 in the G12A mutant of RzB. It certainly has far-reaching consequences for RNA catalysis since tautomerization provides a way of increasing the catalytic repertoire of RNA, and natural ribozymes may use this strategy to catalyze chemical reactions. However, more experiments are needed to probe the G12 divalent metal binding site, potential tautomerization of the nucleobase at position 12. We are currently in the process of testing type II hammerhead ribozyme that contains a variety of unnatural nucleobases at position 12. These nucleobases include 2-aminopurine, 2,6-diaminopurine, 6-thioguanine, and adenine. Divalent metal titrations, and pH-rate profiles of these mutants will provide further insights into the role of G12 in general base catalysis, and the role G12 divalent metal in nucleobase tautomerization, and shifting the pK_a of G12.

CHAPTER 5. HUNTING FOR FUNCTIONAL SELF CLEAVING RIBOZYMES

5.1 Declaration of collaborative work

The work mentioned in this chapter is the result of a collaborative effort in the Golden lab. Most of the work mentioned in this chapter was performed by Aamir Mir including establishing protein purification protocol, method development for 2', 3' - cyclic phosphate capture, preparation of the libraries described in this chapter, and the analysis of RNA-seq results. Samantha Lee helped purify *At* ligase and Tpt1 for library preparation, purified the total RNA from *Anopheles gambiae*, and *Drosophila melanogaster* for MiSeq library preparation. She also purified the total RNA from *Aedes aegypti* for HiSeq library preparation, and performed the ligation reactions. Aamir Mir wrote the chapter and prepared all the figures mentioned in this chapter. I would also like to acknowledge Dr. Elizabeth Tran, Siwen Wang and Yu-Hsuan Lai for providing critical feedback regarding the RNA-seq library preparation protocol and analysis of the RNA-seq results.

5.2 Introduction

The majority of the known ribozymes catalyze the cleavage of a phosphodiester bond in RNA. The cleavage reaction produces a 5’ product with 2’, 3’ –cyclic phosphate at its 3’ end and a 3’ product that contains a 5’ hydroxyl (*162*). These self-cleaving ribozymes were originally discovered in viroids as part of their rolling circle replication mechanism (*162*). As discussed in Chapter 2, several self-cleaving ribozymes are known, such as the hammerhead, the hepatitis delta virus (HDV) and the hairpin ribozymes. An important property of these ribozymes is that they have a conserved secondary structure but not the primary structure and can adopt a variety of complex folds (*216*). The diverse structures of these self-cleaving ribozymes are responsible for different catalytic strategies these self-cleaving ribozymes utilize for cleaving the phosphodiester bond. These strategies have already been discussed in Chapter 2-4. However, RNA in the primordial world may have required even more diversity in the strategies employed by ribozymes for catalysis. To identify these catalytic strategies used by ribozymes, we need to discover novel structures of self-cleaving ribozymes. A high-throughput approach is needed to find more natural self-cleaving ribozymes.

The earliest self-cleaving ribozymes were all discovered in viral RNAs in the late 1980s. Then surprisingly no ribozymes were discovered until Breaker and colleagues discovered *glmS* ribozyme, twister ribozyme, pistol ribozyme and many others (*176, 194, 203*). These ribozymes were discovered using bioinformatics based approaches. Structural and genetic similarities to known ribozymes were exploited to find these new

ribozymes. While bioinformatics approaches are invaluable for discovery, they tend to be biased towards existing motifs belonging to self-cleaving ribozymes (*176, 194*). This bias could be one of the reasons why very few self-cleaving ribozymes have been discovered, and underscores the need for an unbiased experimental approach to find novel self-cleaving ribozymes.

Moreover, many bioinformatics based searches performed using sequence motifs of known ribozymes have revealed thousands of potential ribozyme motifs belonging to existing ribozymes in the available genomes (Table 5.1) (*175, 222, 223*). They are found in all domains of life, including several candidates in humans (*288*). These approaches have identified *Aedes aegypti, Anopheles gambiae,* and *Drosophila melanogaster* as the organisms that harbor many self-cleaving ribozymes. These searches were performed with known consensus sequences and secondary structures of existing ribozymes (*289-291*). While it is remarkable that self-cleaving ribozymes are abundant in nature, it is not clear if these motifs are functional *in-vivo*. Many of the motifs were found to be active *in-vitro*. However, whether they cleave *in-vivo* or even get expressed is unclear. Many of these motifs are found to be in the intergenic and silenced genomic locations. Hence, they would not be expected to be functional inside organisms. Secondly, the physiological roles of these motifs are also not properly understood. As discussed in Chapter 2, some studies have suggested their involvement in gene expression, but the vast majority of them have unknown functions (*203, 205, 206, 292*). Hence, in addition to being unbiased, we need a high-throughput approach that can detect functional self-cleaving ribozymes from the transcriptome of organisms.

Table 5.1. Motifs belonging to some self-cleaving ribozymes are abundant in nature. The table shows the number motifs similar to the shown self-cleaving ribozymes found using bioinformatics searches. Self-cleaving ribozymes that do tend to be abundant are not shown. The table summarizes the results from references (*175, 176, 194, 222, 293*). However, this is not a comprehensive list, and the number of motifs is likely higher. The motifs shown here are also not unique sequences. Multiple copies of the same motif in the same genome are counted multiple times.

Domain	**Hammerhead**	**HDV**	**Twister**	**Hatchet**	**Pistol**	**glmS**
Archaea	2	0	0	0	0	0
Bacteria	162	1	32	1	34	582
Eukaryotes	31102	1291	1163	0	0	0
Schistosoma	25370	267	1051	0	0	0
Insects	4637	120	64	0	0	0
Viruses	25	1	1	0	0	0
Environmental	1209	541	1495	302	875	150

Since the sequence of ribozymes in nature is composed out of four nucleotides, they appear to evolve relatively easily *in-vitro* (*62, 77*). Several artificial ribozymes have been discovered using this approach (*66, 294, 295*). Some of these ribozymes are discussed in Chapter 1, and Table 1.1. Furthermore, this phenomenon is evident in a recent study by Popovic et al. (*61*). They devised several selection strategies for self-cleaving ribozymes and combined them with high-throughput sequencing to find self-cleaving RNA motifs. As expected, the hammerhead motif was one of the most abundant ribozymes, however, they found hundreds of other self-cleaving RNA families. Hence, indicating that self-cleaving ribozymes can readily evolve *in-vitro.* If ribozymes tend to evolve relatively easily *in-vitro*, it can be expected that billions of years of

natural evolutionary processes must have produced many such self-cleaving ribozymes, which are waiting to be discovered.

Since ribozymes appear to evolve relatively easily *in-vitro* and that known self-cleaving ribozymes appear to be widespread in nature, we hypothesize that many more self-cleaving ribozymes remain to be discovered. The initial search for self-cleaving ribozymes will focus on *A. aegypti, A. gambiae* and *D. melanogaster.* As mentioned, these organisms are known to harbor many motifs similar to known self-cleaving ribozymes (*175, 222, 296*). Therefore, they will not only serve to validate our method, but may also increase the chances of discovering new self-cleaving ribozymes.

To hunt for new ribozymes and to understand the biological importance of the known ribozymes, we developed a biochemical screen to select for self-cleaving RNAs within the transcriptomes of the selected organisms. The biochemical screen selects for the RNA products of the self-cleaving ribozymes that contain 2', 3' –cyclic phosphate at the 3' end. Such a screen is more likely to find functional and biologically relevant self-cleaving ribozymes compared to bioinformatics based screens. Moreover, our approach does not require a consensus sequence of known ribozymes, thus increasing our chances of finding previously unknown ribozymes with novel secondary structures. This biochemical approach will be the first study specifically designed to look for functional ribozymes. In addition to self-cleaving ribozymes, data generated from this study can be used to study cleavage by Ire1p, miRNA loading into Dicer, and tRNA cleavage (*297, 298*). It has been reported that Ire1p, involved in protein unfolding response, cleaves an intron from XBP1 pre-mRNA and generates 2', 3' –cyclic phosphates (*298, 299*). Similarly

tRNAs can undergo cleavage in the acceptor region and anticodon loop under stress, and for pre-tRNA processing by tRNA splicing endonuclease (*297, 300, 301*). Moreover, there are other proteins that can cleave RNAs to generate products with 2', 3' -cyclic phosphates including angiogenin, RNase L, RNase T, GCN4, and others (*302-304*). Lastly, U6 snRNA also contains a stable 2', 3' -cyclic phosphate RNA at its 3' end which is important for U6 snRNP assembly (*305*). Hence, looking for 2', 3' –cyclic phosphate containing RNAs can provide insights into other important processes in the cell. Lastly, the discovery of new ribozymes can potentially expand the limited catalytic toolkit available to RNA. It is likely that new self-cleaving ribozymes will use different catalytic strategies to cleave the phosphodiester bond. Expansion of RNA catalytic potential will not only explain some of the roles self-cleaving ribozymes are playing in modern biology, but also illuminate the catalytic strategies used by RNA in the primordial world.

5.3 Methods

5.3.1 Preparation of ribozymes

Two ribozymes used in this study are the HDV ribozyme and the hammerhead ribozyme. The HDV ribozyme produces a 5' cleavage product of 3 nt whereas the hammerhead ribozyme gives 5' cleavage product of 7 nt. Both the HDV and the hammerhead ribozymes contain two separate strands: ribozyme strand and a separate substrate strand. They were designed based on the previously reported sequences (*178,*

186, 306). Chemically synthesized, DY-547 labeled strand of HDV ribozyme and the hammerhead ribozyme containing the scissile phosphate ('substrate') was purchased from Thermo Scientific. The ribozyme strands were produced using large scale *in vitro* transcription reactions and purified using denaturing acrylamide gels as described previously (*177*).

5.3.2 Purification of *A. thaliana* tRNA ligase and *S. cerevisiae* Tpt1.

At tRNA ligase was purified using a previously mentioned protocol (*297*). The plasmids expressing *At* tRNA ligase were also acquired from the same group. The *At* tRNA ligase is fused to an N-terminal hexahistidine tag in a pET-28a plasmid. The Tpt1 is fused to N-terminal hexahistine tag and a C-terminal Strep tag in a pET-53 plasmid. These plasmids were transformed into BL21 (DE3) RIPL cells per the instructions of Agilent Technologies. Single colonies were first used to inoculate 50 mL LB containing 34 µg/mL chloramphenicol and 50 µg/mL of either ampicillin (pET-53) or kanamycin (pET-28a). These cultures were grown overnight, and then mixed into 750 mL of LB containing the same antibiotics. These cultures were grown at 37 °C for additional 2-3 hours until OD_{600} was measured to be 0.6. At this point, the cultures were cooled in ice for 30 minutes, and then ethanol and isopropyl β-D-1 thiogalactopyranoside (IPTG) were added to final concentrations of 2 % and 0.4 mM respectively. The cultures were then transferred to a 17 °C shaking incubator for approximately 20 hours. Following that, the cultures were centrifuged at 6000 g for 20 minutes to collect the cell pellet which was

then frozen at -80 °C. The cell pellets were then thawed on ice followed by the addition of sonication buffer (50 mM Tris-HCL at pH 7.5, 250 mM NaCl and 10 % sucrose). To lyse the cells, 0.2 mg/mL lysozyme was added and the solution rotated at 4 °C for 30 min. This solution was then pulse sonicated for 30 seconds three times on ice. Then 0.1 % Triton X-100 was added followed by centrifugation at 15000 rpm. Following centrifugation, the supernatants were decanted into 3-5 mL of pre-equilibrated Nickel agarose (Qiagen) solution, and the mixtures gently rotated at 4 °C for 1-2 hours. Once the binding was complete, the beads were washed twice with 15 mL of wash buffer (50 mM Tris-HCL at pH 7.5, 250 mM NaCl, 25 mM imidazole and 10 % glycerol). Then, the beads were transferred to a fritted chromatography column, and eluted using the elution buffer (50 mM Tris-HCL at pH 7.5, 250 mM NaCl, 300 mM imidazole and 10 % glycerol) into 10 fractions. The fractions were then analyzed on an SDS page gel to identify the fractions containing the desired proteins. These fractions were pooled and dialyzed against dialysis buffer 1 (50 mM Tris-HCl at pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 0.4 mM DTT and 10 % glycerol) to remove the excess salts. The proteins were then further purified using heparin sepharose. First the proteins were allowed to bind 2-3 mL of heparin sepharose (GE) for 1-2 hours. This was followed by two wash steps where 15 mL of wash buffer 2 (50 mM Tris-HCL at pH 7.5, 150 mM NaCl, 1 mM EDTA, 5 mM DTT and 10 % glycerol) was used to wash the beads. Then the beads were again transferred to columns and the proteins eluted using a second elution buffer (50 mM Tris-HCL at pH 7.5, 500 mM NaCl, 1 mM EDTA, 5 mM DTT and 10 % glycerol). The pooled fractions were analyzed on an SDS page gel, and the fractions that showed high protein

levels were pooled and dialyzed against dialysis buffer 2 (50 mM Tris-HCl at pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 0.4 mM DTT and 50 % glycerol). The proteins were then stored at -20 °C.

5.3.3 Preparation of RNAs with 2', 3' -cyclic phosphate ends

In order to generate RNAs with 2' 3' –cyclic phosphate ends, ribozyme reactions of the hammerhead ribozyme and the HDV ribozyme were performed. For both the hammerhead ribozyme and the HDV ribozyme, 50 µL reactions were set up which included 10 µM substrate and ribozyme strands and 50 mM Tris-HCL (pH 7.5). Reactions were heated to 90 ^{o}C for 2 min and cooled at room temperature for 10 min before 5 mM Mg^{2+} was added to initiate the cleavage of the substrate strands. The ribozyme mixtures containing cleaved products were ethanol precipitated and resuspended in water for subsequent use.

5.3.4 Ligation reactions

Ligation of the RNAs containing 2' 3' –cyclic phosphate termini to a preadenylated adapter 1 (5'- rAppCTGTAGGCACCATCAATGAT-Cy3) and preadenylated TruSeq adapter (5'- rAppNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGTCGCC ddC) were performed according to a previously described protocol (*297*). 1 µM ribozyme

reaction and 3 µM adapter oligo were incubated with 1 µM *At* tRNA ligase at 30 ^{o}C in the presence of 14 % polyethylene glycol and 1X tRNA ligase buffer. The 20 µL reactions were allowed to proceed for 3 hours before 10 µL of quench buffer were added. The reactions were separated on a 6-10 % denaturing polyacrylamide gels containing 6 M Urea. The gels were scanned on the Typhoon Imager by detecting the emission of Cy3 before being analyzed on ImageQuant 5.1 (Molecular Dynamics).

5.3.5 RNA extractions from *D. melanogaster, A. aegypti and A. gambiae*.

Total RNA was extracted from three organisms; *D. melanogaster, A. aegypti*, and *A. gambiae*. All organisms were harvested during larval phase of their development. For RNA extraction, Trizol reagent (Life Technologies) was used to homogenize the larvae in a Dounce homogenizer. Following homogenization, the aqueous phase which contains the total RNA was subjected to Zymo Research's Direct-zol^{TM} RNA MiniPrep Plus RNA purification kit as indicated by the manufacturer.

5.3.6 Detection of known RNAs with 2', 3' -cyclic phosphates using PCR

As mentioned in the introduction, a few RNAs in the cells contain a 2', 3' -cyclic phosphate at their 3' ends. Apart from the products of self-cleaving ribozymes, U6 snRNAs also contain a stable 2', 3' -cyclic phosphate. To show that such RNAs can be

detected by ligating them to a known adapter using *At* ligase, we decided to detect U6 snRNAs in the three organisms. In addition, a previously reported hammerhead ribozyme from *A. aegypti* was also the subject of this investigation (*222*).

Ligation reactions were assembled by adding 2 µg of total RNA to 3 µM adapter DNA oligo while the other ingredients were kept the same. For one of the reactions, RNA was heated to 90 ^{o}C for 2 min and cooled at room temperature for 10 min to ensure proper folding. After a 3 hr incubation at 30 ^{o}C, ligation reactions were separated on a 6 % denaturing polyacrylamide gel (7M urea) and ligated RNA (>20 nt) was excised and purified. The purified RNA was incubated for 3 hours with 10 µM Tpt1 in 1X Tpt1 buffer (20 mM Tris-HCl at pH 7.5, 5 mM $MgCl_2$, 2.5 mM spermidine, 0.1 mM DTT, 0.4% Triton X-100 and 1 mM NAD^{+}).

500 ng – 1 µg of Tpt1 RNA was used for the first-strand cDNA synthesis using Superscript III Reverse Transcriptase (Invitrogen). An adapter complimentary primer was used for cDNA synthesis. 2 uL of the reverse transcription reaction mixture was used for subsequent PCR analysis. To amplify U6 snRNA and the hammerhead ribozyme in *A. aegypti,* forward primers complimentary to these RNAs were used. Adapter complimentary primer served as the reverse primer. PCR amplification was performed by following the GoTaq Flexi DNA Polymerase (Promega) protocol. The amplified DNA was separated on a 1.8 % agarose gel and stained with Ethidium Bromide. The amplified DNA was excised out, purified, and cloned into pUC-19 vector, which was sent for low-throughput sequencing.

5.3.7 Library preparation for RNA-sequencing for self-cleaving ribozymes (Ribozyme Seq)

200 µg of total RNA from *A. aegypti* was incubated with 10 pmols of preadenylated Truseq adapter in a 20 uL reaction as described in prior ligation protocols with one exception; the Truseq adapter was also biotinylated on the 3' end. The ligation reactions were separated on a 6 % acrylamide gel and RNA fragments greater than 60 nt were excised and purified. Magnetic streptavidin beads (Dynabeads MyOne Streptavidin T1) were used according to the manufacturer's protocol to enrich for RNAs that were ligated to the Truseq adapter. Ligated RNAs were treated with Tpt1 and chemically fragmented using RNA fragmentation kit (Life Technologies). Following fragmentation, RNA was ethanol precipitated and reverse transcribed using Truseq adapter complimentary primer. Thermostable group II intron reverse transcriptase (TGIRT) was used for cDNA synthesis (*307*). Single stranded cDNA was then treated with shrimp alkaline phosphatase to remove the phosphorylated ends. The ethanol precipitated mixture was then treated with 10 mM sodium hydroxide to hydrolyze the RNA. Alternatively, 1 uL of RNase A and 1 uL of RNase H was used to degrade the free RNA and RNA-DNA hybrids. The cDNAs were then ligated to the Truseq index adapters (5'-NNNNNGATCGGAAGAGCACACGTCTGAACTCCAGTCACXXXXXXATCTCGTATGCCGTCTTCTGCTT3ddC) using T4 RNA ligase 1 (NEB). Index barcodes from Illumina index adapter 1-12 were used for this study. Ligated cDNAs were then separated on a 6 % denaturing acrylamide gel to select for cDNAs between ~100 nt and 400 nt. The purified cDNAs

were then PCR amplified using Phusion DNA polymerase (NEB). The amplified cDNA library was separated on a 6 % native acrylamide gel. DNA bands migrating slower than the adapter dimers were excised and purified. Final cDNA library was analyzed on Bioanalyzer and sequenced using 2 x 150 MiSeq.

5.3.8 Computational Analysis of the RNA-sequencing data.

The raw reads from the RNA-seq experiment were first filtered based on quality (less than Phred-20) followed by the removal of adapter sequences. Following processing, any reads that were less than 30 nt were also removed. These analyses were accomplished using the FASTX-toolkit and Trimmomatic (*308-310*). These processed reads were then mapped to the reference genomes using either Bowtie 2 or Burrows-Wheeler Aligner (BWA) (*311, 312*). The genomic assemblies used for *H. sapiens*, *A. aegypti*, *D. melanogaster*, and *A. gambiae* were GRCh38, AaegL3, BDGP6, and AgamP4 respectively. The genomic assemblies and annotation file (gff3) were downloaded from the Ensembl database (*313*).

The mapped reads were then analyzed using a number of different programs including HTSeq, Bedtools, Bioconductor, and BioPython (*314, 315*). Using these programs, first, the number of reads mapping to different regions were obtained, and then the mapped reads were used to generate a fasta formatted file that contained sequences belonging to all the peaks in the dataset. A peak was defined as any region that contained one or more mapped read. The extracted sequences were of length 500

nt, where 250 nt belonged to upstream region of the highest point of the peak and 250 nt to downstream region of the peak. These extracted sequences were then subjected to further analysis.

The extracted sequences were manipulated to filter out RNAs that are known to have 2', 3' –cyclic phosphates. Examples of such RNAs include U6 snRNA, tRNAs, and existing self-cleaving ribozymes. To do so, the genomic locations of extracted sequences were matched to available databases of these RNAs. Furthermore, to predict if the extracted sequences bore any resemblance to known RNAs, bioinformatics tools called Infernal and RNArobo were used to search for the secondary of structures of known RNAs (*316, 317*). Infernal uses covariance models of known RNAs to search for similar RNAs whereas RNArobo uses a structure key. Covariance models were downloaded from the Rfam database (*318*). Covariance models were used for U6 snRNA, tRNAs, CoTc ribozyme, CPEB3 ribozyme, glmS ribozyme, hairpin ribozyme, HDV ribozyme, twister ribozyme, twister sister ribozyme, VS ribozyme, pistol ribozyme, hatchet ribozyme and hammerhead ribozymes (type 1, 2, 3, 9 & 10). Various structure keys for the hammerhead ribozyme, HDV ribozyme, twister ribozyme, twister sister ribozyme, pistol ribozyme, hatchet ribozyme, and tRNA were used in this analysis. The structure keys reflected the conserved secondary structures of these RNAs.

After the known capped RNAs had been identified, the peaks were ranked based on their likelihood of being RNAs with cyclic phosphate ends. For this purpose, peaks that showed abrupt endings were selected. To identify such peaks, coverage per base was first calculated, and then regions where a sharp drop-off was observed were

selected. The sharp drop-offs were then ranked according to their magnitudes. The cleavage sites were then mapped based on the location of the sharp drop-offs.

5.3.9 Testing for self-cleavage activity

Self-cleavage activity was tested using a previously described protocol (*319*). First, primers were designed to amplify the sequence of a potential ribozyme from the genomic DNA. The primers were designed such that the final amplified DNA contained XbaI and HindIII sites on either ends, EarI site at the 3' end of the ribozyme sequence, and the T7 promoter sequence at the 5' end of the potential ribozyme. Taq (Promega) or Vent polymerase (NEB) was used to amplify the candidate ribozyme sequences. To prepare the DNA for run-off transcription, EarI was used to cut the DNA fragment followed by phenol chloroform extraction. Then, 100 uL *in-vitro* transcription reactions were performed in 5 mM Mg^{2+} to transcribe RNAs belonging to candidate ribozymes as previously described (*319*). The resulting RNAs were purified using a 6 % denaturing acrylamide gel. Next, the RNAs were tested for self-cleavage activity by preparing 18 µL reactions containing 0.5 µM RNA and 50 mM Tris-HCl (pH 7.5). These reactions were heated to 90 °C to refold the RNA and then cooled for 10 minutes at room temperature. Then, 2 µL of 100 mM $MgCl_2$ was added to initiate cleavage and the reactions were incubated at 37 °C for 1-2 hours. Reactions were then separated on 6-10 % acrylamide gels containing 6 M urea. Gels were then stained with SYBR-Gold (Thermo Fisher) and analyzed on the Typhoon Imager using an excitation wavelength of 495 nm.

5.4 Results and Discussion

5.4.1 *At* tRNA ligase can successfully ligate preadenylated DNA adapters to RNAs containing 2', 3' -cyclic phosphates

The products of ribozyme cleavage reactions are 5'-hydroxyl and 2', 3' –cyclic phosphate. In order to capture the products of the cleavage reaction, we utilized a previously described method of using *At* tRNA ligase to ligate 2', 3' –cyclic phosphate to a known adapter (*297*). The ability of *At* tRNA ligase to perform such chemistry was tested on two different substrates of different ribozymes. As can be seen in Figure 5.1, both RNAs were successfully ligated to an adapter.

The HDV ribozyme has been described previously, and it produces a 5' cleavage product which is 3 nt long (*186*). The mobility of the product, however is slower than the substrate which can be attributed to change in charge-mass ratio of the cleaved product. Lanes 6-10 in Figure 5.1A confirm that when 5'-cleaved HDV product is added to a mixture containing a preadenylated adapter, *At* tRNA ligase can successfully ligate the two oligos. Ligation reaction takes approximately 3 hrs to reach completion. Similarly, the hammerhead ribozyme's 5' cleavage product is 6 nt long and *At* tRNA ligase is able to ligate it to the adapter only when the reaction contains cleaved products, adapter, and tRNA ligase (Figure 5.1B). The hammerhead ribozyme product seems to be prone to hydrolysis of the cyclic phosphate to a 3'-phosphate leading the

appearance of a new band below the cleaved product. As part of its mechanism, *At* tRNA ligase can only accept intermediates with a 2’ phosphate (*320*).

The same procedure was used to test whether the preadenylated Truseq universal adapter can be ligated to the products of the hammerhead ribozyme. Figure 5.1C shows that when the Truseq adapter is incubated with the cleaved hammerhead ribozyme substrate in the presence of *At* tRNA ligase, a high molecular weight band appears which is consistent with adapter being ligated to the hammerhead ribozyme product. These results suggest that *At* ligase can be used to select for RNAs that contain a 2’, 3’ -cyclic phosphate at their 3’ end.

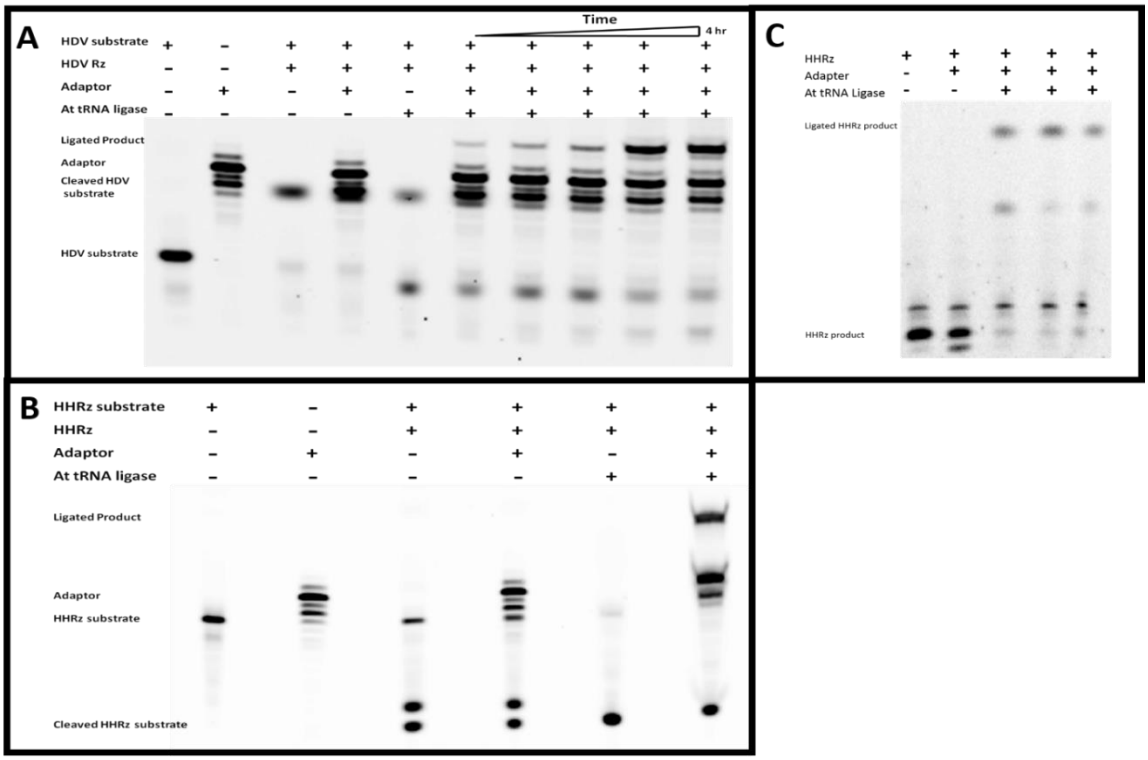

Figure 5.1. *At* tRNA ligase ligates RNAs with 2'-3' cyclic phosphates to preadenylated DNA adapters.
2' 3' –cyclic phosphates containing RNAs were generated using either an HDV ribozyme (**A**) or HHRz (**B & C**). HDV ribozyme reaction produces 2',3'-cyclic phosphate containing RNA of lengths 3 nt. HHRz generates a product of 6 nt. The 2' 3' –cyclic phosphate containing product in (**A**), (**B**), and (**C**) is DY547 labelled and so is the adapter DNA in **A** and **B**. In **C,** the Truseq Universal adapter was used. Ligation reactions containing adapter DNA, ribozyme cleavage products, and *At* tRNA ligase were incubated at 30 ^{o}C for 3 hours. Reaction mixtures were separated on a denaturing 6% polyacrylamide gel and visualized using Typhoon Imager.

5.4.2 2', 3' -cyclic phosphate containing RNAs within total RNAs of *D. melanogaster* and *Aedes aegypti* can be successfully detected by PCR analysis

In order to show that our biochemical screen is functional, we decided to look for RNAs containing 2', 3' –cyclic phosphates in two organisms which are readily

available. U6 snRNA is known to have a stable 2', 3' –cyclic phosphate at its 3' end. Therefore, it would provide us with a positive control for the biochemical screen that we wish to employ. Figure 5.2A demonstrates that a band consistent with the size of ligated U6 snRNA was amplified. Sequencing confirmed that the sequence of amplified DNA fragment matched that of U6 snRNA attached to the DNA adapter. Similarly, U6 snRNA was also detected in *A. aegypti* larvae (Figure 5.2A).

To further ensure that our method can detect cleaved products of self-cleaving ribozymes, we detected the hammerhead ribozyme in the total RNA isolated from the larvae of *A. aegypti.* Previously, it was revealed that the *A. aegypti* genome has numerous occurrences of the hammerhead ribozyme-like motifs (*321*). Using PCR analysis, we looked for one such hammerhead ribozyme. The PCR amplification of this the hammerhead ribozyme revealed a band on the gel consistent with the length of the cleaved the hammerhead ribozyme RNA ligated to an adapter (Figure 5.2B). The DNA fragment was then confirmed by sequencing (not shown).

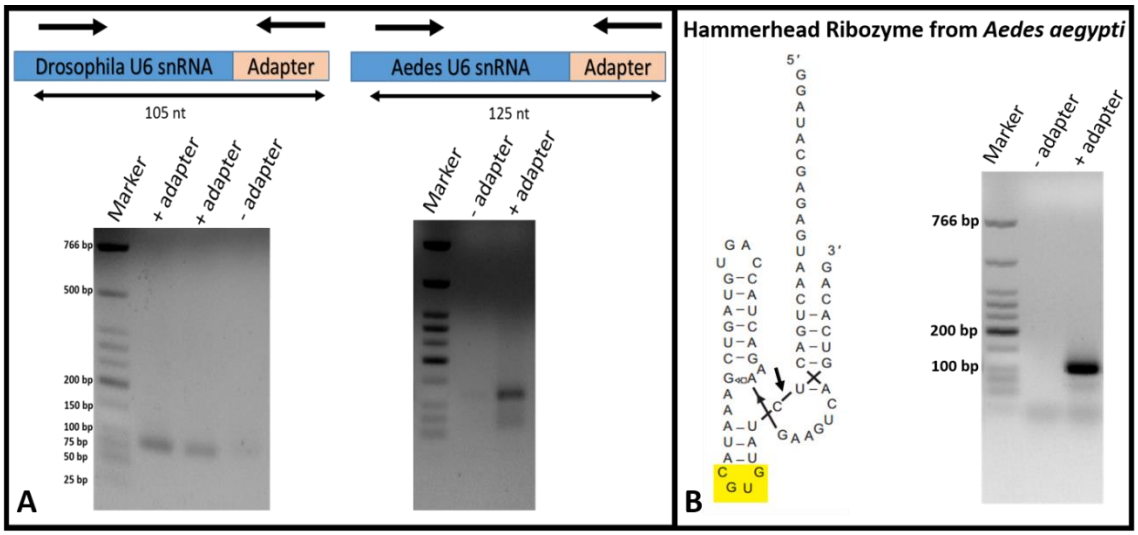

Figure 5.2. *At* tRNA ligase is capable of ligating 2' 3' –cyclic phosphate containing RNAs present in *D. melanogaster* and *Aedes aegypti*.
Total RNA was extracted from *D. melanogaster* and *A. aegypti* and incubated with adapter DNA 1 and *At* tRNA ligase. The resulting ligated RNAs were gel purified, reverse transcribed, and then subjected to PCR analysis. Primers were used to amplify ligated U6 SNRNAs (**A**) or ligated cleaved HHRz (**B**). Gels were visualized using Ethidium bromide and visualized under UV light.

5.4.3 Analysis of the MiSeq RNA sequencing results

Three different organisms were used to generate cDNA libraries for MiSeq sequencing: *A. aegypti*, *D. melanogaster*, and *A. gambiae*. The details of the protocols are mentioned in the methods. In addition, the dataset from Schutz et al., was also used for comparison (*297*). Prior to sequencing the libraries were analyzed on Bioanalyzer (Agilent) and the traces generated are shown in Figure 5.3. After size selection of the DNA fragments between 150 and 400 bp the library still contained a number of adapter and primer dimers. These dimers were located around 150 bp had to be removed using AMPure (Beckman Coulter) purification. While the adapter dimers removal was

successful, there was a noticeable loss of DNA fragments greater than 150 bp especially in *A. gambiae*, and *D. melanogaster* (Figure 5.3, right). This suggested that AMPure purification and native gel purification might not be the best strategies for adapter removal. Therefore, for future libraries, a denaturing polyacrylamide gel will be used for size selection. If necessary, multiple rounds of denaturing purification will be used instead of AMPure purification.

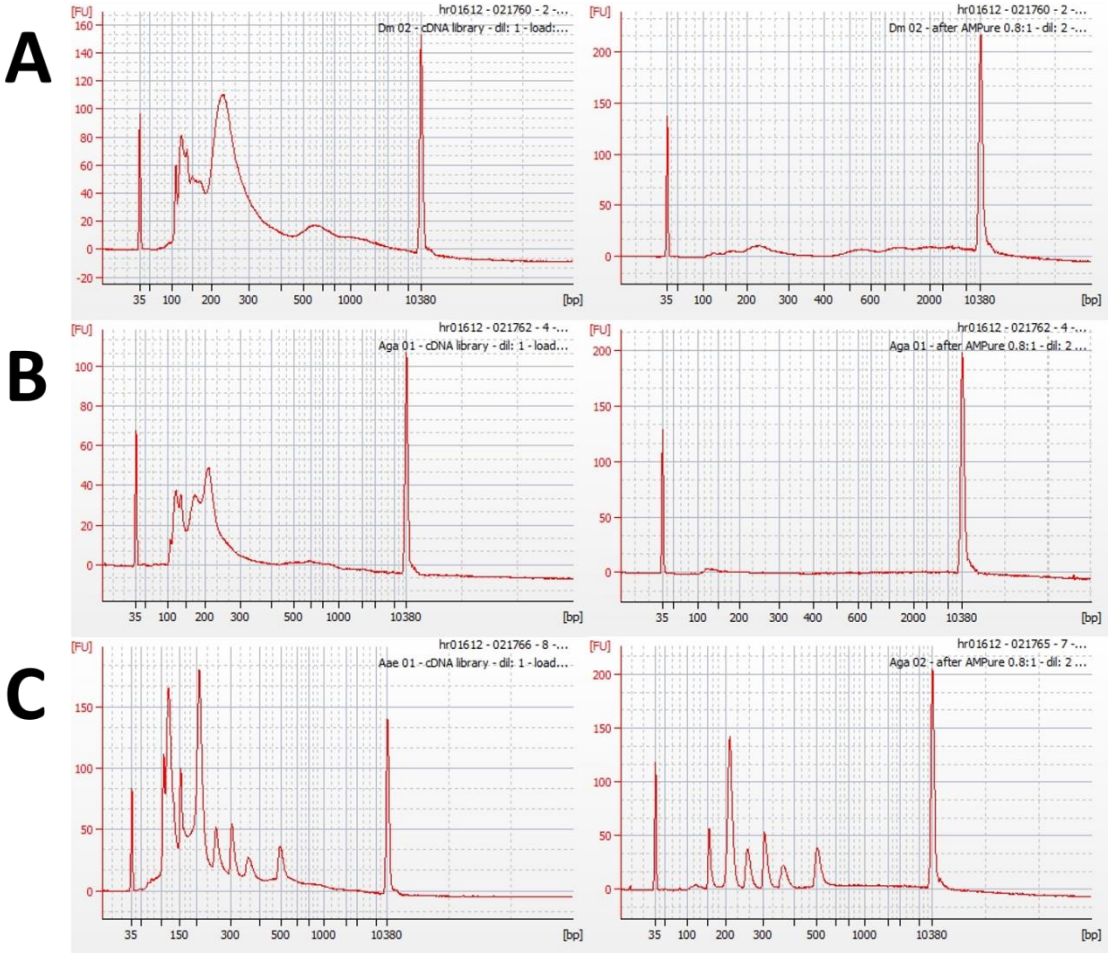

Figure 5.3. Bioanalyzer traces of the MiSeq libraries.
The size distribution of the DNA fragments in the MiSeq libraries from *D. melanogaster* (**A**), *A. gambiae* (**B**), and *A. aegypti* (**C**) are shown. Libraries prior to AMPure purification are shown on the **left** and after AMPure purification are shown on the **right**.

The number of reads generated from the sequencing runs for these RNAs are summarized in Table 5.2. The number of total reads generated from the MiSeq run were all less than 1 million. Almost ~50% of these reads contained adapter and primers sequences, which suggests that the RNA-seq library preparation protocol needs to be improved further. The problems may lie in the final selection step where a native gel was used to excise DNA fragments between ~150 bp and 400 bp. Purification through multiple rounds of separation on a denaturing polyacrylamide gel might yield better results. Alignment of the reads to reference genomes was performed through Bowtie 2. However, for *A. aegypti,* BWA yielded better alignment compared to Bowtie 2. For *D. melanogaster* and *A. aegypti*, paired-end alignment was performed. The percentage of reads that aligned in *A.* aegypti, D*. melanogaster*, and *A. gambiae* were 33 %, 17 % and 14 % respectively. The *H. sapiens* dataset was generated from HiSeq sequencing and generated only single end reads, which were 36 bp long. The majority of these reads successfully mapped to the HG38 assembly of the reference genome. About 78 % of the reads successfully aligned to the reference genome. Higher success, in this case might be due to a cleaner library preparation protocol, or due to shorter single end reads. In the MiSeq libraries, the majority of the reads mapped to regions that were annotated as repeat regions. More than 90 % of the mapped reads in *A. gambiae* and *D. melanogaster* aligned to repeat regions, whereas 41 % of the mapped reads in *A. aegypti* belonged to repeat regions. Whether these repeat regions harbor self-cleaving ribozymes or if this is a result of incorrect mapping remains to be seen.

Table 5.2. The number reads related to MiSeq experiments, and downstream processing.
The reads belonging to *D. melanogaster*, *A. aegypti*, and *A. gambiae* were obtained from this study, whereas the *H. sapiens* reads were obtained from a different study which used Illumina 1x36 HiSeq (*297*). QC stands for quality control processing of the raw reads as mentioned in the methods.

	D. melanogaster	***A. gambiae***	***A. aegypti***	***H. sapiens***
Run type	MiSeq 2x150	MiSeq 2x150	MiSeq 2x150	HiSeq 1x36
Genome assembly	BDGP6	AgamP4	AaegL3	GRCh38
Aligner	Bowtie 2	Bowtie 2	BWA	Bowtie 2
Total reads	737922	769584	230160	5716254
Total reads after QC	42086	77034	99188	4472289
Total aligned reads	6164	10792	32791	3469713

5.4.3.1 RNAs containing 2', 3' -cyclic phosphates were enriched over other RNAs

To validate whether our library preparation protocol specifically selected for RNAs containing 2', 3' -cyclic phosphates, the number of reads mapping to U6 snRNAs were analyzed. U6 snRNAs are known to contain 2', 3' -cyclic phosphates at their 3' ends, and are important for their function inside cells (*305*). As internal controls, the number of reads mapping to U5 and U2 snRNAs were also calculated. U5, U2, and U6 are all part of the spliceosomal complex, and are expected to be present in the same molar concentrations within the cells (*322*). If the library preparation protocol developed here specifically targets RNAs containing 2', 3' -cyclic phosphate containing RNAs, then U6 snRNA would be expected to be enriched over U5 and U2 snRNAs. As Figure 5.4 shows, more reads map to U6 snRNAs compared to U5 and U2 snRNAs. The

number of reads mapping to U6 snRNA in *D. melanogaster*, *A. gambiae*, and *A. aegypti* are 200, 150, and 457 respectively. On the other hand, the number of reads mapping to U5, and U2 snRNAs are less than 10 in all three organisms. These results suggest that RNAs that contain a stable 2', 3' -cyclic phosphate termini are enriched in the libraries that we prepared. As a control, we also calculated the number of reads mapping to U6, U5 and U2 snRNAs in the human brain cells library. Surprisingly, we do not see the enrichment of U6 snRNAs over U5, and U2 snRNAs. This might be because Schutz et al. did not enrich ligated RNAs using streptavidin. Furthermore, this demonstrates the importance of an enrichment step in the library preparation protocol. Since RNAs containing 2', 3' -cyclic phosphates may constitute less than 5 % of the total RNA, it is important to enrich these RNAs, and remove RNAs that do not possess cyclic phosphates.

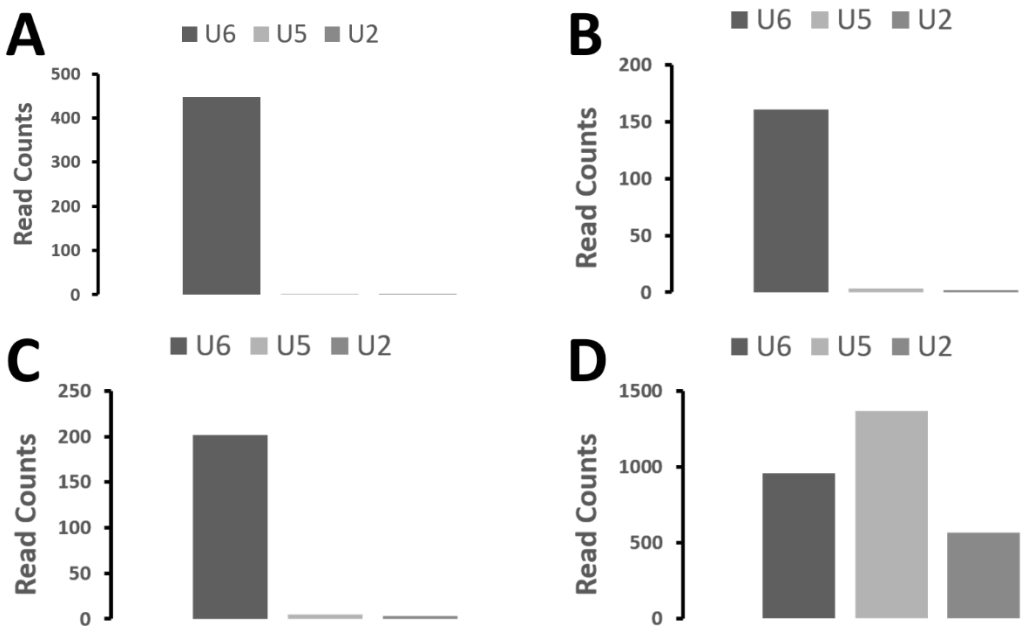

Figure 5.4. Comparison of reads mapping to U6 snRNA vs. U5, and U2 snRNAs.
The number of reads mapping to U6, U5 and U2 snRNAs was calculated for *A. aegypti* (**A**), *A. gambiae* (**B**), *D. melanogaster* (**C**), and *H. sapiens* (**D**). U5 and U2 snRNAs acted as internal controls since they do not generally possess 2', 3' -cyclic phosphate ends.

5.4.3.2 The majority of the previously found self-cleaving ribozymes were not detected in the MiSeq libraries.

To further validate if the library preparation protocol that we developed can detect self-cleaving ribozymes, the number of reads mapping to any of the previously identified self-cleaving ribozymes was calculated. *A. aegypti* is known to contain over 4000 copies of self-cleaving ribozymes (*222*). Similarly, several hundred self-cleaving ribozymes are also found in *A. gambiae (175)*, whereas a few self-cleaving ribozymes have also been found in humans, and *D. melanogaster (323)*. Surprisingly, we were

unable to detect the majority of these self-cleaving ribozymes in all the datasets. As Table 5.3 shows, only 5 previously found sequences belonging to self-cleaving ribozymes were detected. In the case of the *H. sapiens* library, only two previously described ribozymes called the CoTc, and the CPEB3 had one read mapping to them. Failure to detect previously found self-cleaving ribozymes could be attributed to a number of reasons. The most likely reason is that the products of ribozyme cleavage may be present in much lower concentrations within the total RNA compared to RNAs that contain a stable 2', 3' -cyclic phosphate end. U6 snRNA is protected and shielded by the spliceosome complex proteins. However, other cleaved RNAs may be subject to degradation by the cellular ribonucleases. In order to detect these self-cleaving ribozymes, we may need to sequence deeper. For this reason, a new *A. aegypti* library was sequenced using HiSeq to generate several million reads, and will be described later. Another reason could be that these previously found ribozymes are either not transcribed or do not self-cleave *in-vivo*. Many of the previous studies that discovered these self-cleaving ribozymes only tested them *in-vitro*; whether they cleave *in-vivo* is unclear.

Table 5.3. Number of previously found self-cleaving ribozyme motifs found in the MiSeq sequencing results.
The table shows the number motifs to which one or more reads mapped.

Organism	Number of motifs detected
D. melanogaster	0
A. gambiae	0
A. aegypti	5
H. sapiens	2

5.4.3.3 Several peaks in the RNA-seq datasets are predicted to be known self-cleaving ribozymes

While we were unable to find the majority of the previously described sequences of the self-cleaving ribozymes, we hypothesized that we may be able to find new motifs belonging to the existing self-cleaving ribozymes in the RNA-seq datasets. To predict motifs that resemble known self-cleaving ribozymes, we used both structure based and conservation based computational searches. RNABob and RNArobo use structure based search, whereas Infernal uses covariance models built from multiple sequence alignments of existing self-cleaving ribozymes to search for similar motifs (*316, 317, 324, 325*). We searched for all nine self-cleaving ribozyme classes and their known variants. As Figure 5.5 shows, the *H. sapiens* library contained the most self-cleaving ribozymes, including over 500 hammerhead ribozyme motifs and 21 HDV ribozyme motifs. Since the *H. sapiens* library did not show an enrichment of U6 snRNA over U5 and U2 snRNAs, many of these predicted ribozymes might be false positives. We tested 13 of these

motifs which had more than 10 reads mapping to them for self-cleavage activity, and all of them failed to self-cleave *in-vitro*. It shows that most of the peaks from *H. sapiens* library might be non-specific RNAs. In *A. aegypti*, 19 motifs belonging to the hammerhead and the HDV ribozymes were found. Lastly, *D. melanogaster* and *A. gambiae* contained 1-2 motifs belonging to the HDV and/or the hammerhead ribozymes. None of these motifs have been found before in these organisms, and hence, are new findings. These motifs are predicted to be self-cleaving ribozymes because our libraries are enriched for RNAs containing 2', 3' -cyclic phosphate containing RNAs. However, they have not been tested for self-cleavage activity *in-vitro*. It is also possible that these RNAs are a result of non-enzymatic RNA cleavage, or non-ribozyme mediated cleavage, such as, via ribonucleases.

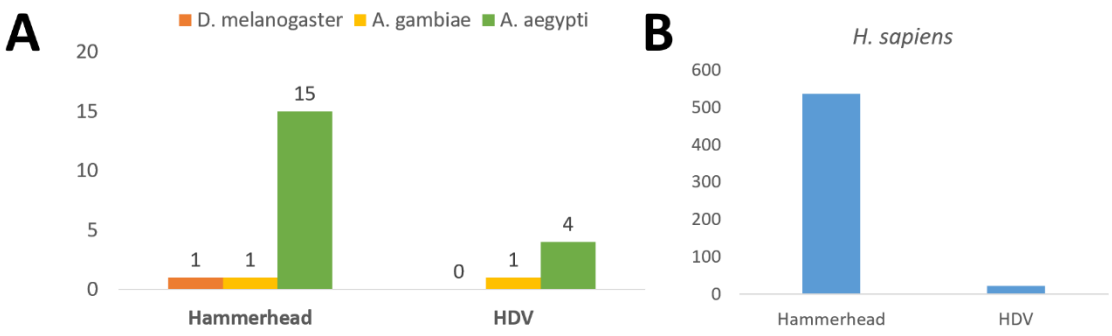

Figure 5.5. Predicted known self-cleaving ribozymes.
Using RNArobo, and Infernal programs, extracted peaks were searched to find motifs similar to known self-cleaving ribozymes as described in methods. Motifs found in *A. aegypti*, *A. gambiae* and *D. melanogaster* are shown in **A** whereas motifs found in *H. sapiens* are shown in **B**.

Eikbush et al., and others have previously reported that many *Drosophila* species tend to have HDV ribozymes in the 5’ UTR regions of R2 retrotransposons (*296, 323*). These R2 elements are found inserted within the 28S rRNAs. In *D. melanogaster*, 12 different peaks corresponded to R2 retrotransposons. Of these, 4 showed some similarities to the HDV ribozyme. These sequences contained the HDV ribozyme core, but were unable to fold like an HDV ribozyme. Hence, we hypothesize that these might be different variants of the HDV ribozyme, and therefore will need to be tested for self-cleavage activity.

5.4.3.4 Overview of the top 100 peaks resulting from self-cleavage of RNA

Our library preparation protocol is able to specifically select for RNAs containing cyclic phosphates on their 3’ ends. However, it is possible that other abundant RNAs within the transcriptome may also be present in our sequencing libraries. To filter such RNAs and specifically detect RNAs that resulted from ligation of 2’, 3’ -cyclic phosphate containing RNAs to Truseq adapter, we only selected peaks that showed a sharp drop-off on one end (Figure 5.6A). The sharp drop-off position signifies that the Truseq adapter ligated to RNA at that location. The sharp drop-off point is where the RNA is likely to have a 2’, 3’ -cyclic phosphate. This is evident from Figure 5.6A where number of reads mapping to U6 snRNA end abruptly at the 3’ end of the U6 snRNA. The 3’ end of the U6 snRNA contains the 2’, 3’ -cyclic phosphate. To select peaks that show abrupt

endings, an algorithm was developed using shell scripting (Figure 5.6B). This algorithm detected such sharp drop-offs in the read coverage and ranked them according to the magnitudes of the drop-offs.

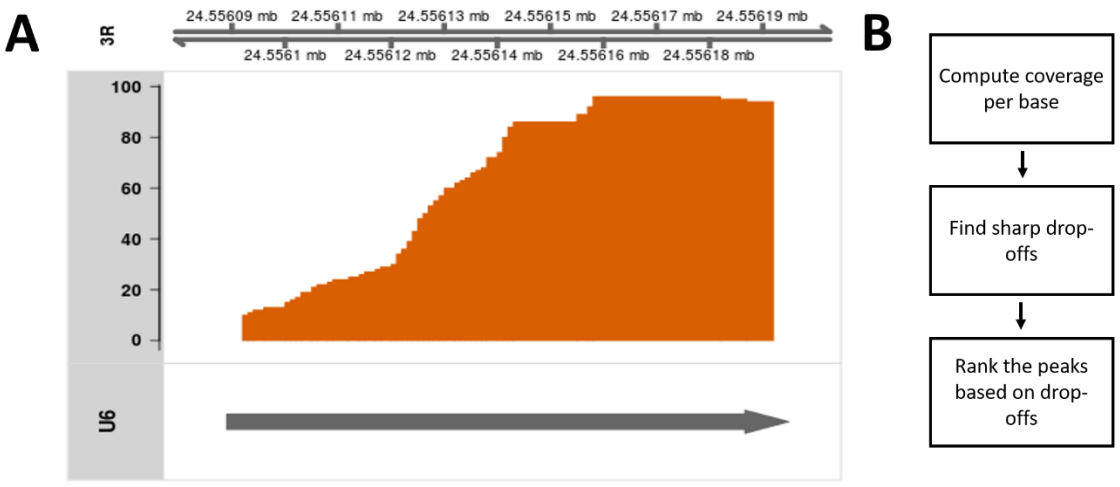

Figure 5.6. Peaks resulting from RNAs containing 2', 3' -cyclic phosphates.
A. The coverage graph of the U6 snRNA from *D. melanogaster* is shown as an example of a peak resulting from RNAs containing 2', 3' -cyclic phosphate ends. Such peaks are characterized by sharp drop-offs. Top panel shows the chromosomal location and size of the peak, and the bottom panel shown the name and the orientation of the annotation. Y-axis of the coverage graph shows the number of reads. **B**. A simple shell script was written to detect sharp drop-offs similar the one shown in **A**. The script included the steps shown here.

The top 100 peaks containing sharp drop-offs are summarized in Table 5.4. These top peaks were subjected to BLAST (Basic Local Alignment Search Tool) searches to find any similarities to known RNAs within these organisms (*326*). Many of these top peaks belong to either rRNAs or intergenic and unannotated regions. It is unclear whether the sequences belonging to rRNAs are due to ribozyme cleavage or non-specific RNA

cleavage. Even non-specific cleavage in rRNAs can generate top peaks using our RNA-seq protocol due to the overabundance of rRNA in cells. However, the majority of the top peaks from *A. aegypti* did not have any BLAST matches. It demonstrates that there is a large variation between different assemblies of the *A. aegypti* genomes. The two top peaks from *D. melanogaster* and *A. gambiae* are shown in Figure 5.7, and their corresponding secondary structures are shown in Figure 5.8. These peaks might be self-cleaving RNAs, since the shape of peaks is very similar to the U6 snRNA peak. In *A. gambiae*, both the peaks are annotated as repeat regions but BLAST search reveals that the sequences resemble rRNA sequences. It is interesting to note that one of the peaks from Figure 5.7A (right) came from three different RNAs containing 2', 3' -cyclic phosphates. These might be three different ribozymes embedded next to each other. Ribozymes do tend to get inserted into repeat regions, and several times are also found embedded close to each other. Both the top peaks from *D. melanogaster* came from chromosome X, and were annotated as rRNAs (Figure 5.7B). However, the peak corresponding to pre-rRNA of 28S rRNA might be an important candidate for self-cleaving ribozyme because it may have a role in processing of the rRNA. The top two peaks from *A. aegypti* are not annotated with any feature. However, one of the peaks in Figure 5.7C (left) is found 300 bp upstream of the 5.8S rRNA. However, a BLAST search against the national center for biotechnology information (NCBI) nucleotide database showed a greater than 99 % sequence similarity to sequences from the genomes of a few bacterial species including *Limnohabitans* and *Acidovorax (327)*. A general BLAST search for most of the 80 top peaks from *A. aegypti* revealed a similar result (not

shown). Either these sequences are unidentified parasitic sequences that inserted themselves in the *A. aegypti* genome or they may belong to microbes in *A. aegypti* gut or environmental microbiome. The peak on the right in Figure 5.7B had an 86 % sequence similarity to a long-interspersed retrotransposon (LINE) called Juan-A (*328, 329*).

To further ensure that the top peaks resulted from self-cleavage activity of RNA, the locations of the cleavage sites within the predicted secondary structures of the RNAs from the top peaks were identified (Figure 5.8). The secondary structures were predicted using mfold, and only the structures with the lowest free energies were considered (*330*). This analysis showed that the predicted cleavage sites of these peaks are located within the non-Watson-Crick base paired regions of the RNA (Figure 5.8). Nucleobases unconstrained from the Watson-Crick base pairing allow the RNA to adopt the in-line geometry which is important for RNA to self-cleave (*164*). It is to be seen whether these top peaks resulted from self-cleavage activity of ribozymes or cleavage from other processes. They will need to be tested for self-cleavage activity *in-vitro*.

Table 5.4. An overview of the top 100 peaks that have sharp drop offs.
The algorithm used to identify these peaks is shown in Figure 5.6. After the peaks were identified, the sequences corresponding to the peaks were subjected to BLAST searches. In this table peaks are organized by the top BLAST matches.

	D. melanogaster	*A. gambiae*	*A. aegypti*	*H. sapiens*
No BLAST match	-	-	80	-
rRNA	34	71	0	26
ncRNA	0	0	3	0
mRNA	0	28	8	15
Predicted Ribozymes	1	0	0	0
Other/Unannotated	65	1	9	59

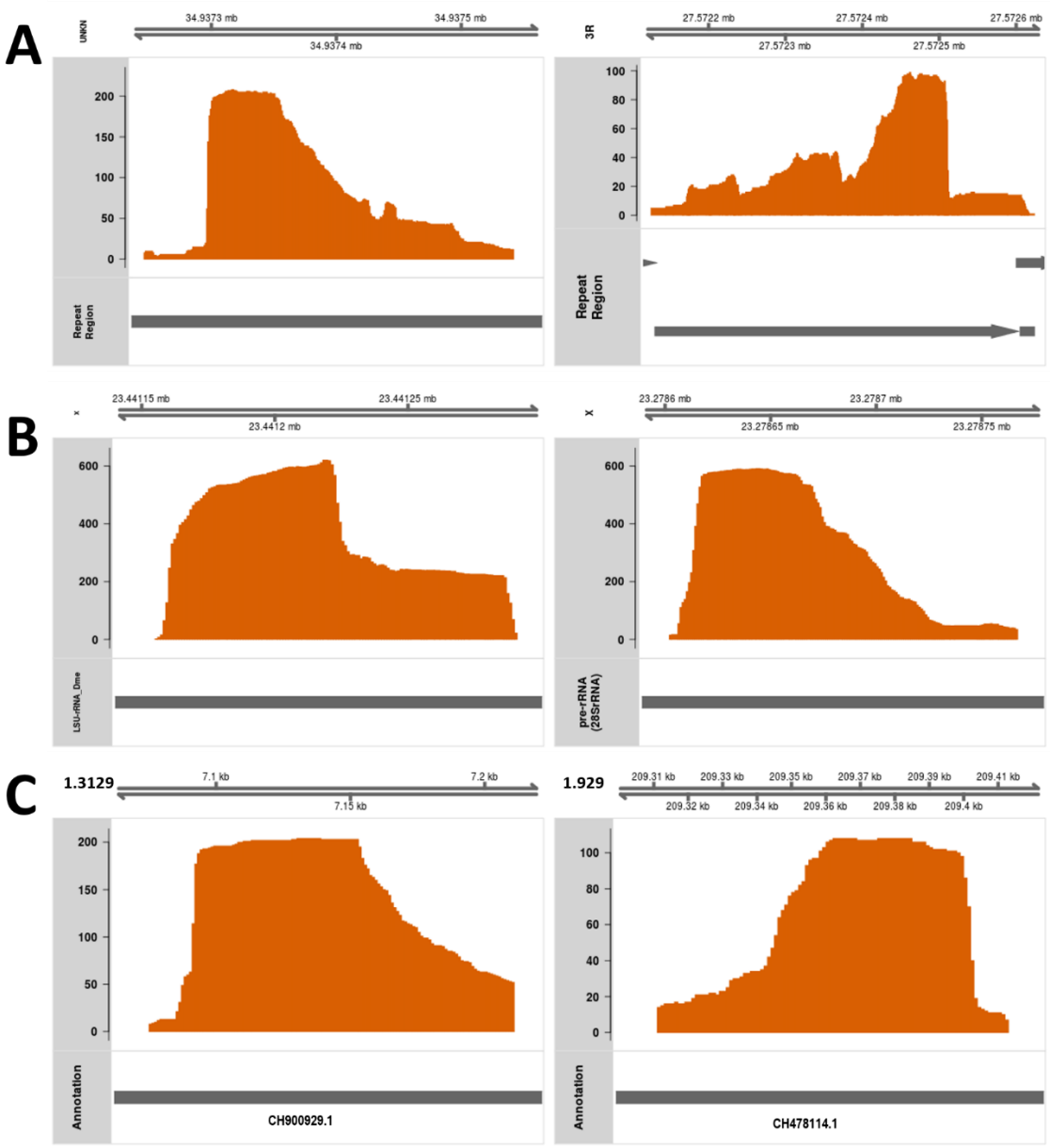

Figure 5.7. Top two peaks corresponding to potential ribozymes.
The figure shows the coverage graphs of the top two peaks generated from sharp drop-off algorithm analysis for *A. gambiae* (**A**) *D. melanogaster* (**B**) and *A. aegypti* (**C**). In A, the peak on the left came from chromosome annotated as UNKN, and the peak from the right came from chromosome 3R, whereas in B, both peaks are from chromosome X. In C, the peaks belong to supercontigs 3129 and 929 respectively. Top panels in each graph show the chromosomal locations and sizes of the peak, and the bottom panels shown the names and the orientations of the annotations. Y-axes of the coverage graphs show the total number of reads.

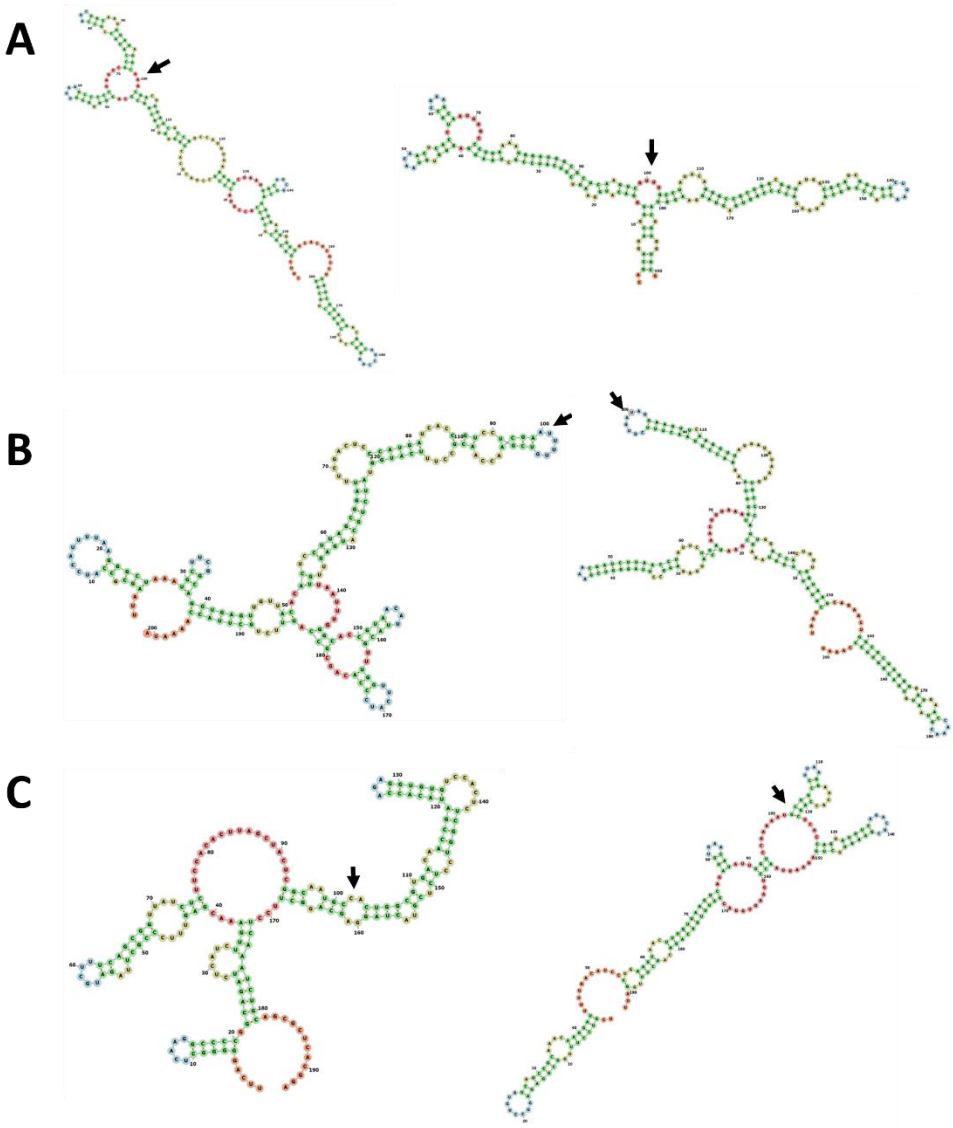

Figure 5.8. The cleavage sites of the RNAs belonging to top two peaks are located in single stranded regions.
Secondary structures of the top two peaks from *A. gambiae* (**A**) *D. melanogaster* (**B**) and *A. aegypti* (**C**) were predicted using mfold (*330*) and the resulting structure was drawn using forna (*331*). Arrows denote the site of cleavage based on the position of sharp drop-offs.

5.4.4 Preparation of *A. aegypti* libraries for HiSeq sequencing

The MiSeq libraries for *A. aegypti*, *D. melanogaster*, and *A. gambiae* did not contain many of the known self-cleaving ribozymes. Hence, we decided to make a new library for *A. aegypti* for 2x100 HiSeq sequencing. The objective was to sequence deeper to identify self-cleaving RNAs, as they may be present at very low concentrations. The library preparation protocol for this library was the same as before with one exception; the final size selection was performed using denaturing polyacrylamide gel, and the DNA fragments greater than ~200 bp were excised. Below 200 bp, several adapter and primer dimers were observed. This method of size exclusion appears to eliminate most of the adapter dimers (Figure 5.9). Two final libraries were prepared. In one of them, the upper size limit was ~500 bp, which means the DNA fragments between 200 bp and 500 bp were excised. The second library contained DNA fragments greater than 200 bp. No upper limit was established for the second library. This second library shows a much lower amount of primer and adapter dimers compared to the first library (Figure 5.8B). Hence, we proceeded with deep sequencing of the second *A. aegypti* library. At the time of writing this thesis, the RNA-seq results had not been obtained.

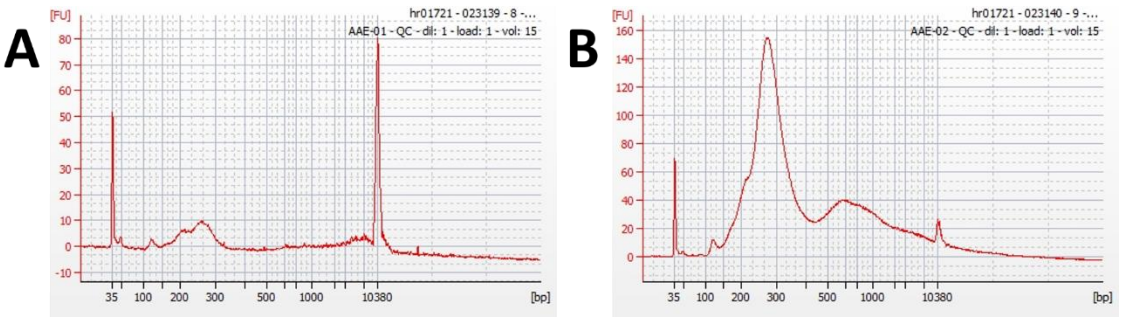

Figure 5.9. Bioanalyzer traces of *A. aegypti* libraries prepared for HiSeq sequencing. **A**. Final size selection after PCR enrichment included excision of DNA fragments between ~200-500 bp. **B**. Final size selection involved excision of DNA fragments greater than 200 bp. It is clear that the second library contains a lot more DNA averaging in length around 300 bp, and it also contains less than 5 % of primer dimers (~125 bp). For these reasons library B was used for HiSeq sequencing.

5.4.5 Conclusions

We have established a reliable library preparation method to capture RNAs that contain 2', 3' -cyclic phosphates. While the original protocol was based on the study by Schutz et al., significant changes have been made to the method (*297*). Some of the major changes include significantly increasing the amount of starting total RNA, enrichment for ligated RNAs, using a more thermostable reverse transcriptase, and paired end sequencing with longer reads. These changes have enabled us to enrich for RNAs containing 2', 3' -cyclic RNAs present in the total RNAs. An example of such an RNA is U6 snRNA. U6 snRNA is one of the top peaks in the MiSeq libraries. Several previously described self-cleaving ribozymes were also predicted to be among the RNAs in our libraries. However, most of the previously found self-cleaving ribozymes were not

found in our libraries, suggesting that either they are not functional *in-vivo* or they are present in very low concentrations. While deeper sequencing may be able to detect these self-cleaving ribozymes, further changes to the RNA-seq library preparation protocol may be necessary. It is likely that cellular machinery uses different mechanisms to recycle or further process ribozyme-cleaved RNAs. This may explain why protected RNAs containing 2', 3' -cyclic phosphates are present in higher concentrations, and unprotected RNAs such as the cleaved RNAs were barely detected. Hence, another important step may be needed that enriches for these ribozymes cleaved RNAs by using mutant organisms. The mutations in these organisms should be aimed at cellular machinery that processes RNAs containing 2', 3' -cyclic phosphates. Such an approach will stabilize the cleaved RNAs concentrations, and may allow easier detection. Good candidates for such mutations involve the exosome complex and/or other 3'-5' exoribonucleases. Another approach is to induce self-cleavage of ribozyme sequences in the total RNA prior to *At* tRNA ligation. Cleavage can be induced by refolding total RNA by heating and cooling, followed by addition of divalent and monovalent metals. Such an approach may not find functional ribozymes since the cleavage is induced, however, it may identify most RNAs that can self-cleave *in-vitro*. A recent study showed that RNAs that only contain a 3'-phopshate can be depleted through treatment with calf intestinal phosphatase (CIP) and T4 polynucleotide kinase (T4 PNK) followed by periodate oxidation (*332*). A similar method could be used to deplete RNAs that do not contain 2', 3' -cyclic phosphates. Such an approach can remove majority of the RNAs with 3'-phopshates, and may improve the ligation efficiency of RNAs containing 2',3' –cyclic

phosphates to the Truseq adapter using *At* tRNA ligase. Lastly, self-cleavage testing of the predicted known self-cleaving ribozymes, and detection of previously identified motifs of self-cleaving ribozymes will inevitably allow to develop smarter algorithms to discover new self-cleaving ribozymes from the RNA-seq datasets. To design a smarter algorithm, parameters, such as the shape of the peaks that belong to previously identified or predicted known self-cleaving ribozymes, their genomic locations, and the total number reads mapping to them can be used to discover novel RNAs that can self-cleave.

REFERENCES

REFERENCES

1. Gilbert, W. (1986) Origin of life: The RNA world, *Nature 319*, 618-618.
2. Crick, F. H. C. (1968) The origin of the genetic code, *Journal of Molecular Biology 38*, 367-379.
3. Orgel, L. E. (1968) Evolution of the genetic apparatus, *Journal of Molecular Biology 38*, 381-393.
4. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena, *Cell 31*, 147-157.
5. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme, *Cell 35*, 849-857.
6. Lehman, N. (2015) The RNA world: 4,000,000,050 years old, *Life (Basel) 5*, 1583-1586.
7. Alberts B, J. A., Lewis J, et al. . (2002) The RNA world and the Origins of Life, In *Molecular Biology of the Cell* 4 ed., New York: Garland Science.
8. Altman, S. (2011) Ribonuclease P, *Philosophical Transactions of the Royal Society B: Biological Sciences 366*, 2936-2941.
9. Fox, G. E. (2010) Origin and Evolution of the Ribosome, *Cold Spring Harbor Perspectives in Biology 2*, a003483.
10. Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., and Yusupov, M. (2012) One core, two shells: bacterial and eukaryotic ribosomes, *Nat Struct Mol Biol 19*, 560-567.
11. Steitz, T. A., and Moore, P. B. RNA, the first macromolecular catalyst: the ribosome is a ribozyme, *Trends in Biochemical Sciences 28*, 411-418.
12. Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) The Structural Basis of Ribosome Activity in Peptide Bond Synthesis, *Science 289*, 920-930.
13. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution, *Science 289*, 905-920.
14. White, H. B., 3rd. (1976) Coenzymes as fossils of an earlier metabolic state, *J Mol Evol 7*, 101-104.
15. Connell, G. J., and Christian, E. L. (1993) Utilization of cofactors expands metabolism in a new RNA world, *Orig Life Evol Biosph 23*, 291-297.

16. Keefe, A. D., Newton, G. L., and Miller, S. L. (1995) A possible prebiotic synthesis of pantetheine, a precursor to coenzyme A, *Nature 373*, 683-685.
17. Jauker, M., Griesser, H., and Richert, C. (2015) Spontaneous Formation of RNA Strands, Peptidyl RNA, and Cofactors, *Angewandte Chemie (International Ed. in English) 54*, 14564-14569.
18. Roth, A., and Breaker, R. R. (1998) An amino acid as a cofactor for a catalytic polynucleotide, *Proceedings of the National Academy of Sciences 95*, 6027-6031.
19. Macheroux, P., Kappes, B., and Ealick, S. E. (2011) Flavogenomics--a genomic and structural view of flavin-dependent proteins, *The FEBS Journal 278*, 2625-2634.
20. Kwok, C. K., Tang, Y., Assmann, S. M., and Bevilacqua, P. C. The RNA structurome: transcriptome-wide structure probing with next-generation sequencing, *Trends in Biochemical Sciences 40*, 221-232.
21. Weinberg, Z., Perreault, J., Meyer, M. M., and Breaker, R. R. (2009) Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis, *Nature 462*, 656-659.
22. Westhof, E. (2010) The amazing world of bacterial structured RNAs, *Genome Biol 11*, 108.
23. PRODY, G. A., BAKOS, J. T., BUZAYAN, J. M., SCHNEIDER, I. R., and BRUENING, G. (1986) Autolytic Processing of Dimeric Plant Virus Satellite RNA, *Science 231*, 1577-1580.
24. Breaker, R. R. (2012) Riboswitches and the RNA world, *Cold Spring Harbor Perspectives in Biology 4*, a003566.
25. Serganov, A., and Nudler, E. A Decade of Riboswitches, *Cell 152*, 17-24.
26. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A., and Breaker, R. R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme, *Nature 428*, 281-286.
27. Leslie E, O. (2004) Prebiotic Chemistry and the Origin of the RNA world, *Critical Reviews in Biochemistry and Molecular Biology 39*, 99-123.
28. Butlerow, A. (1861) Bildung einer zuckerartigen Substanz durch Synthese, *Justus Liebigs Annalen der Chemie 120*, 295-298.
29. Ritson, D., and Sutherland, J. D. (2012) Prebiotic synthesis of simple sugars by photoredox systems chemistry, *Nat Chem 4*, 895-899.
30. Miller, S. L. (1953) A Production of Amino Acids Under Possible Primitive Earth Conditions, *Science 117*, 528-529.
31. Parker, E. T., Cleaves, H. J., Dworkin, J. P., Glavin, D. P., Callahan, M., Aubrey, A., Lazcano, A., and Bada, J. L. (2011) Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark discharge experiment, *Proceedings of the National Academy of Sciences 108*, 5526-5531.
32. Johnson, A. P., Cleaves, H. J., Dworkin, J. P., Glavin, D. P., Lazcano, A., and Bada, J. L. (2008) The Miller Volcanic Spark Discharge Experiment, *Science 322*, 404-404.

33. Yuasa, S., Flory, D., Basile, B., and Oro, J. (1984) Abiotic synthesis of purines and other heterocyclic compounds by the action of electrical discharges, *J Mol Evol 21*, 76-80.
34. Oró, J., and Kimball, A. P. (1961) Synthesis of purines under possible primitive earth conditions. I. Adenine from hydrogen cyanide, *Archives of Biochemistry and Biophysics 94*, 217-227.
35. Callahan, M. P., Smith, K. E., Cleaves, H. J., Ruzicka, J., Stern, J. C., Glavin, D. P., House, C. H., and Dworkin, J. P. (2011) Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases, *Proceedings of the National Academy of Sciences 108*, 13995-13998.
36. Sephton, M. A. (2002) Organic compounds in carbonaceous meteorites, *Nat Prod Rep 19*, 292-311.
37. Robertson, M. P., and Joyce, G. F. (2012) The Origins of the RNA world, *Cold Spring Harbor Perspectives in Biology 4*.
38. Powner, M. W., Gerland, B., and Sutherland, J. D. (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions, *Nature 459*, 239-242.
39. Pasek, M. A., Harnmeijer, J. P., Buick, R., Gull, M., and Atlas, Z. (2013) Evidence for reactive reduced phosphorus species in the early Archean ocean, *Proceedings of the National Academy of Sciences 110*, 10089-10094.
40. Higgs, P. G., and Lehman, N. (2015) The RNA world: molecular cooperation at the origins of life, *Nat Rev Genet 16*, 7-17.
41. Suzuki, K., Hatase, K., Nishiyama, D., Kawasaki, T., and Soai, K. (2010) Spontaneous absolute asymmetric synthesis promoted by achiral amines in conjunction with asymmetric autocatalysis, *Journal of Systems Chemistry 1*, 5.
42. Mislow, K. (2003) Absolute asymmetric synthesis: A commentary, *Collect Czechoslov Chem Commun 68*.
43. Prakash, T. P., Roberts, C., and Switzer, C. (1997) Activity of 2′,5′-Linked RNA in the Template-Directed Oligomerization of Mononucleotides, *Angewandte Chemie International Edition in English 36*, 1522-1523.
44. Trevino, S. G., Zhang, N., Elenko, M. P., Lupták, A., and Szostak, J. W. (2011) Evolution of functional nucleic acids in the presence of nonheritable backbone heterogeneity, *Proceedings of the National Academy of Sciences 108*, 13492-13497.
45. Leu, K., Kervio, E., Obermayer, B., Turk-MacLeod, R. M., Yuan, C., Luevano, J.-M., Chen, E., Gerland, U., Richert, C., and Chen, I. A. (2013) Cascade of Reduced Speed and Accuracy after Errors in Enzyme-Free Copying of Nucleic Acid Sequences, *Journal of the American Chemical Society 135*, 354-366.
46. Zhang, S., Zhang, N., Blain, J. C., and Szostak, J. W. (2013) Synthesis of N3'-P5'-linked phosphoramidate DNA by nonenzymatic template-directed primer extension, *J Am Chem Soc 135*, 924-932.

47. Cafferty, B. J., Gállego, I., Chen, M. C., Farley, K. I., Eritja, R., and Hud, N. V. (2013) Efficient Self-Assembly in Water of Long Noncovalent Polymers by Nucleobase Analogues, *Journal of the American Chemical Society 135*, 2447-2450.
48. Benner, S. A. (2010) Defining Life, *Astrobiology 10*, 1021-1030.
49. Mutschler, H., Wochner, A., and Holliger, P. (2015) Freeze–thaw cycles as drivers of complex ribozyme assembly, *Nat Chem 7*, 502-508.
50. Kaddour, H., and Sahai, N. (2014) Synergism and mutualism in non-enzymatic RNA polymerization, *Life (Basel) 4*, 598-620.
51. Wu, T., and Orgel, L. E. (1992) Nonenzymic template-directed synthesis on hairpin oligonucleotides. 2. Templates containing cytidine and guanosine residues, *Journal of the American Chemical Society 114*, 5496-5501.
52. Rembold, H., and Orgel, L. E. (1994) Single-strand regions of poly(G) act as templates for oligo(C) synthesis, *Journal of Molecular Evolution 38*, 205-210.
53. Deck, C., Jauker, M., and Richert, C. (2011) Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA, *Nat Chem 3*, 603-608.
54. Bowler, F. R., Chan, C. K. W., Duffy, C. D., Gerland, B., Islam, S., Powner, M. W., Sutherland, J. D., and Xu, J. (2013) Prebiotically plausible oligoribonucleotide ligation facilitated by chemoselective acetylation, *Nat Chem 5*, 383-389.
55. Kuruvilla, E., Schuster, G. B., and Hud, N. V. (2013) Enhanced Nonenzymatic Ligation of Homopurine Miniduplexes: Support for Greater Base Stacking in a Pre-RNA world, *ChemBioChem 14*, 45-48.
56. Gao, K., and Orgel, L. E. (2000) Polyphosphorylation and non-enzymatic template-directed ligation of oligonucleotides, *Orig Life Evol Biosph 30*, 45-51.
57. Ralser, M. (2014) The RNA world and the origin of metabolic enzymes, *Biochemical Society Transactions 42*, 985-988.
58. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E., and Bartel, D. P. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension, *Science 292*, 1319-1325.
59. Attwater, J., Wochner, A., and Holliger, P. (2013) In-ice evolution of RNA polymerase ribozyme activity, *Nat Chem 5*, 1011-1018.
60. Salehi-Ashtiani, K., and Szostak, J. W. (2001) In vitro evolution suggests multiple origins for the hammerhead ribozyme, *Nature 414*, 82-84.
61. Popović, M., Fliss, P. S., and Ditzler, M. A. (2015) In vitro evolution of distinct self-cleaving ribozymes in diverse environments, *Nucleic Acids Research 43*, 7070-7082.
62. Tang, J., and Breaker, R. R. (2000) Structural diversity of self-cleaving ribozymes, *Proceedings of the National Academy of Sciences 97*, 5784-5789.
63. Pan, T., and Uhlenbeck, O. C. (1992) In vitro selection of RNAs that undergo autolytic cleavage with lead(2+), *Biochemistry 31*, 3887-3895.
64. Athavale, S. S., Petrov, A. S., Hsiao, C., Watkins, D., Prickett, C. D., Gossett, J. J., Lie, L., Bowman, J. C., O'Neill, E., Bernier, C. R., Hud, N. V., Wartell, R. M., Harvey, S. C., and Williams, L. D. (2012) RNA Folding and Catalysis Mediated by Iron (II), *PLoS One 7*, e38024.

65. Bartel, D. P., and Szostak, J. W. (1993) Isolation of new ribozymes from a large pool of random sequences [see comment], *Science 261*, 1411-1418.
66. Bagby, S. C., Bergman, N. H., Shechner, D. M., Yen, C., and Bartel, D. P. (2009) A class I ligase ribozyme with reduced Mg2+ dependence: Selection, sequence analysis, and identification of functional tertiary interactions, *RNA 15*, 2129-2146.
67. Robertson, M. P., Hesselberth, J. R., and Ellington, A. D. (2001) Optimization and optimality of a short ribozyme ligase that joins non-Watson-Crick base pairings, *RNA 7*, 513-523.
68. Bergman, N. H., Lau, N. C., Lehnert, V., Westhof, E., and Bartel, D. P. (2004) The three-dimensional architecture of the class I ligase ribozyme, *RNA 10*, 176-184.
69. Vlassov, A. V., Johnston, B. H., Landweber, L. F., and Kazakov, S. A. (2004) Ligation activity of fragmented ribozymes in frozen solution: implications for the RNA world, *Nucleic Acids Research 32*, 2966-2974.
70. Paul, N., and Joyce, G. F. (2002) A self-replicating ligase ribozyme, *Proceedings of the National Academy of Sciences 99*, 12733-12740.
71. Ferretti, A. C., and Joyce, G. F. (2013) Kinetic Properties of an RNA Enzyme That Undergoes Self-Sustained Exponential Amplification, *Biochemistry 52*, 1227-1235.
72. Unrau, P. J., and Bartel, D. P. (1998) RNA-catalysed nucleotide synthesis, *Nature 395*, 260-263.
73. Martin, L., Unrau, P., and Müller, U. (2015) RNA Synthesis by in Vitro Selected Ribozymes for Recreating an RNA world, *Life 5*, 247.
74. Chapple, K. E., Bartel, D. P., and Unrau, P. J. (2003) Combinatorial minimization and secondary structure determination of a nucleotide synthase ribozyme, *RNA 9*, 1208-1220.
75. Lau, M. W. L., Cadieux, K. E. C., and Unrau, P. J. (2004) Isolation of Fast Purine Nucleotide Synthase Ribozymes, *Journal of the American Chemical Society 126*, 15686-15693.
76. Moretti, J. E., and Müller, U. F. (2014) A ribozyme that triphosphorylates RNA 5′-hydroxyl groups, *Nucleic Acids Research 42*, 4767-4778.
77. Williams, K. P., Ciafré, S., and Tocchini-Valentini, G. P. (1995) Selection of novel Mg(2+)-dependent self-cleaving ribozymes, *The EMBO journal 14*, 4551-4557.
78. Pan, T., and Uhlenbeck, O. C. (1992) In vitro selection of RNAs that undergo autolytic cleavage with Pb2+, *Biochemistry 31*, 3887-3895.
79. Pan, T., and Uhlenbeck, O. C. (1992) A small metalloribozyme with a two-step mechanism, *Nature 358*, 560-563.
80. Ekland, E. H., Szostak, J. W., and Bartel, D. P. (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences, *Science 269*, 364-370.
81. Chapman, K. B., and Szostak, J. W. (1995) Isolation of a ribozyme with 5'-5' ligase activity, *Chem Biol 2*, 325-333.

82. Lorsch, J. R., and Szostak, J. W. (1994) In vitro evolution of new ribozymes with polynucleotide kinase activity, *Nature 371*, 31-36.
83. Curtis, E. A., and Bartel, D. P. (2005) New catalytic structures from an existing ribozyme, *Nat Struct Mol Biol 12*, 994-1000.
84. Huang, F., and Yarus, M. (1997) 5‘-RNA Self-Capping from Guanosine Diphosphate, *Biochemistry 36*, 6557-6563.
85. Kumar, R. K., and Yarus, M. (2001) RNA-Catalyzed Amino Acid Activation, *Biochemistry 40*, 6998-7004.
86. Huang, F., Bugg, C. W., and Yarus, M. (2000) RNA-Catalyzed CoA, NAD, and FAD Synthesis from Phosphopantetheine, NMN, and FMN, *Biochemistry 39*, 15548-15555.
87. Tuschl, T., Sharp, P. A., and Bartel, D. P. (1998) Selection in vitro of novel ribozymes from a partially randomized U2 and U6 snRNA library, *EMBO J 17*, 2637-2650.
88. Seelig, B., and Jäschke, A. (1999) A small catalytic RNA motif with Diels-Alderase activity, *Chemistry & Biology 6*, 167-176.
89. Tarasow, T. M., Tarasow, S. L., and Eaton, B. E. (1997) RNA-catalysed carbon-carbon bond formation, *Nature 389*, 54-57.
90. Fusz, S., Eisenführ, A., Srivatsan, S. G., Heckel, A., and Famulok, M. (2005) A Ribozyme for the Aldol Reaction, *Chemistry & Biology 12*, 941-950.
91. Tsukiji, S., Pattnaik, S. B., and Suga, H. (2003) An alcohol dehydrogenase ribozyme, *Nat Struct Mol Biol 10*, 713-717.
92. Tsukiji, S., Pattnaik, S. B., and Suga, H. (2004) Reduction of an Aldehyde by a NADH/Zn2+-Dependent Redox Active Ribozyme, *Journal of the American Chemical Society 126*, 5044-5045.
93. Zhang, B., and Cech, T. R. (1997) Peptide bond formation by in vitro selected ribozymes, *Nature 390*, 96-100.
94. Saito, H., Kourouklis, D., and Suga, H. (2001) An in vitro evolved precursor tRNA with aminoacylation activity, *The EMBO Journal 20*, 1797-1806.
95. Lee, N., Bessho, Y., Wei, K., Szostak, J. W., and Suga, H. (2000) Ribozyme-catalyzed tRNA aminoacylation, *Nat Struct Mol Biol 7*, 28-33.
96. Sengle, G., Eisenführ, A., Arora, P. S., Nowick, J. S., and Famulok, M. (2001) Novel RNA catalysts for the Michael reaction, *Chemistry & Biology 8*, 459-473.
97. Baskerville, S., and Bartel, D. P. (2002) A ribozyme that ligates RNA to protein, *Proceedings of the National Academy of Sciences 99*, 9154-9159.
98. Saha, R., Pohorille, A., and Chen, I. A. (2014) Molecular Crowding and Early Evolution, *Origins of Life and Evolution of Biospheres 44*, 319-324.
99. Desai, R., Kilburn, D., Lee, H.-T., and Woodson, S. A. (2014) Increased Ribozyme Activity in Crowded Solutions, *Journal of Biological Chemistry 289*, 2972-2977.
100. Adamala, K., and Szostak, J. W. (2013) Nonenzymatic Template-Directed RNA Synthesis Inside Model Protocells, *Science 342*, 1098.

101. Chen, I. A., and Szostak, J. W. (2004) Membrane growth can generate a transmembrane pH gradient in fatty acid vesicles, *Proc Natl Acad Sci U S A 101*, 7965-7970.
102. Spiegelman, S., Haruna, I., Holland, I. B., Beaudreau, G., and Mills, D. (1965) The synthesis of a self-propagating and infectious nucleic acid with a purified enzyme, *Proceedings of the National Academy of Sciences of the United States of America 54*, 919-927.
103. Oehlenschläger, F., and Eigen, M. (1997) 30 Years Later – a New Approach to Sol Spiegelman's and Leslie Orgel's in vitro EVOLUTIONARY STUDIES Dedicated to Leslie Orgel on the occasion of his 70th birthday, *Origins of life and evolution of the biosphere 27*, 437-457.
104. Bansho, Y., Ichihashi, N., Kazuta, Y., Matsuura, T., Suzuki, H., and Yomo, T. (2012) Importance of parasite RNA species repression for prolonged translation-coupled RNA self-replication, *Chem Biol 19*, 478-487.
105. Ichihashi, N., Usui, K., Kazuta, Y., Sunami, T., Matsuura, T., and Yomo, T. (2013) Darwinian evolution in a translation-coupled RNA replication system within a cell-like compartment, *Nature Communications 4*, 2494.
106. Scotti, M. M., and Swanson, M. S. (2016) RNA mis-splicing in disease, *Nat Rev Genet 17*, 19-32.
107. Abelson, J., Trotta, C. R., and Li, H. (1998) tRNA Splicing, *Journal of Biological Chemistry 273*, 12685-12688.
108. Matera, A. G., and Wang, Z. (2014) A day in the life of the spliceosome, *Nat Rev Mol Cell Biol 15*, 108-121.
109. Irimia, M., and Roy, S. W. (2014) Origin of Spliceosomal Introns and Alternative Splicing, *Cold Spring Harbor Perspectives in Biology 6*.
110. Sharp, P. A. On the origin of RNA splicing and introns, *Cell 42*, 397-400.
111. Hausner, G., Hafez, M., and Edgell, D. R. (2014) Bacterial group I introns: mobile RNA catalysts, *Mobile DNA 5*, 8.
112. Haugen, P., Simon, D. M., and Bhattacharya, D. The natural history of group I introns, *Trends in Genetics 21*, 111-119.
113. Golden, B. L., Kim, H., and Chase, E. (2005) Crystal structure of a phage Twort group I ribozyme–product complex, *Nat Struct Mol Biol 12*, 82-89.
114. Golden, B. L. (2013) Catalysis by RNA, Structural Themes: Group I Introns A2 - Lennarz, William J, In *Encyclopedia of Biological Chemistry* (Lane, M. D., Ed.), pp 397-400, Academic Press, Waltham.
115. Stahley, M. R., and Strobel, S. A. (2006) RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis, *Current Opinion in Structural Biology 16*, 319-326.
116. Michel, F., Hanna, M., Green, R., Bartel, D. P., and Szostak, J. W. (1989) The guanosine binding site of the Tetrahymena ribozyme, *Nature 342*, 391-395.
117. Stahley, M. R., and Strobel, S. A. (2005) Structural Evidence for a Two-Metal-Ion Mechanism of Group I Intron Splicing, *Science 309*, 1587.

118. Hougland, J. L., Kravchuk, A. V., Herschlag, D., and Piccirilli, J. A. (2005) Functional Identification of Catalytic Metal Ion Binding Sites within RNA, *PLoS Biol 3*, e277.
119. Shan, S.-o., Yoshida, A., Sun, S., Piccirilli, J. A., and Herschlag, D. (1999) Three metal ions at the active site of the Tetrahymena group I ribozyme, *Proceedings of the National Academy of Sciences of the United States of America 96*, 12299-12304.
120. Lambowitz, A. M., and Zimmerly, S. (2011) Group II Introns: Mobile Ribozymes that Invade DNA, *Cold Spring Harbor Perspectives in Biology 3*, a003616.
121. Zimmerly, S., and Semper, C. (2015) Evolution of group II introns, *Mobile DNA 6*, 7.
122. Qu, G., Kaushal, P. S., Wang, J., Shigematsu, H., Piazza, C. L., Agrawal, R. K., Belfort, M., and Wang, H.-W. (2016) Structure of a group II intron in complex with its reverse transcriptase, *Nat Struct Mol Biol 23*, 549-557.
123. Marcia, M., and Pyle, Anna M. Visualizing Group II Intron Catalysis through the Stages of Splicing, *Cell 151*, 497-507.
124. Gordon, P. M., Fong, R., and Piccirilli, J. A. (2007) A second divalent metal ion in the group II intron reaction center, *Chem Biol 14*, 607-612.
125. de Lencastre, A., Hamill, S., and Pyle, A. M. (2005) A single active-site region for a group II intron, *Nat Struct Mol Biol 12*, 626-627.
126. Sontheimer, E. J., Gordon, P. M., and Piccirilli, J. A. (1999) Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome, *Genes Dev 13*, 1729-1741.
127. Casalino, L., Palermo, G., Rothlisberger, U., and Magistrato, A. (2016) Who Activates the Nucleophile in Ribozyme Catalysis? An Answer from the Splicing Mechanism of Group II Introns, *Journal of the American Chemical Society 138*, 10374-10377.
128. Chalamcharla, V. R., Curcio, M. J., and Belfort, M. (2010) Nuclear expression of a group II intron is consistent with spliceosomal intron ancestry, *Genes & Development 24*, 827-836.
129. Wan, R., Yan, C., Bai, R., Huang, G., and Shi, Y. (2016) Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution, *Science 353*, 895.
130. Butcher, S. E. (2011) The spliceosome and its metal ions, *Met Ions Life Sci 9*, 235-251.
131. Valadkhan, S., Mohammadi, A., Wachtel, C., and Manley, J. L. (2007) Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing, *RNA 13*, 2300-2311.
132. Valadkhan, S., Mohammadi, A., Jaladat, Y., and Geisler, S. (2009) Protein-free small nuclear RNAs catalyze a two-step splicing reaction, *Proc Natl Acad Sci U S A 106*, 11901-11906.
133. Butcher, S. E. (2009) The spliceosome as ribozyme hypothesis takes a second step, *Proceedings of the National Academy of Sciences of the United States of America 106*, 12211-12212.

134. Wahl, M. C., Will, C. L., and Lührmann, R. (2009) The Spliceosome: Design Principles of a Dynamic RNP Machine, *Cell 136*, 701-718.
135. Fica, S. M., Tuttle, N., Novak, T., Li, N. S., Lu, J., Koodathingal, P., Dai, Q., Staley, J. P., and Piccirilli, J. A. (2013) RNA catalyses nuclear pre-mRNA splicing, *Nature 503*, 229-234.
136. Fica, S. M., Mefford, M. A., Piccirilli, J. A., and Staley, J. P. (2014) Evidence for a group II intron–like catalytic triplex in the spliceosome, *Nat Struct Mol Biol 21*, 464-471.
137. Freed, E. F., Bleichert, F., Dutca, L. M., and Baserga, S. J. (2010) When ribosomes go bad: diseases of ribosome biogenesis, *Molecular bioSystems 6*, 481-493.
138. Cech, T. R. (2000) The Ribosome Is a Ribozyme, *Science 289*, 878.
139. Colussi, T. M., Costantino, D. A., Zhu, J., Donohue, J. P., Korostelev, A. A., Jaafar, Z. A., Plank, T.-D. M., Noller, H. F., and Kieft, J. S. (2015) Initiation of translation in bacteria by a structured eukaryotic IRES RNA, *Nature 519*, 110-113.
140. Leung, E. K. Y., Suslov, N., Tuttle, N., Sengupta, R., and Piccirilli, J. A. (2011) The Mechanism of Peptidyl Transfer Catalysis by the Ribosome, *Annual Review of Biochemistry 80*, 527-555.
141. Muth, G. W., Chen, L., Kosek, A. B., and Strobel, S. A. (2001) pH-dependent conformational flexibility within the ribosomal peptidyl transferase center, *RNA 7*, 1403-1415.
142. Polacek, N., Gaynor, M., Yassin, A., and Mankin, A. S. (2001) Ribosomal peptidyl transferase can withstand mutations at the putative catalytic nucleotide, *Nature 411*, 498-501.
143. Martin Schmeing, T., Huang, K. S., Strobel, S. A., and Steitz, T. A. (2005) An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA, *Nature 438*, 520-524.
144. Youngman, E. M., Brunelle, J. L., Kochaniak, A. B., and Green, R. The Active Site of the Ribosome Is Composed of Two Layers of Conserved Nucleotides with Distinct Roles in Peptide Bond Formation and Peptide Release, *Cell 117*, 589-599.
145. Sharma, P. K., Xiang, Y., Kato, M., and Warshel, A. (2005) What Are the Roles of Substrate-Assisted Catalysis and Proximity Effects in Peptide Bond Formation by the Ribosome?, *Biochemistry 44*, 11307-11314.
146. Trobro, S., and Åqvist, J. (2005) Mechanism of peptide bond synthesis on the ribosome, *Proceedings of the National Academy of Sciences of the United States of America 102*, 12395-12400.
147. Nielsen, H., Westhof, E., and Johansen, S. (2005) An mRNA Is Capped by a 2', 5' Lariat Catalyzed by a Group I-Like Ribozyme, *Science 309*, 1584.
148. Nielsen, H., Beckert, B., Masquida, B., and Johansen, S. D. (2007) Chapter 12 The GIR1 Branching Ribozyme, In *Ribozymes and RNA Catalysis*, pp 229-252, The Royal Society of Chemistry.

149. Meyer, M., Nielsen, H., Oliéric, V., Roblin, P., Johansen, S. D., Westhof, E., and Masquida, B. (2014) Speciation of a group I intron into a lariat capping ribozyme, *Proceedings of the National Academy of Sciences 111*, 7659-7664.
150. Tuschl, T., Sharp, P. A., and Bartel, D. P. (2001) A ribozyme selected from variants of U6 snRNA promotes 2',5'-branch formation, *RNA 7*, 29-43.
151. Einvik, C., Nielsen, H., Westhof, E., Michel, F., and Johansen, S. (1998) Group I-like ribozymes with a novel core organization perform obligate sequential hydrolytic cleavages at two processing sites, *RNA 4*, 530-541.
152. Kirsebom, L. A., and Trobro, S. (2009) RNase P RNA-mediated cleavage, *IUBMB Life 61*, 189-200.
153. Walker, S. C., and Engelke, D. R. (2006) Ribonuclease P: the evolution of an ancient RNA enzyme, *Crit Rev Biochem Mol Biol 41*, 77-102.
154. Mondragon, A. (2013) Structural studies of RNase P, *Annu Rev Biophys 42*, 537-557.
155. Alifano, P., Rivellini, F., Piscitelli, C., Arraiano, C. M., Bruni, C. B., and Carlomagno, M. S. (1994) Ribonuclease E provides substrates for ribonuclease P-dependent processing of a polycistronic mRNA, *Genes & Development 8*, 3021-3031.
156. Engelke, D. R., and Fierke, C. A. (2015) The evolution of RNase P, *RNA 21*, 517-518.
157. Jarrous, N., and Gopalan, V. (2010) Archaeal/Eukaryal RNase P: subunits, functions and RNA diversification, *Nucleic Acids Research 38*, 7885-7894.
158. Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T., and Mondragon, A. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA, *Nature 468*, 784-789.
159. Crary, S. M., Kurz, J. C., and Fierke, C. A. (2002) Specific phosphorothioate substitutions probe the active site of Bacillus subtilis ribonuclease P, *RNA 8*, 933-947.
160. Cassano, A. G., Anderson, V. E., and Harris, M. E. (2004) Analysis of Solvent Nucleophile Isotope Effects: Evidence for Concerted Mechanisms and Nucleophilic Activation by Metal Coordination in Nonenzymatic and Ribozyme-Catalyzed Phosphodiester Hydrolysis, *Biochemistry 43*, 10547-10559.
161. Zahler, N. H., Sun, L., Christian, E. L., and Harris, M. E. (2005) The pre-tRNA nucleotide base and 2'-hydroxyl at N(-1) contribute to fidelity in tRNA processing by RNase P, *J Mol Biol 345*, 969-985.
162. Ferré-D'Amaré, A. R., and Scott, W. G. (2010) Small Self-cleaving Ribozymes, *Cold Spring Harbor Perspectives in Biology 2*.
163. McCown, P., Winkler, W., and Breaker, R. (2012) Mechanism and Distribution of glmS Ribozymes, In *Ribozymes* (Hartig, J. S., Ed.), pp 113-129, Humana Press.
164. Emilsson, G. M., Nakamura, S., Roth, A., and Breaker, R. R. (2003) Ribozyme speed limits, *RNA 9*, 907-918.
165. Thompson, J. E., Kutateladze, T. G., Schuster, M. C., Venegas, F. D., Messmore, J. M., and Raines, R. T. (1995) Limits to Catalysis by Ribonuclease A, *Bioorganic chemistry 23*, 471-481.

166. Raines, R. T. (2004) Active Site of Ribonuclease A, In *Artificial Nucleases* (Zenkova, M. A., Ed.), pp 19-32, Springer Berlin Heidelberg, Berlin, Heidelberg.
167. Buzayan, J. M., Gerlach, W. L., and Bruening, G. (1986) Non-enzymatic cleavage and ligation of RNAs complementary to a plant virus satellite RNA, *Nature 323*, 349-353.
168. Fedor, M. J. (1999) Tertiary structure stabilization promotes hairpin ribozyme ligation, *Biochemistry 38*, 11040-11050.
169. Müller, S., Appel, B., Krellenberg, T., and Petkovic, S. (2012) The many faces of the hairpin ribozyme: Structural and functional variants of a small catalytic rna, *IUBMB Life 64*, 36-47.
170. Rupert, P. B., Massey, A. P., Sigurdsson, S. T., Ferré, D., and Amaré, A. R. (2002) Transition State Stabilization by a Catalytic RNA, *Science 298*, 1421.
171. Symons, R. H. (1997) Plant pathogenic RNAs and RNA catalysis, *Nucleic Acids Research 25*, 2683-2689.
172. Kath-Schorr, S., Wilson, T. J., Li, N.-S., Lu, J., Piccirilli, J. A., and Lilley, D. M. J. (2012) General Acid–Base Catalysis Mediated by Nucleobases in the Hairpin Ribozyme, *Journal of the American Chemical Society 134*, 16717-16724.
173. Guo, M., Spitale, R. C., Volpini, R., Krucinska, J., Cristalli, G., Carey, P. R., and Wedekind, J. E. (2009) Direct Raman Measurement of an Elevated Base pKa in the Active Site of a Small Ribozyme in a Precatalytic Conformation, *Journal of the American Chemical Society 131*, 12908-12909.
174. Perrotta, A. T., and Been, M. D. (1991) A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA, *Nature 350*, 434-436.
175. Webb, C.-H. T., Riccitelli, N. J., Ruminski, D. J., and Lupták, A. (2009) Widespread Occurrence of Self-Cleaving Ribozymes, *Science 326*, 953.
176. Weinberg, Z., Kim, P. B., Chen, T. H., Li, S., Harris, K. A., Lunse, C. E., and Breaker, R. R. (2015) New classes of self-cleaving ribozymes revealed by comparative genomics analysis, *Nat Chem Biol 11*, 606-610.
177. Chen, J.-H., Yajima, R., Chadalavada, D. M., Chase, E., Bevilacqua, P. C., and Golden, B. L. (2010) 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*, 6508-6518.
178. Schürer, H., Lang, K., Schuster, J., and Mörl, M. (2002) A universal method to produce in vitro transcripts with homogeneous 3′ ends, *Nucleic Acids Research 30*, e56.
179. Nakano, S.-i., Chadalavada, D. M., and Bevilacqua, P. C. (2000) General Acid-Base Catalysis in the Mechanism of a Hepatitis Delta Virus Ribozyme, *Science 287*, 1493-1497.
180. Nakano, S.-i., and Bevilacqua, P. C. (2007) Mechanistic Characterization of the HDV Genomic Ribozyme: A Mutant of the C41 Motif Provides Insight into the Positioning and Thermodynamic Linkage of Metal Ions and Protons, *Biochemistry 46*, 3001-3012.

181. Chen, J. H., Gong, B., Bevilacqua, P. C., Carey, P. R., and Golden, B. L. (2009) A catalytic metal ion interacts with the cleavage Site G.U wobble in the HDV ribozyme, *Biochemistry 48*, 1498-1507.
182. Ferre-D'Amare, A. R., Zhou, K., and Doudna, J. A. (1998) Crystal structure of a hepatitis delta virus ribozyme, *Nature 395*, 567-574.
183. Perrotta, A. T., Shih, I. h., and Been, M. D. (1999) Imidazole Rescue of a Cytosine Mutation in a Self-Cleaving Ribozyme, *Science 286*, 123.
184. Gong, B., Chen, J.-H., Chase, E., Chadalavada, D. M., Yajima, R., Golden, B. L., Bevilacqua, P. C., and Carey, P. R. (2007) Direct Measurement of a pKa near Neutrality for the Catalytic Cytosine in the Genomic HDV Ribozyme Using Raman Crystallography, *Journal of the American Chemical Society 129*, 13335-13342.
185. Das, S. R., and Piccirilli, J. A. (2005) General acid catalysis by the hepatitis delta virus ribozyme, *Nat Chem Biol 1*, 45-52.
186. Chen, J., Ganguly, A., Miswan, Z., Hammes-Schiffer, S., Bevilacqua, P. C., and Golden, B. L. (2013) Identification of the Catalytic Mg2+ Ion in the Hepatitis Delta Virus Ribozyme, *Biochemistry 52*, 557-567.
187. Golden, B. L. (2011) Two Distinct Catalytic Strategies in the Hepatitis Delta Virus Ribozyme Cleavage Reaction, *Biochemistry 50*, 9424-9433.
188. Bevilacqua, P. C., and Yajima, R. (2006) Nucleobase catalysis in ribozyme mechanism, *Curr Opin Chem Biol 10*, 455-464.
189. Wilson, T. J., and Lilley, D. M. J. (2011) 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.
190. Saville, B. J., and Collins, R. A. A site-specific self-cleavage reaction performed by a novel RNA in neurospora mitochondria, *Cell 61*, 685-696.
191. Doudna, J. A., and Cech, T. R. (2002) The chemical repertoire of natural ribozymes, *Nature 418*, 222-228.
192. Suslov, N. B., DasGupta, S., Huang, H., Fuller, J. R., Lilley, D. M. J., Rice, P. A., and Piccirilli, J. A. (2015) Crystal structure of the Varkud satellite ribozyme, *Nat Chem Biol 11*, 840-846.
193. Wilson, T. J., Li, N.-S., Lu, J., Frederiksen, J. K., Piccirilli, J. A., and Lilley, D. M. J. (2010) Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme, *Proceedings of the National Academy of Sciences 107*, 11751-11756.
194. Roth, A., Weinberg, Z., Chen, A. G. Y., Kim, P. B., Ames, T. D., and Breaker, R. R. (2014) A widespread self-cleaving ribozyme class is revealed by bioinformatics, *Nat Chem Biol 10*, 56-60.
195. Liu, Y., Wilson, T. J., McPhee, S. A., and Lilley, D. M. J. (2014) Crystal structure and mechanistic investigation of the twister ribozyme, *Nat Chem Biol 10*, 739-744.
196. Eiler, D., Wang, J., and Steitz, T. A. (2014) Structural basis for the fast self-cleavage reaction catalyzed by the twister ribozyme, *Proceedings of the National Academy of Sciences of the United States of America 111*, 13028-13033.

197. Ren, A., Košutić, M., Rajashankar, K. R., Frener, M., Santner, T., Westhof, E., Micura, R., and Patel, D. J. (2014) In-line alignment and Mg2+ coordination at the cleavage site of the env22 twister ribozyme, *Nature Communications 5*, 5534.
198. Wilson, T. J., Liu, Y., Domnick, C., Kath-Schorr, S., and Lilley, D. M. J. (2016) The Novel Chemical Mechanism of the Twister Ribozyme, *Journal of the American Chemical Society 138*, 6151-6162.

199. Košutić, M., Neuner, S., Ren, A., Flür, S., Wunderlich, C., Mairhofer, E., Vušurović, N., Seikowski, J., Breuker, K., Höbartner, C., Patel, D. J., Kreutz, C., and Micura, R. (2015) A Mini-Twister Variant and Impact of Residues/Cations on the Phosphodiester Cleavage of this Ribozyme Class, *Angewandte Chemie 127*, 15343-15348.
200. Ren, A., Vusurovic, N., Gebetsberger, J., Gao, P., Juen, M., Kreutz, C., Micura, R., and Patel, D. J. (2016) Pistol ribozyme adopts a pseudoknot fold facilitating site-specific in-line cleavage, *Nat Chem Biol 12*, 702-708.
201. Harris, K. A., Lünse, C. E., Li, S., Brewer, K. I., and Breaker, R. R. (2015) Biochemical analysis of pistol self-cleaving ribozymes, *RNA 21*, 1852-1858.
202. Li, S., Lünse, C. E., Harris, K. A., and Breaker, R. R. (2015) Biochemical analysis of hatchet self-cleaving ribozymes, *RNA 21*, 1845-1851.
203. Barrick, J. E., Corbino, K. A., Winkler, W. C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., Wickiser, J. K., and Breaker, R. R. (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control, *Proceedings of the National Academy of Sciences of the United States of America 101*, 6421-6426.
204. Martick, M., Horan, L. H., Noller, H. F., and Scott, W. G. (2008) A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA, *Nature 454*, 899-902.
205. Salehi-Ashtiani, K., Lupták, A., Litovchick, A., and Szostak, J. W. (2006) A Genomewide Search for Ribozymes Reveals an HDV-Like Sequence in the Human CPEB3 Gene, *Science 313*, 1788-1792.
206. Teixeira, A., Tahiri-Alaoui, A., West, S., Thomas, B., Ramadass, A., Martianov, I., Dye, M., James, W., Proudfoot, N. J., and Akoulitchev, A. (2004) Autocatalytic RNA cleavage in the human [beta]-globin pre-mRNA promotes transcription termination, *Nature 432*, 526-530.
207. Milewski, S. (2002) Glucosamine-6-phosphate synthase--the multi-facets enzyme, *Biochim Biophys Acta 1597*, 173-192.
208. Fantus, I. G., Goldberg, H. J., Whiteside, C. I., and Topic, D. (2006) The Hexosamine Biosynthesis Pathway, In *The Diabetic Kidney* (Cortes, P., and Mogensen, C. E., Eds.), pp 117-133, Humana Press, Totowa, NJ.
209. Ferré-D'Amaré, A. R. (2010) The glmS ribozyme: use of a small molecule coenzyme by a gene-regulatory RNA, *Quarterly reviews of biophysics 43*, 423-447.

210. Collins, J. A., Irnov, I., Baker, S., and Winkler, W. C. (2007) Mechanism of mRNA destabilization by the glmS ribozyme, *Genes & Development 21*, 3356-3368.
211. Klein, D. J., Ferré, D., and Amaré, A. R. (2006) <div xmlns="http://www.w3.org/1999/xhtml">Structural Basis of glmS Ribozyme Activation by Glucosamine-6-Phosphate</div&gt, *Science 313*, 1752.
212. McCarthy, T. J., Plog, M. A., Floy, S. A., Jansen, J. A., Soukup, J. K., and Soukup, G. A. (2005) Ligand Requirements for glmS Ribozyme Self-Cleavage, *Chemistry & Biology 12*, 1221-1226.
213. Klein, D. J., Been, M. D., and Ferré-D'Amaré, A. R. (2007) Essential Role of an Active-Site Guanine in glmS Ribozyme Catalysis, *Journal of the American Chemical Society 129*, 14858-14859.
214. Zhang, S., Stevens, D. R., Goyal, P., Bingaman, J. L., Bevilacqua, P. C., and Hammes-Schiffer, S. (2016) Assessing the Potential Effects of Active Site Mg2+ Ions in the glmS Ribozyme-Cofactor Complex, *J Phys Chem Lett 7*, 3984-3988.
215. Chadalavada, D. M., Gratton, E. A., and Bevilacqua, P. C. (2010) The human HDV-like CPEB3 ribozyme is intrinsically fast-reacting, *Biochemistry 49*, 5321-5330.
216. Talini, G., Branciamore, S., and Gallori, E. (2011) Ribozymes: Flexible molecular devices at work, *Biochimie 93*, 1998-2005.
217. Kolev, N. G., Hartland, E. I., and Huber, P. W. (2008) A manganese-dependent ribozyme in the 3'-untranslated region of Xenopus Vg1 mRNA, *Nucleic Acids Res 36*, 5530-5539.
218. Soukup, G. A., and Breaker, R. R. (1999) Relationship between internucleotide linkage geometry and the stability of RNA, *RNA 5*, 1308-1325.
219. Mir, A., Chen, J., Robinson, K., Lendy, E., Goodman, J., Neau, D., and Golden, B. L. (2015) Two Divalent Metal Ions and Conformational Changes Play Roles in the Hammerhead Ribozyme Cleavage Reaction, *Biochemistry 54*, 6369-6381.
220. UHLENBECK, O. C. (1987) A small catalytic oligoribonucleotide, *Nature 328*, 596−600.
221. de la Peña, M., and García-Robles, I. (2010) Ubiquitous presence of the hammerhead ribozyme motif along the tree of life, *RNA 16*, 1943-1950.
222. Perreault, J., Weinberg, Z., Roth, A., Popescu, O., Chartrand, P., Ferbeyre, G., and Breaker, R. R. (2011) Identification of Hammerhead Ribozymes in All Domains of Life Reveals Novel Structural Variations, *PLoS Comput Biol 7*, e1002031.
223. Seehafer, C., Kalweit, A., Steger, G., Gräf, S., and Hammann, C. (2011) From alpaca to zebrafish: Hammerhead ribozymes wherever you look, *RNA 17*, 21-26.
224. Cervera, A., and De la Peña, M. (2014) Eukaryotic Penelope-Like Retroelements Encode Hammerhead Ribozyme Motifs, *Molecular Biology and Evolution 31*, 2941-2947.
225. Suga, H., Cowan, J. A., and Szostak, J. W. (1998) Unusual Metal Ion Catalysis in an Acyl-Transferase Ribozyme, *Biochemistry 37*, 10118-10125.

226. De Young, M. B., Siwkowski, A., and Hampel, A. (1997) Determination of Catalytic Parameters for Hairpin Ribozymes, In *Ribozyme Protocols* (Turner, P. C., Ed.), pp 209-220, Humana Press, Totowa, NJ.
227. Hampel, A., and Cowan, J. A. (1997) A unique mechanism for RNA catalysis: the role of metal cofactors in hairpin ribozyme cleavage, *Chemistry & Biology 4*, 513-517.
228. Nesbitt, S., Hegg, L. A., and Fedor, M. J. (1997) An unusual pH-independent and metal-ion-independent mechanism for hairpin ribozyme catalysis, *Chemistry & Biology 4*, 619-630.
229. Klawuhn, K., Jansen, J. A., Souchek, J., Soukup, G. A., and Soukup, J. K. (2010) Analysis of Metal Ion Dependence in glmS Ribozyme Self-Cleavage and Coenzyme Binding, *ChemBioChem 11*, 2567-2571.
230. Brooks, K. M., and Hampel, K. J. (2011) Rapid Steps in the glmS Ribozyme Catalytic Pathway: Cation and Ligand Requirements, *Biochemistry 50*, 2424-2433.
231. Gong, B., Klein, D. J., Ferré-D'Amaré, A. R., and Carey, P. R. (2011) The glmS Ribozyme Tunes the Catalytically Critical pKa of Its Coenzyme Glucosamine-6-phosphate, *Journal of the American Chemical Society 133*, 14188-14191.
232. Liberman, J. A., Guo, M., Jenkins, J. L., Krucinska, J., Chen, Y., Carey, P. R., and Wedekind, J. E. (2012) A Transition-State Interaction Shifts Nucleobase Ionization toward Neutrality To Facilitate Small Ribozyme Catalysis, *Journal of the American Chemical Society 134*, 16933-16936.
233. Nakano, S.-i., Proctor, D. J., and Bevilacqua, P. C. (2001) 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.
234. Roychowdhury-Saha, M., and Burke, D. H. (2007) Distinct reaction pathway promoted by non-divalent-metal cations in a tertiary stabilized hammerhead ribozyme, *RNA 13*, 841-848.
235. Cerrone-Szakal, A. L., Chadalavada, D. M., Golden, B. L., and Bevilacqua, P. C. (2008) Mechanistic characterization of the HDV genomic ribozyme: The cleavage site base pair plays a structural role in facilitating catalysis, *RNA 14*, 1746-1760.
236. Veeraraghavan, N., Ganguly, A., Chen, J.-H., Bevilacqua, P. C., Hammes-Schiffer, S., and Golden, B. L. (2011) Metal Binding Motif in the Active Site of the HDV Ribozyme Binds Divalent and Monovalent Ions, *Biochemistry 50*, 2672-2682.
237. Han, J., and Burke, J. M. (2005) Model for General Acid−Base Catalysis by the Hammerhead Ribozyme: pH−Activity Relationships of G8 and G12 Variants at the Putative Active Site, *Biochemistry 44*, 7864-7870.
238. Martick, M., and Scott, W. G. (2006) Tertiary Contacts Distant from the Active Site Prime a Ribozyme for Catalysis, *Cell 126*, 309-320.
239. Lambert, D., Heckman, J. E., and Burke, J. M. (2006) Three Conserved Guanosines Approach the Reaction Site in Native and Minimal Hammerhead Ribozymes, *Biochemistry 45*, 7140-7147.

240. Anderson, M., Schultz, E. P., Martick, M., and Scott, W. G. (2013) Active-Site Monovalent Cations Revealed in a 1.55-Å-Resolution Hammerhead Ribozyme Structure, *Journal of Molecular Biology 425*, 3790-3798.
241. Bevilacqua, P. C. (2003) Mechanistic Considerations for General Acid−Base Catalysis by RNA: Revisiting the Mechanism of the Hairpin Ribozyme, *Biochemistry 42*, 2259-2265.
242. Wang, S., Karbstein, K., Peracchi, A., Beigelman, L., and Herschlag, D. (1999) Identification of the Hammerhead Ribozyme Metal Ion Binding Site Responsible for Rescue of the Deleterious Effect of a Cleavage Site Phosphorothioate, *Biochemistry 38*, 14363-14378.
243. O'Rear, J. L., Wang, S., Feig, A. L., Beigelman, L., Uhlenbeck, O. C., and Herschlag, D. (2001) Comparison of the hammerhead cleavage reactions stimulated by monovalent and divalent cations, *RNA 7*, 537-545.
244. Scott, E. C., and Uhlenbeck, O. C. (1999) A re-investigation of the thio effect at the hammerhead cleavage site, *Nucleic Acids Research 27*, 479-484.
245. Ward, W. L., and DeRose, V. J. (2012) Ground-state coordination of a catalytic metal to the scissile phosphate of a tertiary-stabilized Hammerhead ribozyme, *RNA 18*, 16-23.
246. Murray, J. B., Seyhan, A. A., Walter, N. G., Burke, J. M., and Scott, W. G. (1998) The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone, *Chemistry & Biology 5*, 587-595.
247. Roychowdhury-Saha, M., and Burke, D. H. (2006) Extraordinary rates of transition metal ion-mediated ribozyme catalysis, *RNA 12*, 1846-1852.
248. Boots, J. L., Canny, M. D., Azimi, E., and Pardi, A. (2008) Metal ion specificities for folding and cleavage activity in the Schistosoma hammerhead ribozyme, *RNA 14*, 2212-2222.
249. KISSELEVA, N., KHVOROVA, A., WESTHOF, E., and SCHIEMANN, O. (2005) Binding of manganese(II) to a tertiary stabilized hammerhead ribozyme as studied by electron paramagnetic resonance spectroscopy, *RNA 11*, 1-6.
250. Vogt, M., Lahiri, S., Hoogstraten, C. G., Britt, R. D., and DeRose, V. J. (2006) Coordination Environment of a Site-Bound Metal Ion in the Hammerhead Ribozyme Determined by 15N and 2H ESEEM Spectroscopy, *Journal of the American Chemical Society 128*, 16764-16770.
251. Martick, M., Lee, T.-S., York, D. M., and Scott, W. G. (2008) Solvent Structure and Hammerhead Ribozyme Catalysis, *Chemistry & Biology 15*, 332-342.
252. Lee, T.-S., López, C. S., Giambaşu, G. M., Martick, M., Scott, W. G., and York, D. M. (2008) Role of Mg2+ in Hammerhead Ribozyme Catalysis from Molecular Simulation, *Journal of the American Chemical Society 130*, 3053-3064.
253. Lee, T.-S., Giambaşu, G. M., Sosa, C. P., Martick, M., Scott, W. G., and York, D. M. (2009) Threshold occupancy and specific cation binding modes in the hammerhead ribozyme active site are required for active conformation, *Journal of molecular biology 388*, 195-206.

254. Wong, K.-Y., Lee, T.-S., and York, D. M. (2011) Active Participation of the Mg2+ Ion in the Reaction Coordinate of RNA Self-Cleavage Catalyzed by the Hammerhead Ribozyme, *Journal of Chemical Theory and Computation 7*, 1-3.
255. SAKSMERPROME, V., ROYCHOWDHURY-SAHA, M., JAYASENA, S., KHVOROVA, A., and BURKE, D. H. (2004) Artificial tertiary motifs stabilize trans-cleaving hammerhead ribozymes under conditions of submillimolar divalent ions and high temperatures, *RNA 10*, 1916-1924.
256. Chi, Y.-I., Martick, M., Lares, M., Kim, R., Scott, W. G., and Kim, S.-H. (2008) Capturing Hammerhead Ribozyme Structures in Action by Modulating General Base Catalysis, *PLoS Biol 6*, e234.
257. Golden, B. L. (2007) Preparation and Crystallization of RNA, In *Macromolecular Crystallography Protocols: Volume 1, Preparation and Crystallization of Macromolecules* (Walker, J. M., and Doublié, S., Eds.), pp 239-257, Humana Press, Totowa, NJ.
258. Otwinowski, Z., and Minor, W. (1997) [20] Processing of X-ray diffraction data collected in oscillation mode, In *Methods in Enzymology*, pp 307-326, Academic Press.
259. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallographica Section D 66*, 213-221.
260. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallographica Section D 66*, 486-501.
261. Schultz, E. P., Vasquez, E. E., and Scott, W. G. (2014) Structural and catalytic effects of an invariant purine substitution in the hammerhead ribozyme: implications for the mechanism of acid-base catalysis, *Acta Crystallographica Section D 70*, 2256-2263.
262. Thomas, J. M., and Perrin, D. M. (2008) Probing General Base Catalysis in the Hammerhead Ribozyme, *Journal of the American Chemical Society 130*, 15467-15475.
263. Peracchi, A., Beigelman, L., Usman, N., and Herschlag, D. (1996) Rescue of abasic hammerhead ribozymes by exogenous addition of specific bases, *Proceedings of the National Academy of Sciences of the United States of America 93*, 11522-11527.
264. Blount, K. F., and Uhlenbeck, O. C. (2005) The Structure-Function Dilemma of the Hammerhead Ribozyme, *Annual Review of Biophysics and Biomolecular Structure 34*, 415-440.
265. Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) Three-dimensional structure of a hammerhead ribozyme, *Nature 372*, 68-74.

266. Scott, W. G., Finch, J. T., and Klug, A. (1995) The crystal structure of an AII-RNAhammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage, *Cell 81*, 991-1002.
267. De la Peña, M., Gago, S., and Flores, R. (2003) Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity, *The EMBO Journal 22*, 5561-5570.
268. Khvorova, A., Lescoute, A., Westhof, E., and Jayasena, S. D. (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity, *Nat Struct Mol Biol 10*, 708-712.
269. Canny, M. D., Jucker, F. M., Kellogg, E., Khvorova, A., Jayasena, S. D., and Pardi, A. (2004) Fast Cleavage Kinetics of a Natural Hammerhead Ribozyme, *Journal of the American Chemical Society 126*, 10848-10849.
270. Nelson, J. A., and Uhlenbeck, O. C. (2006) When to Believe What You See, *Molecular Cell 23*, 447-450.
271. Nelson, J. A., and Uhlenbeck, O. C. (2008) Hammerhead redux: Does the new structure fit the old biochemical data?, *RNA 14*, 605-615.
272. Burke, D. H., and Greathouse, S. T. (2005) Low-magnesium, trans-cleavage activity by type III, tertiary stabilized hammerhead ribozymes with stem 1 discontinuities, *BMC Biochemistry 6*, 1-9.
273. Gilbert, S. D., Reyes, F. E., Edwards, A. L., and Batey, R. T. (2009) Adaptive Ligand Binding by the Purine Riboswitch in the Recognition of Guanine and Adenine Analogs, *Structure 17*, 857-868.
274. Bebenek, K., Pedersen, L. C., and Kunkel, T. A. (2011) Replication infidelity via a mismatch with Watson–Crick geometry, *Proceedings of the National Academy of Sciences 108*, 1862-1867.
275. Wang, W., Hellinga, H. W., and Beese, L. S. (2011) Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis, *Proceedings of the National Academy of Sciences 108*, 17644-17648.
276. Ai, H., Chen, J., and Zhang, C. (2012) Amino–Imino Adenine Tautomerism Induced by the Cooperative Effect between Metal Ion and H2O/NH3, *The Journal of Physical Chemistry B 116*, 13624-13636.
277. Buskiewicz, I. A., and Burke, J. M. (2012) Folding of the hammerhead ribozyme: Pyrrolo-cytosine fluorescence separates core folding from global folding and reveals a pH-dependent conformational change, *RNA 18*, 434-448.
278. Singh, V., Fedeles, B. I., and Essigmann, J. M. (2015) Role of tautomerism in RNA biochemistry, *RNA 21*, 1-13.
279. Ward, W. L., Plakos, K., and DeRose, V. J. (2014) Nucleic Acid Catalysis: Metals, Nucleobases, and Other Cofactors, *Chemical Reviews 114*, 4318-4342.
280. Mir, A., and Golden, B. L. (2016) Two Active Site Divalent Ions in the Crystal Structure of the Hammerhead Ribozyme Bound to a Transition State Analogue, *Biochemistry 55*, 633-636.
281. McKay, D. B. (1996) Structure and function of the hammerhead ribozyme: an unfinished story, *RNA 2*, 395-403.

282. Slim, G., and Gait, M. J. (1992) The role of the exocyclic amino groups of conserved purines in hammerhead ribozyme cleavage, *Biochemical and Biophysical Research Communications 183*, 605-609.
283. Fu, D. J., Rajur, S. B., and McLaughlin, L. W. (1993) Importance of specific guanosine N7-nitrogens and purine amino groups for efficient cleavage by a hammerhead ribozyme, *Biochemistry 32*, 10629-10637.
284. Knöll, R., Bald, R., and Fürste, J. P. (1997) Complete identification of nonbridging phosphate oxygens involved in hammerhead cleavage, *RNA 3*, 132-140.
285. Williams, D. M., Pieken, W. A., and Eckstein, F. (1992) Function of specific 2'-hydroxyl groups of guanosines in a hammerhead ribozyme probed by 2' modifications, *Proceedings of the National Academy of Sciences of the United States of America 89*, 918-921.
286. Thomas, J. M., and Perrin, D. M. (2009) Probing General Acid Catalysis in the Hammerhead Ribozyme, *Journal of the American Chemical Society 131*, 1135-1143.
287. Huheey, J. E. (1983) *Inorganic Chemistry*, 3 ed., Harper & Row, New York.
288. de la Peña, M., and García-Robles, I. (2010) Intronic hammerhead ribozymes are ultraconserved in the human genome, *EMBO Reports 11*, 711-716.
289. Riccitelli, N. J., Delwart, E., and Luptak, A. (2014) Identification of minimal HDV-like ribozymes with unique divalent metal ion dependence in the human microbiome, *Biochemistry 53*.
290. Riccitelli, N. J., and Lupták, A. (2010) Computational discovery of folded RNA domains in genomes and in vitro selected libraries, *Methods (San Diego, Calif.) 52*, 133-140.
291. Jimenez, R. M., Delwart, E., and Lupták, A. (2011) Structure-based Search Reveals Hammerhead Ribozymes in the Human Microbiome, *The Journal of Biological Chemistry 286*, 7737-7743.
292. Fong, N., Ohman, M., and Bentley, D. L. (2009) Fast ribozyme cleavage releases transcripts from RNA polymerase II and aborts co-transcriptional pre-mRNA processing, *Nat Struct Mol Biol 16*, 916-922.
293. McCown, P. J., Roth, A., and Breaker, R. R. (2011) An expanded collection and refined consensus model of glmS ribozymes, *RNA 17*, 728-736.
294. Lincoln, T. A., and Joyce, G. F. (2009) Self-Sustained Replication of an RNA Enzyme, *Science 323*, 1229-1232.
295. Ohuchi, M., Murakami, H., and Suga, H. (2007) The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus, *Current Opinion in Chemical Biology 11*, 537-542.
296. Eickbush, D. G., and Eickbush, T. H. (2010) R2 Retrotransposons Encode a Self-Cleaving Ribozyme for Processing from an rRNA Cotranscript, *Molecular and Cellular Biology 30*, 3142-3150.
297. Schutz, K., Hesselberth, J. R., and Fields, S. (2010) Capture and sequence analysis of RNAs with terminal 2′,3′-cyclic phosphates, *RNA 16*, 621-631.

298. Gonzalez, T. N., Sidrauski, C., Dörfler, S., and Walter, P. (1999) Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway, *The EMBO Journal 18*, 3119-3132.
299. Sidrauski, C., and Walter, P. (1997) The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response, *Cell 90*, 1031-1039.

300. Trotta, C. R., Miao, F., Arn, E. A., Stevens, S. W., Ho, C. K., Rauhut, R., and Abelson, J. N. (1997) The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases, *Cell 89*, 849-858.
301. Yamasaki, S., Ivanov, P., Hu, G. F., and Anderson, P. (2009) Angiogenin cleaves tRNA and promotes stress-induced translational repression, *J Cell Biol 185*, 35-42.
302. Luhtala, N., and Parker, R. (2010) T2 Family ribonucleases: ancient enzymes with diverse roles, *Trends Biochem Sci 35*, 253-259.
303. Cooper, D. A., Jha, B. K., Silverman, R. H., Hesselberth, J. R., and Barton, D. J. (2014) Ribonuclease L and metal-ion–independent endoribonuclease cleavage sites in host and viral RNAs, *Nucleic Acids Research 42*, 5202-5216.
304. Nikolaev, Y., Deillon, C., Hoffmann, S. R., Bigler, L., Friess, S., Zenobi, R., Pervushin, K., Hunziker, P., and Gutte, B. (2010) The leucine zipper domains of the transcription factors GCN4 and c-Jun have ribonuclease activity, *PLoS One 5*, e10765.
305. Licht, K., Medenbach, J., Luhrmann, R., Kambach, C., and Bindereif, A. (2008) 3'-cyclic phosphorylation of U6 snRNA leads to recruitment of recycling factor p110 through LSm proteins, *RNA 14*, 1532-1538.
306. Brown, T. S., Chadalavada, D. M., and Bevilacqua, P. C. (2004) Design of a Highly Reactive HDV Ribozyme Sequence Uncovers Facilitation of RNA Folding by Alternative Pairings and Physiological Ionic Strength, *Journal of Molecular Biology 341*, 695-712.
307. Mohr, S., Ghanem, E., Smith, W., Sheeter, D., Qin, Y., King, O., Polioudakis, D., Iyer, V. R., Hunicke-Smith, S., Swamy, S., Kuersten, S., and Lambowitz, A. M. (2013) Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing, *RNA 19*, 958-970.
308. Bolger, A. M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics 30*, 2114-2120.
309. (2010) FASTX-Toolkit.
310. Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., and Nekrutenko, A. (2010) Manipulation of FASTQ data with Galaxy, *Bioinformatics 26*, 1783-1785.
311. Langmead, B., and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2, *Nat Meth 9*, 357-359.

312. Li, H., and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform, *Bioinformatics 26*, 589-595.
313. Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F. J., Maurel, T., McLaren, W., Murphy, D. N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H. S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S. P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S. J., Cunningham, F., Aken, B. L., Zerbino, D. R., and Flicek, P. (2016) Ensembl 2016, *Nucleic Acids Research 44*, D710-D716.
314. Anders, S., Pyl, P. T., and Huber, W. (2014) HTSeq - A Python framework to work with high-throughput sequencing data, *bioRxiv*.
315. Quinlan, A. R., and Hall, I. M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics 26*, 841-842.
316. Nawrocki, E. P., and Eddy, S. R. (2013) Infernal 1.1: 100-fold faster RNA homology searches, *Bioinformatics 29*, 2933-2935.
317. Rampášek, L., Jimenez, R. M., Lupták, A., Vinař, T., and Brejová, B. (2016) RNA motif search with data-driven element ordering, *BMC Bioinformatics 17*, 216.
318. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. R. (2003) Rfam: an RNA family database, *Nucleic Acids Res 31*.
319. Golden, B. L., and Cech, T. R. (1996) Conformational switches involved in orchestrating the successive steps of group I RNA splicing, *Biochemistry 35*, 3754-3763.
320. Remus, B. S., and Shuman, S. (2013) A kinetic framework for tRNA ligase and enforcement of a 2′-phosphate requirement for ligation highlights the design logic of an RNA repair machine, *RNA 19*, 659-669.
321. Perreault, J., Weinberg, Z., Roth, A., Popescu, O., Chartrand, P., Ferbeyre, G., and Breaker, R. R. (2011) Identification of Hammerhead Ribozymes in All Domains of Life Reveals Novel Structural Variations, *PLoS Computational Biology 7*, 1-13.
322. Stevens, S. W., Ryan, D. E., Ge, H. Y., Moore, R. E., Young, M. K., Lee, T. D., and Abelson, J. Composition and Functional Characterization of the Yeast Spliceosomal Penta-snRNP, *Molecular Cell 9*, 31-44.
323. Ruminski, D. J., Webb, C.-H. T., Riccitelli, N. J., and Lupták, A. (2011) Processing and translation initiation of non-long terminal repeat retrotransposons by hepatitis delta virus (HDV)-like self-cleaving ribozymes, *J Biol Chem 286*.
324. Nawrocki, E. P., Kolbe, D. L., and Eddy, S. R. (2009) Infernal 1.0: inference of RNA alignments, *Bioinformatics 25*.
325. Eddy SR. RNABob: a program to search for RNA secondary structure motifs in sequence databases. 1996. unpublished.
326. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool, *J Mol Biol 215*, 403-410.

327. National Library of Medicine (US), N. C. f. B. I. (2002) PART 1 The Databases, In *The NCBI handbook [Internet]*, Bethesda (MD).
328. Biedler, J. K., and Tu, Z. (2007) The Juan non-LTR retrotransposon in mosquitoes: genomic impact, vertical transmission and indications of recent and widespread activity, *BMC Evolutionary Biology 7*, 112.
329. Arensburger, P., Hice, R. H., Wright, J. A., Craig, N. L., and Atkinson, P. W. (2011) The mosquito Aedes aegypti has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs, *BMC Genomics 12*, 606-606.
330. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Research 31*, 3406-3415.
331. Kerpedjiev, P., Hammer, S., and Hofacker, I. L. (2015) Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams, *Bioinformatics 31*, 3377-3379.
332. Honda, S., Morichika, K., and Kirino, Y. (2016) Selective amplification and sequencing of cyclic phosphate-containing RNAs by the cP-RNA-seq method, *Nat. Protocols 11*, 476-489.

VITA

VITA

Aamir Mir

Department of Biochemistry, Purdue University

Education:

2011-2016 Ph.D., Major: Biochemistry, Concentration: Computational Life Sciences, Purdue University, West Lafayette, IN

2007-2011 B.A., Major: Interdisciplinary major in Biology and Chemistry, Lawrence University, Appleton, WI

Research Training:

2011-2016 Graduate Research Assistant, Advisor: Dr. Barbara Golden, Department of Biochemistry, Purdue University, West Lafayette, IN

2010-2011 Undergraduate Research Assistant, Advisor: Dr. Kim Dickson, Lawrence University, Appleton, WI

Teaching Training:

2016 Graduate Teaching Assistant, Professor: Dr. Orla Hart, Department of Biochemistry, Purdue University, West Lafayette, IN

2012-2016 Mentor for Undergraduate and Graduate Students, Department of Biochemistry, Purdue University, West Lafayette, IN

2010 Undergraduate Teaching Assistant, Professor: Dr. Brad Rence, Lawrence University, Appleton, WI

Service:

2013-2016 Senator, Purdue Graduate Student Government, Purdue University, West Lafayette, IN

2014-2016 Coordinator, Graduate-Postdoc Seminar Series, Department of Biochemistry, Purdue University, West Lafayette, IN

2013-2015 Chair and Member, Student Invited Seminar Committee, Department of Biochemistry, Purdue University, West Lafayette, IN

Awards and Fellowships:

2016 Bisland Dissertation Fellowship, Purdue University, West Lafayette, IN

2015 Purdue Graduate Student Professional Grant
Purdue University, West Lafayette, IN
2015-2016 Bird Stair Fellowship,
Purdue University, West Lafayette, IN
2015 Purdue Graduate Student Travel Grant
Purdue University, West Lafayette, IN
2015 RNA Society Travel Fellowship, 2015 RNA Society Meeting,
Madison, WI
2011 Howard and Helen Russell Award for Excellence in Biological Research,
Lawrence University, Appleton, WI

Scientific Meetings:

2016 Poster Presentation: "Revisiting the catalytic mechanism of the hammerhead ribozyme", *Rustbelt RNA Meeting,* Cleveland, OH.
2015 Invited Oral Presentation: "Revisiting the catalytic mechanism of the hammerhead ribozyme", *The 20th Annual Meeting of the RNA Society,* Madison, WI.
2015 Poster Presentation: "Revisiting the catalytic mechanism of the hammerhead ribozyme", *The 20th Annual Meeting of the RNA Society,* Madison, WI.
2015 Poster Presentation: "Revisiting the catalytic mechanism of the hammerhead ribozyme", *Midwest Enzyme Chemistry Conference,* Chicago, IL.
2014 Poster Presentation: "Revisiting the catalytic mechanism of the hammerhead ribozyme", *Midwest Enzyme Chemistry Conference,* Evanston, IL.
2013 Poster Presentation: "The catalytic mechanism of the hammerhead ribozyme", *Rustbelt RNA Meeting,* Cleveland, OH.
2012 Invited Oral Presentation: "Role of G12 and divalent metals in the catalytic mechanism of the hammerhead ribozyme", *Biochemistry Annual Research Retreat,* Turkey Run, IN

Publications:

1- **Mir A**, Chen J, Robinson K, Lendy E, Goodman J, Neau D, Golden B. Two divalent metal ions and conformational changes play roles in the hammerhead ribozyme cleavage reaction. *Biochemistry (ACS Publications)* 2015
2- **Mir A,** Golden B. Two active site divalent ions in the crystal structure of the hammerhead ribozyme bound to a transition state analogue. *Biochemistry (ACS Publications)* 2016